Exploring transport processes across the symbiotic interface of amoebal host and early-stage photosynthetic organelle in *Paulinella chromatophora*

Inaugural-Dissertation

Zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Linda Oberleitner

aus Herdecke, Deutschland

Düsseldorf, Dezember 2020

aus dem Institut für Mikrobielle Zellbiologie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Prof. Dr. Eva C. M. Nowack
2. Prof. Dr. Lutz Schmitt

Tag der mündlichen Prüfung: 07.04.2021

Eidesstattliche Versicherung

Ich versichere an Eides Statt, dass die Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist. Die Dissertation wurde in ihrer jetzigen oder ähnlichen Form noch bei keiner anderen Hochschule eingereicht. Ich habe zuvor keine erfolglosen Promotionsversuche unternommen.

_____ _____

Ort, Datum Linda Oberleitner

Die Untersuchungen zur vorliegenden Arbeit wurden von Juli 2016 bis August 2020 in Düsseldorf an der Heinrich-Heine-Universität am Institut für Mikrobielle Zellbiologie unter der Betreuung von Frau Prof. Dr. Eva C. M. Nowack durchgeführt.

Teile dieser Arbeit wurden veröffentlicht in:

Oberleitner L, Poschmann G, Macorano L, Schott-Verdugo S, Gohlke H, Stühler K, and Nowack ECM (2020). The puzzle of metabolite exchange and identification of putative octotrico peptide repeat expression regulators in the nascent photosynthetic organelles of *Paulinella chromatophora*. *Front. Microbiol.* 11:607182.

Summary

The endosymbiotic acquisition of mitochondria and plastids > 1 billion years ago was central for the evolution of eukaryotic life. However, the complex processes that enable organelle integration remain poorly understood as mitochondria and plastids, owing to their ancient origin, can provide only limited insights into early stages of organellogenesis. The photosynthetic organelles in amoeba of the genus *Paulinella,* called “chromatophores”, evolved independently from plastids ~100 million years ago from a cyanobacterium. Thus, chromatophores offer the possibility to determine common rules and to reveal the degrees of freedom during early steps of organelle evolution. It has been shown earlier that massive import of nuclear encoded proteins into chromatophores has already established and a chromatophore transit peptide (crTP) has been identified. However, little is known about the chromatophore protein import route and the factors involved. Protein import machineries comparable to the translocons present in mitochondria and plastids are absent and the crTP differs fundamentally from other known N-terminal targeting sequences. To learn more about the function of the crTP, mature N-termini of nuclear encoded proteins were determined upon import into chromatophores via high-efficiency undecanal-based N-termini enrichment followed by mass spectrometry. The analysis revealed specific cleavage of the crTP and enabled revaluation of the current working model for protein import into chromatophores that involves protein trafficking through the Golgi. Interestingly, chromatophores also show tight metabolic integration into the host cell, while at the same time chromatophore-encoded metabolite transporters are strikingly rare. These findings imply that similar to the situation in mitochondria and plastids, nuclear encoded transporters were recruited to establish metabolic connectivity. To identify such transporters, enriched insoluble protein fractions of isolated chromatophores were analyzed by mass spectrometry. In a second approach, selected nuclear encoded candidate chromatophore envelope transporters were characterized in heterologous systems and specific antibodies were generated to evaluate the subcellular localization of a putative amino acid transporter. The results imply that, unexpectedly, nuclear encoded transporters are not inserted into the chromatophore inner envelope. Instead, several expanded groups of short chromatophore-targeted orphan proteins were identified that might be involved in modulating membrane permeability, suggesting that the mechanism generating metabolic connectivity of the chromatophore fundamentally differs from the one in mitochondria and plastids. Yet another critical factor for organellogenesis appears to be the establishment of nuclear control over organellar gene expression. Chromatophore-localized biosynthetic pathways as well as multiprotein complexes include proteins encoded in either the organellar or nuclear genome, suggesting coordination of gene expression levels between both. This study identified an expanded family of chromatophore-targeted helical repeat proteins that resemble known eukaryotic organelle-targeted expression regulators, suggesting their convergent evolution.

Zusammenfassung

Der endosymbiotische Erwerb von Mitochondrien und Plastiden vor > 1 Milliarde Jahren war zentral für die Evolution eukaryotischen Lebens. Da diese Organellen jedoch, bedingt durch ihre frühe Entstehung, nur eingeschränkt Einblicke in frühe Stadien der Organellogenese bieten können, sind die komplexen Prozesse, die die Integration eines Organells ermöglichen, noch weitgehend unverstanden. Die photosynthetischen Organellen in Amöben der Gattung *Paulinella*, „Chromatophoren" genannt, entstanden unabhängig von Plastiden vor ~100 Millionen Jahren aus einem Cyanobakterium. Daher bieten Chromatophoren die Möglichkeit, sowohl allgemeine Regeln, als auch Freiheitsgrade früher Stadien der Organellenevolution zu ermitteln. Bisher konnte gezeigt werden, dass ein massiver Import von kernkodierten Proteinen in Chromatophoren bereits stattfindet und ein Chromatophoren-Transitpeptid (crTP) wurde identifiziert. Über den Proteinimportweg und involvierte Faktoren ist jedoch einstweilen wenig bekannt. Proteinimportmaschinerien, die vergleichbar mit den Translocons in Mitochondrien und Plastiden wären sind nicht vorhanden und das crTP unterscheidet sich fundamental von anderen N-terminalen Targeting-Sequenzen. Um mehr über die Funktion des crTPs zu erfahren, wurden die N-Termini kernkodierter Proteine nach dem Import in die Chromatophoren durch „hocheffiziente Undecanal-basierte N-Termini-Anreicherung" und anschließende massenspektrometrische Analyse bestimmt. Die Analyse offenbarte eine definierte Spaltung des crTPs und ermöglichte die Weiterentwicklung des aktuellen Modells für Proteinimport in Chromatophoren, welches Proteinmigration durch den Golgiapparat vorsieht. Interessanterweise sind Chromatophoren auch metabolisch stark in die Wirtszelle integriert, während chromatophorenkodierte Metabolittransporter gleichzeitig bemerkenswert selten sind. Diese Erkenntnis legt nahe, dass, wie in Mitochondrien und Plastiden, kernkodierte Transporter rekrutiert wurden, um metabolische Konnektivität zu etablieren. Um solche Transporter zu identifizieren, wurden unlösliche Proteinfraktionen isolierter Chromatophoren angereichert und massenspektrometrisch analysiert. In einem zweiten Ansatz wurden ausgesuchte kernkodierte Transporter in heterologen Systemen charakterisiert und spezifische Antikörper wurden hergestellt, um die subzelluläre Lokalisation eines mutmaßlichen Aminosäuretransporters zu bestimmen. Die Ergebnisse zeigen unerwarteterweise, dass kernkodierte Transporter nicht in die innere Chromatophorenhüllmembran eingesetzt werden. Stattdessen wurden kurze Proteine in Chromatophoren identifiziert, die Mitglieder größerer Gruppen zu sein scheinen und in die Modulierung der Membranpermeabilität involviert sein könnten, was auf einen zu Mitochondrien und Plastiden fundamental unterschiedlichen Mechanismus zur Herstellung metabolischer Konnektivität hinweist. Einen weiteren kritischen Faktor bei der Organellogenese stellt die Etablierung der Kontrolle des Zellkerns über die Genexpression im Organell dar. Chromatophorenlokalisierte Biosynthesewege und Multi-Proteinkomplexe setzen sich sowohl aus chromatophorenkodierten, als auch aus kernkodierten Proteinen zusammen, was auf eine Koordinierung der Genexpression beider Genome hinweist. Durch Massenspektrometrie konnte eine Familie kernkodierter „Helical Repeat Proteine" in Chromatophoren identifiziert werden, die bekannten eukaryotischen Genexpressionsregulatoren ähneln, was auf eine konvergente Evolution hinweist.

List of Abbreviations

%	Percentage	iMet	Initial methionine residue of proteins
aa	Amino acids	*in silico*	By using computational simulation
Amp	Ampicillin	*in vitro*	Independent of a living organism
AMP	Antimicrobial peptide	*in vivo*	In a living organism
APS	Ammoniumpersulfate	Kan	Kanamycin
BSA	Bovine serum albumin	kb	Kilobases
Cam	Chloramphenicol	kDa	Kilodaltons
cDNA	Complementary DNA	LC-MS/MS	Liquid chromatography coupled to tandem mass spectrometry
CL	Chromatophore lysate	MES	2-(N-morpholino)ethanesulfonic acid
CM	Chromatophore insoluble fraction	mRNA	Messenger RNA
crTP	Chromatophore transit peptide	MS	Mass spectrometry (usually refers to LC-MS/MS)
cTP	Chloroplast transit peptide	mTP	Mitochondrial transit peptide
C-terminal	Carboxyterminal	N-terminal	Aminoterminal
dH_2O	Distilled water	nt	Nucleotides
DMSO	Dimethylsulphoxide	OM	Outer membrane
DNA	Deoxyribonucleic acid	ORF	Open reading frame
dNTP	Deoxynucleoside triphosphate	P_i	Inorganic phosphate (PO_4^{3-})
DTT	Dithiothreitol	PEG	Polyethylene glycol
EDTA	Ethylenediaminetetraacetic acid	PM	*Paulinella* whole cell insoluble fraction
EGTA	Ethyleneglycoldiaminetetraacetic acid	RNA	Ribonucleic acid
EM	Electron microscopy	RT	Room temperature
ER	Endoplasmic reticulum	SDS	Sodiumdodecylsulfate
EtOH	Ethanol	SL	Spliced leader
et al.	*Et alii* (and others)	SP	Signal peptide
EGT	Endosymbiotic gene transfer	SPP	Stromal processing peptidase
FL	Full-length	Spec	Spectinomycin
Gbp	Giga base pairs	TEM	Transmission electron microscopy
gDNA	Genomic DNA	TEMED	Tetramethylethylenediamine
GFP	Green fluorescent protein	TMH	Transmembrane helix
h	Hour	Tris	Tris(hydroxymethyl)aminomethane
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid	v/v	Volume per volume
HGT	Horizontal/lateral gene transfer	w/v	Weight per volume
HPLC	High pressure liquid chromatography	WT	Wild-type
IM	Inner membrane	x g	x fold gravitational force

Table of Contents

1 Introduction

1.1 Origin of Plastids: A Cyanobacterial Endosymbiont Enables Eukaryotic Photosynthesis

Oxygenic photosynthesis represents an ancient bioenergetic concept for the light energy-driven synthesis of energy-rich organic sugars utilizing the inorganic compounds CO_2 and water. Established by cyanobacteria probably > 2.3 billion years ago, it became by far the most successful concept for primary production (Fischer *et al.*, 2016; Rasmussen *et al.*, 2013). As a by-product oxygenic photosynthesis reshaped the planets early anoxic atmosphere to an O_2-rich atmosphere allowing for the evolution of aerobic respiration and thus for the complex multicellular consumers we associate with modern biology (Fischer *et al.*, 2016; Raymond & Segre, 2006).

However, today oxygenic photosynthesis is not restricted to cyanobacteria, but is also performed by a myriad of eukaryotes, ranging in their complexity from unicellular algae to trees. Plastids are organelles that carry out the photosynthetic reactions in eukaryotes and provide the whole cell with energy-rich compounds, thus enabling photoautotrophic lifestyle. Endosymbiotic theory postulates that the eukaryotic semi-autonomous organelles – plastids and mitochondria – both are derived from formerly free-living prokaryotes that initially became endosymbiotic to eukaryotic hosts, an event called "primary endosymbiosis" (Mereschkowsky, 1905; Sagan/Margulis, 1967). Although aspects of the theory are still discussed controversially, it has been largely accepted (Gray, 2017; Martin *et al.*, 2015). Accordingly, plastids apparently evolved from a cyanobacterium of the β-clade which became endosymbiotic to a heterotrophic eukaryotic host cell > 1 billion years ago (Marin *et al.*, 2005; Mereschkowsky, 1905; Rodríguez-Ezpeleta *et al.*, 2005), although the timing is debated (McFadden, 2014). This single event of primary endosymbiosis gave rise to three eukaryotic photoautotrophic lineages which together comprise the Archaeplastida: Glaucophytes, rhodophytes (red algae) and Viridiplantae (green algae and land plants)(Gould *et al.*, 2008; Rodríguez-Ezpeleta *et al.*, 2005). As remnant characteristics of their cyanobacterial ancestors, all chloroplasts retain a circular haploid genome (Ris and Plaut, 1962; Timmis *et al.*, 2004) and a clearly prokaryotic gene expression machinery is still recognizable (Barkan, 2011). Also, plastids are surrounded by two envelope membranes that are derived from the cyanobacterial inner and outer cell membranes, respectively (Cavalier-Smith, 2000). The morphology of glaucophyte plastids even still resembles cyanobacteria as they retained carboxysome-like bodies, phycobilisomes, a peptidoglycan wall and concentric thylakoid architecture (Jackson *et al.*, 2015).

Once established, photoautotrophy spread across various eukaryotic lineages by the engulfment of photosynthetic unicellular eukaryotes by heterotrophic unicellular eukaryotes. This process, called secondary endosymbiosis – or tertiary endosymbiosis, when engulfment occurred in a serial fashion – strongly contributed to today´s fascinating diversity of photoautotrophs, including chimeric eukaryotes with chloroplasts surrounded by three or even four membranes (Gould *et al.*, 2008; Keeling, 2010).

1.1.1 Massive Endosymbiotic Genome Reduction Requires Import of Nuclear Encoded Proteins into Plastids

Endosymbiotic lifestyle is usually accompanied by a drastic reduction in genome size and coding capacity in the endosymbiont (McCutcheon, 2016). Accordingly, chloroplast genomes range in coding capacity between 60 and 200 genes, which is at all events very reduced compared to genomes of free-living cyanobacteria (2000-5000 genes) (Lee *et al.*, 2016; Timmis *et al.*, 2004). Many genes were lost completely as the encoded products became dispensable in an intracellular environment while other genes were transferred to the host´s nuclear genome in a process called Endosymbiotic Gene Transfer (EGT). The transfer of genetic material is enabled whenever lysis of single plastids may occasionally occur and seems to be an ongoing process in species with several plastids per cell (Lister *et al.*, 2003; Stegemann *et al.*, 2003). Following transfer, a newly acquired gene can conceivably even get expressed by the eukaryotic machinery and the protein might get retargeted post-translationally to plastids, where the original copy is redundant and thus eventually lost. Even though this series of events bears numerous obstacles, e.g. functional expression of a prokaryotic gene in a eukaryotic environment and targeting to / import into an organelle enveloped by two membranes, apparently at least 400 EGT-derived proteins are encoded in the nucleus, while fulfilling their function in the plastid in the model plant *Arabidopsis thaliana* (hereafter *Arabidopsis*; Qiu *et al.*, 2013). However, we know today that the major part of the plastid proteome is not comprised by EGT-derived proteins of cyanobacterial ancestry, but by host proteins that got re-targeted to plastids or – to a minor extent – by proteins with bacterial ancestry that were integrated into the nuclear genome via Horizontal Gene Transfer (HGT) (Qiu *et al.*, 2013).

The targeting and import of an estimated 2000-4000 nuclear encoded and cytosolically synthesized proteins (in *Arabidopsis*; Leister, 2016) to / into chloroplasts is usually tied to the presence of cleavable N-terminal sequences on these proteins, the so called Chloroplast Transit Peptides (cTPs). They generally are comprised of 50-70 amino acids that share no consensus sequence or motifs. Though, they are overall rich in hydroxylated, basic, and hydrophobic amino acids, with the first 15 amino acids being specifically rich in serine (Bhushan *et al.*, 2006). Also, cTPs are believed to be unstructured with some reports about the formation of amphipathic α-helices when a membranous environment is

mimicked (Bruce, 2001; Wienk *et al.*, 1999). Thus, physicochemical and structural characteristics might provide cTP function (Garg and Gould, 2016).

For actual translocation of cTP-containing pre-proteins across the chloroplast envelope the cTP is recognized by a complex multi-component machinery: The TOC / TIC complex (for Translocon of the Outer / Inner Chloroplast envelope). The protein-conducting β-barrel outer envelope pore is represented by Toc75. The protein is nuclear encoded in plants but apparently originated from its cyanobacterial ancestor Omp85, a pore that facilitates insertion of membrane proteins into outer membranes of Gram-negative bacteria (Bodył *et al.*, 2009; Gross and Bhattacharya, 2009b; Knopp *et al.*, 2020). The composition of the TIC complex has been disputed, as Tic110, Tic20, and Tic21 each have been proposed to form the protein-conducting channel in the inner envelope (Heins *et al.,* 2002; Kovács-Bogdán *et al.,* 2011; Teng *et al.,* 2006; but see Bölter and Soll, 2016; Duy *et al.,* 2007; Nakai, 2018; Richardson and Schnell, 2019; Gross and Bhattacharya, 2009). For import into the plastid stroma, unfolded pre-proteins are bound by specific chaperones in the cytosol, then guided to the translocon and recognized by membrane-bound receptors. Successive funneling of the unfolded pre-proteins through Toc75 and Tic110/20/21 requires energy derived from ATP hydrolysis as well as the activity of several molecular chaperones and proteins involved in redox regulation of the machinery (Gross and Bhattacharya, 2009b). Upon import, the cTP is cleaved by Stromal Processing Peptidase (SPP) (Teixeira and Glaser, 2013). Mature stromal proteins can now fold into their functional conformation, while proteins designated to thylakoids are further imported into the sub-compartment via the bacterial Sec- or Tat-pathways. This usually requires a bipartite targeting signal consisting of a bacterial signal peptide C-terminal to the cTP (Bodył *et al.*, 2009).

Noteworthy, a fraction of proteins found in plastids seem to lack a cTP (Armbruster *et al.*, 2009; Kleffmann *et al.*, 2004). Signal-anchored or tail-anchored single-spanning outer membrane proteins (where the transmembrane domain is located in the very N-terminus or C-terminus, respectively) as well as many β-barrel integral outer membrane proteins represent well known examples (Kim *et al.*, 2019), while the extent of cTP independent import is under debate for inner envelope proteins and stromal proteins. Experiments on few proteins indicate that TOC / TIC-independent import via vesicle trafficking from the endomembrane system is utilized for delivery to the plastid surface (Kaneko *et al.*, 2016; Kitajima *et al.*, 2009; Nanjo *et al.*, 2006; Villarejo *et al.*, 2005). However, the exact mechanism and especially how passage across the inner envelope is accomplished remain speculative (Baslam *et al.*, 2016).

Last but not least it should be noted that the mechanism for protein import into mitochondria shares remarkable similarity to the one that operates in plastids. Proteins destined for import carry a Mitochondrial Targeting Peptide (mTP) that resembles cTPs in amino acid composition and secondary

structure. The Translocons of the Outer / Inner Mitochondrial envelope (forming the TOM / TIM complex) display several structural and functional commonalities with TOC / TIC, although the systems do not share homology. Thus, a convergent trajectory for the establishment of protein import into these organelles might be assumed (Gross and Bhattacharya, 2009a; Sommer and Schleiff, 2014).

1.1.2 Nuclear Encoded Transporters Control Metabolic Fluxes across the Plastid Envelope

The stable integration of an endosymbiont into a host cell apparently demands a high degree of metabolic connectivity between both partners. Modern chloroplasts provide the cell with numerous compounds like most amino acids, fatty acids and glycerolipids, carotenoids and riboflavin, as well as some hormones and secondary metabolites (Chen *et al.*, 2018). However, their primary function is carbon assimilation. Thus, it can be assumed that chloroplasts are the result of an endosymbiosis that was initiated primarily by the host cell to exploit photosynthesis as a comfortable source of energy in form of reduced carbon compounds.

In line with that, the most prominent plastid envelope transporters for sugar-phosphates are host-encoded in all three lineages of Archaeplastida. UhpC, a glucose-6-phosphate / phosphate exchanger, is suggested to represent the earliest facilitator for the flux of carbon from endosymbiont to host (Facchinelli *et al.*, 2013b; Price *et al.*, 2012). The protein likely was acquired via HGT from intracellular parasites (Chlamydiales or Rickettsiales) that were present in the same host cell as the young cyanobacterial endosymbiont (Karkar *et al.*, 2015). However, today the more effective transporters of the plastidic phosphate translocator family represent the most prominent carbon exchangers in rhodophytes and chlorophytes and UhpC was even lost in land plants (Nowack and Weber, 2018). Phylogenetic analysis showed that plastidic phosphate translocators are related to Nucleotide-Sugar Translocators (NSTs) of the eukaryotic endomembrane system, indicating that such a transporter was re-targeted to the endosymbiont´s inner envelope and initially enabled the export of ADP-glucose in exchange for AMP (Colleoni *et al.*, 2010; Weber *et al.*, 2006). The family expanded and today in higher plants comprises the main carbon shuttle Triose-phosphate / Phosphate Translocator (TPT) (Heldt and Rapley, 1970) as well as the specialized Glucose-6-phosphate / Phosphate Translocator (GPT), Xylulose-5-phosphate / Phosphate Translocator (XPT), and Phospho*enol*pyruvate / Phosphate Translocator (PPT) (Weber *et al.*, 2006). GPT and XPT facilitate the plastidial import of sugar-phosphates that serve as substrates for the oxidative pentose phosphate pathway as well as starch synthesis or the Calvin-Benson cycle, respectively, while PPT imports phospho*enol*pyruvate that is used in the shikimate pathway (Weber and Linka, 2011). The homo-exchange of carbon compounds for inorganic phosphate (P_i) guarantees a balance in the phosphate content of the plastid and the cytosol and ensures a

constant provision of phosphate to ATP synthesis by the photosynthetic light reactions (Nowack and Weber, 2018).

Another transporter that is present in all archaeplastidial lineages and thus was established early in endosymbiosis is the Nucleoside Triphosphate Transporter (NTT) (Linka *et al.*, 2003; Nowack and Weber, 2018). It provides the plastid with ATP in exchange for ADP and P_i whenever the demand for ATP cannot be covered by the photosynthetic light reactions, e.g. at night. That was important as storage polysaccharides were synthesized and stored in the cytosol in the form of glycogen in the archaeplastidial ancestor. Today starch represents the default storage polysaccharide found in Archaeplastida and still is exclusively found in the cytosol in rhodophytes and glaucophytes (Cenci *et al.*, 2014). Noteworthy, like UhpC, also NTT represents an HGT from an intracellular chlamydia-like parasite (Karkar *et al.*, 2015).

Generally HGT is believed to have strongly contributed to the integration of plastids and the relatively huge proportion of genes of chlamydial affiliation in archaeplastidial genomes (Karkar *et al.*, 2015; Qiu *et al.*, 2013) lead to the proposal of the so-called ménage à trois hypothesis (Cenci *et al.*, 2017; Facchinelli *et al.*, 2013a). It proposes that the cyanobacterial endosymbiont entered its host cell simultaneously with a chlamydia-like parasite and that both prokaryotes were initially surrounded by a host-derived vesicle. Ancient carbon flux might have been established with the help of chlamydial HGTs, e.g. leading to the integration of glucose-6-phosphate exporter UhpC into the cyanobacterial plasma membrane. Further flux from the vesicle that houses both prokaryotes might have been facilitated by a nucleotide-sugar transporter, representing the aforementioned ancient eukaryotic ADP-glucose exporter.

However, today the typical plant plastid inner envelope permeome comprises an estimated 100-150 transporters, many of them being metabolite transporters and almost all of them being nuclear encoded (Mehrshahi *et al.*, 2013; Weber *et al.*, 2005). Noteworthy, more than half of the 37 well-studied examples are also phylogenetically host-derived (Karkar *et al.*, 2015). Although most of the putative envelope transporters are not functionally characterized, functions proposed by sequence similarity and transporter classification point to the presence of transporters involved in ion homeostasis (Finazzi *et al.*, 2015) as well as transport of sugar-phosphates, carboxylates, and nucleotides or cofactors that cannot be synthesized in plastids like nicotinamide adenine dinucleotide, S-adenosylmethionine, or folates. Correspondingly, several nuclear encoded porins that show broader substrate specificities are known to permit passage of solutes across the chloroplast outer envelope (Breuers *et al.*, 2011; Goetze *et al.*, 2015; Wang *et al.*, 2013). In contrast, the plant plastidial genome does not encode metabolite transporters anymore, indicating that the metabolic flux between a plant cell and its chloroplast is under nuclear control.

1.1.3 Nuclear Encoded Factors Control Expression of Plastid-encoded Proteins

As mentioned before, some proteins remain to be encoded in the residual plastidial genome. As a result, many protein complexes are of dual genetic origin, i.e. they are composed of subunits that are either encoded in the nuclear or the plastidial genome. Nevertheless, these complexes have to be assembled in stochiometric amounts. Also, coordination of nuclear and plastidial gene expression is essential to respond to changes in environmental conditions and – in the case of plants – to control the differentiation of different kinds of plastids (Kusnetsov, 2018). Thus, integration of an organelle does require well-regulated interactive gene expression. Several nuclear encoded factors are known in plants and algae that provide control over gene expression on different levels. For example, transcriptional regulation is provided by the function of nuclear encoded plastid-targeted sigma factors and transcription factors (Kusnetsov, 2018). However, organellar protein levels are determined in first line by post-transcriptional mechanisms that influence RNA processing and stability. Such mechanisms include cleavage of polycistronic plastidial RNAs, intron splicing, polyadenylation, and cytosine to uracil RNA editing (del Campo, 2009). Among others, nuclear encoded helical repeat proteins like those of the Pentatrico Peptide Repeat (PPR) family are predominantly involved in regulation of these processes not only in plastids but also in mitochondria (Barkan and Small, 2014). PPR-proteins are characterized by stretches of tandem degenerate 35 aa sequence motifs, with each motif folding into two α-helices causing the motif stretches to form a super helix (Yin *et al.*, 2013). The structure enables sequence-specific binding to single-stranded RNA, allowing for the stabilization of or factor recruitment to specific mRNAs (Yan *et al.*, 2019). According to their specificity, the family of PPR-proteins comprises around 150 individual plastid-targeted members in *Arabidopsis* (Colcombet *et al.*, 2013). Noteworthy, PPR-proteins do not seem to represent the major group of helical repeat-containing plastid gene expression regulators in all photosynthetic eukaryotes. For example, the family of Octotrico Peptide Repeat (OPR) proteins (containing 38 aa motifs) dominates in the model green algae *Chlamydomonas reinhardtii* (hereafter *Chlamydomonas*) with around 30 putatively plastid-targeted members (Eberhard *et al.*, 2011) compared to only around 5 plastid-targeted PPR-proteins (Tourasse *et al.*, 2013). Similarly, further families of helical repeat proteins as the HAlf-a-Tetratricopeptide repeat (HAT; 34 aa motifs) and Mitochondrial TERmination Factor (mTERF; 30 aa motifs) proteins are known to be involved in gene expression in both the eukaryotic organelles, plastids and mitochondria (Hammani *et al.*, 2014; Kleine, 2012; Preker and Keller, 1998). Thus, different helical repeat proteins appear to represent factors ubiquitously expressed in eukaryotic genomes on the purpose of providing control over organellar gene expression (Hammani *et al.*, 2014).

1.2 *Paulinella chromatophora* as a Model to Study Organellogenesis

The early steps leading to the acquisition of eukaryotic organelles mitochondria and plastids still remain miraculous, although a lot of effort has been made to understand the processes involved, as well as the environmental and physiological requirements, and the timescale. The reason is obvious: The event in question is exceptionally rare as it apparently happened only twice in the history of life and additionally is tremendously ancient as it happened more than 1.5 and 1 billion years ago, respectively (Embley and Martin, 2006; McFadden, 2014).

Fortunately, the discovery of a much more recent organelle in the protist *Paulinella chromatophora* (Fig. 1.2-1A) provides the unique opportunity to study early organellogenesis in progress. *P. chromatophora* represents a filose thecate amoeba of rhizarian affiliation that is globally found in sediments of freshwater ponds and lakes (Bhattacharya *et al.*, 1995). It was first described 115 years ago by Robert Lauterborn who ranked it "among the most interesting representatives of its division" (Lauterborn, 1895; Melkonian and Mollenhauer, 2005). Especially the unusual, large, green, sausage shaped "inclusions" present in each cell attracted his attention and lead him to the following considerations on their nature: "Either they represent autonomous cyanobacterial symbionts or organelles of the amoebal cell body" (Lauterborn, 1895). Today, the nature of the inclusions, referred to as "chromatophores", has been studied extensively and phylogenetic and experimental data suggest that the chromatophore indeed constitutes an early stage photosynthetic organelle (Marin *et al.*, 2005; Nowack and Grossman, 2012; Singer *et al.*, 2017). Thus, *P. chromatophora* is unique as it represents so far the only known organism besides mitochondria and plastids that established organelles of primary endosymbiotic origin (Fig. 1.2-1B) (Gabr *et al.*, 2020; Nowack, 2014). Apparently, the chromatophore was established around 90 – 140 million years ago, making it a "recent" organelle in terms of evolution and compared to plastids that are approximately ten times older (Delaye *et al.*, 2016; McFadden, 2014). Therefore, *P. chromatophora* (and its two known photoautotrophic sister species *P. longichromatophora* [Kim and Park, 2016] and *P. micropora* [Lhee *et al.*, 2017]) can provide insight into the mechanisms that enable a host cell to gain control over its endosymbiont and the series of events that drive its metabolic and genomic integration.

Several phylogenetic analyses indicate that cyanobacteria of the *Synechococcus/Prochlorococcus/Cyanobium* clade (also known as α-cyanobacteria) represent the chromatophores closest relatives, underpinning their independent origin from plastids that are monophyletic to β-cyanobacteria (Marin *et al.*, 2005; 2007; Yoon *et al.*, 2006; 2009). As remnants of their cyanobacterial origin, chromatophores possess carboxysomes, phycobilisomes, concentric thylakoid architecture, and a peptidoglycan wall that is flanked by an inner and an outer envelope membrane (Kies, 1974). The provenance of the latter might be either a host membrane (e.g. a digestive

vacuole [Nomura and Ishida, 2016; Sato *et al.*, 2020]) or the cyanobacterial outer membrane or a mixture of both. However, although they resemble cyanobacterial ultrastructure, chromatophores appear to be extremely large (15-20 µm long) when compared to their free-living cyanobacterial relatives that are at least ten times smaller. The cell cycles of host cell and chromatophore are tightly linked and the number of chromatophores is restricted to two per cell. Following mitosis, one chromatophore is passed on to the daughter cell (see time lapse video in Nomura *et al.*, 2014), where it elongates and undergoes binary fission. Outside of their host cell chromatophores are not able to survive for longer times.

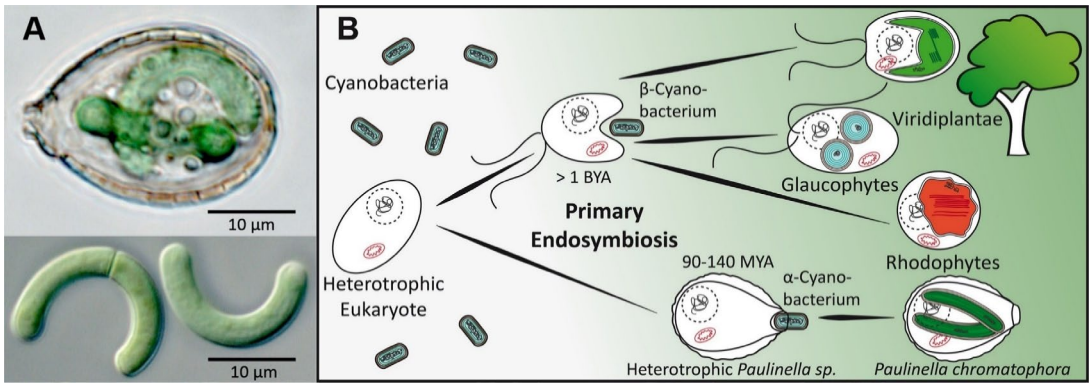


Figure 1.2-1 *Paulinella chromatophora* as a model to study primary acquisition of photosynthesis by eukaryotes. A Microscopic picture of a *P. chromatophora* cell (top) and isolated chromatophores (bottom) taken from Oberleitner *et al.*, 2019. **B** Acquisition of photosynthesis by eukaryotes. Engulfment of a cyanobacterium by a heterotrophic eukaryote > 1 billion years ago gave rise to three eukaryotic lineages harbouring primary plastids (Archaeplastida [upper branch]): Glaucophytes, rhodophytes (red algae), and Viridiplantae (green algae and land plants). The endosymbiosis of a cyanobacterium with an ancestral heterotrophic *Paulinella sp.* was established much more recently and led to the evolution of the chromatophore (lower branch). The presence of two envelope membranes and a vestigial genome attest the cyanobacterial origin of all plastids and chromatophores. Glaucophyte plastids and chromatophores additionally retained a peptidoglycan wall (brown) and concentric thylakoids. The figure was made according to Gould *et al.*, 2008.

1.2.1 Endosymbiotic Genome Reduction and Genome Complementarity

The complete chromatophore genome has been sequenced. It comprises 878 protein-coding genes, indicating a genome reduction by approx. two thirds when compared to free-living α-cyanobacteria, but being still five to ten times larger than typical plastid genomes (Nowack *et al.*, 2008; Timmis *et al.*, 2004). Thus, the chromatophore already underwent reductive genome evolution due to its intracellular lifestyle. In line with that, many genes of unknown function that are believed to be relevant only when free-living cyanobacteria face specific environmental changes are missing on the chromatophore genome. Apart from that, also essential functions were lost either partially or

completely, e.g. several pathways for the synthesis of amino acids like leucine, threonine, proline, arginine, histidine, aromatic amino acids (partially lost), as well as methionine, serine, aspartate, asparagine, and alanine (completely lost) (Nowack *et al.*, 2016). Also, the cofactors nicotinamide adenine dinucleotide, riboflavine, thiamine, biotin, cobalamine, pantothenate, and coenzyme A cannot be synthesized according to the chromatophore´s gene repertoire. Finally, all genes encoding for enzymes involved in TriCarboxylic Acid (TCA) -cycle are absent in the chromatophore genome (Nowack *et al.*, 2008).

Assembly of the *P. chromatophora* nuclear genome failed so far due to its huge size of an estimated 10 Gbp, but Illumina HiSeq data generated from cDNA provides a high-quality transcriptome (Nowack *et al.*, 2016). Analysis of the protein models derived from the transcript reads revealed that functions encoded in the nuclear genome and in the chromatophore genome – to some extend – seem to complement each other, e.g. in the case of the serine- and methionine biosynthetic pathways (Nowack *et al.*, 2016). Both are entirely nuclear encoded and were lost from the chromatophore genome. In contrast, the arginine biosynthetic pathway appears chimeric, involving several exclusively nuclear encoded and exclusively chromatophore-encoded genes. In other pathways then again, like those for proline- or leucine biosynthesis, only a single gene is missing in the chromatophore, while a nuclear copy is present.

At least 58 genes of α-cyanobacterial origin have been identified in the *P. chromatophora* nuclear genome and thus, potentially represent EGTs derived from the chromatophore. Most of these genes play a role in photosynthesis or light adaptation, e.g. three small photosystem I subunits PsaE, PsaK1 and PsaK2 (Nakayama and Ishida, 2009; Nowack *et al.*, 2011; Reyes-Prieto *et al.*, 2010). However, HGT seems to predominate over EGT in the nuclear genome, as at least 169 genes derived from other bacteria were identified. The accumulation of bacterial HGTs might be explained by the phagothrophic lifestyle of *P. chromatophora´s* non-photosynthetic ancestor. It was further hypothesized that HGT-derived genes might enable the replacement of chromatophore-encoded genes by nuclear encoded genes in cases where no eukaryotic gene can provide the required function, e.g. because it is a prokaryote-specific function. Peptidoglycan synthesis represents such a function and in fact one single gene (*murF,* encoding D-Ala-D-Ala ligase) is missing from the chromatophore-encoded pathway, while an HGT-derived *murF*-copy is encoded in the nucleus (Nowack *et al.*, 2016). Other examples are DNA ligase (LigA) and DNA polymerase I (PolA), two enzymes involved in bacterial DNA replication that can be found in the nuclear genome, but not in the chromatophore genome.

Taken together, these observations indicate that proteins encoded in the nucleus – by either repurposed host-derived genes or genes acquired via EGT or HGT – seem to make their way back to the chromatophore where they contribute to its genetic, physiological, and metabolic functions.

1.2.2 Import of Nuclear Encoded Proteins into the Chromatophore

Indeed, the presence of nuclear encoded proteins in chromatophores has been demonstrated. Experimental evidence for the chromatophore localization of the aforementioned photosystem I subunit PsaE came from Immunogold analyses using specific antibodies raised against *P. chromatophora* PsaE and the incorporation of PsaE, PsaK1, and PsaK2 in the photosystem I complex was confirmed by compositional analysis of the isolated complex (Nowack and Grossman, 2012). Furthermore, autoradiography of isolated radiolabeled photosystem I subunits that were synthesized in the presence of chloramphenicol (an inhibitor of prokaryotic translation) demonstrated that PsaE, PsaK1, and PsaK2 are synthesized in the amoebal cytosol by eukaryotic ribosomes, demonstrating that protein (not mRNA) is shuttled (Nowack and Grossman, 2012). On this occasion it has to be pointed out that so far, no method has been established for the genetic manipulation of *P. chromatophora*. Thus, expression of recombinant tagged fusion proteins, which is widely used for determination of protein subcellular localization, is not an option. Nevertheless, the extent of protein import into chromatophores was addressed experimentally only recently, namely applying protein Mass Spectrometry (MS; Singer *et al.,* 2017). The comparison of semi-quantitative shotgun proteomic data from isolated chromatophores and whole *P. chromatophora* cells allowed for the identification of 207 nuclear encoded proteins that appear to be enriched in chromatophores. Those proteins are regarded import candidates and can be classified by their length into short import candidates (< 95 aa) and long import candidates (> 125 aa).

The expanded group of short import candidates is partially comprised of small proteins involved in photosynthesis and light protection, e.g. the aforementioned photosystem I subunits PsaE and PsaK. However, the largest proportion of short import candidates consists of orphan proteins that do not have homologs in other organisms, but comprise conspicuous cysteine motifs (CxxC or CxxxxC) and/or stretches of positively charged amino acids that are characteristic of AntiMicrobial Peptides (AMPs). AMPs are well known eukaryotic effectors that attack pathogenic bacteria either by self-translocation across the membranes to subsequently attack intracellular targets or by formation of pores in the bacterial cell envelope membranes, thereby dissipating the proton motive force (Brogden, 2005). The purpose is to ultimately kill harmful bacteria. Apart from that, so called symbiotic AMPs are produced by host cells in diverse symbiotic systems where their purpose is obviously not to kill intracellular symbionts (Mergaert *et al.*, 2017). Instead, these AMPs seem to enable the host cell to gain control over its endosymbiont, e.g. over its cell division (Login *et al.*, 2011; Van de Velde *et al.*, 2010). Apart from that, also a function of pore-forming symbiotic AMPs in metabolite exchange is discussed (Mergaert *et al.*, 2006; 2017).

Long import candidates usually represent enzymes with functions involved in metabolism or genetic information processing. Many of these seem to specifically fill gaps in chromatophore-encoded pathways, e.g. pyrroline-5-carboxylate reductase (ProC) or 3-isopropylmalate dehydrogenase (LeuB) that catalyze the only enzymatic reactions in proline and leucine biosynthesis, respectively, for which the corresponding gene is missing from the chromatophore genome (Fig. 1.2-2; Singer *et al.,* 2017). The arginine biosynthetic pathway in the chromatophore appears to be chimeric, since three proteins involved are nuclear encoded and imported (bifunctional glutamate/ornithine acetyltransferase ArgJ, N-acetylglutamate kinase ArgB, N-acetyl-gamma-glutamyl-phosphate reductase ArgC), two are chromatophore-encoded (ArgD and ArgF) and further two are nuclear encoded and likely reside in the cell´s mitochondria (ArgG and ArgH) as they possess predicted mTPs. Thus, as assumed earlier, massive protein import into chromatophores indeed enables the host cell to intervene in its endosymbiont´s metabolism.

An intriguing feature of long import candidates (for which N-terminal full-length sequence information is available) is an N-terminal extension of approximately 200 aa. The sequence is highly conserved among import candidates, but shows no similarity to other proteins or domains. A hydrophobic α-helix in the very N-terminus, that is sometimes interpreted as a TransMembrane Helix (TMH) depending on the algorithm used for prediction, represents a conspicuous common structural element. In analogy to cTPs and mTPs that effect protein import into plastids or mitochondria, the sequence was termed Chromatophore Transit Peptide (crTP) and accordingly is believed to facilitate targeting to / import into chromatophores. Owing to the crTP´s marked primary sequence conservation, further 252 long import candidates that have not been MS-identified in the chromatophore proteome could be *in silico* predicted (Singer *et al.*, 2017). Among these are also some of the aforementioned HGT-derived proteins, like MurF (involved in peptidoglycan synthesis) and LigA (involved in DNA replication). Whether the crTP is cleaved off during / after translocation like mTPs and cTPs is not known, but mass spectrometric identification of crTP-derived peptides in chromatophore-derived samples speaks against a full cleavage (Singer *et al.*, 2017). Also the crTP-mediated mechanism for protein transport is still entirely unknown. The only TOC / TIC components found in *P. chromatophora* are chromatophore-encoded homologs of the putative inner envelope protein conducting pore Tic21 (but see Duy *et al.,* 2007), the regulatory components Tic32 and Tic62, and several components of a putative stromal molecular motor responsible for pulling imported proteins into the organelle stroma (Gagat and Mackiewicz, 2014). Consequently, a simplified TIC could translocate proteins across the inner envelope. However, an outer envelope pore like Toc75 or its bacterial homolog Omp85 could not be identified (Bodył *et al.*, 2010; Gagat and Mackiewicz, 2014). Due to the presumable lacking of a protein conducting machinery, it was suggested that vesicular transport via the secretory pathway might be involved in protein transport to the chromatophore intermembrane space (Bodył *et al.*, 2012;

Mackiewicz *et al.*, 2012b,a). The idea was supported by the finding that photosystem I subunit PsaE was detected in the Golgi (Nowack and Grossman, 2012). Notably, PsaE represents a short imported protein and accordingly lacks a crTP. However, a *P. chromatophora* crTP was able to target yellow fluorescent protein to plant plastids *in vivo*, indicating that common components or principles work in plastid- and chromatophore protein import (Singer *et al.*, 2017).

In summary, the combination of the experimental and *in silico* approaches extended the list of *P. chromatophora* chromatophore import candidates to altogether 432 proteins, comprising approx. one third of the known chromatophore proteome. Thus, in addition to endosymbiotic genome reduction also massive protein import is accomplished in *P. chromatophora* in support to calling the chromatophore indeed an organelle and not just an endosymbiont.

1.2.3 Expected Metabolic Fluxes across the Chromatophore Envelope

Protein import allows the chromatophore to produce several essential compounds, even though some or all of the enzymes involved are no longer chromatophore-encoded. However, still extensive metabolite transport is demanded for the following reasons:

(i) Protein import likely does not (yet?) facilitate biosynthesis of all essential metabolites, as the set of imported proteins identified via proteomics and crTP-predictions still leaves some incomplete or entirely missing pathways. Examples are serine- and alanine biosynthesis (entirely missing) or threonine-, cysteine-, and methionine biosynthesis (partially missing) (Fig. 1.2-2; Singer *et al.,* 2017). In those cases it seems likely, that end products or intermediates are produced outside of the chromatophore and that they are subsequently transported across the chromatophore envelope membranes.

(ii) Some metabolites can be synthesized only in the chromatophore, but not in the host cell. For example, only the chromatophore possesses the complete set of enzymes required for biosynthesis of branched chain amino acids (leucine, isoleucine, valine; Fig. 1.2-2) and lysine (Singer *et al.*, 2017). Thus, these metabolites would have to be exported from the chromatophore. Also citrulline is likely exported and used for arginine biosynthesis in mitochondria. Subsequently, arginine could be shuttled back into chromatophores (Fig. 1.2-2). Also export of fatty acids should be taken into assumption, as indicated recently by ^{14}C bicarbonate labeling experiments (Sato *et al.*, 2020).

(iii) The chromatophore represents *P. chromatophora´s* only source of energy in form of reduced carbon. The amoeboid cell completely quit heterotrophy and is now inevitably depending on photosynthesis and carbon fixation in the chromatophore (Fig. 1.2-2). Accordingly, efficient export of photosynthetically fixed carbon has been *in silico* predicted (Valadez-Cano *et al.*, 2017) and

experimentally verified in ^{14}C bicarbonate labeling experiments (Kies and Kremer, 1979; Sato *et al.*, 2020). However, the identity of the transported compound(s) is unknown but export of hexoses, sugar-phosphates, or triose-phosphates is likely. On the other hand, storage polysaccharides have been detected by Lugol´s iodine staining in the host cytosol only (E. Nowack, personal observation) indicating that the chromatophore has to be provided with carbon compounds resulting from polysaccharide catabolism at night (Fig. 1.2-2).

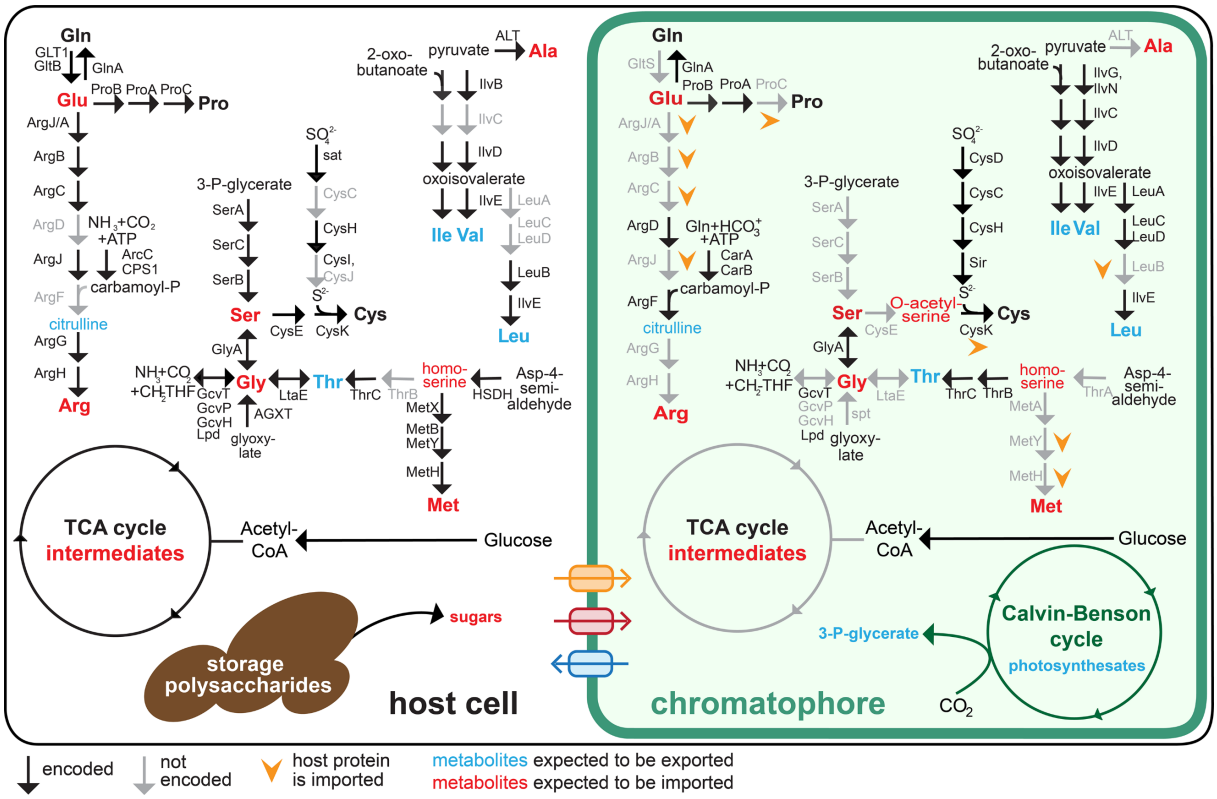


Figure 1.2-2 Metabolic complementarity of *P. chromatophora* host cell and chromatophore calls for metabolite transport. Synthesis pathways for selected amino acids are shown in detail. Glycolysis, TCA cycle, and Calvin-Benson cycle are outlined. Pale lettering/arrows indicate that the gene is missing in the chromatophore genome (Nowack *et al.*, 2008) or in nuclear transcriptome data (Nowack *et al.*, 2016). Some functions that are no longer encoded in the chromatophore genome are compensated by import of nuclear encoded proteins (Singer *et al.*, 2017). Many pathways are chimeric, i.e. they comprise chromatophore-encoded and nuclear encoded proteins. Numerous metabolites are expected to be shuttled across the chromatophore envelope according to presence of a pathway for their synthesis in either compartment. The respective transporters/mechanisms that facilitate metabolite exchange (pink and blue) or protein import (orange) are currently unknown.

However, the repertoire of chromatophore-encoded transporters has not been studied in detail, but seems very limited and comprises mostly ion transporters (Nowack *et al.*, 2008). Especially importers for nitrogen compounds, transporters for amino acids, and a dedicated carbon exporter are missing. Though, assuming that the evolution of the chromatophore recapitulates plastid evolution to some extent, the presence of nuclear encoded metabolite transporters in chromatophore envelope

membranes is highly expected. The metabolic integration of plastids (and mitochondria) requires extensive metabolite transport. As mentioned before (see 1.1.2), a plethora of host-derived metabolite transporters is inserted into the plastid outer and inner envelopes (Breuers *et al.*, 2011; Weber and Linka, 2011). Among them are well-studied examples like TPT that constitutes the export route for reduced carbon in plastids of plants and algae (Heldt and Rapley, 1970; Price *et al.*, 2012), but also an estimated 100 more transporters for which substrate(s) and function are only predicted (Mehrshahi *et al.*, 2013; Weber *et al.*, 2005). However, such nuclear encoded transporters have not been identified so far in *P. chromatophora´s* chromatophore, neither by mass spectrometric analysis of the chromatophore lysate proteome, nor was a crTP *in silico* predicted on a transporter (Singer *et al.*, 2017). However, typical transporters represent highly hydrophobic integral membrane proteins of low abundance, rendering their detection in non-targeted proteomic analyses challenging (Antelo-Varela *et al.*, 2019). Also, it can be speculated that the import route for such proteins might be crTP-independent, therefore impairing their *in silico* prediction. Such alternative targeting mechanisms are well known from mitochondria, where the insertion of most solute carriers into the inner envelope membrane is independent of an N-terminal mTP (Ferramosca and Zara, 2013). Thus, nuclear encoded transporters residing in the chromatophore might have easily escaped detection in previous approaches.

1.2.4 Expression of Chromatophore-encoded Proteins under Nuclear Control

Although many nuclear encoded proteins appear to be imported into chromatophores, the larger fraction of the chromatophore proteome is still chromatophore-encoded. Thus, establishment of nuclear control over organellar gene expression should be considered an important issue during organellogenesis supporting (i) adjustment of the chromatophore to the physiological state of the host cell, and (ii) assembly of chromatophore-localized protein complexes composed of subunits encoded in either the organellar or nuclear genome in stoichiometric amounts (Hammani *et al.*, 2014; Woodson and Chory, 2008). Presence of functional protein complexes of dual genetic origin has already been shown in *P. chromatophora* (photosystem I; Nowack and Grossman, 2012). The difference in copy numbers between chromatophore and nuclear genome (~100 vs. one or two copies; Nowack *et al.,* 2016) calls for coordination of gene expression between nucleus and chromatophore. Accordingly, a large number of proteins annotated as transcription factors have already been identified among chromatophore-targeted proteins (Singer *et al.*, 2017). However, nuclear encoded factors involved in post-transcriptional regulation comparable to the huge arsenal of helical repeat proteins that is targeted to plastids and mitochondria (see 1.1.3) have not been identified so far. Thus, the mechanisms by which gene expression coordination is achieved in *P. chromatophora* remain to be discovered.

1.3 Aims of this Thesis

Although hundreds of nuclear encoded proteins appear to be imported into chromatophores, little is known about the import route and the factors involved. Absence of most TOC / TIC homologs and the unique structure of the crTP suggest that protein import into the nascent chromatophore differs fundamentally from the mechanisms operating in ancient eukaryotic organelles. This study aims at contributing to a better understanding of the mechanism underlying protein import into chromatophores by investigating the fate of the crTP upon import.

Another focus of this study lies on metabolite shuttling between the *P. chromatophora* host cell and the chromatophore. Metabolic capacities of the *P. chromatophora* chromatophore and host cell are complementary resulting in the need for extensive exchange of metabolites such as sugars, amino acids, and cofactors across the two envelope membranes that surround the chromatophore. However, the nature of the transporters underlying the deduced solute transport processes across the chromatophore envelope is unknown. In plants and algae, transport across the plastid inner membrane is mediated by a large set of host-derived transporters that are highly specific for their substrates while transport across the plastid outer membrane is enabled largely by (semi-)selective pores formed by nuclear encoded β-barrel proteins. This study aims at testing the hypothesis that nuclear encoded transporters were recruited to establish metabolic connectivity between organelle and host cell, also in *P. chromatophora*.

Besides the establishment of metabolic connectivity, the evolution of nuclear control over organellar gene expression can be regarded another cornerstone in organellogenesis. Previously, a large number of proteins annotated as transcription factors was identified among chromatophore-targeted proteins. However, in plants and algae, expanded families of helical repeat proteins are targeted to organelles in order to modulate expression of the remaining organelle-encoded genes. Actually, helical repeat proteins appear to represent ubiquitous nuclear factors involved in regulation of organellar gene expression. Thus, the identification of organelle-targeted expression regulators in *P. chromatophora* represents another aim of this study.

Different experimental approaches are employed in this study including N-terminomics, microbiological and biochemical analyses of transporters, and mass spectrometric analysis of the chromatophore insoluble proteome. The results obtained may add valuable knowledge to determine common rules and reveal fascinating differences regarding hallmarks of organellogenesis, as they provide insights in the shuttling of proteins as well as metabolites across membranes of the symbiotic interface and the establishment of host control over organellar gene expression.

2 Results

2.1 Insights in Processing of Protein N-termini in the Chromatophore

Although hundreds of nuclear encoded proteins are apparently targeted to and imported into the chromatophore, the import route, mechanism, and involved factors remain elusive. For short imported proteins (< 95 aa) vesicular transport via the secretory pathway might be involved in transport to the intermembrane space as photosystem I subunit PsaE was detected in the Golgi immunologically (Bodył *et al.*, 2012; Mackiewicz *et al.*, 2012b,a; Nowack and Grossman, 2012). However, such direct evidence for the subcellular localization of further, especially long imported proteins is still lacking. The presence of the conserved putative chromatophore Transit Peptide (crTP) exclusively on long proteins suggests that even different routes/mechanisms might be exploited for different classes of proteins. Assuming that the crTP mediates protein import it seems conceivable that specialized regions within the 200 aa conserved sequence are involved in individual import steps, e.g. in targeting to the organelle, crossing the outer and inner envelope membranes and possibly also the thylakoid membrane, respectively.

In plastids, protein import into the stroma requires a Chloroplast Transit Peptide (cTP) which is cleaved off by Stromal Processing Peptidase (SPP) resulting in a mature protein that possesses a new N-terminus. In proteins that are supposed to localize to thylakoids the cTP is followed by a bacterial signal peptide that enables further transport into the thylakoid lumen via the Sec- or Tat- pathways. This bipartite architecture requires another cleavage by Thylakoidal Processing Peptidase (TPP) finally generating mature N-termini of luminal proteins. Additionally, many plastid proteins require N-terminal modifications, e.g. N-terminal methionine excision by methionine aminopeptidases or N-acetylation by N-acetyltransferases.

However, nothing is known so far about (possibly import-related) post-translational modifications in *P. chromatophora*. To learn more about the post-translational fate of the crTP and to possibly draw conclusions on its function(s) in protein import the HUNTER (High-efficiency Undecanal-based N-Termini EnRichment) method (Weng *et al.*, 2019) was applied here to explore the chromatophore N-terminome.

2.1.1 Application of the HUNTER Method for Determination of Protein N-termini in the Chromatophore

The HUNTER method is based on *in vitro* whole proteome dimethylation of protein N-termini, followed by tryptic digest and undecanal labeling of non-modified tryptic (internal) peptides. Internal peptides are then depleted from the sample due to the highly hydrophobic undecanal tag. This allows for the

identification of mainly N-terminal peptides in subsequent mass spectrometric analysis and thus enables mapping of protein N-termini. Naturally occurring acetylated N-terminal peptides are also recovered by the method and are even particularly valuable, as N-terminal acetylation happens exclusively *in vivo*. In contrast, owed to the applied method, dimethylated peptides can occasionally result from protein degradation during sample preparation. Additionally, tryptic peptides carrying N-terminal glutamates or glutamines might cyclize to pyro-glutamate spontaneously and thus escape undecanal labeling and depletion. Therefore pyro-Glu peptides are considered artifacts in this study and not used for the determination of protein N-termini. If several redundant N-terminal peptides have been identified that differ only by their N-terminal modification (i.e. acetylated and dimethylated versions of a peptide) or only by their C-termini (i.e. by differential trypsin cleavage site skipping), these peptides are counted towards one N-terminus (referred to as “unique N-terminus”). Multiple unique N-termini might be identified for a protein, representing either different proteoforms or break down products of that protein. However, for some proteins a principal N-terminus can be identified due to its much higher intensity and/or much higher number of spectral counts. Chromatophore lysate (triplicates of 200-250 µg protein each in HEPES-buffered 6M Guanidine-HCl pH 7.5) was analyzed as described in 4.2.7.

As depicted in figure 2.1-1A, 335 acetylated or dimethylated N-terminal peptides corresponding to chromatophore-encoded proteins and 121 peptides corresponding to nuclear encoded proteins could be identified in chromatophore lysates. These peptides enabled determination of 367 unique protein N-termini, 105 of which represent N-termini of nuclear encoded proteins. Most of these N-termini have been mapped exclusively by one or several dimethylated peptide(s) (> 70 % CE and > 50 % NE), while a minor fraction was mapped by acetylated peptide(s) exclusively. Only for some N-termini both versions of peptides were present in the samples. Together the N-termini determined in this study correspond to 129 chromatophore-encoded and 75 nuclear encoded proteins (Tab. S1). Of the 53 nuclear encoded proteins for which 5´-full-length sequence information is available (5´-FL-proteins), 30 possess a crTP, 7 are long but lack a crTP, 14 are short, and for 2 the length cannot be determined (Fig. 2.1-1B). The number of unique N-termini exceeds the number of corresponding proteins as for 45 chromatophore-encoded and 18 nuclear encoded proteins multiple possible N-termini were identified.

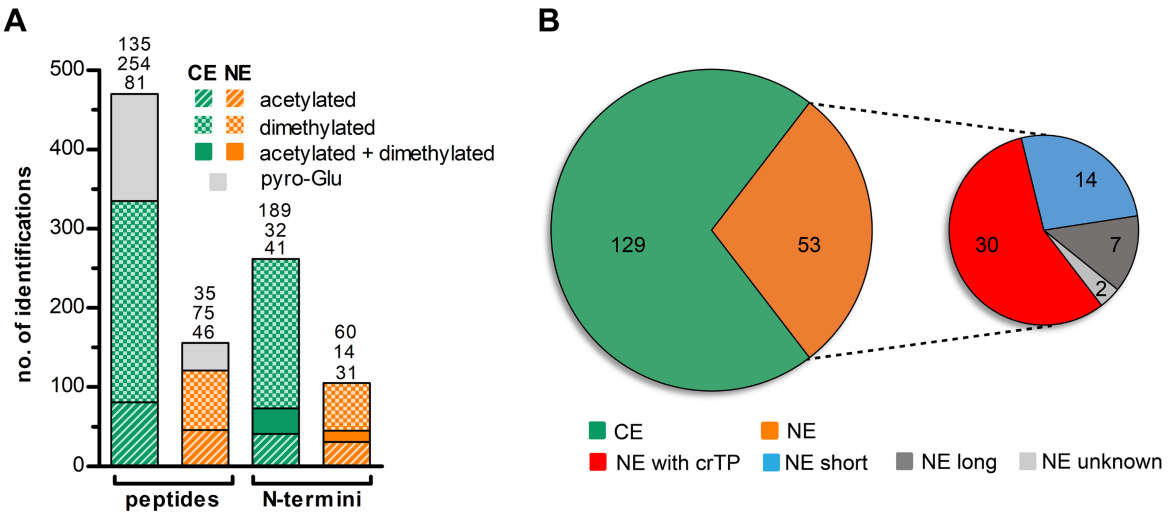


Figure 2.1-1 N-termini identified in the chromatophore. A Total number of validated N-terminal peptides and number of corresponding unique protein N-termini. N-terminal peptides are either N-terminally acetylated or dimethylated. Tryptic peptides carrying N-terminal pyro-glutamate might represent artifacts and are not considered for the determination of protein N-termini. A color code indicates how many N-termini were determined exclusively by acetylated, exclusively by dimethylated, or by both kinds of peptides for chromatophore-encoded (CE) and nuclear encoded (NE) proteins. **B** Numbers and classification of proteins for which N-termini could be identified. Only proteins for which full-length 5´-sequence information is available are represented. CE, chromatophore-encoded; NE, nuclear encoded; NE short, proteins shorter than 95 aa; NE long, proteins longer than 125 aa but lacking a crTP; NE unknown, length cannot be determined due to lacking 3´-sequence information.

2.1.2 N-terminal Processing of Chromatophore-encoded Proteins

For most of the chromatophore-encoded proteins (83 of 129) canonical N-termini have been determined, i.e. the protein is either not processed (22 proteins) or only the very first amino acid is removed (55 proteins) or both proteoforms were identified for a protein (6 proteins) (Fig. 2.1-2A). For 28 of these proteins however, peptide(s) matching to the canonical N-terminus plus further peptide(s) corresponding to at least one non-canonical proteoform were identified and for 46 proteins exclusively non-canonical N-termini have been identified. The non-canonical N-termini usually map to the N-terminal 20 or 100 aa of protein models (Fig. 2.1-2A). A correlation between the identification of multiple N-termini and high protein abundance (as deduced from an independent shotgun MS analysis on whole cells and chromatophores [see section 2.2.3; Oberleitner *et al.*, 2020]) can be observed sometimes, e.g. 12 N-termini have been identified for phycobilisome linker polypeptide or 9 for RubisCO large subunit. However, this is not always the case as besides protein abundance also other factors (protein lifetime, availability of trypsin cleavage sites, physicochemical properties of generated peptides) affect the number of N-terminal peptides that can be detected by MS. Notably, often several

N-termini identified for a protein are located within a range of only 10 aa, e.g. 8 of the 12 unique N-termini identified for phycobilisome linker polypeptide are located between position 19 to 28 (Fig. A6.2-1). The phenomenon has been observed earlier and is regarded a result of either sloppy cleavage specificity and/or additional processing steps by aminopeptidases (Rowland *et al.*, 2015). The average number of unique N-termini per protein identified in this study is two for chromatophore-encoded proteins.

However, some proteoforms might represent biologically relevant processing products, while others may represent protein degradation products that have emerged *in vivo* or during sample preparation. Non-canonical N-termini generally show lower intensities (Fig. A6.2-2) and for some proteins, degradation products might be distinguished from a true canonical N-terminus, as the intensity of peptides mapping to the canonical N-terminus is much higher than peptide intensities of putative degradation products (see examples in Fig. A6.2-1). Contradicting that, the occurrence of N-acetylations on such N-termini indicates that they rather represent products of proteolytic processing, as degradation products are usually not acetylated. However, while acetylated N-terminal peptides were identified for approx. 40-50 % of the canonical proteoforms and of the non-canonical proteoforms of bin 03-20, the fraction of acetylated non-canonical proteoforms decreases as the distance of the identified N-termini from the theoretical start of the protein models increases (Fig. 2.1-2A). This might indicate that the fraction of N-termini resulting from protein degradation during sample preparation is higher for N-termini located at positions > 20 aa from the start methionine. Notably, for eight of the proteins for which only non-canonical N-termini could be determined, canonical N-termini might be present as a matter of fact, but these proteins do not feature trypsin-cleavable arginine residues in positions 8 to 41 from the start methionine, abolishing the generation of detectable canonical N-terminal peptides by trypsin cleavage (peptides of 8 to 40 aa can be MS-identified according to the applied settings, see 4.2.7.3). Further N-termini that presumably result from cleavage at methionine also might be misinterpreted as non-canonical due to mis-annotation of the proteins correct translation initiation site and actually represent canonical N-termini (grey arrows in Fig. A6.2-2B). Finally, for four proteins of bin 21-100 the identified processing site matches a predicted signal peptide cleavage site (orange arrows in Fig. A6.2-2B).

Generally, the distribution of amino acids occurring in position P1 (denoting the residue preceding the identified N-terminus) and position P1´ (denoting the N-terminal residue of the identified N-terminus) is neither random nor does it reflect the overall frequencies of amino acids in the predicted chromatophore-encoded proteome (Fig. 2.1-2B and A6.2-2). The most common amino acids in P1 positions are methionine, arginine, asparagine, alanine, and serine (Fig. 2.1-2B). Methionine in the P1 position is usually related to iMet excision in canonical N-termini (Fig. 2.1-2B and A6.2-2). In non-

canonical cleavage sites, the most common amino acids in the P1 position are arginine (23 % of sites), asparagine, alanine, and serine (together 31 % of sites; Fig. 2.1-2B and A6.2-2).

Methionine, serine, threonine, and alanine represent the most common amino acids in P1´positions (Fig. 2.1-2B) and the respective N-terminal peptides are also highly abundant, pointing to a stabilizing effect of these N-terminal amino acids (Fig. A6.2-2). Consequently, proteins usually show methionine (31 %), serine, alanine, threonine or valine (together 49 %) residues at their canonical N-termini, while serine, threonine, alanine and glycine are frequently present at P1´positions in non-canonical N-termini (Fig. 2.1-2B).

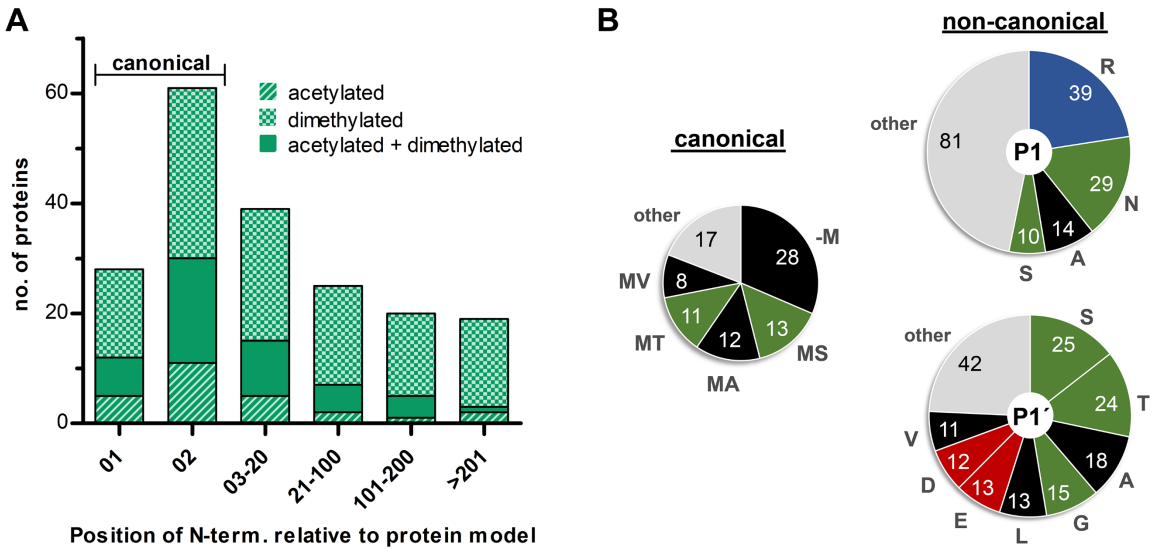


Figure 2.1-2 Position and amino acid composition of processing sites in chromatophore-encoded proteins. **A** Position of identified N-termini with respect to protein models. The number of proteins is depicted for which one or several N-terminal peptide(s) matching the respective protein model at the indicated positional ranges have been identified. A color code indicates how many proteins have N-termini identified exclusively by acetylated, exclusively by dimethylated, or by both kinds of peptides. **B** Amino acid frequencies at P1 and P1´positions of unique N-termini. Total numbers of the most frequent amino acids present at P1 (preceding the identified N-terminus) and P1´ (N-terminal residue of the identified N-terminus) positions of canonical or non-canonical N-termini are shown. Black, hydrophobic; green, polar amino acids and glycine; red, negatively charged; blue, positively charged.

2.1.3 N-terminal Processing of Imported Proteins

2.1.3.1 Processing of the crTP

N-terminal peptides could be identified for 75 nuclear encoded proteins (53 FL-proteins; Fig. 2.1-1, Tab. S1). Of these, 35 (30 FL-proteins) possess a crTP and represent import candidates that were either MS-identified earlier (24 proteins) or were predicted import candidates (11 proteins) for which presence in the chromatophore could be verified in this experiment for the first time (Tab. S1). No canonical N-termini have been identified for crTP-proteins. Instead, 24 of the FL-crTPs are processed in a positional range from 36-70 aa downstream of the start methionine of the protein models (Fig. 2.1-3A and D, denoted site 1). Given that an approximate range for the position of the processing site can be deduced from alignments also for crTP-proteins that lack 5´-full-length sequence information the number even increases to 29 proteins processed in the 36-70 range. As observed already for chromatophore-encoded proteins, multiple cleavages often occur in close proximity, i.e. the generated proteoforms only differ by one to five amino acids (Fig. 2.1-3D). The average number of unique N-termini per protein is 1.8 for crTP-proteins. A correlation between a higher number of unique N-termini per protein and high intensity of the corresponding N-terminal peptides was not observed. Also the proteins for which multiple N-termini have been identified here do not seem to be more abundant in the cell, as deduced from transcriptomic data (looking at Reads Per Kilobase Million in Nowack *et al.,* 2016) or an independent shotgun MS analysis on whole cells and chromatophores (see section 2.2.3; Oberleitner *et al.*, 2020).

Although processing site 1 is located between two conserved regions in a non-alignable region with no primary sequence conservation between individual crTPs, a preference for certain amino acids around the site becomes apparent (Fig. 2.1-3B). The region is overall rich in serine and glycine. Upstream of the processing site (pos -1 to -10) positive charges are more prevalent and phenylalanine is present at the P1 position in 25 % of all cases. Downstream of the site (pos 1 to 10) negative charges are frequent and this region is also comparably rich in serine, glycine, alanine and proline. Serine is present in the P1´position in 40 % of all N-termini identified, followed by alanine (19 %), phenylalanine (13 %), and isoleucine (8 %). Notably, isoleucine is always acetylated, while serine and alanine are sometimes acetylated (30-40 % of peptides), and phenylalanine is never acetylated. When multiple processing sites have been identified in close proximity for a protein, the corresponding peptides usually differ in relative intensity (Fig. A6.2-3). This might either indicate preferential cleavage between certain residues by the responsible peptidase or the amino acid composition of the newly generated N-terminus influences protein stability. However, no obvious prevalence for a certain N-terminal amino acid, acetylation status or relative site position was observed among N-termini of high relative or absolute intensity.

When compared to cTP processing sites of proteins imported into plastids some unambiguous similarities become apparent, e.g. the distribution of charges upstream and downstream of the processing site, the prevalence of serine and alanine at the P1´ position, and an overall high frequency of serine (Fig. 2.1-3C). However, in contrast to the cTP processing site, glycine is overall more common around the crTP processing site and the occurrence of aromatic phenylalanine at the P1 and P1´positions clearly is a distinguishing feature.

However, not all proteins are processed at the same position. Two proteins have additional processing sites approx. 60 aa and 110 aa downstream of the crTP (range >211), respectively. Further six crTPs are not cleaved at all at the 36-70 range but elsewhere in the sequence, notably in regions of high sequence conservation (Fig. 2.1-3D). Three of these are processed at the same alignment position (range 176-210, site 2) which marks the very end of the conserved crTP, producing N-termini starting with aromatic amino acids tyrosine or phenylalanine.

Noteworthy, a short motif involved in protein sorting at the *trans* Golgi can be found in the very N-terminus of 29 of the 31 crTPs for which full-length sequence information is available. The YxxΦ sorting signal (where “Y” stands for tyrosine, “Φ” for a bulky, hydrophobic amino acid, and “x” for any amino acid) is usually located at the cytoplasmic tail of membrane spanning protein cargoes and is known to bind to AP1, a clathrin adaptor protein complex that couples cargo recruitment to coated vesicle budding (Park and Guo, 2014). Noteworthy, the two crTPs lacking the signal possess the sequence [DE]LxxPLL instead. The sequence is not identical but still similar to the motif [DE]xxxL[LI], another sorting signal bound by AP1 (Park and Guo, 2014).

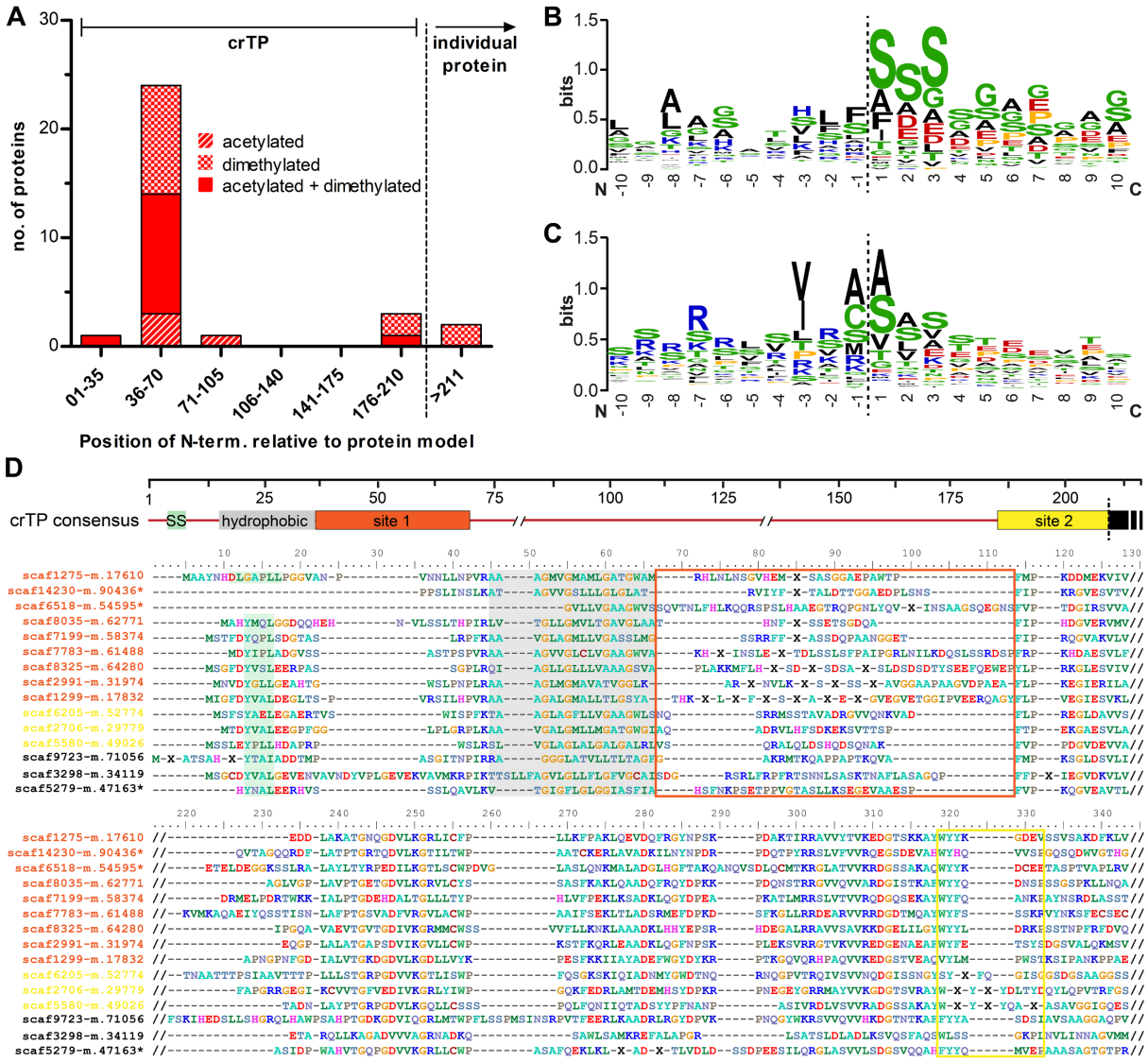


Figure 2.1-3 Position and amino acid composition of processing sites in the crTP of long imported proteins. **A** Position of identified N-termini with respect to protein models. The number of crTP-containing proteins is depicted for which N-terminal peptide(s) matching the respective protein model at the indicated positional ranges have been identified. A color code indicates how many proteins have N-termini identified exclusively by acetylated, exclusively by dimethylated, or by both kinds of peptides in the given positional range. Only proteins for which 5´-full-length sequence information is available have been considered. Positions 1-210 of long imported proteins contain the conserved regions typical for a crTP. **B** Amino acid frequencies around processing sites (dashed line) in the crTP. The sequence logo is based on 48 unique N-termini that map to the 36-70 positional range of 29 individual protein models. **C** Amino acid frequencies around processing sites of the cTP of *Arabidopsis thaliana*. The sequence logo is based on the best-ranked N-termini of 162 nuclear encoded stromal proteins according to Rowland *et al.,* 2015. Black, hydrophobic; green, polar amino acids and glycine; red, negatively charged; blue, positively charged; yellow, proline. **D** Alignment of processed crTPs. Processing sites are represented by X. Nine out of 29 crTPs are shown that are processed in a positional range of 36-70 aa relative to their protein models (site 1), three crTPs are processed at their very C-terminus (site 2), and further three crTPs are processed elsewhere. Alignment positions corresponding to the ranges highlighted in the crTP consensus sketch are marked by colored boxes. The hydrophobic region preceding site 1 is highlighted in grey. The putative sorting signals YxxΦ and [DE]LxxPLL are highlighted in green. Proteins lacking 5´-full-length sequence information are marked by asterisks.

2.1.3.2 Processing of Short Proteins

N-terminal peptides could be identified for 14 short proteins (< 95 aa), 7 of which were MS-identified as import candidates in independent experiments, while further 7 represent new identifications (Tab. S1). Four more proteins have not been classified as short proteins due to lacking evidence for full-length 5´-sequence information, but likely represent short proteins because processing occurs at the putative start methionine. Additionally, two of the proteins have short orthologs annotated as high light inducible protein or 40S ribosomal protein S28, respectively.

For most of the proteins only one N-terminal peptide has been identified and that usually refers to a canonical N-terminus. In most cases (15 out of 18 proteins) the initial methionine is removed and one protein remains entirely unprocessed. Interestingly, the only two short proteins for which non-canonical N-termini have been identified also represent the only proteins that possess a predicted TMH. Cleavage occurs among the N-terminal 10 aa between alanine and a negatively charged residue in both cases.

2.1.3.3 Other Nuclear Encoded Proteins

N-terminal peptides have been identified for seven long proteins that do not possess a crTP (Fig. 2.1-1B, Tab. S1). Interestingly, one of the long proteins annotated as gamma-glutamylcyclotransferase and possessing an N-terminal TMH was identified as an import candidate in an independent experiment (scaffold10322-m.73921, see section 2.2.3). A single acetylated peptide corresponding to the unprocessed protein N-terminus (Tab. S1) points to either import of the complete protein or attachment/insertion of the TMH to/into the chromatophore envelope.

Further seven proteins are likely long, as the full-length sequence is not known but the available sequence fragment is already far longer than short proteins. Twelve proteins cannot be classified, as only short fragments of sequence information are available. However, four of these proteins might be short, as mentioned above. Some of these orphan proteins likely represent contaminations resulting from residual host cell material on the isolated chromatophores, e.g. 40S or 60S ribosomal proteins or actin related proteins (Tab. S1). Others have been detected before exclusively in whole cell samples (Tab. S1). However, for most of the proteins it cannot be determined whether they represent import candidates or contaminating host proteins.

2.2 Transporters of the Chromatophore Envelope

Extensive metabolite shuttling is expected between the *P. chromatophora* host cell and the chromatophore (see 1.2.3), but only relatively few transporters seem to be encoded in the chromatophore genome. However, chromatophore-encoded transport capacities have not been analyzed in detail and the presence of nuclear encoded transporters in chromatophores has not been examined so far. Thus, in a first step the chromatophore genome was carefully screened for possible envelope transporters and compared to the permeome in cyanobacterial plasma membranes and envelopes of fully integrated plastids. Integrating these information, eight nuclear encoded candidate chromatophore envelope transporters were selected and checked for their substrate specificities in heterologous systems. Specific antibodies were generated to evaluate the subcellular localization of the most promising candidate. In a second approach, the chromatophore insoluble proteome was screened for the presence of nuclear encoded metabolite transporters.

2.2.1 Paucity of Chromatophore-encoded Solute Transporters

Chromatophore-encoded transporters appear to be rare. So far, only several secondary active transporters for inorganic ions, one ABC-type multidrug efflux system, and two ABC-type uptake systems for cobalt and inorganic phosphate, respectively, have been recognized (Nowack *et al.*, 2008).

In contrast, 89 putative cell envelope membrane transporters and multi-subunit transport systems could be identified in the scaffold-level genome assembly of the free-living α-cyanobacterium *Synechococcus* sp. WH5701 (hereafter *Synechococcus*; Fig. 2.2-1, Tab. S2) which represents a close relative of the chromatophore (Marin *et al.*, 2007). In the fully sequenced genome of the model cyanobacterium *Synechocystis* sp. PCC6803 (hereafter *Synechocystis*) even more than 100 putative envelope transporters could be identified (Fig. 2.2-1, Tab. S2; compare Paulsen *et al.*, 2000). Due to the photoautotrophic lifestyle of cyanobacteria, most of their plasma membrane transporters are involved in import of ions, metals and inorganic anions like bicarbonate, phosphate, sulfate, and nitrate. The cyanobacterial outer membrane apparently contains several selective porins, e.g. for the export of lipopolysaccharides or secretion of glycolipids, but primarily for the uptake of solutes (Huang *et al.*, 2004; Nicolaisen *et al.*, 2009; Qiu *et al.*, 2018). Interestingly, cyanobacteria are also capable of the uptake of amino acids (Quintero *et al.*, 2001) and some can even use sugars in periods of heterotrophic lifestyle (Rippka *et al.*, 1979; Stal and Moezelaar, 1997). However, the repertoire of such transporters in the plasma membrane of cyanobacteria is rather narrow, when compared to heterotrophic bacteria (Paulsen *et al.*, 2000).

In contrast, exchangers of sugar-phosphates but also metabolites like carboxylates or ATP/ADP are well characterized in the *Arabidopsis thaliana* plastid envelope (Facchinelli and Weber, 2011). The typical plastid inner membrane permeome comprises an estimated 100-150 transporters, the majority of them being of the single subunit secondary active or channel type (Mehrshahi *et al.*, 2013; Weber *et al.*, 2005; Fig. 2.2-1 and Tab. S2). 37 of these transporters have been confidently assigned functions and many of them transport metabolites (Karkar *et al.*, 2015). Correspondingly, several porins are known to permit passage of solutes across the chloroplast outer envelope (Breuers *et al.*, 2011; Goetze *et al.*, 2015; Harsman *et al.*, 2016; Wang *et al.*, 2013). Transporters are usually encoded in the nuclear genome and post-translationally inserted into the organellar envelopes, while plastids of plants and algae usually do not encode for transporters in their own genomes anymore.

To gain deeper understanding of the metabolite transport capabilities still encoded in the chromatophore genome the set of chromatophore-encoded putative transporters was carefully revaluated. Re-annotation of the data from 2008 (Nowack *et al.*, 2008) using BlastP (Altschul *et al.*, 1990) and application of the automatic transporter annotation tool TransAAP (Elbourne *et al.*, 2017) lead to the identification of further chromatophore-encoded transport systems or system components. Comparison to entries in the Transporter Classification Database (TCDB; Saier *et al.,* 2016) allowed for evaluation or revaluation of their putative function (Tab. 2.2-1).

Table 2.2-1 Transporters and transport systems identified in the chromatophore genome. Each transporter was assigned to a family according to the transporter classification system (Saier *et al.*, 2016). The putative substrates and subcellular localization were inferred from experimental data on cyanobacterial orthologs or from other well-studied members of the respective family, if available. IM, inner membrane; TM, thylakoid membranes.

# TCDB	Family	Transporter/ Transport system	Putative substrate	Putative localization	Locus tag	References
Channels						
1.A.23	The Small Conductance Mechanosensitive Ion Channel (MscS) Family	Exchange channel	Ions	IM	PCC0817	(Rasmussen, 2016)
1.A.26	The Mg^{2+} Transporter-E (MgtE) Family	Import channel MgtE	Mg^{2+}	IM	PCC0450	(Pohland and Schneider, 2019)
1.A.112	The Cyclin M Mg^{2+} Exporter (CNNM) Family	Export channel	Mg^{2+}?	IM	PCC0210 / PCC0504	(Giménez-Mascarell *et al.*, 2019)
Secondary active transporters						
2.A.4	The Cation Diffusion Facilitator (CDF) Family	Antiporter	Cd^{2+}/ Zn^{2+}/ Co^{2+} (Cu^{2+}/ Ni^{2+}) : H^+	IM	PCC0013	(Kolaj-Robin *et al.*, 2015)
2.A.19	The Ca^{2+} : Cation Antiporter (CaCA) Family	Antiporter	Ca^{2+}/Na^+/K^+ : H^+/Cations	IM	PCC0003	(Waditee *et al.*, 2004)
2.A.38	The K^+ Transporter (Trk) Family	Importer KtrABE	K^+	IM	PCC0715 PCC0665 PCC0666	(Matsuda *et al.*, 2004; Zulkifli *et al.*, 2010)
2.A.47	The Divalent Anion : Na^+ Symporter (DASS) Family	Symporter	Oxyanions (sulfate or phosphate) or carboxylates (di- or tricarboxylates) and Na^+	IM	PCC0664	(Gisin *et al.*, 2010; Mancusso *et al.*, 2012; Pootakham *et al.*, 2010; Rhie *et al.*, 2014; Weber and Flügge, 2002)
2.A.53	The Sulfate Permease (SulP) Family	Symporter BicA	Bicarbonate and Na^+	IM	PCC0004	(Price *et al.*, 2004)
2.A.66	The Multidrug / Oligosaccharidyl-lipid / Polysaccharide (MOP) Flippase Superfamily	Flippase MurJ	Peptidoglycan-lipid II	IM	PCC0513	(Kumar *et al.*, 2019)
2.A.37	The Monovalent Cation : Proton Antiporter-2 (CPA2) Family	Antiporter NhaS3	Na^+ : H^+	TM	PCC0474	(Tsunekawa *et al.*, 2009)
2.A.7.3	The 10 TMS Drug / Metabolite Exporter (DME) Family	Exporter	Amino acids / metabolites?	unclear	PCC0734	(Jack *et al.*, 2001; Rouanet and Nasser, 2001; Zakataeva *et al.*, 2006)
2.A.88	Vitamin Uptake Transporter / Energy-coupling factor (VUT or ECF) Family	Importer BioY	Biotin	unclear	PCC0253	(Eitinger *et al.*, 2011; Finkenwirth *et al.*, 2013)
Primary active transporters						
3.A.1.7	The Phosphate Uptake Transporter (PhoT) Family	ABC2-type importer PstSACB	Phosphate	IM	PCC0161 PCC0162 PCC0163 PCC0395	(Reyes-Prieto *et al.*, 2010; Valadez-Cano *et al.*, 2017)

3.A.1. 25/26	The Biotin Uptake Transporter (BioMNY) Family or Putative Thiamine Uptake Transporter (ThiW) Family	Putative ECF-type importer	Co^{2+}/vitamins?	IM	PCC0300 PCC0049 PCC0834 PCC0705	(Eitinger *et al.*, 2011; Rodionov *et al.*, 2009)
3.A.1. 106/135	Lipid Exporter (LipidE) Family or Drug Exporter-4 (DrugE4) Family	Putative heterodimeric ABC1-type efflux transporter	Glycolate?/ glycolipid?/ multidrug?	IM	PCC0921 PCC0227	(Braakman *et al.*, 2017)
3.A.1.203	Peroxysomal Fatty Acyl CoA Transporter (P-FAT) Family	Putative homodimeric ABCD-type importer	Fatty acyl CoA?/very long chain fatty acids?/ antimicrobial peptides?	IM	PCC0669	(Guefrachi *et al.*, 2015; Linka and Esser, 2012)
3.A.1.105	Drug Exporter-1 (DrugE1) Family	Heterodimeric ABC2-type efflux transporter CmrAB	Multidrug	TM	PCC0749 PCC0750	(Menéndez *et al.*, 2007)
3.A.3	The P-type ATPase (P-ATPase) Superfamily	P-type ATPase CtaA	Cu^{2+}	unclear	PCC0303	(Huertas *et al.*, 2014; Raimunda *et al.*, 2011; Seigneurin-Berny *et al.*, 2006)
Transporters of unknown biochemical mechanism						
9.A.4	The YggT or Fanciful K^+ Uptake-B (FkuB; YggT) Family	Putative uptake transporter YggT	K^+?	IM	PCC0175 / PCC0672	(Ito *et al.*, 2009; Nakamura *et al.*, 1996)
9.A.8	The Ferrous Iron Uptake (FeoB) Family	Uptake transporter FeoB	Fe^{2+}	IM	PCC0239	(Sestok *et al.*, 2018)
Putative transporters						
9.B.67	O-antigen Polymerase (OAP) Family	Protein involved in bicarbonate uptake IctB	Bicarbonate?	IM	PCC0529	(Bonfil *et al.*, 1998; Shibata *et al.*, 2002)
9.B.73	Chloroplast Envelope / Cyanobacterial Membrane Protein (CemA) Family	Extrusion protein PxcA	H^+?	IM	PCC0044	(Katoh *et al.*, 1996; Rolland *et al.*, 1997; Sasaki *et al.*, 1993)
9.B.14	The Heme Handling Protein (HHP) Family	Putative exporter	Heme	TM	PCC0266	(Sutherland *et al.*, 2018)

Several putative ion transporters (PCC0210, PCC0504, PCC0175, PCC0672), a biotin uptake transporter (BioY, PCC0253), a second putatively heterodimeric multidrug efflux ABC-transporter (PCC0227, PCC0921), and a putatively homodimeric ABC-type importer for either fatty acids or antimicrobial peptides (PCC0669) represent new identifications in the chromatophore genome. Altogether, only 25 transporters were identified of which 19 putatively localize to the inner envelope as judged from the localization of orthologs from other species (Tab. 2.2-1). Substrates of most of these transporters are – according to annotation – restricted to inorganic ions (e.g. Na^+, K^+, Fe^{2+}, Mg^{2+}, PO_4^{2-}, HCO_3^-). Orthologs for ion- and metal transporters, the Na^+-dependent bicarbonate uptake transporter BicA, or the phosphate ABC-transporter PstSACB are chromatophore-encoded. However, orthologs of well-known cyanobacterial uptake systems for nitrogen and sulfur compounds, e.g. nitrate (Omata *et al.*, 1993), ammonium (Montesinos *et al.*, 1998), urea (Valladares *et al.*, 2002), or sulfate (Laudenbach and Grossman, 1991) are missing on the chromatophore genome (Fig. 2.2-1). Only one DME-family (10 TMS Drug/Metabolite Exporter) transporter (PCC0734) could potentially be involved in metabolite export and one DASS-family (Divalent Anion : Na^+ Symporter) transporter (PCC0664) could facilitate import of either di-/tricarboxylates or sulfate via a Na^+-symport mechanism. However, due to the multitude of possible substrates for members of both of these families (Jack *et al.*, 2001; Markovich, 2012) actual substrate specificities of these transporters cannot be predicted based on sequence similarity only. Altogether, in comparison to the permeome found in free-living cyanobacteria, deduced gene loss from the chromatophore genome likely concerns in the order of 70 solute transporters (Fig. 2.2-1). In line with the chromatophore´s poor set of inner envelope transporters, chromatophore-encoded β-barrel outer envelope pores could not be identified at all.

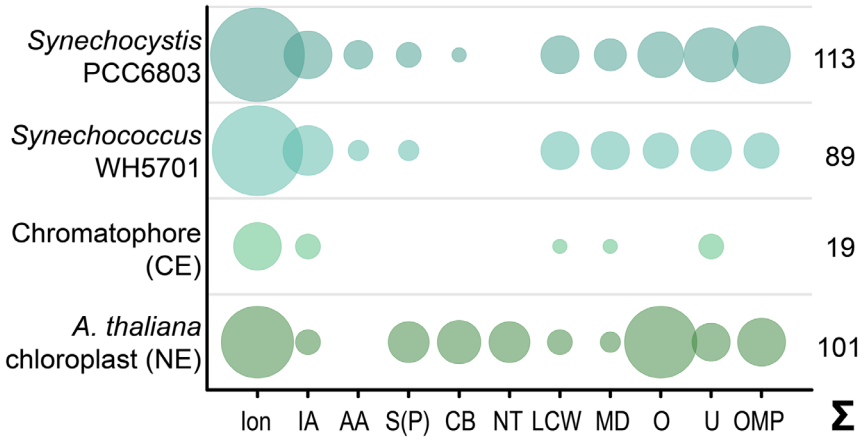


Figure 2.2-1 Predicted solute transport capacities of the chromatophore, its close free-living relative *Synechococcus* sp. WH5701, the model cyanobacterium *Synechocystis* sp. PCC6803, and the fully integrated plastid of *Arabidopsis thaliana*. Only transport systems for which experimental evidence suggests localization to the plasma membrane or the organellar envelope are shown. CE, chromatophore-encoded; NE, nuclear encoded; Ion, ion/metals; IA, inorganic anions (phosphate, sulfate, nitrate, bicarbonate); AA, amino acids; S(P), sugars (hexoses, oligosaccharides) or sugar-phosphates; CB, mono-/di-/tricarboxylates; NT, nucleotides; LCW, lipids and lipopolysaccharides; MD, multidrug; O, other; U, unknown; OMP, outer membrane pores; Σ, total predicted transporters.

As described earlier, the plastid envelope permeome severely differs from the ancestral cyanobacterial permeome, reflecting its functional specialization on metabolite exchange with a host cell. As depicted, also in *P. chromatophora* the highly expected exchange of metabolites (especially amino acids and triose- or hexose-phosphates) between chromatophore and amoebal host cell most likely cannot be accomplished by organelle-encoded transporters. Thus, nuclear encoded transporters might in analogy to the situation in plastids also function in the chromatophore.

2.2.2 A Bottom-Up Approach for the Identification of Nuclear Encoded Chromatophore Envelope Transporters: Characterization of Candidate Transporters

The deduced demand for extensive metabolite exchange between the chromatophore and its host cell in combination with the apparently insufficient transport capabilities encoded in the chromatophore genome led to the hypothesis that nuclear encoded transporters are inserted into the chromatophore envelope. However, no such transporters have been identified so far. This study aimed at identifying nuclear encoded, chromatophore-targeted metabolite transporters. Since *P. chromatophora* cannot be genetically manipulated so far, the subcellular localization of a transporter cannot be determined easily, e.g. by the expression of recombinant, tagged versions of that transporter. Therefore, in a bottom-up approach, candidate transporters were expressed heterologously and tested for their

ability to transport substrates for which shuttling across the chromatophore envelope is expected. In a second step, localization of candidate transporters to the chromatophore envelope has to be verified *in vivo*, making use of specific antibodies raised against the candidate proteins.

2.2.2.1 Candidate Nuclear Encoded Chromatophore Envelope Transporters

Generation of specific antibodies for a transporter to be detected immunohistochemically is a time consuming and expensive process. Since a plethora of transporters are encoded in the nuclear genome, while only very few of them can be tested for their subcellular localization, the following decision parameters were considered: (i) The transporter should comprise only one single subunit. Multi-component systems are less likely to be targeted to an organelle, as the vast majority of plastid envelope transporters is represented by single subunit channels, permeases or secondary active transporters (Tab. S2). (ii) The transporter should transport substrates that are most likely imported to or exported from the chromatophore. Amino acids, sugars and sugar-phosphates are highly expected to be shuttled across the chromatophore envelope (see Fig. 1.2-2). However, prediction of the substrates transported by a transporter is challenging based on primary sequence information only and should therefore be verified experimentally. Therefore, the eight transporters listed in table 2.2-2 were chosen for characterization of their substrate specificities in heterologous systems.

Table 2.2-2 Nuclear encoded metabolite transporters that were selected as candidate chromatophore envelope transporters. The family according to the transporter classification system, the putative mode of transport, and the predicted number of transmembrane helices (TMHs) are indicated. Putative substrates were inferred from experimental data on well-studied members of the respective family. BGATr was very likely acquired via HGT according to Nowack *et al.*, 2016, while all other transporters are likely host-derived.

Given name	Protein model	Transporter classification	Mode of transport	TMHs	Putative substrates	Origin	References
BGATr	scaf6082-m.52069	2.A.25.1 The Alanine or Glycine : Cation Symporter (AGCS) Family	Na^+-coupled secondary active symport	11	Alanine or glycine and Na^+ or H^+	γ-proteobacterial	(Bualuang *et al.*, 2015; Ma *et al.*, 2019; MacLeod and MacLeod, 1986; Moore and Leigh, 2005)
NAATr1	scaf6842-m.56439	2.A.18.6 The Amino Acid/Auxin Permease (AAAP) Family	Na^+-coupled secondary active symport coupled with H^+ antiport	10	Glycine, alanine, methionine, cysteine, serine, asparagine, glutamine, proline, threonine, histidine and Na^+	Eukaryotic	(Hatanaka *et al.*, 2000; Mackenzie *et al.*, 2003; Nakanishi *et al.*, 2001)
NAATr2	scaf5339-m.47536*	2.A.18.9 The Amino Acid/Auxin Permease (AAAP) Family	Na^+-coupled secondary active symport	11	Glutamine, arginine, asparagine phenylalanine, leucine, isoleucine, tryptophan, methionine, tyrosine and Na^+	Eukaryotic	(Rebsamen *et al.*, 2015; Wyant, 2017)
STr1	See appendix (A6.3-1 STr1)	2.A.123.1 The Sweet; PQ-loop; Saliva; MtN3 (Sweet) Family	Bidirectional permease	7	Glucose, fructose, sucrose	Eukaryotic	(Eom *et al.*, 2015; Klemens *et al.*, 2013)
STr2	scaf3336-m.34386	2.A.1.1 The Major Facilitator Superfamily (MFS)	H^+-coupled secondary active symport	12	Hexoses, pentoses, tetroses, sugar-alcohols	Eukaryotic	(Reinders *et al.*, 2005; Weber *et al.*, 2000)
STr3	scaf3197-m.33427*	2.A.1.1 The Major Facilitator Superfamily (MFS)	Bidirectional permease	12	Glucose, galactose, fructose, dehydroacetic acid, glucosamine	Eukaryotic	(Long and Cheeseman, 2015)
PTr1	scaf6509-m.54544	2.A.7.9 The Triose-phosphate Transporter (TPT) Family	P_i-coupled secondary active antiport	9	Phosphorylated C_6, C_5-, and C_3-compounds (Glucose-6-P, PEP, triose-P, xylulose-5-P, glycerate-P) and P_i	Eukaryotic	(Facchinelli and Weber, 2011; Weber and Linka, 2011)
PTr4	See appendix (A6.3-1 PTr4)	2.A.7.9 The Triose-phosphate Transporter (TPT) Family	P_i-coupled secondary active antiport	9	Phosphorylated C_6, C_5-, and C_3-compounds (Glucose-6-P, PEP, triose-P, xylulose-5-P, glycerate-P) and P_i	Eukaryotic	(Facchinelli and Weber, 2011; Weber and Linka, 2011)

* The protein model was N-terminally trimmed to the first methionine encoded downstream of the spliced leader in the corresponding transcript.

2.2.2.2 Verification of Substrate Specificities using Complementation Assays

Eight nuclear encoded transporters were tested for their capability in transporting metabolites that are most likely imported to or exported from the chromatophore. For this purpose, the corresponding ORFs were amplified from *P. chromatophora* cDNA and cloned into different vectors for direct expression or expression from a defined locus after genomic integration. Expression was induced in different transporter knockout strains, each of them unable to import specific metabolites across their plasma membranes. The mutant strains used include an *E. coli* strain unable to import hexose-6-phosphates (Baba *et al.*, 2006), a yeast strain unable to import hexoses (Wieczorke *et al.*, 1999), another yeast strain unable to import proline, arginine, citrulline, GABA, aspartate, and glutamate (Besnard *et al.*, 2016; Fischer *et al.*, 2002), and a cyanobacterial strain unable to import neutral amino acids (Eisenhut *et al.*, 2007)(see 4.1.3). When grown on agar plates supplemented with one of those metabolites as the sole source of either carbon or nitrogen, only such cells should be able to form colonies, which express a *P. chromatophora* transporter able to facilitate import of the given carbon- or nitrogen compound. As several amino acids (e.g. glycine, phenylalanine) are toxic to the cyanobacterial strain used when applied at high concentrations, here the uptake of these amino acids is displayed by growth inhibition. Thus, in general complementation of the knockout mutant phenotype (i.e. complete or partial restorage of the wild type phenotype) on a specific substrate by a *P. chromatophora* transporter indicates that the given substrate indeed represents a possible natural substrate of that transporter *in vivo*. Such complementation experiments offer some notable advantages when compared to biochemical methods for functional characterization of transporters, e.g. they are more sensitive and might also function when only low amounts of functional protein are present in the cells (Trevisson *et al.*, 2009).

The *E. coli* Δ*uhpT* hexose-6-phosphate transporter deletion strain allows for complementation experiments on hexose-6-phosphates as the sole source of carbon. However, the two putative phosphate translocators PcPTr1 and PcPTr4 failed to import glucose-6-phosphate or fructose-6-phosphate into *E. coli* cells in the described experimental setup (Fig. 2.2-2). The growth behavior of cells expressing the transporters equals that of the negative control strain carrying only the empty vector. However, in the positive control strain the Δ*uhpT* phenotype is rescued by expression of EcUhpT, enabling growth on hexose-6-phosphates equal to that on glucose. Supply of higher substrate concentrations (up to 2 % (w/v) carbon source) lead to similar results. Higher concentrations of IPTG (> 10 µM) further impaired growth of the PcPTr1- or PcPTr4-expressing strains even on glucose, indicating that strong expression of the heterologous membrane proteins has toxic effects on *E. coli* (data not shown).

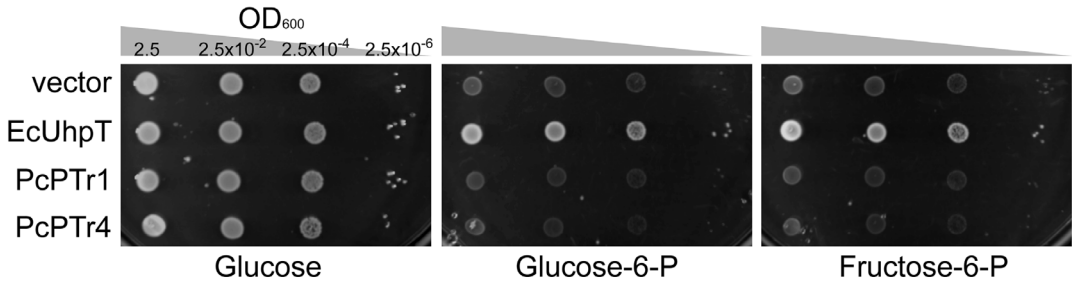


Figure 2.2-2 *P. chromatophora* putative phosphate translocators fail to import hexose-6-phosphates. Serial dilutions of an *E. coli uhpT*-knockout strain carrying vectors for expression of *P. chromatophora* putative phosphate translocators (PcPTr1 or PcPTr4) or the *E. coli* phosphate translocator EcUhpT or carrying the empty vector were dripped on M9 minimal medium supplemented with 0.2 % (w/v) glucose or hexose-6-phosphates as the sole carbon sources. Expression was induced with 1 µM IPTG from vector pTAC-MAT-Tag-2 and supported by vector pRARE (providing tRNAs for translation of rare codons in *E. coli*). Plates contained the antibiotics ampicillin and chloramphenicol and were incubated for 2 d at 28°C.

Also, the three putative sugar transporters PcSTr1, PcSTr2, and PcSTr3 failed to import hexoses into *S. cerevisiae* cells of the multiple hexose transporter knockout strain EBY.VW4000 (Fig. 2.2-3). Cells expressing the transporters grew worse on hexoses than the negative control strain carrying only the empty vector, but behave like the negative control when the disaccharide maltose is provided. Thus, accumulation of membrane proteins might be harmful to cells (under selective conditions) which is a well-known problem regarding membrane protein overexpression (Wagner *et al.*, 2007). Homologous expression of hexose transporter ScHxt7 in the hexose-uptake mutant rescues the phenotype on hexoses and even enhances growth on maltose. Supply of higher substrate concentrations (up to 10 % (w/v) carbon source) lead to similar results (data not shown).

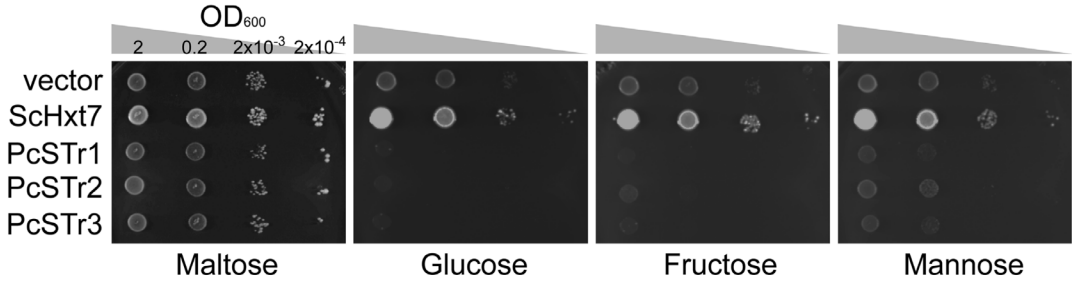


Figure 2.2-3 *P. chromatophora* putative sugar transporters fail to import hexoses. Serial dilutions of an *S. cerevisiae* multiple hexose transporter knockout strain carrying vectors for expression of *P. chromatophora* putative sugar transporters (PcSTr1, PcSTr2, or PcSTr3) or the *S. cerevisiae* hexose transporter ScHxt7 or carrying the empty expression vector were dripped on YEP medium supplemented with 0.1 % (w/v) maltose or 0.2 % (w/v) hexoses as the sole carbon sources. Expression was under a constitutively active promoter. Plates contained the antibiotic G-418 and were incubated for 3 d at 28°C.

Two different expression systems were applied to test *P. chromatophora* amino acid transporters for their transport capabilities. The yeast multiple amino acid transporter knockout mutant strain 22Δ8AA was utilized for complementation experiments on the transport of the neutral amino acid proline as well as citrulline and arginine which are highly expected to be transported across the chromatophore envelope (see Fig. 1.2-2). The strain is reported to also allow for testing of γ-aminobutyric acid (GABA), aspartate, and glutamate uptake which was attempted here although there is no indication for a function in transport of acidic amino acids by the *P. chromatophora* transporters tested. The *Synechocystis ΔnatB* neutral amino acid transporter mutant was used for complementation experiments on the transport of several neutral amino acids representing predicted substrates for all three tested *P. chromatophora* transporters.

However, in *S. cerevisiae* none of the three transporters seemed to import any of the tested amino acids (Fig. 2.2-4A). The mutant strain as well as the complemented mutant were only able to grow when nitrogen was supplied in form of ammonium sulfate, while strain 23344c (the wildtype parental of 22Δ8AA) could import amino acids and utilize them as nitrogen source. Noteworthy, the complementation experiment was not evaluable for aspartate and glutamate as the mutant unexpectedly grew like the wildtype on these substrates when supplied at 1 mM concentrations (data not shown). As the transporters tested are predicted to be Na^+ symporters the experiment was repeated with and without supplementing 100 mM NaCl but revealed the same result. However, the inability of *P. chromatophora* transporters to import amino acids in a heterologous system can have several reasons and might not be connected to a true inability of the tested transporters to transport the given substrates. Also low expression, aggregation, aberrant protein folding and post-translational processing, or mistargeting to the endomembrane system might abolish transporter function. As the transporters carry N-terminal V5-tags, their expression and distribution in the yeast cells was evaluated via immunofluorescence microscopy (Fig. 2.2-4B). The *S. cerevisiae* transporter ScHxt7 is functionally expressed (see Fig. 2.2-3) and thus represents an example for sufficient expression and localization to the plasma membrane. PcBGATr showed a fluorescence pattern very similar to ScHxt7 but the transporter seemed to be present also in intracellular membrane structures in some cells. However, PcBGATr likely properly localized to the plasma membrane at least in a fraction of the cells observed. PcNAATr1 gave only a very weak signal and could be observed exclusively in intracellular structures (probably aggregates). PcNAATr2 might to some extent localize to the plasma membrane but is to a large extent mistargeted to intracellular membranes that might represent the cortical endoplasmic reticulum. The results indicate that a positive outcome of the complementation experiments performed with these respective strains was abolished already due to absence of the transporter in the yeast plasma membrane.

A cyanobacterial expression system might avert the problem of subcellular mistargeting because it lacks an endomembrane system (except thylakoids) and the chromatophore inner envelope is believed to still share characteristics of the cyanobacterial plasma membrane. Thus, a cyanobacterium might resemble the putative native environment of the *P. chromatophora* transporters and enhance proper targeting. Strain *ΔnatB* is a *Synechocystis* sp. PCC6803 mutant lacking the periplasmic substrate-binding protein (NatB) of the Nat ABC-transporter for neutral amino acids and is thus unable to import external neutral amino acids. Since the cyanobacterium is sensitive to several amino acids when supplied externally (Eisenhut *et al.*, 2007; Labarre *et al.*, 1987) successful import of the tested amino acids is noticeable by growth retardation or cell death. Phenylalanine and ethionine (a toxic analog of methionine) both strongly inhibited growth of the wildtype, but not of the mutant strain as it is not able to import these neutral amino acids (Fig. 2.2-4C). However, the complemented mutants behaved like the non-complemented mutant indicating that the amino acids are not imported by any of the tested *P. chromatophora* transporters. Concentrations ranging from 30 to 120 µM for ethionine and from 250 to 750 µM for phenylalanine were tested and showed comparable results (data not shown). The experiments were also repeated with and without supplementation of 100 mM NaCl to the medium and revealed similar results. In contrast, glycine import seems to be facilitated by all three *P. chromatophora* transporters. The complemented mutants show the same growth behavior as wildtype cells. PcBGATr and PcNAATr2 also further reduce colony growth in the wildtype background. The experiment was performed with two biological replicates of each strain and the effect was present at glycine concentrations ranging from 20 to 30 mM (data not shown). Noteworthy, the result was not reproducible when NaCl was added to the medium elevating the Na^+ concentration from ca. 20 mM in standard BG11 to 120 mM in supplemented BG11. At high NaCl concentration all strains showed the same appearance and growth behavior on glycine, regardless of the presence or absence of a functional Nat transporter, rendering the experiment unevaluable (data not shown). The combination of mild osmotic stress and the toxic effect of glycine might have led to an unforeseen metabolic cross-reaction in these cells.

PcBGATr represents a particularly interesting chromatophore envelope transporter candidate due to its bacterial phylogenetic origin (Tab. 2.2-2) and was therefore further analyzed. To validate the observed ability of PcBGATr for glycine transport by another experimental approach and to screen PcBGATr for transport of more neutral amino acids, the internal accumulation of amino acids supplied to liquid cultures for a period of 25 h was measured via HPLC (Fig. 2.2-4D). As expected, externally applied glycine was taken up by wildtype cells and even accumulated until it constituted 30-50 % of all intracellular amino acids quantified. This represents a drastic increase when compared to the relative glycine abundance in untreated cells, which is about 1 % of total amino acids. Glycine was also taken up to some extent by the NatB-mutant but accumulated to much lower levels of only around 15 % of

total amino acids. A comparable difference between wildtype and NatB-mutant was detected earlier in a similar experiment (Eisenhut *et al.*, 2007). However, contrary to the results from complementation experiments PcBGATr did not show any effect on glycine uptake in the mutant background. The result was the same in four independent experiments and also when external glycine concentration was lowered from 30 to 5 mM. However, an effect was visible in wildtype cells, where glycine accumulation was approx. 50 % higher in PcBGATr-expressing cells than in control cells. The result was reproducible with two biological replicates in two individual experiments. It has to be considered that a strong but unpredictable impact on cell metabolism upon the observed enormous intracellular accumulation of glycine might have pleiotropic effects that influence the outcome of the experiment.

The HPLC method also allowed for screening of PcBGATr for a function in transport of several other neutral amino acids (Fig. 2.2-4D). Aiming to reduce the sample number, the effect of PcBGATr was only tested in the mutant background (*ΔnatB*). In untreated cells levels of all tested amino acids were about as low as the values measured for glycine, while glutamate, aspartate, and the combined peak signal for arginine/glutamine together made up approx. 75-90 % of total intracellular amino acids (data not shown). Comparable relative abundances have been determined earlier in untreated *Synechocystis* (Kiyota *et al.*, 2014). All externally applied amino acids severely accumulated in the cells, but to a different extent. However, a weak effect of PcBGATr is only visible for serine, indicating that this amino acid might represent a possible substrate of the transporter, though this possibility was not further tested due to time restrictions.

Uptake experiments using ^{14}C-labelled glycine have been applied as a third method to further analyze possible PcBGATr-mediated glycine import (Fig. 2.2-4E). *Synechocystis* cells were incubated with the radioactive substrate for different timespans ranging from 1 to 30 min. The substrate was then removed and only intracellular ^{14}C-glycine was measured in a scintillation counter. The experiment revealed a clear accumulation of glycine in wildtype cells due to action of the cyanobacterial Nat transport system during the observed timespan, although saturation of the transport reaction facilitated by Nat was not achieved under the experimental conditions and clearly requires a more extensive timescale and/or lower substrate concentration. Comparably only very little glycine entered NatB-knockout cells even after 30 min. PcBGATr-expressing mutant cells, however, behaved like control cells (considering that the counts measured at 8 min represent an outlier). Thus, the positive result from complementation experiments indicating PcBGATr´s ability to transport glycine could not be verified when applying different experimental setups. However, it should be noted that no Na^+ ions were added to the transport mix in this experiment but might be necessary for glycine symport according to PcBGATr´s annotation (Tab. 2.2-2). Thus, the experiment might be repeated at moderate Na^+ concentrations around 20 mM, emulating the conditions present in BG11 medium that was used in the other experiments.

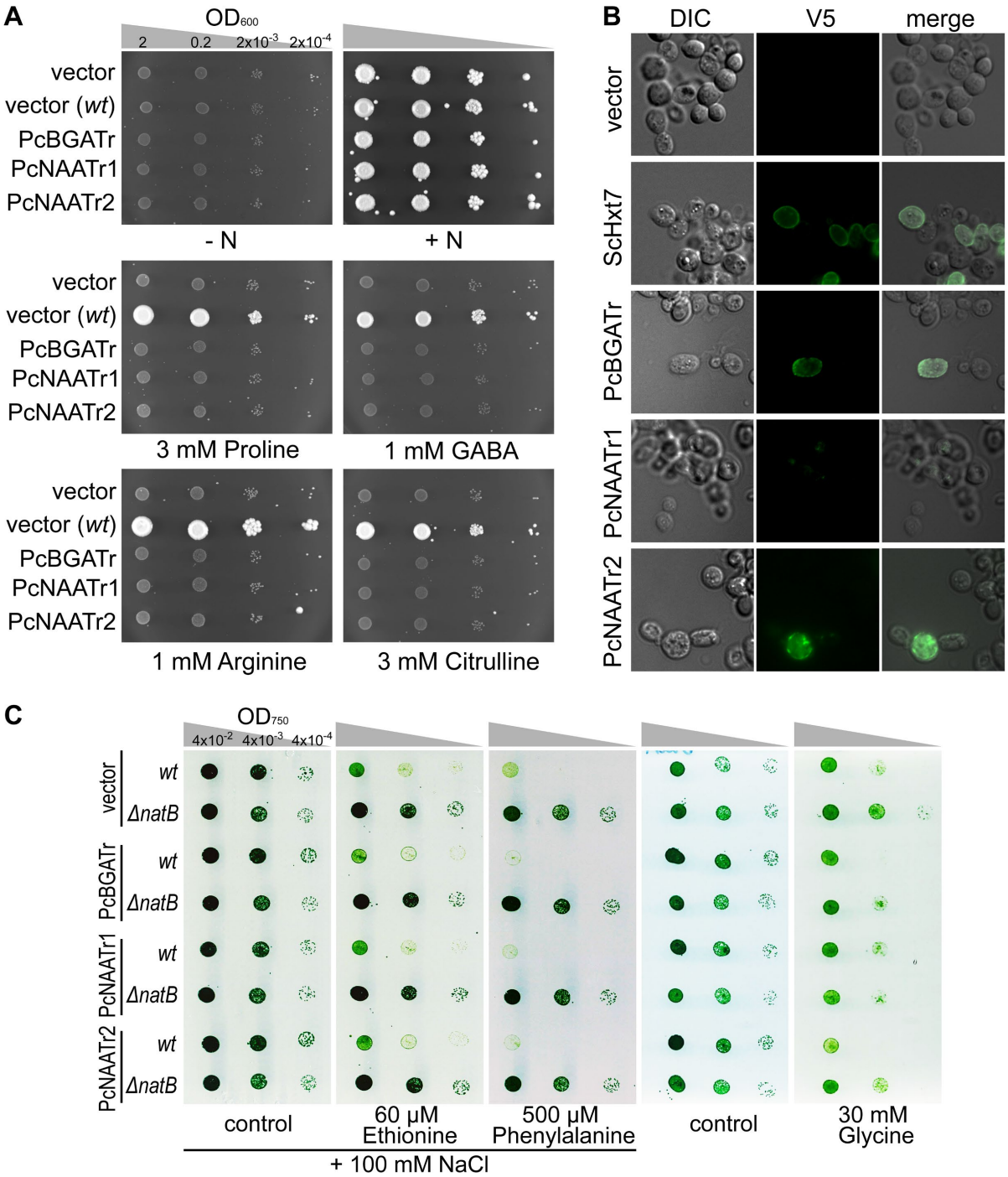

A
OD600
2
0.2
2x10-3
2x10-4
vector
vector (wt)
PcBGATr
PcNAATr1
PcNAATr2
- N
+ N
3 mM Proline
1 mM GABA
1 mM Arginine
3 mM Citrulline
B
DIC
V5
merge
vector
ScHxt7
PcBGATr
PcNAATr1
PcNAATr2
C
OD750
4x10-2
4x10-3
4x10-4
wt
ΔnatB
control
60 µM Ethionine
500 µM Phenylalanine
30 mM Glycine
+ 100 mM NaCl
D
intracellular AA [% of total AA]
vector
PcBGATr
Gly
Ala
Met
Ser
Trp
Val
Ile
Leu
E
glycine/cells [pmol/OD750]
ΔnatB-PcBGATr
time [min]

Figure 2.2-4 Characterization of *P. chromatophora* putative amino acid transporters. A Serial dilutions of an *S. cerevisiae* multiple amino acid transporter knockout strain carrying vectors for expression of *P. chromatophora* putative amino acid transporters (PcBGATr, PcNAATr1, or PcNAATr2) or the empty expression vector were dripped on either SD medium (-N), SD medium supplemented with 0.5 g/l ammonium sulfate (+N), or with 1-3 mM amino acids as the sole nitrogen sources. A wildtype strain (*wt*) was used as a positive control. Expression was under a methionine-repressible promoter. Cells were grown in the presence of 100 mM NaCl and in the absence of uracil and methionine for 9 d at 28°C. **B** Subcellular localization of V5-tagged *P. chromatophora* amino acid transporters expressed in *S. cerevisiae* was visualized via immunofluorescence microscopy. Strains are the same as in A but the V5-tagged *S. cerevisiae* hexose transporter ScHxt7 was used as a positive control for localization to the plasma membrane. DIC, differential interference contrast. **C** Serial dilutions of a *Synechocystis* neutral amino acid transporter knockout strain (*ΔnatB*) carrying genomically integrated ORFs of *P. chromatophora* putative amino acid transporters (PcBGATr, PcNAATr1, or PcNAATr2) or only a spectinomycin resistance gene (vector) were dripped on either BG11 medium or BG11 medium supplemented with the cytotoxic methionine analog ethionine or the amino acids phenylalanine or glycine, which are toxic to the cells (when present at high concentrations). The medium was usually supplemented with 100 mM NaCl, but not when glycine was used in the experiment. Expression occurred from the *psbAII* locus. The wildtype strain (*wt*) was used as a positive control, i.e. the cells do not survive as toxic amino acids are actively imported. Cells were grown in the presence of spectinomycin for 16 d at approx. 30 µE and 28°C. **D** Intracellular amino acid concentrations in *Synechocystis* neutral amino acid transporter knockout strain (*ΔnatB*) carrying genomically integrated ORF of *P. chromatophora* putative amino acid transporter PcBGATr or only a spectinomycin resistance gene (vector) after 25 h of incubation with and without 30 mM glycine (+/-) or with 30 mM alanine, methionine, serine, tryptophan, valine, isoleucine, or leucine in BG11 medium at 150 rpm, approx. 30 µE, and 28°C in the presence of spectinomycin. Intracellular amino acids were extracted, derivatized, and their abundance was determined via HPLC. In the case of glycine, bars show the mean of two biological replicates and also values measured for wild type cells and cells not treated with amino acids are shown. **E** Uptake of ^{14}C-glycine by *Synechocystis* wildtype (*wt*) or neutral amino acid transporter knockout strain (*ΔnatB*) carrying genomically integrated ORF of *P. chromatophora* putative amino acid transporter PcBGATr or only a spectinomycin resistance gene (*wt* and *ΔnatB*). Cells were incubated in ^{14}C-glycine containing Tricine buffer at pH 8.1 and 28°C for the indicated timespan, washed, and intracellular ^{14}C radioactivity was measured in a scintillation counter.

2.2.2.3 Immunodetection of BGATr in *P. chromatophora* Cell Fractions

Experiments on cyanobacterial cells expressing PcBGATr, PcNAATr1, and PcNAATr2 indicated that the *P. chromatophora* putative alanine or glycine : Na^+ symporter and the two neutral amino acid permeases might indeed transport glycine (see Fig. 2.2-4AC,D), an amino acid that probably has to be imported into the chromatophore (see Fig. 1.2-2). However, it is not known whether the nuclear encoded transporters are present at the chromatophore envelope to facilitate such transport. Since *P. chromatophora* currently cannot be genetically manipulated and consequently does not allow for

the expression of tagged fusion proteins, peptide antibodies were generated against one transporter to evaluate its subcellular localization. BGATr was chosen due to its interesting phylogenetic affiliation that indicates its acquisition via lateral gene transfer (Nowack *et al.*, 2016).

The raw serum of a rabbit immunized with two BGATr-peptides (see 4.2.2.8) contains antibodies that bound to several proteins in *P. chromatophora* whole cell lysate and chromatophore lysate (Fig. 2.2-5A). A faint background signal is also visible when pre-immune serum was used. A band approx. fitting the predicted molecular weight of the BGATr protein of 59 kDa was detected in both samples but appeared more intense in the chromatophore. However, the experiment was repeated, and the band showed equal intensity in both samples this time (data not shown).

To increase specificity for BGATr the α-BGATr antibody was affinity-purified from the immune serum making use of the two peptides that were used for immunization (see 4.2.2.8). The purified antibody binds to a specific band between 55 and 70 kDa (Fig. 2.2-5B). However, very slight additional bands appeared at 70 kDa and between 25 and 35 kDa when a longer exposure time was used to detect the chemiluminescent signal and even more bands appeared after storage of the immune serum at -20°C or -80°C for several months (data not shown).

To make sure that the prominent band detected by α-BGATr indeed represents BGATr protein, the antibody was tested on membrane fractions of *Synechocystis* cells expressing BGATr (Fig. 2.2-5C). A faint band could be detected in samples containing *P. chromatophora* BGATr and was not present in samples derived from wild type cyanobacteria. The result was reproducible. However, the band appears at a slightly higher molecular weight than the respective band in *P. chromatophora* lysate, which might be indicative for different post-translational modifications on the protein in prokaryotic *Synechocystis* and eukaryotic *P. chromatophora*. Noteworthy, presence of PcBGATr in *Synechocystis* membrane fractions could be verified via mass spectrometry (test run, data not shown).

Finally, chromatophore lysate was tested for the presence of the putative BGATr band (Fig. 2.2-5D). The antibody detected a strong signal in *P. chromatophora* lysate but only a very faint signal at approx. the same position in chromatophore lysate. The experiment was repeated several times and the signal never appeared stronger in chromatophore lysate than in whole cell lysate. The result suggests that BGATr is likely not inserted into the chromatophore inner envelope, as chromatophore-targeted proteins should appear enriched in the chromatophore fraction compared to whole cell lysate.

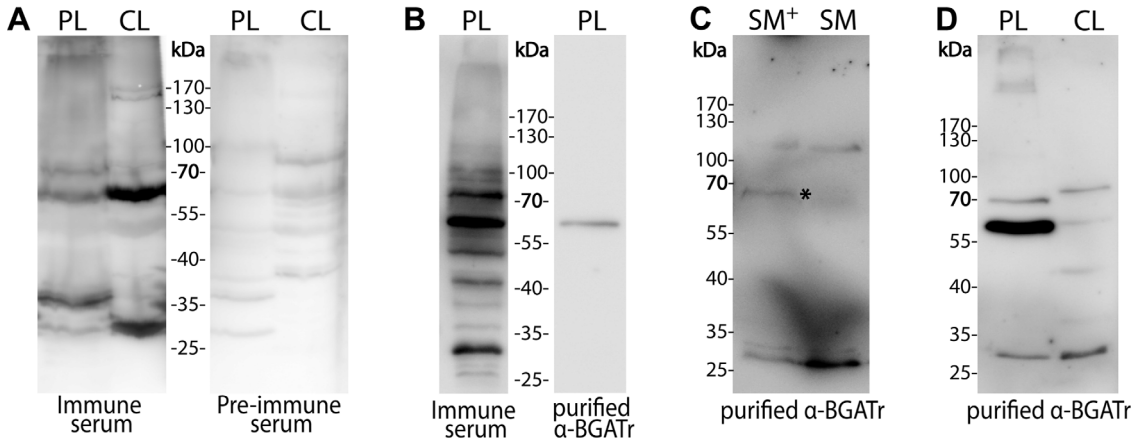


Figure 2.2-5 BGATr is likely not present in the chromatophore. A Antibodies in α-BGATr immune serum bound to multiple bands on 50 µg *P. chromatophora* whole cell lysate (PL) and chromatophore lysate (CL). Very faint bands were also visible when the pre-immune serum was used. **B** A distinct band was detected by affinity-purified α-BGATr antibody (diluted 1:7) on whole cell lysate. **C** Affinity-purified α-BGATr antibody detected a distinct band (*) in 150,000 x g membrane fractions from *Synechocystis* sp. PCC6803 expressing BGATr (SM^+), but not in wild type membranes (SM). Membranes were solubilized in sample buffer plus 6.5 M urea, 1 % TritonX-100, and 1 % Na-deoxycholate. **D** A strong band is detected in *P. chromatophora* whole cell lysate by affinity-purified α-BGATr antibody. However, a slight band can also be observed in chromatophore lysate at approx. the same position. Unless otherwise stated, protein was extracted from cells or chromatophores through TCA-precipitation and 20 µg lysate protein in 8 M urea buffer was loaded per lane. All samples were solubilized at 37°C for 30-60 min in sample buffer containing 150 mM DTT and 3 % SDS. 10 % SDS-polyacrylamide gels were run in TGS buffer. Serum was used in a 1:50 dilution, α-BGATr antibody was used in a 1:16 dilution unless otherwise stated. Chemiluminescence on membranes in A and B was detected exactly under the same conditions (e.g. exposure time), respectively. The predicted molecular weight for BGATr is 59 kDa.

An attempt to detect BGATr in embedded, sectioned *P. chromatophora* cells via Immunogold Electron Microscopy (EM) failed (Fig. A6.3-2), probably because of the very low sensitivity of the α-BGATr antibody that was used here in a 1:1 dilution (usually, the concentration of antibody used in Immunogold EM is recommended to be a hundred times the concentration used in western blots). Gold particles appeared not enriched in a certain structure of the cell but were distributed uniformly across the cells. Much fewer particles were present in the negative control (i.e. sections treated with the secondary antibody only), excluding non-specific binding of the secondary antibody and indicating non-specific binding of α-BGATr, probably due to the high concentration used.

2.2.3 A Top-Down Approach for the Identification of Nuclear Encoded Chromatophore Envelope Transporters: Proteomics

In order to complement the bottom-up experimental approach for the identification of nuclear encoded chromatophore-targeted metabolite transporters a targeted proteomic approach was applied. The existing chromatophore lysate proteomic dataset (Singer *et al.*, 2017) was revaluated with a focus on membrane proteins and a second mass spectrometric analysis was carried out, this time on chromatophore insoluble protein and membrane fractions.

2.2.3.1 Revaluation of the Singer Proteomic Dataset

The scarcity of chromatophore-encoded solute transporters suggested that in *P. chromatophora*, as in plastids, nuclear encoded transport systems establish metabolic connectivity of the chromatophore. However, among the 432 import candidates identified previously by Singer *et al.* 2017, only three proteins contained more than one predicted TMH (Tab. 2.2-3). One of these proteins (identified by *in silico* prediction) contains two TMHs, only one of which is predicted with high confidence. Of the other two proteins (identified by MS), one is short and contains two predicted TMHs; the other contains eight predicted TMHs. However, this latter protein was identified with one peptide only and shows no BlastP hits against the NCBI nr database, whereas an alternative ORF (in the reverse complement) shows similarity to an NAD-dependent epimerase/dehydratase. Therefore, this latter protein likely represents a false positive (a false discovery rate of 1 % was accepted in this analysis).

Table 2.2-3: Previously identified import candidates do not comprise nuclear encoded solute transporters. The table lists numbers of proteins previously identified to be imported into the chromatophore by *in silico* prediction (based on presence of a crTP), liquid chromatography coupled to tandem MS (LC-MS/MS), and total (Singer *et al.*, 2017). Proteins are sorted by the number of predicted TMHs that are located outside of the conserved crTP-sequence.

	0 TMH	1 TMH	>1 TMH
***In silico* predicted (crTP)**	289	3	1
LC-MS/MS identified	194	11	2
Total	416	13	3

The absence of multi-spanning transmembrane proteins among import candidates could have two reasons. (i) Analogical to the mTP-independent insertion of many nuclear encoded carriers into the mitochondrial inner membrane (Ferramosca and Zara, 2013), these proteins might use a crTP-independent import route, impairing their *in silico* prediction as import candidates. (ii) transmembrane

proteins are often underrepresented in LC-MS analyses owing to low abundance levels (compare Antelo-Varela *et al.,* 2019; Maaß *et al.,* 2011, 2014; Muntel *et al.,* 2014), few trypsin cleavage sites, as well as unfavorable retention and ionization properties. Accordingly, MS analysis of the chromatophore soluble proteome identified 47 % of the soluble but only 21 % of TMH-containing proteins predicted in the chromatophore genome. Among these were transmembrane subunits of only three of the 25 chromatophore-encoded transport systems indicating low recovery of transmembrane proteins.

2.2.3.2 Enrichment of the Chromatophore Insoluble Fraction

To enhance identification of transmembrane proteins, such proteins were enriched by collecting the insoluble fractions from isolated chromatophores (CM samples) and intact *P. chromatophora* cells (PM samples). Comparison of CM and PM samples to chromatophore lysates (CL samples) by SDS-PAGE revealed distinct banding patterns between the three samples and high reproducibility between three biological replicates (Fig. 2.2-6A). Two consecutive high-salt and high-pH treatments effectively depleted soluble and membrane attached proteins from the samples as deduced from the respective banding patterns in collected washing fractions W1 and W2 (Fig. 2.2-6B). Further enrichment of membrane proteins or separation of outer membrane, inner membrane, and thylakoids was not feasible owing the slow growth of *P. chromatophora* (one cell division every six days) and low yield of chromatophore isolations. Also, electron microscopic analysis of isolated chromatophores suggested that the outer membrane is lost during chromatophore isolation (Fig. 2.2-6C; compare Kies, 1974; Sato *et al.*, 2020).

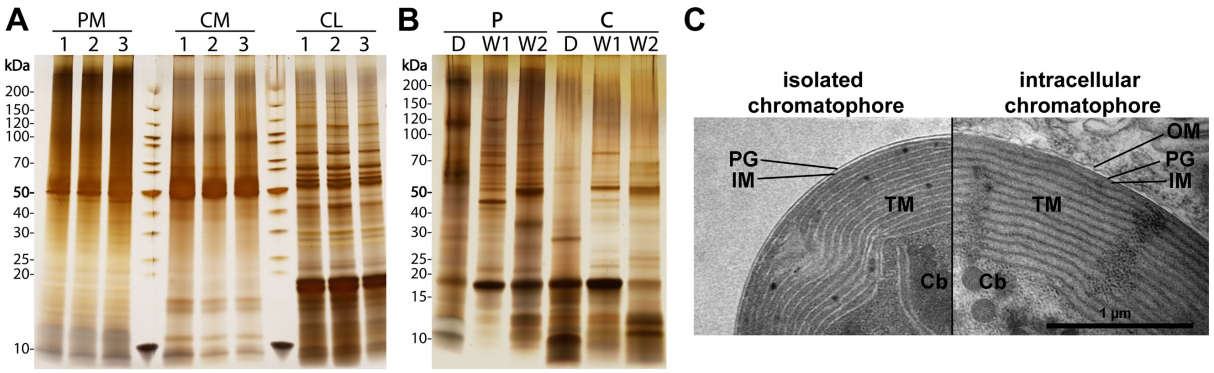


Figure 2.2-6 Cell fractionation for enrichment of the *P. chromatophora* insoluble proteome. A Resolved protein from three replicates (1, 2, 3) of each, chromatophore lysates (CL) as well as high-salt and carbonate-washed *P. chromatophora* membranes (PM) and chromatophore membranes (CM). **B** Cell/organelle debris (D), high salt wash fractions (W1), and high pH carbonate wash fractions (W2) collected during fractionation of whole cells (P) and isolated chromatophores (C). 1 µg of protein per lane was resolved on 4-20 % polyacrylamide gels and silver stained. **C** TEM micrographs of Epon embedded isolated chromatophore (left) and chromatophore in the context of a *P. chromatophora* cell (right). The outer envelope membrane (OM) observed in intact cells was lost during the isolation process. IM, inner envelope membrane; PG, peptidoglycan; TM, thylakoid membranes; Cb, carboxysomes.

2.2.3.3 The Chromatophore Insoluble Proteome

To further boost identification of transmembrane proteins by MS two approaches were tested. (i) The reduction of heterogeneity in the sample via de-glycosylation of proteins before in gel digestion and (ii) protein digestion with chymotrypsin instead of trypsin (i.e. cleavage C-terminal to aromatic amino acids which are abundant instead of positively charged amino acids which are rare in TMHs). However, de-glycosylation or chymotrypsin-mediated digestion decreased protein identification by 24 % and 74 %, respectively (Fig. 2.2-7A), and only two transmembrane proteins were identified exclusively by either method compared to the standard trypsin protocol (Fig. 2.2-7B). Thus, the following work relied on trypsin digestion of enriched transmembrane proteins.

Two consecutive, independent MS analyses of three replicates of each, CM, PM, and CL samples led to the identification of 1,886 nuclear and 555 chromatophore-encoded proteins over all fractions (Tab. 2.2-4 and S3). Although most chromatophore-localized transmembrane proteins were also identified in CL samples (Tab. 2.2-4), individual transmembrane proteins were clearly enriched in CM compared to CL samples (Fig. 2.2-7C).

Table 2.2-4 Proteins identified in this study by LC-MS/MS. Numbers of chromatophore-encoded (CE) and nuclear encoded (NE) proteins identified in at least one out of two independent MS experiments with ≥3 spectral counts in chromatophore-derived samples (i.e. CM+CL) or whole cell membranes (PM). The number of predicted TMHs (outside of the crTP) is indicated. For proteins identified in CM samples, total number of proteins and number of proteins enriched in CM as compared to PM samples (in brackets) is indicated separately.

	All proteins		**1 TMH**		**>1 TMH**	
	CE	**NE**	**CE**	**NE**	**CE**	**NE**
Chromatophore insoluble (CM)	533 (506)	297 (236)	28 (24)	20 (13)	70 (67)	5 (2)
Chromatophore soluble (CL)	551	354	28	25	67	7
Chromatophore total	555	361	28	27	70	7
Whole cell insoluble (PM)	385	1691	24	209	50	175
Total	555	1886	28	218	70	179

In CM samples, 98/213 (or 46 %) of the chromatophore-encoded transmembrane proteins were identified, representing a gain of 118 % compared to the previous analysis by Singer *et al.*, 2017 (Fig. 2.2-7D). In particular, of the 25 chromatophore-encoded solute transport systems, 18 (or 72 %) were identified with at least one subunit, and 15 (or 60 %) were identified with their transmembrane subunit (Fig. 2.2-7E) while the previous study identified only five of these transporters (three with their transmembrane subunit). Highest intensities (representing a rough estimation for protein abundances) were found in CM samples for an ABC-transporter annotated as multidrug importer of the P-FAT family (level 4 to 5 [see 4.2.8.3 and Fig. 2.2-7E], placing the transporter among the 10 % most abundant proteins in CM). Also the bicarbonate transporter BicA, two multidrug efflux ABC-transporters, and a NhaS3 proton/sodium antiporter were found in the upper tiers of abundance levels (level 3 to 4, placing them among the 30 % most abundant proteins in CM). The remaining transporters showed moderate to low abundance levels (Fig. 2.2-7E).

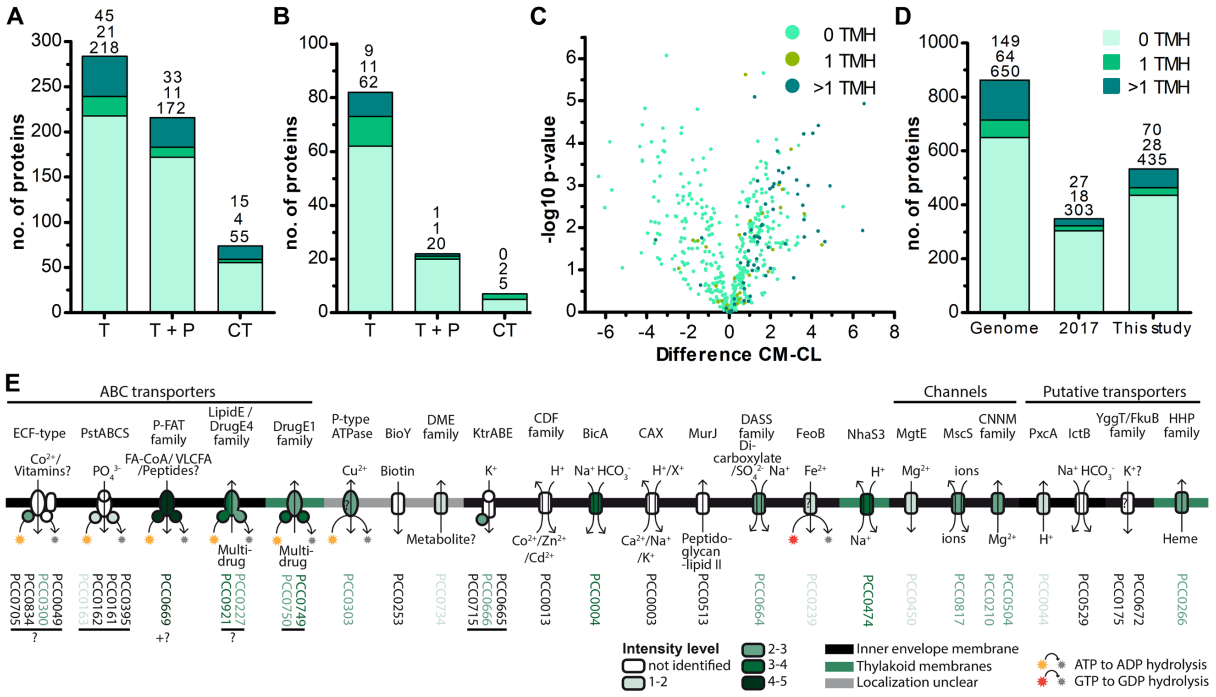


Figure 2.2-7 Increased recovery of transmembrane proteins by MS analysis of enriched insoluble chromatophore proteins. A Total chromatophore-encoded proteins identified in PM samples by LC-MS/MS after treatment with trypsin (T), with the amidase PNGase F prior to trypsin cleavage (T+P), or with chymotrypsin (CT) and **B** proteins exclusively identified after respective treatments. The number of predicted TMHs is indicated by a color code. **C** Enrichment of chromatophore-encoded membrane proteins in MS experiment 1 (see Fig. A6.3-3 for MS experiment 2). The difference between the mean log2-transformed normalized intensities of individual proteins in CM and CL samples ($\overline{log2(normInt_{CM})} - \overline{log2(normInt_{CL})}$; Difference) is plotted against significance (-log10 p-values in Student´s t-test) for proteins identified in all three triplicates of either CM or CL or both. A positive difference value indicates protein enrichment in CM, negative values indicate depletion in CM compared to CL samples. Values for proteins detected only in one sample have been imputed and are only shown when their difference is significant. The number of predicted TMHs is indicated by a color code. **D** Numbers of proteins encoded on the chromatophore genome (Genome) and chromatophore-encoded proteins identified with ≥3 spectral counts in chromatophore-derived samples in a previous (2017; Singer *et al.*, 2017) and current (This study) proteome analysis. The number of predicted TMHs is indicated by a color code. **E** Detection of chromatophore-encoded transport systems. Annotation or TCDB-family, predicted mode of transport, substrates, and probable subcellular localization are provided. For each protein, the mean normalized intensity in CM (over both MS experiments) is indicated by a color code (see also Tab. S2).

Determination of nuclear encoded proteins enriched in CM compared to PM samples led to the identification of 188 high confidence (HC) and further 48 low confidence (LC; see 4.2.8.3 and Fig. A6.3-4) import candidates (Fig. 2.2-8A, Tab. S4). Of these 236 import candidates, 159 were known import candidates (Singer *et al.*, 2017; Fig. 2.2-8B, Tab. S4), with 46 proteins now experimentally confirming import candidates previously only predicted *in silico*. 77 proteins represent new import candidates, mostly lacking N-terminal sequence information (42 proteins) or representing short import candidates (22 proteins).

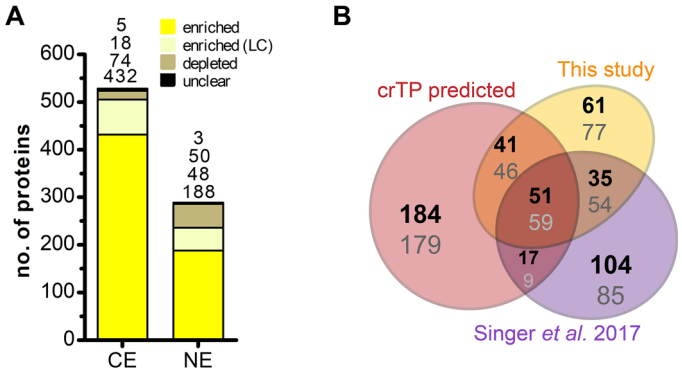


Figure 2.2-8 Import candidates. A Chromatophore-encoded (CE) and nuclear encoded (NE) proteins enriched in CM compared to PM samples. Yellow, proteins enriched with high confidence; light yellow, proteins enriched with low confidence (LC); brown, proteins depleted in CM; black, proteins classified as "unclear" (see 4.2.8.3 and Fig. A6.3-4). Only proteins identified with ≥3 spectral counts in the chromatophore samples in at least one out of two independent MS experiments were considered. **B** Numbers of newly identified import candidates in this study (see A; yellow), previously MS-identified import candidates (Singer *et al.,* 2017; purple), and *in silico* predicted import candidates (crTP predicted in Singer *et al.,* 2017; red). Numbers in bold indicate distribution of proteins considering only HC import candidates, numbers in grey considering all import candidates.

2.2.3.4 No Multi-spanning Transmembrane Proteins Appear to be Imported into the Chromatophore

Nucleus-encoded multi-spanning transmembrane proteins appeared invariably depleted in chromatophores (Figs. 2.2-9A, B, C). Only two of 236 import candidates were multi-spanning transmembrane proteins according to the TMHMM prediction server (Tab. 2.2-4). However, one of these (scaffold1608-m.20717, with seven predicted TMHs, arrowhead in Fig. 2.2-9B) was identified by only one hepta-peptide and shows no similarity to other proteins in the NCBI nr database whereas an overlapping ORF (in another reading frame) encodes a peroxidase that was MS-identified in Singer *et al.* 2017 likely classifying the protein as a false positive. For the other import candidate (scaffold18898-m.107131, with two predicted TMHs and an enrichment level close to 0, arrowhead in Fig. 2.2-9C) a full-length transcript sequence is missing precluding determination of the correct start codon. Thus,

this protein might represent in fact a short import candidate with a single TMH. Of the three nuclear encoded multi-spanning transmembrane proteins that were present but appeared depleted in CM compared to PM samples (Tab. 2.2-4), two were annotated as mitochondrial NAD(P) transhydrogenase and mitochondrial ATP/ADP translocase, suggesting a mild contamination of CM samples with mitochondrial membrane material.

In comparison, 70 chromatophore-encoded multi-spanning transmembrane proteins were identified in CM samples, and 67 of these appeared enriched in CM samples. In PM samples, 50 chromatophore- and 175 nuclear encoded multi-spanning transmembrane proteins were found (Fig. 2.2-9A, Tab. 2.2-4). Notably, also the nuclear encoded putative metabolite transporters STr3 and NAATr2 (see 2.2.2) were among these identifications, speaking against their presence in the chromatophore.

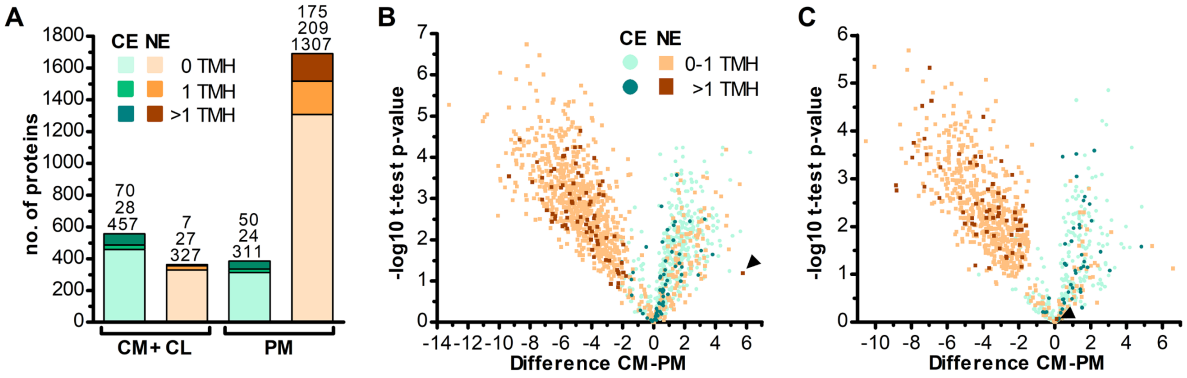


Figure 2.2-9 No evidence for import of nuclear encoded multi-spanning transmembrane proteins into chromatophores. Identification and enrichment of chromatophore-encoded (CE, green) and nuclear encoded (NE, orange) proteins. The number of predicted TMHs (located outside of the crTP) is indicated by a color code. **A** Proteins with 0, 1 or >1 predicted TMHs identified in at least one out of two independent MS experiments with ≥3 spectral counts in chromatophores (CM + CL samples) or whole *P. chromatophora* cells (PM samples). Total numbers of proteins are given above each bar. **B and C** The difference of intensities of individual proteins between CM and PM samples ($\overline{log2(normInt_{CM})} - \overline{log2(normInt_{PM})}$; Difference) is plotted against significance (-log10 p-values in Student´s t-test) for proteins detected with ≥3 spectral counts in the chromatophore samples (for proteins detected in CM only or CM and PM) or in whole cell samples (proteins detected in PM only). Values for proteins detected only in one sample have been imputed and are only shown when their difference is significant. Data from **B** MS experiment 1 and **C** MS experiment 2 are shown separately. Scaffold1608-m.20717 and scaffold18898-m.107131 (see text) are marked by arrowheads in B and C, respectively. In both analyses, among the proteins enriched in CM (Difference CM-PM > 0), the proportion of identified multi-spanning transmembrane proteins encoded in the chromatophore (49 of 409 in B; 39 of 134 in C) as compared to the nucleus (0 of 132 excluding the false positive in B; 1 of 54 in C) is significantly higher (both: p-value = 0.002, Fishers's Exact Test).

To test for the robustness of TMH predictions obtained by TMHMM, import candidates were re-analyzed with a second TMH prediction tool (Consensus Constrained TOPology prediction (CCTOP); Tab. S4). Although the exact positions or lengths of individual helices were slightly altered in many cases, overall the predictions were largely congruent between the two prediction tools. For 480 out of 508 import candidates, predicted numbers of TMHs were essentially identical between TMHMM and CCTOP; CCTOP predicted 23 additional import candidates with a single TMH, and four additional import candidates with two or three TMHs outside of the crTPs (with three out of four proteins showing a rather low reliability score of the prediction of < 65). Importantly, also CCTOP results did not yield any evidence for the insertion of classical nuclear encoded transporters (i.e. proteins with ≥4 TMHs) into the chromatophore inner membrane. The remaining text refers to TMHMM predictions.

2.2.4 Targeting of Single-spanning Transmembrane Proteins and Antimicrobial Peptide-like Proteins to the Chromatophore

In contrast to the striking lack of multi-spanning transmembrane proteins, there were 13 (5 HC and 8 LC) single-spanning transmembrane proteins (containing one TMH outside of the crTP) among the identified import candidates (Tab. 2.2-4). Three of these proteins contain a TMH close to their C-terminus and likely represent tail-anchored proteins. One of these proteins is long and annotated as low-density lipoprotein receptor-related protein 2-like, the other two (with N-terminal sequence information missing) as polyubiquitin. However, most import candidates with one TMH (10 proteins) represent short proteins. These short import candidates included two high light-inducible proteins (i.e. thylakoid-localized cyanobacterial proteins involved in light acclimation of the cell, Zhang *et al.*, 2017). The remaining eight proteins are orphan proteins lacking detectable homologs in other species (BlastP against NCBI nr database, cutoff e^{-03}); all of these contain a TMH with a large percentage of small amino acids (26-45 % glycine, alanine, serine). The TMH is especially rich in glycine (10-24 %, as compared to a median of 8.2 % glycine in TMHs of single- or multi-spanning transmembrane proteins, Saidijam, Azizpour, and Patching 2018) and is located close to the negatively charged protein N-terminus (Fig. 2.2-10A).

In the previous proteome analysis by Singer *et al.*, 2017, short orphan proteins represented the largest group of MS-identified import candidates (1/3 of total). However, most of these proteins did not possess predicted TMHs. Based on the occurrence of specific cysteine motifs (CxxC, CxxxxC) and stretches of positively charged amino acids these short proteins were described as antimicrobial peptide (AMP)-like proteins (Singer *et al.*, 2017). Indeed, antimicrobial activity of four different proteins was verified experimentally in confrontation assays using purified recombinant protein on *E. coli*, *S. cerevisiae*, and the cyanobacterium *Synechococcus elongatus* (A. Singer and L. Oberleitner,

unpublished). Including the eight TMH-containing proteins (see above), the current study identified further 19 short orphan import candidates that show no similarity or – only few of them – that show similarity only to hypothetical proteins in other species. Scrutiny of all 88 short orphan import candidates (resulting from both studies together) revealed that these short import candidates form at least four distinct groups (Fig. 2.2-10A). The TMH-containing proteins are denoted group 1 (10 proteins). Members of group 2 (12 proteins) contain a conserved motif of unknown function that occurs also in bacterial proteins that often possess domains pointing towards DNA processing-related functions (Fig. 2.2-10A, B). Members of group 3 (10 proteins) contain another conserved motif of unknown function that encompasses two cysteine-motifs (CxxxxC and CxxC). Members of group 4 (30 proteins) show either one or two CxxC mini motifs (one of these is often CPxCG) but no further sequence conservation. The remaining 26 short orphan import candidates have no obvious common characteristics, but several appear to have a propensity to form amphipathic helices (Fig. 2.2-10A).

Screening a large nuclear *P. chromatophora* transcriptome dataset (Nowack *et al.*, 2016) revealed additional putative members of groups 1 to 3 (Fig. 2.2-10A and A6.3-5): further 53 translated transcripts represent short proteins with a predicted TMH in the N-terminal 2/3 of the sequence that is rich (> 20 %) in small amino acids and have an N-terminus with a net charge ≤ 0. Notably, the TMHs of > 90 % of all group 1 proteins comprise at least one (small)xxx(small) motif (where “small” stands for glycine, alanine, or serine and “x” for any amino acid) which can promote oligomerization of single-spanning transmembrane proteins (Teese and Langosch, 2015). Furthermore, many of these putative group 1 short import candidates are predicted to have antimicrobial activity and/or pore-lining residues (Tab. S5). Further 192 and 28 translated transcripts contain the conserved motifs of group 2 or 3, respectively. Importantly, all MS-identified members of these extended protein groups were identified in chromatophore-derived samples in this and the previous analysis.

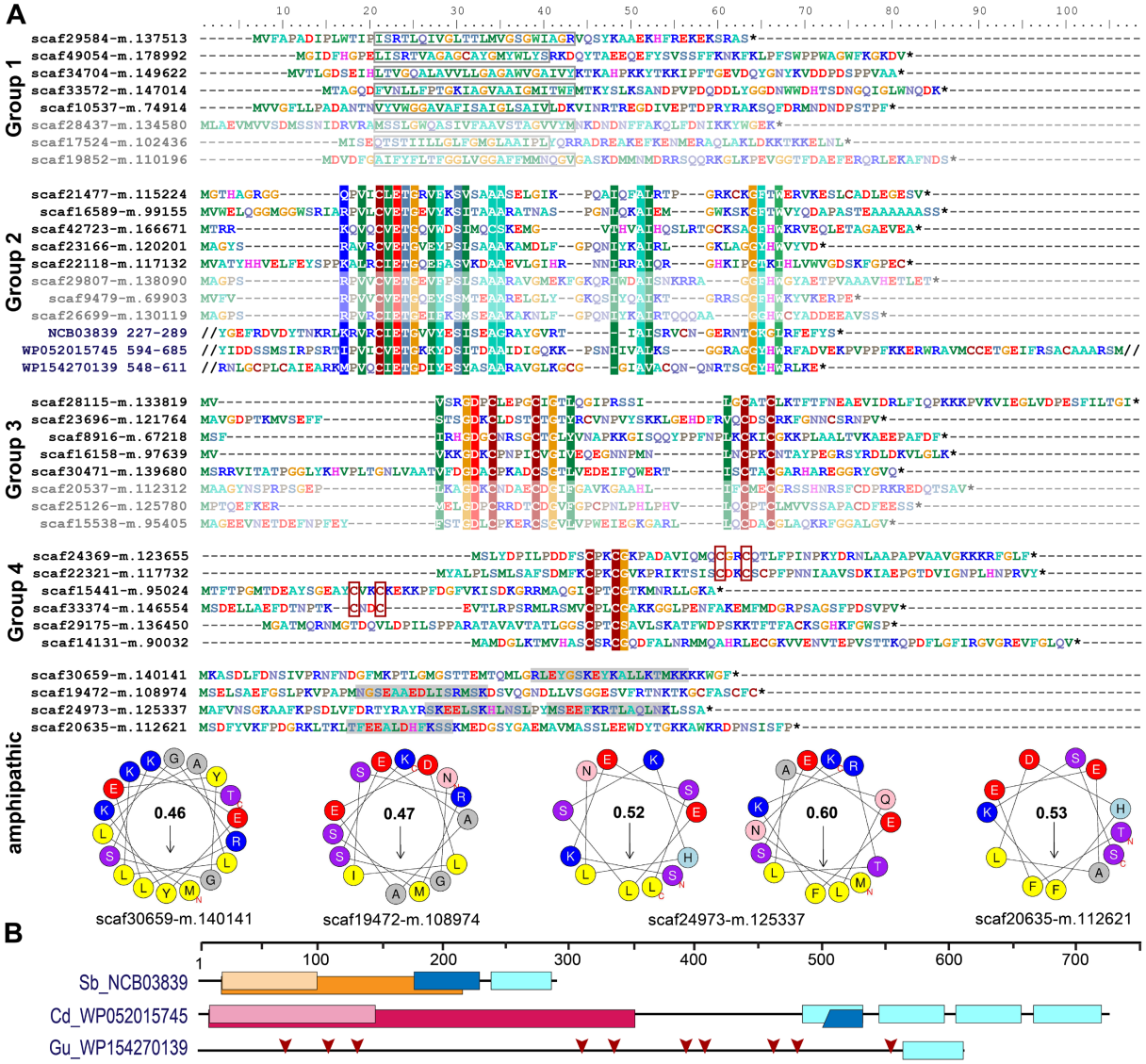


Figure 2.2-10 Short orphan import candidates form distinct groups. A For each group, representative MS-identified proteins (bright colors) and, if applicable, similar proteins identified among translated nuclear transcripts (pale colors) are displayed. Group 1: boxes indicate position of the predicted TMH. Group 2-4: colored background indicates ≥70 % amino acid identity over alignments containing all MS-identified proteins of the respective group. The conserved sequence motif in group 2 was identified in diverse bacterial proteins (three examples with their NCBI accession number and amino acid positions are provided). Group 4: CxxC motifs are highlighted. Amphipathic: some short import candidates that do not belong to group 1 to 4 feature amphipathic helices. Corresponding wheel diagrams and hydrophobic moments of sequences containing at least 70 % residues with an α-helix-probability > 50 % according to NetSurfP (highlighted in grey) are provided below. **B** Domain structure of bacterial proteins shown in A. Light blue boxes, conserved group 2 sequence motif; orange, group I intron endonuclease domain; light orange, GIY-YIG excision nuclease domain; pink, superfamily II DNA or RNA helicase domain (SSL2); light pink, DEXH-box helicase domain of DEAD-like helicase restriction enzyme family proteins; blue, DNA-binding motif found in homing endonucleases and related proteins (NUMOD); red arrows, individual CxxC motifs. Sb, *Spirochaetia bacterium*; Cd, *Clostridioides difficile*; Gu, *Gordonibacter urolithinfaciens*.

2.3 Identification of a Novel Class of Nuclear Encoded Putative Regulators for Chromatophore Gene Expression

In this study, the chromatophore insoluble proteome was investigated not only to identify hydrophobic membrane proteins, but also nuclear encoded proteins involved in chromatophore gene expression regulation. Such proteins can be enriched by ultracentrifugation due to their potential binding to DNA or RNAs in high molecular weight polysome structures. Of the 235 import candidates (excluding the false positive, see 2.2.3.4) identified in this study via mass spectrometry, 76 proteins represent new import candidates (Fig. 2.3-8B). Indeed, a particularly large number of newly MS-identified import candidates (24 proteins) fall into the category "genetic information processing" (Fig. 2.3.-1). Among these proteins an expanded group of 10 RNA-binding or RAP domain-containing proteins (where RAP stands for RNA binding domain abundant in APIcomplexans, Lee and Hong, 2004) stood out.

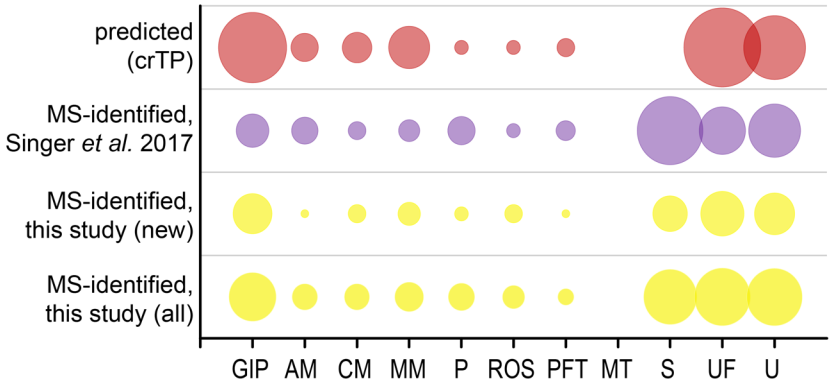


Figure 2.3-1 Functional categories of import candidates. Functions of import candidates identified in this study (yellow), previously MS-identified import candidates (Singer *et al.,* 2017; purple), and *in silico* predicted import candidates (crTP predicted in Singer *et al.,* 2017; red). "New" import candidates were MS-identified in this study, but not in Singer *et al.,* 2017. GIP, genetic information processing; AM, amino acid metabolism; CM, carbohydrate metabolism; MM, miscellaneous metabolism; P, photosynthesis and light protection; ROS, response to oxidative stress; PFT, protein folding and transport; MT, metabolite transport; S, short proteins (< 95 aa) without functional annotation/homologs; UF, unspecific function; U, unknown function.

2.3.1 An Expanded Family of Octotrico Peptide Repeat Putative Expression Regulators is Targeted to the Chromatophore

These RNA-binding proteins encompass, in addition to the crTP, from N- to C-terminus a variable region of 0-320 aa followed by a ~105 aa long conserved region (CR1), and 2-13 repeats of a degenerate 38 aa motif with the most conserved residues being xxxPxxxxLxxxxxxxxxxxxxFxxQxxxxxLNAxAKL. The repeat stretch is often followed by a 110 aa long conserved region (CR2) and the 60 aa long RAP domain (Fig. 2.3-2). This domain organization resembles the one of organelle-targeted Octotrico Peptide Repeat

(OPR; i.e. 38 aa peptide repeat) gene expression regulators in green algae and plants (Fig. 2.3-2B, D) and repeat-containing T3SS effector proteins described from symbiotic or pathogenic bacteria (Fig. 2.3-2B, E, F). All repeat motifs share the prediction to fold into two antiparallel α-helices. 3D-structure prediction suggests folding of the α-helical repeats in *P. chromatophora* OPR proteins into a super helix (or α-solenoid) structure (Fig. 2.3-2G).

Screening the complete *P. chromatophora* transcriptome identified OPR proteins as part of an expanded protein family containing at least 101 members with 1-13 individual OPR motifs (Tab. S6). Besides the 12 chromatophore-localized OPR proteins identified by MS (Fig. 2.3-2A), further 12 OPR proteins for which full-length N-terminal sequence information was available were identified in the transcriptome. Seven of these proteins contained a crTP (Fig. 2.3-2A), the remaining five a mitochondrial targeting signal.

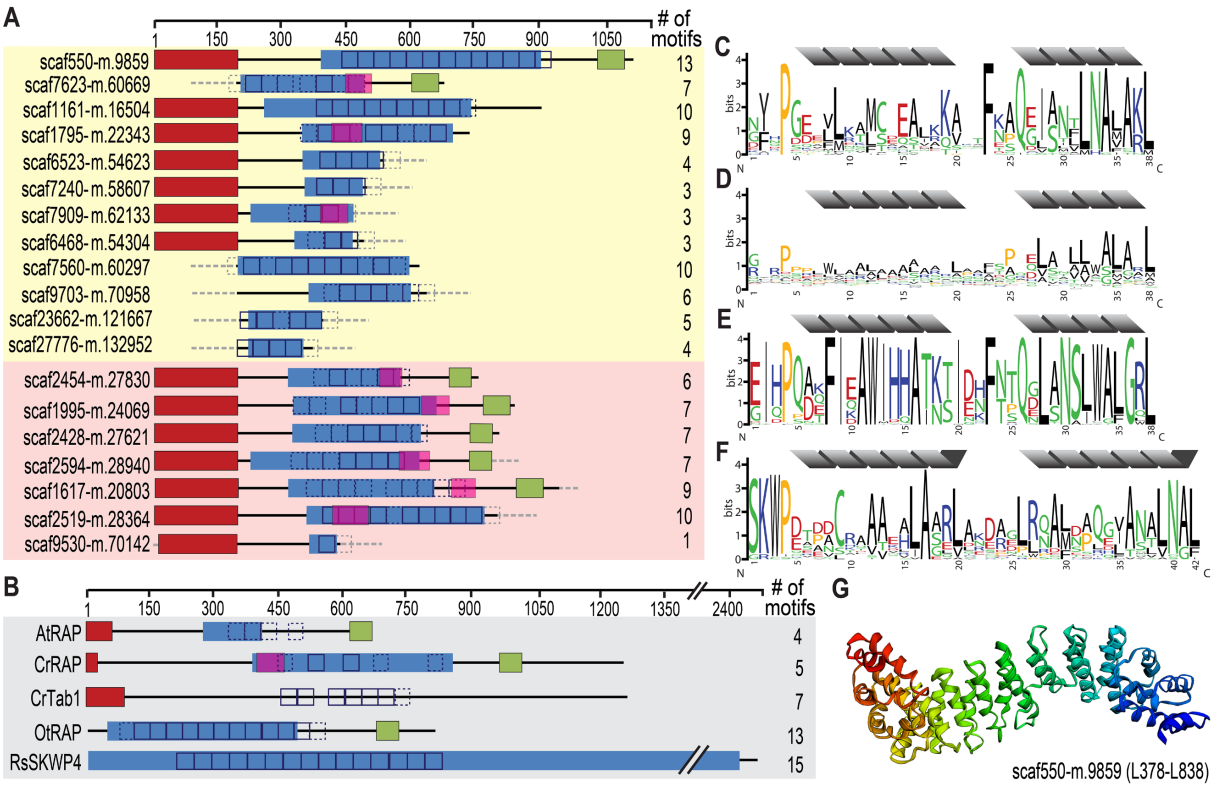


Figure 2.3-2 Identification of an expanded family of putative OPR expression regulators targeted to the chromatophore (and mitochondrion) in *P. chromatophora.* **A** Domain structure of 12 OPR-containing import candidates identified by MS (yellow background) and further 7 predicted import candidates with a similar domain structure (red background). The number of motif repeats identified in individual proteins is indicated. **B** Domain structure and motif repeats in (putative) expression regulators from other organisms. AtRAP, *Arabidopsis thaliana* RAP domain-containing protein, NP_850176.1, (Kleinknecht *et al.*, 2014); CrTab1, *Chlamydomonas reinhardtii* PsaB expression regulator, ADY68544.1, (Rahire *et al.*, 2012). OtRAP, *Orientia tsutsugamushi* uncharacterized RAP domain-containing protein, KJV97331.1, and RsSKWP4, *Ralstonia soleraceum* RipS4-family effector, AXW63421.1, (Mukaihara and Tamura, 2009) appear as the highest scoring BlastP/DELTA Blast hits (in the NCBI nr database) for *P. chromatophora* OPR proteins. **C** 38-aa repetitive motif found in *P. chromatophora* import candidates. **D** OPR motif found in *Chlamydomonas* expression regulators (designed according to Cline *et al.*, 2017). **E** Motif derived from *O. tsutsugamushi* OPR proteins. **F** 42-aa SKWP motif derived from RipS-family effectors in *R. soleraceum, Xanthomonas euvesicatoria,* and *Mesorhizobium loti* (Mukaihara and Tamura, 2009; Okazaki *et al.*, 2008; Teper *et al.*, 2016). Individual repeats are predicted to fold into two α-helices (grey). Red, targeting signal (crTP for *P. chromatophora* proteins, cTP for AtRAP and CrTab1, mTP for CrRAP); blue, PRK09169-multidomain (Pssm-ID 236394); pink, FAST-kinase like domain (Pssm-ID 310980); green, RAP domain (Pssm-ID 312021); boxes, individual repeats of the motifs shown in C-F ($p<e^{-20}$; $p<e^{-10}$ for CrRAP and CrTab1); dashed boxes, weak motif repeats ($p<e^{-10}$; $p<e^{-7}$ for CrRAP and CrTab1); grey dashed boxes/lines, sequence information incomplete. **G** Predicted 3D-structure of the OPR-containing region in scaffold550-m.9859.

3 Discussion

3.1 Insights in Processing of Protein N-termini in the Chromatophore and Consequences for Protein Import

This work, for the first time, sheds light on N-terminal processing and modification of chromatophore-encoded and chromatophore-targeted proteins. The HUNTER method allowed to gain an overview of some common N-terminal modifications including N-terminal methionine excision and N-terminal acetylation. Also, this work delivers the first clues on the fate of the crTP upon import into the chromatophore and enables the proposal of an improved working model for crTP-mediated protein import.

3.1.1 Characteristics of the Chromatophore-encoded N-terminome

N-termini for 129 chromatophore-encoded proteins, representing 15 % of the predicted chromatophore-encoded proteome, could be identified (Fig. 2.1-1, Tab. S1). Most of these proteins (at least 60 %) have been shown here to be present as canonical proteoforms, with around 20 % of all proteins (or ~30 % of canonical proteoforms) possessing their initial methionine (iMet) and around 40 % of all proteins (or ~70 % of canonical proteoforms) being targets of N-terminal methionine excision by Met-aminopeptidase(s) (Tab. S1). In chromatophores, the iMet is preferentially removed when serine, alanine, threonine, or valine are the penultimate residues (Fig. 2.1-2B). The chromatophore-encoded protein PCC0019 might be responsible for iMet excision, as it shows 62 % similarity to the Met-aminopeptidase MatC from *Synechocystis* sp. 6803 (sll0555; Atanassova *et al.,* 2003; Drath, Baier, and Forchhammer 2009). Early studies on *Synechocystis* sp. 6803 used 2D-electrophoresis and N-terminal sequencing to get an insight into the cyanobacterial N-terminome. The analysis revealed, that 80 % of all proteins have canonical N-termini, with 50 % of all proteins (or 62 % of canonical proteoforms) losing the iMet, preferentially when the penultimate amino acid is either serine, alanine or threonine (Sazuka *et al.*, 1999). A meta-analysis of available prokaryotic proteomes revealed that in the cyanobacterial species analyzed even 63-70 % of canonical protein N-termini are subjected to iMet excision, preferentially when serine, alanine, threonine, or – to a minor extent – proline and valine represent the penultimate residues (Bonissone *et al.*, 2013). In fully integrated plant plastids, approx. 90 % of the few remaining organelle-encoded proteins have been reported to possess canonical N-termini, with > 40 % of all proteins being targets of N-terminal methionine excision (especially when penultimate residues are alanine, serine, threonine, valine, or glycine), while approx. 30 % of all proteins retain the iMet (Giglione *et al.*, 2004; 2015; Rowland *et al.*, 2015). Thus, based on the current knowledge, the frequency and specificity of iMet excision is still comparable in free-living

cyanobacteria and integrated chloroplasts, keeping in mind that different methods were used for data acquisition in the examples described here. Accordingly, it is not surprising that a similar situation is found in chromatophores. N-terminal methionine excision is an evolutionary conserved co-translational process generally restricted to proteins that possess small, uncharged amino acids (alanine, serine, threonine, valine, proline, cysteine, or glycine) in their penultimate positions (Varland *et al.*, 2015). However, the preferences vary between organisms (Bonissone *et al.*, 2013). Functionally, iMet excision might be linking identity of the protein N-terminus to its *in vivo* stability it what is known as the "N-end rule pathway" (Dissmeyer *et al.*, 2018). This conserved protein degradation process is based on the recognition of N-terminal residues with the abovementioned residues representing typical stabilizing N-termini. In recent years, regulation of protein stability via the N-end rule pathway has emerged as an important regulator of developmental processes and responses to environmental cues in plants and existence of a prokaryote-type pathway in chloroplasts has been suggested (Dissmeyer *et al.*, 2018; Rowland *et al.*, 2015). Further experimentation will be needed to determine to what extent also chromatophore proteostasis is governed by an N-end rule.

The data presented here also reveal that canonical N-termini exposing methionine, threonine, alanine, and valine are frequently acetylated on chromatophore-encoded proteins (Fig. A6.2-2A). Altogether, approx. half of the canonical N-termini have been identified as an acetylated form (Fig. 2.1-2A). In cyanobacteria however, N-terminal threonine, valine, serine, and alanine residues might become acetylated following iMet excision (Bonissone *et al.*, 2013), but the overall frequency of N-terminal acetylation is generally low in bacteria (< 5 %; Schmidt *et al.,* 2016; Soppa 2010; Yang *et al.,* 2014; Kouyianou *et al.,* 2012). In contrast, many of the few plastid-encoded proteins appear to be acetylated (Giglione and Meinnel, 2001; Huesgen *et al.*, 2013; Rowland *et al.*, 2015). Hence, the frequency of N-acetylation in the chromatophore seems to be more comparable to plastids, indicating involvement of host-derived factors in this process. The functional consequences of N-terminal acetylation are context-dependent and include a function as either a stabilizing modification or a conditional degradation signal that is recognized by factors of the abovementioned N-end rule pathway (Dissmeyer *et al.*, 2018). Additionally, N-acetylation seems to affect protein folding or aggregation, influence a proteins subcellular localization, and enhance protein-protein interactions (Linster and Wirtz, 2018; Varland *et al.*, 2015). Thus, N-terminal acetylation might be another factor involved in organelle integration, e.g. by adjusting organelle function to the physiological state of the cell, as was shown for plant ATP-synthase subunit Ɛ (Hoshiyasu *et al.*, 2013). In chromatophores, acetylation seems to have a stabilizing effect on proteins possessing N-terminal valine or isoleucine and a destabilizing effect on proteins showing methionine, serine, glycine or leucine at their N-termini, as judged from intensities of the corresponding N-terminal peptides (Fig. A6.2-2). However, a chromatophore-encoded N-acetyltransferase could not be identified and orthologs of the common eukaryotic

ribosomal N-acetyltransferases (Linster and Wirtz, 2018) do not seem to be imported into chromatophores, i.e. they were identified only in whole cell lysate in proteomic studies (NatA, NatC; Oberleitner *et al.*, 2020; Singer *et al.*, 2017) and do not possess a crTP (NatA, NatB, NatC, NatD, NatE). In *Arabidopsis* chloroplasts the nuclear encoded N-acetyltransferase NatG has become specialized for post-translational N-terminal acetylation of nuclear encoded but also plastid-encoded proteins (Dinh *et al.*, 2015). Another nuclear-encoded, but co-translationally acting plastidial N-acetyltransferase was discovered only recently in *Chlamydomonas* (Westrich *et al.*, forthcoming). Notably, in contrast to the situation in cyanobacteria (Bonissone *et al.*, 2013), serine is only rarely acetylated on chromatophore-encoded proteins (Fig. A6.2-2). Knowing that N-acetyltransferases usually possess individual substrate specificities (Linster and Wirtz, 2018), this might point to activity of an as yet unknown N-acetyltransferase of non-cyanobacterial origin in chromatophores.

Notably, for one third of the chromatophore-encoded proteins for which a canonical N-terminus has been identified, additional non-canonical N-terminal peptides have been detected and 36 % of all proteins do not seem to be present as their canonical forms at all. As they have lower intensities (Fig. A6.2-2), are less frequently acetylated (Fig. 2.1-2) and (more often) possess potentially destabilizing N-terminal residues (Fig. A6.2-2), many of these N-terminal peptides might correspond to break down products of the respective proteins. Others might result from alternative translational start sites and incorrectly predicted protein models or stem from proteins that actually possess canonical N-termini that could not be detected due to unfavorable sequence features (see 2.1.2). However, they may as well represent N-termini of functionally relevant alternative proteoforms. Few N-termini are regarded clearly a result of signal peptide cleavage, likely related to Sec-dependent thylakoid import (Fig. A6.2-2, Tab. S1). A chromatophore-encoded signal peptidase (PCC0690) that shows 42 % similarity to LepB1 from *Synechocystis* sp. 6803 (sll0716) might be responsible. Notably, the rate of signal cleavage appears low when compared to *Synechocystis* sp. 6803, where almost 20 % of all proteins have been found to be processed upon import into thylakoids (Sazuka *et al.*, 1999). However, it has to be considered that cleavage of targeting sequences might not be predicted properly using the available bioinformatic tools (TargetP, SignalP), as reliability of predictions is expected to decrease when applied to a non-model organism. Thus, the significance of most chromatophore-encoded proteoforms identified in this study currently remains unclear. Importantly, comparably high rates of non-canonical proteolytic products have been reported when comparable methods (i.e. Terminal Amine Isotopic Labeling of Substrates [TAILS] or COmbined FRActional DIagonal Chromatography [COFRADIC]) were used for determination of protein N-termini in plastids (38 %, Huesgen *et al.,* 2013; 40 %, Rowland *et al.,* 2015) or photosynthetic bacteria (35-60 % depending on cell fraction, Kouyianou *et al.,* 2012). As observed in chromatophores (Fig. 2.1-2), many of the cleavages generating such N-termini occur C-terminal to an arginine (or asparagine) residue and create N-termini starting with threonine or serine

(Berry *et al.*, 2017; Kouyianou *et al.*, 2012; Rowland *et al.*, 2015). However, albeit these unexplained non-canonical N-termini account for a high percentage of all N-termini identified in many N-terminome datasets, they are often classified as unknown proteolytic cleavages or degraded proteins and are excluded from further analysis, thus going unreported (Berry *et al.*, 2016). Nevertheless, different mechanisms for the generation of functional proteoforms from one pre-protein are already known (reviewed in Lange and Overall, 2013), e.g. N-terminal trimming of one or several amino acids by exo- or endo-peptidases can form a new protein species with altered function; zymogens require cleavage of a pro-peptide to be activated; cleavage of proteins can generate two shorter stable protein species, both able to exert potentially different activity. Similar mechanisms remain to be discovered experimentally in the future in chromatophores.

Taken together the results obtained in this study render a potential role of host-derived post-translationally acting factors in dynamic regulation of the chromatophore-encoded chromatophore proteome conceivable. Especially N-terminal acetylation might provide a control mechanism universally present in organelles.

3.1.2 N-terminal Processing of the crTP and Proposed Working Model for Protein Import

29 of the 35 crTP proteins for which N-terminal peptides could be identified are processed at the non-alignable region following the N-terminal hydrophobic α-helix, thereby removing it from mature chromatophore proteins (Tab. S1, Fig. 2.1-3A, D). This finding is rather unexpected, as cleavage of the complete targeting peptide is usually required for protein function in organelles (Richter *et al.*, 2005). Amino acid composition around the cleavage site shares some features – like charge distribution and prevalence of serine and alanine in the generated N-termini – with the cTP cleavage sites in many nuclear encoded chloroplast proteins in *Arabidopsis* (Fig. 2.1-3B,C; Rowland *et al.*, 2015), but also in cyanelles of the glaucophyte alga *Cyanophora paradoxa* (Köhler *et al.*, 2015) or in the diatom *Thalassiosira pseudonana* that harbors complex plastids (Huesgen *et al.*, 2013). However, presence of phenylalanine in the P1 as well as the P1´position seems to be a feature exclusive to crTP cleavage. Anyway, targeting peptide cleavage sites in plastids and mitochondria are generally only loosely conserved and the respective processing peptidases rather recognize charge distributions or structural properties at vicinity of the actual sites. The same might be true for the chromatophore processing peptidase(s). It also has to be considered, that a possible crTP peptidase recognition/cleavage motif might be obscured by additional processing steps, e.g. to generate stable N-termini. However, cleavage of cTPs on thousands of different proteins is achieved by a single nuclear encoded M16-type metallopeptidase (Stromal Processing Peptidase, SPP) in plants (Richter and Lamppa, 1998). It seems to be required even for the import process and removal of cTPs is essential for protein function and

plant survival (Zhong *et al.*, 2003). However, putative M16-type processing peptidase with similarity to SPP or related cyanobacterial proteins (Richter *et al.*, 2005) could not be identified upon chromatophore-encoded proteins or import candidates.

At the very N-terminus of crTP proteins, two short motifs (YxxΦ and [DE]LxxPLL) have been discovered (Fig. 2.1-3D) that potentially function as sorting signals for cargo selection into coated vesicles at the *trans* Golgi (Park and Guo, 2014). As the motifs are small, they might occur just by chance and not possess any functional role. However, their localization at the extreme N-terminus (the standard localization for such sorting signals), as well as the fact that those three proteins that do not possess a YxxΦ motif show a [DE]LxxPLL dileucine motif instead, points to functional relevance of the motifs. Notably, the dileucine motif found here differs from the typical motif [DE]xxxL[LI] by one extra residue between the acidic residue and the leucines. However, functionality of comparable motif variants (e.g. EAAAALL or EAAAAPLL) has been proven in human cells (Kozik *et al.*, 2010).

Thus, based on the experimental data generated in this and earlier studies (Nowack and Grossman, 2012; Singer *et al.*, 2017) I propose a working model for crTP-mediated protein trafficking via the *trans* Golgi network (Fig. 3.1-1). Vesicle fusion with the chromatophore outer envelope membrane releases proteins into the inter membrane space, where crTP cleavage might be facilitated by specific peptidases. Several different candidate mechanisms might facilitate passage of proteins across the chromatophore inner envelope membrane.

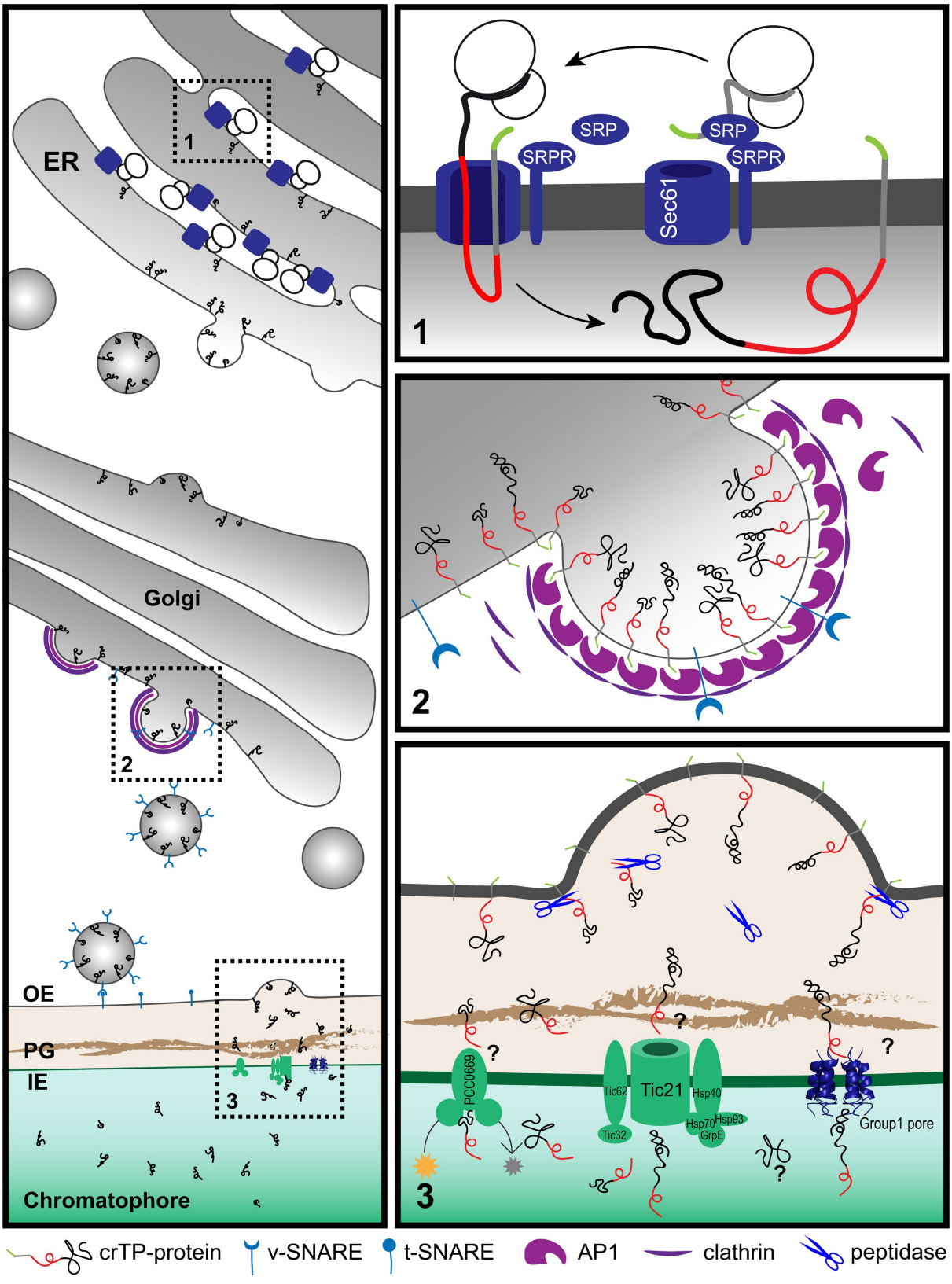


Figure 3.1-1 Working model for crTP-mediated import of long proteins. In a first step, co-translational import of crTP-proteins into the ER lumen via the SRP-system (1) might be triggered by the N-terminal hydrophobic α-helix (grey) in their crTPs. The helix escapes cleavage by signal peptidases, but instead anchors proteins to the membrane while travelling to the *trans* Golgi. High concentration of crTP-proteins at the *trans* Golgi membrane encourages vesicle budding: The putative N-terminal sorting signals YxxΦ and [DE]LxxPLL (green) are recognized

by the clathrin adaptor protein complex AP1, which recruits clathrin units (and numerous other factors) to the membrane initiating vesicle formation (2). Along with the cargo proteins, also dedicated v-SNAREs are incorporated into the vesicle membrane. The clathrin coat is lost once the vesicle pinched off the Golgi. As soon as a vesicle approaches the chromatophore, v-SNAREs bind to their t-SNARE counterparts that decorate the host-derived chromatophore outer membrane. Endocytosis is initiated and cargo crTP-proteins enter the chromatophore intermembrane space (3). A specific endonuclease cleaves the crTPs, thereby releasing soluble proteins. The C-terminal regions of the crTP (red) might facilitate transport across the chromatophore inner membrane via a highly-expressed chromatophore-encoded ABC-transporter (PCC0669, left; see 2.2.3.3), a simplified TIC-like translocon (middle; Gagat and Mackiewicz 2014), or oligomeric pores formed by host-derived group 1 short proteins (right; see 2.2.4).

However, the proposed model definitely requires experimental verification and still leaves a lot of open questions. One concerns the function of the sequence-conserved hydrophobic helix. The helix might trigger co-translational ER import. However, it is not recognized as a signal peptide by common prediction algorithms and must escape cleavage upon ER import, as the motif N-terminal to the helix – according to the model – is required for vesicle formation in the Golgi. Thus, the helix has to be recognized by Signal Recognition Particle (SRP), but at the same time has to escape cleavage by signal peptidase. Based on the experimental data it is currently not possible to tell when exactly cleavage happens. Further questions arise regarding protein sorting at the *trans* Golgi. Do crTP proteins form clusters at specific sites in the Golgi membrane, triggered by interaction of the conserved helix with specific lipid species and/or proteinaceous factors? Do the helices of individual crTP proteins interact? Or are the C-terminal domains of the crTP involved in such interaction processes? Do crTPs interact with co-travelling short imported proteins? How does presence of crTP-proteins recruit chromatophore-specific v-SNAREs to the budding vesicle?

Following vesicle transport to the chromatophore and vesicle fusion with the chromatophore outer membrane, the crTP might be cleaved C-terminal to the hydrophobic helix by an unknown, periplasm-localized processing peptidase to release individual soluble proteins from the membrane. The conserved C-terminal fraction of the crTP might facilitate protein transport across the inner chromatophore membrane. It might be recognized by a chromatophore-encoded ABC-transporter, e.g. PCC0669 that is present in considerably large amounts in the chromatophore (see Fig. 2.2-7) and shares similarity with bacterial peptide importers (Tab. S2; Domenech *et al.,* 2009; Guefrachi *et al.,* 2015). Also the existence of a simplified TIC-like translocon in the chromatophore has been proposed (Gagat and Mackiewicz, 2014), but function of Tic21 as a protein conducting channel has been impeached (Duy *et al.*, 2007). A third potential option for proteins to travel across the inner membrane was discovered in this study. Group 1 short imported proteins might be able to form multimer arrangements in barrel-stave (Fig. A6.2-6A), or shortly lived toroidal pores that could allow not only

for the passage of metabolites, but also protein. However, assuming a bottleneck pore radius shown in Fig. A6.2-6B most folded proteins would not be able to pass. In general, the question whether proteins arrive at the chromatophore surface in a folded, partially folded, or unfolded state (e.g. by the action of chaperones) currently has to remain unanswered. Import of folded or partially folded proteins into mitochondria is known to involve unfoldase activity that seems to be driven by membrane potential and the chaperone Hsp70 (Matouschek, 2003). However, in the case of the Tat-translocon that is specialized on transport of folded protein, the protein conducting channel is formed by oligomers of the small single-spanning transmembrane protein TatA. The number of subunits contributing to the channel can vary from 20 to 30 subunits, indicating that the radius is dynamically adjusted to fit the size of the substrate to be transported (Gohlke *et al.*, 2005). The proposed mechanism is still a matter of debate (Hamsanathan and Musser, 2018) but it encourages the conceivability of similar fascinating mechanisms in *P. chromatophora*. Notably, a function of the crTP as an autotransporter (i.e. a protein that facilitates its own translocation across a membrane via formation of a pore by a fraction of the protein) is imaginable. However, autotransporters are currently only known to facilitate secretion across the bacterial outer membrane where the pore is usually formed by beta-barrel tertiary structures (Dautin and Bernstein, 2007). Formation of such structures is not predicted for crTPs. Also, the cargo-fraction of the protein is usually released from the pore-forming domain via proteolytic cleavage following translocation. The data presented here indicate, that the largest fraction of the crTP is usually retained on mature chromatophore proteins. However, the whole conserved crTP was cleaved off in three of the 35 identified crTPs (Fig. 2.1-3D). Despite the low number of examples, the cleavage should be regarded as functionally relevant, as it appears at exactly the same position and this position constitutes the exact C-terminal end of the crTP. One of the affected proteins is annotated as metal-dependent protein hydrolase, while the other two are SDR family NAD(P)-dependent oxidoreductases. One of the oxidoreductases is 36 % similar to light-dependent protochlorophyllide oxidoreductase PorA of *Arabidopsis,* a protein that catalyzes light-triggered reduction of protochlorophyllide to chlorophyllide in thylakoids (Gabruk and Mysliwa-Kurdziel, 2015). However, the reason for altered crTP cleavage in these proteins remains unknown, as no obvious commonalities, i.e. sequence features (that differ from site 1 processed crTPs) or functional rationales could be determined. Also the meaning of unique processing sites elsewhere in the crTP that were found on further three proteins (Fig. 2.1-3D) currently remains entirely unknown.

However, the model for crTP-mediated protein import presented here offers several starting points for the verification of individual steps involved. In a first step, presence of crTP-proteins at the Golgi will have to be verified. Application of protein-specific antibodies could provide an answer. These could be used for Immunogold EM or in combination with cell fractionation experiments, where they could provide clues on the presence of a full-length or a processed crTP in compartments of the host cell,

e.g. by western blotting of Golgi-enriched fractions (Parsons *et al.*, 2012). Also, the presence of glycosylations on crTP-proteins could indirectly prove travelling through the Golgi, as this type of post-translational modification is carried out in this compartment in eukaryotes. Glycosylation might easily be detected by comparing the banding pattern of a crTP-protein after western blotting untreated *P. chromatophora* whole cell lysates and lysates that have been treated with de-glycosylating enzymes (e.g. PNGase F). Finally, electron cryotomography could be used to generate ultrahigh resolution images of the symbiotic interface. The proposed fusion of vesicles with the chromatophore outer membrane might become visible at the magnification and resolution enabled by this method (Medeiros *et al.*, 2018). A potential function of the conserved α-helix in crTP-oligomerization could be investigated via a ToxR oligomerization assay, that couples oligomerization of the transmembrane helix to be analyzed to expression of a reporter gene in *E. coli* (Joce *et al.*, 2011). In general, proteinaceous interaction partners of crTP-proteins (at the Golgi or during translocation across the chromatophore inner membrane) might be MS-identified by selective pull-down of complexes formed by crTP-proteins and their interaction partners. To this end, transient complexes could be stabilized through glutaraldehyde cross-linking and separated from whole cell lysates, again using specific antibodies against crTP-proteins. However, state-of-the-art cross-linking mass spectrometric workflows in combination with advanced software tools today even enable direct identification of system-wide protein-protein interactions without the need for selective pull-down of complexes (Götze *et al.*, 2019; Schweppe *et al.*, 2017). As all crTP-proteins should share the same core set of interaction partners while travelling to the chromatophore stroma, mass spectrometric identification of a protein network specifically interacting with the crTP seems realistic.

Taken together, protein import into chromatophores seems to differ fundamentally from canonical plastid import pathways. However, as outlined above some elements might be present also in higher plants and might partially explain how a crTP could facilitate import of yellow fluorescent protein into plastids of tobacco (Singer *et al.*, 2017). Further investigation of protein import into chromatophores surely will provide precious insights into organellogenesis and uncover those elements that are universal to eukaryotes and those that are unique to photoautotrophic *Paulinella* species.

3.2 The Puzzle of Metabolite Exchange Across the Chromatophore Envelope

In an approach to identify nuclear encoded chromatophore-targeted metabolite transporters candidate single-subunit transporters were tested for their potential to transport metabolites which are highly expected to be shuttled across the chromatophore envelope membranes. In a second approach, the chromatophore insoluble proteome was screened for presence of nuclear encoded transporters. Unexpectedly, both experimental approaches led to the result that – in contrast to the

situation in plastids and mitochondria – nuclear encoded transporters are most likely not inserted into the chromatophore envelope. However, the *P. chromatophora* host cell might use different strategies to exchange metabolites with its nascent organelle, e.g. pore-forming peptides.

3.2.1 Nuclear Encoded Multi-Spanning Transmembrane Proteins are not Imported into the Chromatophore

Despite the obvious need for extensive metabolite exchange between the chromatophore and cytoplasm (Fig. 1.2-2; Valadez-Cano *et al.*, 2017), the chromatophore likely lost on the order of 70 solute transporters following symbiosis establishment (Fig. 2.2-1). The remaining transport systems do not appear apt to establish metabolic connectivity (Fig. 2.2-7E). Solely two systems, a DME family and a DASS family transporter, might be involved in metabolite transport. However, their substrate specificities and kinetics will have to be determined experimentally. Furthermore, there are three ABC-transporters for which substrate specificity is unknown. However, the high energy costs associated with their ATP-consuming primary active mode of transport appears to be incongruous with high-throughput metabolite shuttling. Some of these ABC-transporters might have become specialized for protein import instead. In line with this idea, the ABC-half transporter PCC0669 that showed highest ion intensities among all chromatophore-encoded transporters (Fig. 2.2-7E), possesses 33 % similarity to BclA of *Bradyrhizobium* sp., a nitrogen-fixing bacterium harbored by *Aeschynomene* legumes. It functions as an importer for Nodule-specific Cysteine-Rich (NCR) peptides produced by the host plants symbiotic nodule cells (Guefrachi *et al.*, 2015). Internalization of the peptides could provide protection against their destabilizing effects on the bacterial membranes and as well bring the peptides in proximity to their intracellular targets, thereby inducing differentiation of the bacteria into the non-reproducing, polyploid state typical for that kind of symbioses (Guefrachi *et al.*, 2015). Further members of the same TCDB-family are involved in conferring pathogens resistance to antimicrobial peptides, e.g. in *Mycobacterium tuberculosis* (Domenech *et al.*, 2009). However, since other transporters in the family are involved in peroxisomal transport of fatty acids or fatty acyl-CoA (Linka and Esser, 2012), similar substrates could also be transported by PCC0669.

In plastids, insertion of nuclear encoded transporters into the inner membrane compensates for transporters that were lost from the organellar genome in the course of endosymbiotic genome reduction (Facchinelli and Weber, 2011; Fischer, 2011; Karkar *et al.*, 2015). Markedly limited coding capacities for transporters can be universally observed in obligate intracellular pathogens and endosymbionts, probably because of the static nature of their intracellular environment (Ren and Paulsen, 2005). As a consequence, also in more recently established endosymbiotic associations, such as plant sap-feeding insects with nutritional bacterial endosymbionts, multiplication of host

transporters followed by their recruitment to the host/endosymbiont interface apparently was involved in establishing metabolic connectivity (Duncan *et al.*, 2014; Price *et al.*, 2011). However, these transporters localize to the symbiosomal membrane, a host membrane that surrounds bacterial endosymbionts. The mechanism enabling metabolite transport across the symbiont´s outer and inner membranes, with symbiont-encoded transport systems being scarce, is a longstanding unanswered question (Mergaert *et al.*, 2017). Despite the import of hundreds of soluble proteins into the chromatophore*,* proteomic analyses provided no evidence for the insertion of nuclear encoded transporters into the chromatophore inner membrane (or thylakoids). The possibility that such proteins escaped detection for technical reasons appears improbable because: (i) 72 % of the chromatophore-encoded transporters were identified in CM samples. Assuming comparable abundances for nuclear encoded chromatophore-targeted transporters, a large percentage of these proteins should have been detected, too. (ii) More than 100 nuclear encoded transporters or transporter components were detected in comparable amounts of PM samples showing that the applied method is feasible to detect this group of proteins. (iii) inner membrane transporters were repeatedly identified in comparable analyses of cyanobacterial (Baers *et al.*, 2019; Choi *et al.*, 2019; Liberton *et al.*, 2016; Pisareva *et al.*, 2011; Plohnke *et al.*, 2015) or plastidial membrane fractions (Bouchnak *et al.*, 2019; Bräutigam *et al.*, 2008; Simm *et al.*, 2013).

In a bottom-up approach, transporters were expressed in cyanobacterial, bacterial, and eukaryotic heterologous systems, all of which representing specific metabolite transporter knockout strains and allowing for simple complementation experiments. However, most of the experiments performed delivered negative results, i.e. the mutant phenotype could not be complemented by *P. chromatophora* transporters indicating that the substrate provided to the cells was not imported (Fig. 2.2-2-4). Thus, the substrates predicted according to the TCDB might in fact not represent substrates of these transporters. However, several reasons also might have led to a false negative outcome of the experiments: (i) problems in heterologous expression, e.g. insufficient expression, aberrant folding, incorrect intracellular localization, faulty post-translational modifications, functional interference of N-terminal tags, and negative effects on cell viability as well as (ii) inappropriate experimental conditions, e.g. substrate concentration, availability of ions / appropriate pH for secondary active transport, and unforeseen physiological responses of cells.

Indeed, heterologous expression of functional eukaryotic membrane proteins came across with several well-known pitfalls, each of which requiring careful analyzation and optimization (for review see Dilworth *et al.*, 2018). However, optimization of protein expression reaches beyond the scope of this study. Eventually, glycine was the only metabolite for which a complementation experiment delivered a positive result for three *P. chromatophora* transporters (Fig. 2.2-4C). The putative glycine/alanine

transporter BGATr was chosen for further analysis, as it was shown earlier that BGATr might be the result of an HGT from a γ-proteobacterium to the *P. chromatophora* nuclear genome and such transfers in many cases seem to compensate for gene losses in the chromatophore (Nowack *et al.*, 2016). However, biochemical approaches like HPLC analysis of cell extracts and ^{14}C-glycine uptake experiments failed to proof function of BGATr as a glycine transporter (Fig. 2.2-4D, E). Insufficient expression in combination with the shorter timeframe of these experiments (few minutes to 25 h, compared to 16 days for complementation experiments) might be reasons for that. Taken together, an involvement of BGATr in glycine transport across the chromatophore envelope seems conceivable, although the functional characterization of BGATr (e.g. a possible function in serine transport, Fig. 2.2-4D) surely requires further experimental work that goes beyond the scope of this study. However, specific antibodies were generated in order to evaluate the transporters subcellular localization. Western blot experiments revealed that the antibody was relatively specific to BGATr but had to be used at very high concentrations (ratio antibody : blocking solution < 1:16) to produce a sufficient signal (Fig. 2.2-5). When used on *P. chromatophora* fractions, a strong band for BGATr was detected in whole cell lysate but not in chromatophore lysate rendering presence of BGATr in the chromatophore inner envelope unlikely (Fig. 2.2-5D). However, presence of BGATr in the chromatophore outer envelope cannot be ruled out since the outer membrane is lost during chromatophore isolation (Fig. 2.2-6). Hence, it was attempted to detect BGATr in embedded, sectioned *P. chromatophora* cells via Immunogold electron microscopy (Fig. A6.3-2). Detection of BGATr failed however, probably because of the very low sensitivity of the α-BGATr antibody. The exact subcellular localization of BGATr thus remains unknown.

Taken together, the combination of the top-down and bottom-up approaches implies that nuclear encoded transporters are not present at the chromatophore inner membrane. Thus, a general mechanism to insert nuclear encoded multi-spanning transmembrane proteins into chromatophore inner membrane and thylakoids likely has not evolved (yet) in *P. chromatophora* (although a few such proteins might insert spontaneously based on their individual physicochemical properties). It was often suggested, that post-translational migration of highly hydrophobic membrane proteins through the aqueous cytosol might be a challenging task. A cell would either have to introduce mutations that reduce overall hydrophobicity in transmembrane regions or develop factors that prevent hydrophobic proteins from aggregation or mistargeting to the ER (Adams and Palmer, 2003; Oh and Hwang, 2015; Popot and De Vitry, 1990). Thus, import of soluble protein might be more straight-forward to evolve and is established already in an earlier stage of organellogenesis than import of hydrophobic proteins. In line with this, selective retention of membrane protein-encoding genes in the genome of an endosymbiont while soluble protein is imported can be observed in the mealybug/*Moranella* symbiosis for genes involved in peptidoglycan synthesis (Bublitz *et al.*, 2019). The intracellular bacterium

Moranella lost several genes responsible for cytosolic steps of peptidoglycan synthesis that seem to be complemented by bacterial HGTs in the mealybug genome, while those genes encoding periplasmic or membrane-localized proteins are retained in *Moranella*. Presence of a mealybug-encoded protein in the endosymbiont has been verified, however, it is still unclear whether protein or mRNA is transported in this endosymbiosis. Yet, protein import is likely and is as well supposed to be generally restricted to soluble protein.

The findings presented here spotlight the puzzling absence of suitable transporters that would allow metabolite exchange across the chromatophore inner membrane. The protein composition of the chromatophore outer membrane is currently unclear. Its putative host origin and the notion that proteins traffic into the chromatophore likely via the Golgi (Nowack and Grossman, 2012) suggest that nuclear encoded transporters could be targeted to the outer membrane by vesicle fusion. Nonetheless, the conservation of active and secondary active inner membrane transporters on the chromatophore genome (Tab. 2.2-1, Fig. 2.2-7E) strongly implies that the chromatophore inner membrane kept its barrier function and there is an electrochemical gradient across this lipid bilayer, calling for defined transport processes. Thus, metabolic flux between host cell and chromatophore might exploit different mechanisms that are independent of classical membrane transporters.

3.2.2 Antimicrobial Peptides as a Possible Alternative to Transporter Mediated Metabolite Shuttling

In contrast to the absence of multi-spanning transmembrane proteins, numerous short single-spanning transmembrane and AMP-like orphan proteins were identified among chromatophore-targeted proteins. These short import candidates fall into at least four expanded groups, suggesting some degree of functional specialization. Interestingly, expanded arsenals of symbiont-targeted polypeptides convergently evolved in many taxonomically unrelated symbiotic associations and thus seem to represent a powerful strategy to establish host control over bacterial endosymbionts (Mergaert, 2018). It has been suggested that these “symbiotic AMPs” have the ability to self-translocate across or self-insert into endosymbiont membranes and mediate control over various biological processes in the symbionts including translation and septum formation (Farkas *et al.*, 2014; Login *et al.*, 2011; Mergaert, 2018; Van de Velde *et al.*, 2010). Apart from that, also a function of symbiotic AMPs in modulation of membrane permeability and metabolite exchange is discussed (Mergaert *et al.*, 2006; 2017). It was shown, that an AMP (Ag5) is produced in the Alder tree root nodules to facilitate the release of nitrogen-rich amino acids from its bacterial nitrogen-fixing endosymbiont *Frankia alni* (Carro *et al.*, 2015). The peptide triggers the release specifically of glutamine and glutamate by inducing membrane porosity when *Frankia* cells are treated with

concentrations < 1 µM. Higher concentrations harm and ultimately kill the bacterium. The results suggest that – at appropriate concentrations – this and probably also other AMPs act as host effectors that are plugged into the membrane of endosymbionts to form specific channels permitting the release of nutrients by diffusion along the concentration gradient. Symbiotic AMPs might in this case promote only a temporal membrane disruption, thereby allowing the maintenance of a proton motive force and thus survival of the endosymbiont (Wimley, 2010). A similar scenario would be conceivable also in *P. chromatophora*, where at least a fraction of the putative AMPs targeted to the chromatophore could bind to or insert into its envelope.

The discovery of TMH-containing group 1 proteins appears to be of particular interest in the context of metabolite exchange. The frequent occurrence of (small)xxx(small) motifs might indicate the potential of these proteins to oligomerize by allowing for close proximity between interacting TMHs. Such associations are known to be stabilized by interfacial van der Waals interactions and/or hydrogen bonding resulting from the excellent geometric fit between the interacting TMHs (Moore *et al.*, 2008; Teese and Langosch, 2015). The predicted pore-lining residues (Tab. S5) in the TMHs of many of these proteins further suggest that they could form homo- or hetero-oligomeric channels. It has been previously reported that AMPs can arrange in channel-like assemblies which facilitate diffusion along concentration gradients (Rahaman and Lazaridis, 2014; Wang *et al.*, 2016), though the lifetime and selectivity of such arrangements requires further investigation. Given the size of the metabolites to be transported, they would be required to form multimer arrangements in barrel-stave (Fig. A6.3-6), or shortly lived toroidal pores, while maintaining the overall impermeability of the membrane. The formation of such pores still begs the question of how they could maintain a selective metabolite transport. An interesting example in that respect is the VDAC channel of the mitochondrial outer membrane which has been described to follow a stochastic gating mechanism, in which bigger, hence slowly diffusing molecules have higher probability to permeate, while passage of small, fast diffusing molecules (ions) is minimized (Berezhkovskii and Bezrukov, 2018). This selectivity mechanism requires a channel to be fast-gating and predominantly in closed conformation.

An alternative mode of action involves soluble, short import candidates which could interact with the chromatophore envelope membranes via stretches of positively charged amino acids and amphipathic helices (Fig. 2.2-10A), and putatively modulate membrane permeability (Mergaert *et al.*, 2017) in what is known as carpet model (Wimley, 2010). The mechanism by which such an interaction could cause a transient permeabilization is still a matter of debate, although the asymmetric distribution of peptides on the membrane bilayer has been pointed out as plausible reason (Guha *et al.*, 2019). This asymmetric distribution creates an imbalance of mass, charge, surface tension, and lateral pressure. A combination of these factors is hypothesized to lead to stochastic local dissipation events relieving asymmetry by

peptide, and possibly lipid, translocation and concomitantly inducing transient permeability to polar molecules. Further experimental work with the identified proteins could shed light on the potential transport mechanism.

Other short import candidates might also attack targets inside of the chromatophore (e.g. DNA, specific RNA species, the replication or translation machineries). The group 2 sequence motif is found also in hypothetical bacterial proteins which include domains related to DNA processing functions (Fig. 2.2-10B). Thus, group 2 proteins might provide the host with control over aspects of genetic information processing in the chromatophore. The presence of dozens to hundreds of similar proteins in the various groups, points to a functional interdependence or reciprocal control of individual peptides. Indeed, co-occurring AMPs have been shown to synergize in insects, e.g. some AMPs permeabilize membranes to enable entry of other AMPs that have intracellular targets (Rahnamaeian *et al.*, 2015).

However, the ability of group 1 proteins and/or proteins harboring amphipathic helices to form pores will have to be addressed experimentally. Therefore, candidate proteins could be expressed heterologously and purified. Availability of purified recombinant protein enables several options for *in vitro* analysis of their function. For example, their ability to permeabilize biological membranes could be tested in giant unilaminar vesicles of defined lipid compositions (i.e. ideally resembling the chromatophore envelope membranes). A membrane impermeable fluorescent dye (e.g. carboxyfluorescein) can be encapsulated into the vesicles and the release of the dye depending on the concentration of externally added protein could be observed under a confocal fluorescence microscope (Ambroggio *et al.*, 2005). Ideally, a defined structure could be obtained for pores formed by some of the short proteins, e.g. via X-ray crystallography or cryo-electron microscopy, enabling advanced insights into their mode of action and transport potential (Avci *et al.*, 2018). However, the potential of group 1 proteins to oligomerize might at first be addressed without the need to purify recombinant protein by performing blue native PAGE of lysate from *E. coli* cells expressing the protein to be analyzed or applying a ToxR oligomerization assay. The assay couples oligomerization of the peptide or transmembrane helix to be analyzed to expression of a reporter gene (e.g. GFP or β-Galactosidase) in *E. coli* (Joce *et al.*, 2011).

3.3 *P. chromatophora* Octotrico Peptide Repeat Proteins Resemble Ubiquitous Organellar Gene Expression Regulators

Besides the establishment of metabolic connectivity, MS analysis illuminated another cornerstone in organellogenesis, the evolution of nuclear control over organellar gene expression. Previously, a large number of proteins annotated as transcription factors has been identified among chromatophore-targeted proteins (Singer *et al.*, 2017). Here a novel class of chromatophore-targeted helical repeat proteins is described. Helical repeat proteins appear to represent ubiquitous nuclear factors involved in regulation of organellar gene expression (Hammani *et al.*, 2014). These proteins are generally characterized by the presence of degenerate 30-40 aa repeat motifs, each of them containing two antiparallel α-helices. The succession of motifs underpins the formation of a super helix that enables sequence specific binding to nucleic acids.

The *P. chromatophora* nuclear genome encodes at least 101 OPR helical repeat proteins (Tab. S6). OPR proteins have mostly been studied in the green alga *Chlamydomonas*, where 44 OPR genes were identified in the nuclear genome, while that number might be underestimated because of the sequence degeneracy of the motif (Fig. 2.3.1-2D). Almost all of these OPR proteins are predicted to localize to organelles (Eberhard *et al.*, 2011) and five have been shown experimentally to be involved in post-transcriptional steps of chloroplast gene expression, e.g. Tab1 (Fig. 2.3.1-2B) is required for the translation of the *psaB* transcript (Rahire *et al.*, 2012). The only known *Arabidopsis* OPR protein is AtRAP (Kleinknecht *et al.*, 2014; Fig. 2.3.1-2B). Binding of AtRAP to the 5´region of the chloroplast 16S rRNA precursor was shown to be required for rRNA maturation and thus for the assembly of functional ribosomes in the chloroplast. However, another family of helical repeat proteins called Pentatrico Peptide Repeat (PPR, repeats of 35 aa) proteins represents the most prominent family of organelle-targeted helical repeat proteins in land plants. They are universally present in eukaryotes but comprise a particularly large number of around 450 family members in *Arabidopsis* (Lurin *et al.*, 2004). The vast majority of *Arabidopsis* PPR proteins are targeted to either mitochondria or chloroplasts or both organelles (Colcombet *et al.*, 2013), where they have been shown to be involved in RNA-level gene expression regulation, e.g. cleavage, editing, splicing, stability, and translation of target mRNAs (Hammani *et al.*, 2014). However, the *Chlamydomonas* genome does only encode for 14 PPR proteins (Tourasse *et al.*, 2013), indicating that different families of organelle-targeted helical repeat proteins have expanded in different phyla to fulfill similar purposes.

Also the *P. chromatophora* OPR proteins seem to be mostly organelle-targeted. Many of them possess, in addition to the OPR stretches, a Fas-activated serine/threonine (FAST) kinase-like domain (Tian *et al.*, 1995) and a C-terminal RAP domain (Fig. 2.3.1-2A). Interestingly, a FAST kinase-like domain was also found on one third of the aforementioned *Chlamydomonas* OPR proteins and is sometimes

followed by a RAP domain (Eberhard *et al.,* 2011; e.g. CrRAP in Fig. 2.3.1-2B). A similar domain architecture is also present in the *Arabidopsis* OPR protein AtRAP (Kleinknecht *et al.,* 2014; Fig. 2.3.1-2B) and in the FASTK family of vertebrate nuclear encoded regulators of mitochondrial gene expression (Boehm *et al.*, 2016). However, the exact molecular functions of FAST kinase-like and RAP domains as well as the two conserved regions in *P. chromatophora* OPR proteins (CR1 and CR2, Tab. S6) that share no similarity with any known domains remain unknown.

It should be noted, that helical repeat proteins are not only restricted to being regulators of gene expression in organelles of eukaryotes, but include also repeat-containing T3SS effector proteins that were described from several symbiotic as well as pathogenic bacteria. Such effectors are usually secreted into the eukaryotic host cell, e.g. to manipulate its metabolism to convert a favorable microenvironment for the infection process or to modulate host immunity. BLAST-analyses of *P. chromatophora* OPR proteins revealed strong similarity between the OPR-motif and the 42 aa SKWP-motif present in RipS-family effectors (e.g. RsSKWP in Fig. 2.3.1-2B; Fig. 2.3.1-2F) of the plant pathogens *Ralstonia solanacearum* and *Xanthomonas euvesicatoria* (Mukaihara and Tamura, 2009; Peeters *et al.*, 2013; Teper *et al.*, 2016) and also of *Mesorhizobium loti* (Okazaki *et al.*, 2010), a symbiotic nitrogen-fixing bacterium harbored by *Lotus* legumes. An expanded family of OPR proteins showing highly similar motifs even in combination with a RAP domain also seems to be present in the obligate intracellular pathogenic bacterium *Orienta tsutsugamushi*, the causative agent of scrub typhus disease (OtRAP in Fig. 2.3.1-2B; Fig. 2.3.1-2E). All motifs share the presence of a prominent proline as well as accumulation of neutral amino acids (leucine, alanine, asparagine) on the C-terminal end and secondary structure predictions suggest the folding of each motif into two separate α-helices. However, the mode of function of secreted SKWP proteins remains to be investigated and it is also currently unknown whether *O. tsutsugamushi* OPR proteins are indeed secreted.

In conclusion, in parallel to the evolution of mitochondria and plastids, also during chromatophore evolution an expanded family of chromatophore-targeted helical repeat proteins evolved. Based on the similarity of their domain architecture to known organelle-targeted expression regulators, the OPR proteins in *P. chromatophora* likely serve as nuclear factors modulating chromatophore gene expression by direct binding to specific target RNAs. It is thus tempting to speculate that these proteins could have been used before chromatophore acquisition for expression regulation of mitochondrial proteins and that these pre-existing expression regulators were then recruited by crTP acquisition to the chromatophore. However, the RNA-binding ability of *P. chromatophora* OPR proteins, their specific target sequences as well as their ability to modulate expression of chromatophore-encoded proteins remain to be tested experimentally. Transcriptomics approaches might be used to demonstrate an mRNA-level correlation between certain chromatophore-encoded genes and genes encoding OPR

proteins (and other nuclear factors), e.g. in form of a concerted response to different growth conditions including light or nutrient availability. Cycloheximide (an inhibitor of translation specifically on eukaryotic ribosomes) might be used to prohibit production of OPR proteins and by this means uncouple gene expression in the chromatophore from nuclear factors. This would allow to compare short-term mRNA level dynamics in the chromatophore in response to specific growth conditions and in the presence or absence of nuclear gene expression. Alternatively, the CLIP method (Cross Linking and ImmunoPrecipitation) might reveal mRNA targets of selected OPR proteins (Lee and Ule, 2018). The method is based on *in vivo* chemical or photo- crosslinking of protein/RNA complexes. A specific OPR protein can then be immunoprecipitated and the bound RNA can be converted to cDNA and sequenced. However, immunoprecipitation of a specific OPR protein requires antibodies specific for that OPR protein, since tagged protein cannot be expressed in *P. chromatophora*. It might be challenging to generate such specific antibodies, due to the highly similar primary sequence and structure of individual OPR proteins.

3.4 Outlook

This study provided valuable new insights into hallmarks of organellogenesis: Shuttling of protein as well as metabolites across membranes of the symbiotic interface and establishment of host control over organellar gene expression.

(i) A focus of this study concerned the vastly uncharacterized protein import mechanism operating in *P. chromatophora*. Absence of most TOC / TIC homologs and the unique structure of the crTP suggest that protein import into chromatophores differs fundamentally from the mechanisms known from ancient eukaryotic organelles. In this study, the chromatophore N-terminome was investigated by the HUNTER method, revealing two specific processing sites in the crTP and contributing to the development of a working model for protein import into chromatophores. Also, the data for the first time provide insights into N-terminal processing of chromatophore-encoded proteins, showing that N-terminal acetylation in chromatophores is more comparable to N-acetylation in plastids than in cyanobacteria. A universal function of N-acetylation in contributing to the adjustment of organelle function to the physiological state of the cell is conceivable.

(ii) This study further aimed at testing the hypotheses that nuclear encoded transporters are recruited to establish metabolic connectivity between chromatophores and host cell in *P. chromatophora*. Two different experimental approaches provided no evidence for the presence of nuclear encoded metabolite transporters or other nuclear encoded multi-spanning transmembrane proteins in the chromatophore, indicating that a general mechanism to insert nuclear encoded multi-spanning transmembrane proteins into chromatophore membranes likely has not evolved (yet) in *P. chromatophora*. However, several different expanded groups were identified in this analysis among short orphan proteins that have been shown to be chromatophore-targeted earlier. Some of these might be involved in an alternative mechanism for metabolite exchange, providing a fascinating counter model to what is known on metabolite shuttling from ancient organelles. However, the hypotheses proposed here require rigorous experimental testing. The data obtained by careful bioinformatic analysis of chromatophore-encoded transporters and mass spectrometric analysis of the chromatophore insoluble proteome might be of great value for further research on *P. chromatophora*.

(iii) Helical repeat proteins that appear to represent ubiquitous nuclear factors involved in regulation of organellar gene expression were for the first time recognized in the nuclear genome of *P. chromatophora* and detected experimentally in chromatophores. Although the function as gene expression regulators has to be proven experimentally, the presence of structurally similar proteins throughout the kingdom of eukaryotes points to an interesting case of convergent evolution in terms of organelle acquisition.

4 Material and Methods

4.1 Material

4.1.1 Plasmid Vectors

Table 4.1-1A Vector backbones used in this study.

#GPN	Name	Description	Organism(s) [Selection]	Source
GPN019	pTAC-MAT-Tag-2	Vector for IPTG-inducible expression of C-terminal MAT (Metal Affinity Tag) fusion protein under the strong tac promoter (a hybrid of the *E. coli* trp and lac promoters).	*E. coli* [Amp]	Sigma-Aldrich
GPN023	pRARE	Providing tRNA genes for efficient expression of heterologous genes containing rare codons in *E. coli.*	*E. coli* [Cam]	Novagen
GPN028	pRS72K	Multicopy vector for constitutive protein expression under truncated yeast hexose transporter Hxt7 promoter (phxt7), terminator tCYC1	*S. cerevisiae* [G-418] and *E. coli* [Amp]	(Tripp *et al.*, 2017), kindly provided by Institut für Molekulare Biowissenschaften, Goethe-Universität Frankfurt
GPN047	P426M25	Multicopy vector for regulatable protein expression under methionine repressible O-acetyl homoserine Sulphydrylase promoter (pMET25), terminator tCYC1	*S. cerevisiae* [Ura] and *E. coli* [Amp]	(Mumberg *et al.*, 1994), kindly provided by Institut für Molekulare Biowissenschaften, Goethe-Universität Frankfurt
GPN024	pAIISK	Vector for the genomic integration of genes into the *psbAII* locus (*slr1311*) of *Synechocystis* sp. PCC3803.	*Synechocystis* sp. PCC3803 [Spec] and *E. coli* [Amp]	(Lagarde *et al.*, 2000), kindly provided by Institut für Biochemie der Pflanzen

Table 4.1-1B Vectors used/generated in this study.

#GPN	Expressed Inserts	Source
pTAC-MAT-Tag-2		
GPN020	*E. coli* phosphate translocator UhpT (Maloney *et al.*, 1990), His-tagged	Kindly provided by Institut für Biochemie der Pflanzen
GPN021	*P. chromatophora* putative phosphate translocator PTr1, His-tagged	This study
GPN022	*P. chromatophora* putative phosphate translocator PTr4, His-tagged	This study
pRS72K		
GPN029	Yeast hexose transporter HXT7 (Reifenberger *et al.*, 1997)	(Tripp *et al.*, 2017), kindly provided by Institut für Molekulare Bio-wissenschaften, Goethe-Universität Frankfurt
GPN032	*P. chromatophora* putative sugar transporter STr1	This study
GPN033	*P. chromatophora* putative sugar transporter STr2	This study
GPN034	*P. chromatophora* putative sugar transporter STr3	This study
P426M25		
GPN063	Yeast hexose transporter HXT7 (Reifenberger *et al.*, 1997), V5-tagged	This study
GPN064	*P. chromatophora* putative bacterial Gly/Ala transporter BGATr, V5-tagged	This study
GPN065	*P. chromatophora* putative neutral amino acid transporter 1 NAATr1, V5-tagged	This study
GPN066	*P. chromatophora* putative neutral amino acid transporter 2 NAATr2, V5-tagged	This study
pAIISK		
GPN025	*P. chromatophora* putative bacterial Gly/Ala transporter BGATr	Eva Nowack
GPN026	*P. chromatophora* putative neutral amino acid transporter 1 NAATr1	Eva Nowack
GPN027	*P. chromatophora* putative neutral amino acid transporter 2 NAATr2	Eva Nowack

4.1.2 Oligonucleotide Primers

Table 4.1-2 Primers used in this study. Residues initially binding to template DNA are written in capital letters, restriction sites are grey, tags are underlined.

#	Direction	Sequence 5´-3´	Binds at	Purpose
111 436 537	fw	CAGTTCCAATCTGAACATCG / ATCAGAATCCTTGCCCAG / CATGGTGAGCAACGATTGC	Upstream of *psbAII* locus of *Synechocystis* sp. PCC3803	Check for segregation of wild type and mutant genome copies, #111 and #438 for sequencing of pAIISK inserts.
112 437 538	rev	AGGAGAGTGCAATTTGCAG / CAGATGTCGTTGCTGGTTAC / CTCCAATCCACACTGGGAAG	Downstream of *psbAII* locus of *Synechocystis* sp. PCC3803	
438	rev	CTTGATGCCTGCAGGTC	Downstream of pAIISK MCS	
485	fw	CTGCAGTAGCAGGATCG	Inside *BGATr* ORF	Sequencing of *BGATr*
486	fw	GGCGAGTTCCAAGTAGG	Inside *NAATr2* ORF	Sequencing of *NAATr2*
476	fw	ataaacacaaaaacaaaaagttttttaattttaatcaaaaa-ATGTTCTCCGAGCTACCGTG	Start of *STr1* ORF	Integration of *STr1* ORF into pRS72K vector
477	rev	ggagggcgtgaatgtaagcgtgacataactaattacatgactcgag-CTACTGACTGATAGGAATACAAAAATGG	End of *STr1* ORF	
478	fw	ataaacacaaaaacaaaaagttttttaattttaatcaaaaa-ATGCAAAGCAACAAAGCTGAG	Start of *STr2* ORF	Integration of *STr2* ORF into pRS72K vector
479	rev	ggagggcgtgaatgtaagcgtgacataactaattacatgactcgag-TTAGTTCCGAGTAGAATCAAAGTC	End of *STr2* ORF	
480	fw	ataaacacaaaaacaaaaagttttttaattttaatcaaaaa-ATGAGTTCCATGCTACATGTC	Start of *STr3* ORF	Integration of *STr3* ORF into pRS72K vector
481	rev	gcgtgaatgtaagcgtgacataactaattacatgactcgag-CTACCACGCTTCCTCATCTTC	End of *STr3* ORF	
494	fw	CATCAAGAACAAACAAGCTCAAC	pHXT7 promoter	Colony-PCR and sequencing of pRS72K inserts
495	rev	CAGGTTGTCTAACTCCTTCC	tCYC1 terminator	
727	fw	cgattccgaacccgctgctgggcctggatagcacc-TCACAAGACGCTGCTATTGC	Start of *HXT7* ORF	N-terminal addition of 3´-sequence of V5-tag to *HXT7* ORF
743	fw	cgattccgaacccgctgctgggcctggatagcacc-GAAATAGATCGAGTTCTAGGTCC	Start of *BGATr* ORF	N-terminal addition of 3´-sequence of V5-tag to *BGATr* ORF
745	fw	cgattccgaacccgctgctgggcctggatagcacc-TCCGAAGCTGGGCTGCTG	Start of *NAATr1* ORF	N-terminal addition of 3´-sequence of V5-tag to *NAATr1* ORF
747	fw	cgattccgaacccgctgctgggcctggatagcacc-GACGCGATGGATGAGCCG	Start of *NAATr2* ORF	N-terminal addition of 3´-sequence of V5-tag to *NAATr2* ORF
725	fw	gatacatagatacaattctattacccccatccatacatgggcaaac-CGATTCCGAACCCGCTGC	3´-end of V5-tag	Addition of 5´-sequence of V5-tag and integration of *HXT7*, *BGATr*, *NAATr1*, and *NAATr2* ORFs into p426M25 vector
689	rev	ggagggcgtgaatgtaagcgtgacataactaattacatgactcgag-TTATTTGGTGCTGAACATTCTC	End of *HXT7* ORF	
744	rev	gaatgtaagcgtgacataactaattacatgactcgag-CTACGATTCAGTTGTTGTCCTC	End of *BGATr* ORF	
746	rev	gaatgtaagcgtgacataactaattacatgactcgag-CTACTGTGTTGGGTCGTGATCG	End of *NAATr1* ORF	
748	rev	gaatgtaagcgtgacataactaattacatgactcgag-TTATGCAAGATTGCCGACGAC	End of *NAATr2* ORF	
718	fw	AAGGTTAAGTAAAGCGTCTGTTAG	pMET25 promoter	Colony-PCR and sequencing of p426M25 inserts
495	rev	CAGGTTGTCTAACTCCTTCC	tCYC1 terminator	
560	fw	ctcgaggtAGAATCAGGCTCATCCCACC	Start of *PTr1* ORF	Integration of *PTr1* ORF into pTAC using XhoI and EcoRI, addition of C-term. 6xHis tag
573	rev	gaattcttagtgatgatgatgatgatg-AGTAGATGAAGTACTCCAAGAAGCAGAAG	End of *PTr1* ORF	
565	fw	aagcttGTTGGTGTGGGCTGGTGG	Start of *PTr4* ORF	Integration of *PTr4* ORF into pTAC using HindIII and EcoRI, addition of C-term. 6xHis tag
566	rev	gaattcttagtgatgatgatgatgatg-GCTTGCTACCGCCGGAG	End of *PTr4* ORF	
588	fw	GACAATTAATCATCGGCTCG	tac promoter	Sequencing of pTAC inserts
589	rev	TCGTCGAGATCGATCAGTG	downstream of MCS	

4.1.3 Strains

Table 4.1-3A Strains used in this study.

Strain	Genotype	Resistance / auxotrophy	Source
Paulinella chromatophora			
CCAC0185 (axenic)	wt	-	Culture Collection of Algae (CCAC), University of Cologne
Escherichia coli			
Top10	F-, *mcrA, Δ(mrr-hsdRMS-mcrBC), Φ80lacZΔM15, ΔlacX74, recA1, araD139, Δ(ara leu)7697, galU, galK, rpsL, (StrR), endA1, nupG*	Str^R	Invitrogen
JW3641-2	F-, *Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ⁻, rph-1, ΔuhpT771::kan, Δ(rhaD-rhaB)568, hsdR514*	Kan^R	Keio Knockout Collection (Baba *et al.*, 2006), Coli Genetic Stock Center, Yale University
Saccharomyces cerevisiae			
EBY.VW4000	*MATα, leu2-3.112, ura3-52, trp1-289, his3-Δ1, MAL2-8^c, SUC2, hxt1-16Δ::loxP, gal2Δ::(ura3/FOA), stl1Δ::loxP, agt1Δ::loxP, mph2(ydl247w)Δ::loxP, mph3(yjr160c)Δ::loxP*	Ura, Leu, Trp, His	(Solis-Escalante *et al.*, 2015; Wieczorke *et al.*, 1999), kindly provided by Institut für Molekulare Biowissenschaften, Goethe-Universität Frankfurt
22Δ8AA	*MATa, ura3-1, gap-1, put4-1, uga4-1, can1::HisG, lyp/alp::HisG, hip1::HisG, dip5::HisG*	Ura	(Fischer *et al.*, 2002), kindly provided by Institut für Molekulare Physiologie
23344c	*Mata, ura3-1*	Ura	M. Grenson (unpublished), kindly provided by Institut für Molekulare Physiologie
***Synechocystis* sp. PCC6803**			
	wt		Kindly provided by Institut für Biochemie der Pflanzen
Cyano003/004	*slr1311::(Spec+BGATr)*	$Spec^R$	Eva Nowack
Cyano005/006	*slr1311::(Spec+NAATr1)*	$Spec^R$	Eva Nowack
Cyano007/008	*slr1311::(Spec+NAATr2)*	$Spec^R$	Eva Nowack
ΔnatB	*slr0559::Kan*	Kan^R	(Eisenhut *et al.*, 2007), kindly provided by Institut für Biochemie der Pflanzen
ΔnatB Cyano009/010	*slr0559::Kan, slr1311::(Spec+BGATr)*	Kan^R, $Spec^R$	Eva Nowack
ΔnatB Cyano011/012	*slr0559::Kan, slr1311::(Spec+NAATr1)*	Kan^R, $Spec^R$	Eva Nowack
ΔnatB Cyano013/014	*slr0559::Kan, slr1311::(Spec+NAATr2)*	Kan^R, $Spec^R$	Eva Nowack

Table 4.1-3B Strains generated in this study.

Name	Vector / modified locus	Purpose
***Synechocystis* sp. PCC6803**		
Cyano015/016	*slr1311::(Spec)*	Wildtype, vector control for complementation tests and other uptake experiments
Synechocystis* sp. PCC6803 *ΔnatB		
Cyano019/020	*slr1311::(Spec)*	Vector control for complementation tests and other uptake experiments
***E. coli* JW3641-2**		
Ec01	GPN019/23	Vector control for complementation tests
Ec02	GPN020/23	Expressing *EcuhpT*, positive control for complementation tests
Ec03	GPN021/23	Expressing *PcPTr1*, for complementation tests
Ec04	GPN022/23	Expressing *PcPTr4*, for complementation tests
***S. cerevisiae* EBY.VW4000**		
GPN028	GPN028	Vector control for complementation tests

GPN029	GPN029	Expressing *Schxt7*, positive control for complementation tests
GPN032	GPN032	Expressing *PcSTr1*, for complementation tests
GPN033	GPN033	Expressing *PcSTr2*, for complementation tests
GPN034	GPN034	Expressing *PcSTr3*, for complementation tests
***S. cerevisiae* 22Δ8AA**		
Sc3_GPN047	GPN047	Vector control for complementation tests and immunofluorescence microscopy
Sc3_GPN063	GPN063	Expressing V5-*Schxt7*, positive control for immunofluorescence
GPN064	GPN064	Expressing V5-*PcBGATr*, for complementation tests
GPN065	GPN065	Expressing V5-*PcNAATr1*, for complementation tests
GPN066	GPN066	Expressing V5-*PcNAATr2*, for complementation tests
***S. cerevisiae* 23344c**		
Sc4_GPN047	GPN047	Wildtype, vector control for complementation tests

4.1.4 Media

4.1.4.1 WARIS-H Medium for Cultivation of *Paulinella chromatophora*

P. chromatophora was cultivated in WARIS-H liquid medium (McFadden and Melkonian, 1986) modified by addition of Na_2SiO_3 x 9 H_2O to a final concentration of 1.5 mM Na_2SiO_3 and supplemented with triple concentrations of vitamin solution and soil extract, respectively. Ingredients 1-8 were prepared as 1000 x stock solutions and stored at 4°C.

#	Ingredients	Concentration in medium	
1	KNO_3	1 mM	100 mg/l
2	$MgSO_4$ x 7 H_2O	81.10 mM	20 mg/l
3	$(NH_4)_2HPO_4$	0.15 mM	20 mg/l
4	$Ca(NO_3)_2$ x 4 H_2O	0.42 mM	100 mg/l
5	HEPES	1 mM	240 mg/l
6	**P-II-Metals**		
	Na_2-EDTA x 2 H_2O (Titriplex III)	8 µM	3 mg/l
	H_3BO_3	18.40 µM	1 mg/l
	$MnCl_2$ x 4 H_2O	0.70 µM	144 µg/l
	$ZnSO_4$ x 6 H_2O	73 nM	21 µg/l
	$CoCl_2$ x 6 H_2O	16.80 nM	4 µg/l
7	**Fe-EDTA[a]**		
	EDTA (Titriplex II)	17.90 µM	5.20 mg/l
	$FeSO_4$ x 7 H_2O	17.90 µM	5 mg/l
	KOH	54 µM	3 mg/l
8	**Vitamins**		
	Vitamin B12	0.45 nM	0.60 µg/l
	Biotin	12.30 nM	3 µg/l
	Thiamin-HCL	0.90 µM	300 µg
	Niacinamide	2.40 µM	0.30 µg/l
9	Soil extract[b]	3 % (v/v)	30 ml/l
10	Na_2SiO_3 x 9 H_2O	1.5 mM	426 mg/l

[a] For solubilization, EDTA and $FeSO_4$ are heated to 100°C for 30 min. 1 M KOH is added slowly to the cooled mixture with continuous stirring.

[b] 10 g of garden-soil is mixed with 120 ml dH_2O and boiled for 10 min. Afterwards it is centrifuged for 10 min (low speed), and the supernatant is filtered through a series of membrane filters from 1.2-0.1 µm pore size. The remaining filtrate is adjusted to 100 ml with dH_2O. Aliquots are stored frozen.

The pH was adjusted to 7.0 with HCl. The medium was sterilized in an autoclave for 20 min at 121°C.

4.1.4.2 BG11 Medium for Cultivation of *Synechocystis* sp. PCC6803

Synechocystis was cultivated in liquid or on agar-solidified BG11 medium (Rippka *et al.*, 1979) modified by addition of $Na_2S_2O_3$ x 5 H_2O to a final concentration of 12 mM and buffered to pH 8.0 with 10 mM TES-KOH. Ingredients 1-5 were combined to a 100 x stock solution, 6-8 were prepared as 1000 x stock solutions and stored at 4°C.

#	Ingredients	Concentration in medium	
1	$CaCl_2$ x 2 H_2O	0.25 mM	36 mg/l
2	Citric acid	29 µM	6 mg/l
3	$NaNO_3$	18 mM	1.5 g/l
4	$MgSO_4$ x 7 H_2O	0.30 mM	75 mg/l
5	Na_2-EDTA[a]	1.40 µM	0.5 mg/l
6	Na_2CO_3	0.19 mM	20 mg/l
7	K_2HPO_4 x 3 H_2O	0.12 mM	30 mg/l
8	**Trace metals**		
	H_3BO_3	46 µM	2.9 mg/l
	$MnCl_2$ x 4 H_2O	9 µM	1.8 mg/l
	$ZnSO_4$ x 7 H_2O	0.77 µM	0.22 mg/l
	Na_2MoO_4 x 2 H_2O	1.60 µM	0.39 mg/l
	$Co(NO_3)_2$ x 6 H_2O	0.17 µM	49 µg/l
9	TES[b]	10 mM	2.3 g/l
10	$CuSO_4$ x 5 H_2O[c]	0.32 µM	79 µg/l
11	Ferric ammonium citrate[d]	30 µM	6 mg/l
12	$Na_2S_2O_3$ x 5 H_2O	12 mM	3 g/l
	Bacto Agar (optional)[e]	1-1.5 % (w/v)	10-15 g/l

[a] From 0.25M Na_2-EDTA (pH 8.0)
[b] From 1M TES-KOH (pH 8.0)
[c] A 0.8 mM stock was filter sterilized and added to the cooled medium after autoclavation.
[d] A 15 mM stock was filter sterilized and added to the cooled medium after autoclavation.
[e] For solid medium 2 x concentrated Agar is dissolved in water, autoclaved separately and added to 2 x concentrated BG11.

The medium was sterilized in an autoclave for 20 min at 121°C. To obtain selective media Spectinomycin and/or Kanamycin were added to the cooled medium after autoclavation to final concentrations of 30 µg/ml and 50 µg/ml, respectively.

4.1.4.3 Media for Cultivation of *E. coli*

E. coli was cultivated either in liquid- or on agar-solidified LB medium, or on agar-solidified M9 minimal medium.

LB medium			**M9 minimal medium**		
Ingredients	**Concentration in medium**		**Ingredients**	**Concentration in medium**	
Bacto Trypton	1 % (w/v)	10 g/l	Na_2HPO_4 x 2 H_2O	48 mM	8.5 g/l
Bacto Yeast Extract	0.5 % (w/v)	5 g/l	KH_2PO_4	22 mM	3 g/l
NaCl	86 mM	5 g/l	NH_4Cl	20 mM	1 g/l
Bacto Agar (optional)	1.5 % (w/v)	15 g/l	NaCl	8.6 mM	0.5 g/l
			$MgSO_4$ x H_2O	1 mM	140 mg/l
			$CaCl_2$ x 2 H_2O	0.1 mM	15 mg/l
			Carbon source[a]	0.2 % (w/v)	2 g/l
			Bacto Agar	1.5 % (w/v)	15 g/l

[a] stocks of glucose, glucose-6-phosphate, or fructose-6-phosphate were sterile filtered and added to the cooled medium after autoclavation.

The medium was sterilized in an autoclave for 20 min at 121°C. To obtain selective media Ampicillin and Chloramphenicol were added to the cooled medium after autoclavation to final concentrations of 100 µg/ml and 25 µg/ml, respectively.

4.1.4.4 Media for Cultivation of *S. cerevisiae*

S. cerevisiae was cultivated either in liquid- or on agar-solidified YEP medium, or in liquid- or on agar-solidified minimal medium.

YEP medium			**SD minimal medium[b]**		
Ingredients	**Concentration in medium**		**Ingredients**	**Concentration in medium**	
Bacto yeast extract	1 % (w/v)	10 g/l	Difco yeast nitrogen base (w/o amino acids and ammonium sulfate)	0.17 % (w/v)	1.7 g/l
Bacto pepton	2 % (w/v)	20 g/l	Glucose	2 % (w/v)	20 g/l
Carbon source[a]	0.2-2 % (w/v)	2-20 g/l	Nitrogen source	1-4 mM	-
Bacto agar (optional)	2 % (w/v)	20 g/l	Bacto agar (optional)	2 % (w/v)	20 g/l

[a] 10 x stocks of maltose, glucose, fructose or mannose were autoclaved separately and added to the cooled medium after autoclavation to avoid Maillard-reaction of sugar and amino acids.
[b] A 10 x stock solution of yeast nitrogen base and glucose was filter sterilized. Stocks of nitrogen sources were filter sterilized and stored at 4°C. To prepare medium sterile ingredients were added to autoclaved water or agar.

YEP medium was sterilized in an autoclave for 20 min at 121°C. To obtain selective medium G-418 (Geneticin) was added to the cooled medium after autoclavation to a final concentration of 200 µg/ml.

4.1.4.5 Supplements Used for Complementation Experiments

Media supplements were prepared as concentrated stock solutions, sterile filtered, and stored at 4°C until usage.

Nutrient	Manufacturer
Glucose x H_2O	Caelo
D-Fructose 99 %	Sigma Aldrich
D-Mannose	Merck
D-glucose-6-phosphate dipotassium salt x H_2O (98-100 %)	Sigma Aldrich
D-fructose-6-phosphate disodium salt x H_2O 98 %	Sigma Aldrich
Glycine 99 %	Thermo Scientific
DL-Phenylalanine 98 %	TCI
DL-Ethionine 99 %	Acros Organics
L-Proline 99 %	ITW Reagents
L-Citrulline 98 %	Sigma Aldrich
L-Arginine 98 %	Sigma Aldrich
γ-aminobutyric acid 99 %	Sigma Aldrich
L-Glutamate 99 %	Acros Organics
L-Aspartate 98 %	Sigma Aldrich

4.1.5 Software and Web Applications

Program	Purpose	Publisher
Adobe Illustrator	Figure design	Adobe Inc.
Adobe Photoshop	Figure design, post-processing of scans/photos	Adobe Inc.
MS Excel	Processing of proteomic data, bubble diagrams	Microsoft Corp.
MS Word	Thesis layout	Microsoft Corp.
GraphPad Prism 5.03	Plotting and bar diagrams	GraphPad Software Inc.
BioEdit	Fasta sequence editing	(Hall, 1999)
ClustalW	Multiple sequence alignments	(Thompson *et al.*, 1994)
Benchling molecular biology suite	Experimental design and analysis of sequencing data	Benchling Inc.
BLASTP	Database screening for similar proteins	(Altschul *et al.*, 1990)
Blast2GO	High-throughput functional annotation of proteins	(Götz *et al.*, 2008)
DELTA-BLAST	Discovery of functional domains in proteins	(Boratyn *et al.*, 2012)
TMHMM 2.0	Batch prediction of transmembrane helices	(Krogh *et al.*, 2001)
CCTOP	Prediction of transmembrane helices	(Dobson *et al.*, 2015)
MCMBB	Prediction of β-barrel proteins encoded in the chromatophore genome	(Bagos *et al.*, 2004)
TCDB	Classification of transporters	(Saier *et al.*, 2016)
TransAAP / TransportDB	Genome wide identification of transporters	(Elbourne *et al.*, 2017)
MEME 5.0.5	Discovery of sequence motifs	(Bailey and Elkan, 1994)
MAST 5.0.5	Determination of number and position of sequence motifs in proteins	(Bailey and Gribskov, 1998)
FIMO 5.0.5	Screening of the *P. chromatophora* transcriptome for sequence motifs	(Grant *et al.*, 2011)
PatternSearch	Database screening for defined sequence patterns	(Zimmermann *et al.*, 2018)
Phyre2	Tertiary structure prediction	(Kelley *et al.*, 2015)
TopModel	Template-based protein structure prediction	(Mulnaes *et al.*, 2020)
CHDOCK	Modeling of symmetric homo-oligomeric complexes	(Yan and Huang, 2019)

HOLE	Analysis of pore dimensions of channel structural models	(Smart *et al.*, 1996)
Jpred4	Prediction of α-helices in repetitive elements	(Drozdetskiy *et al.*, 2015)
NetSurfP-2.0	Prediction of α-helices in AMP-like proteins	(Klausen *et al.*, 2019)
AmpGram	Prediction of AMP peptides	(Burdukiewicz *et al.*, 2020)
MEMSAT-SVM-pore	Prediction of pore-lining residues	(Nugent and Jones, 2012)
HeliQuest	Visualization of α-helices and hydrophobic moments	(Gautier *et al.*, 2008)
WebLogo	Visualization of sequence motifs	(Crooks *et al.*, 2004)
EzMol	Visualization of 3D-protein models	(Reynolds *et al.*, 2018)
eulerAPE	Calculation of area-proportional Venn-diagrams	(Micallef and Rodgers, 2014)
SUBAcon / SUBA4	*A. thaliana* protein subcellular localization database	(Hooper *et al.*, 2014)
PredAlgo	Prediction of targeting signals in *C. reinhardtii*	(Tardif *et al.*, 2012)
TargetP 2.0	Prediction of targeting signals	(Armenteros *et al.*, 2019)
WoLF PSORT	Prediction of targeting signals	(Horton *et al.*, 2007)
Predotar	Prediction of targeting signals	(Small *et al.*, 2004)

4.2 Methods

4.2.1 Molecular Biology Methods

4.2.1.1 Determination of the Concentration of Nucleic Acids

The concentration of DNA or RNA was determined photometrically with a NanoDrop 2000c spectral photometer (Thermo Fischer Scientific) at 260 nm.

4.2.1.2 Isolation of Total RNA from *P. chromatophora* and Synthesis of cDNA

P. chromatophora cells were harvested by centrifugation at 200 x g for 5 min and were shock-frozen in liquid nitrogen. TRI Reagent (Sigma Aldrich, approx. 10 volumes of the cell pellet) was added to the frozen pellet and the mixture was passed through a 0.8 mm cannula 10 times to lyse the cells. After 10 min incubation chloroform (1/5 volume of TRI Reagent volume) was added and the mixture was vortexed thoroughly for 30 sec. Phase separation was achieved by centrifugation at 12,000 x g for 10 min at 4°C. The upper aqueous phase was collected and RNA was precipitated by the addition of isopropanol (1/2 volume of TRI Reagent volume), incubation for 10 min, and centrifugation at 12,000 x g for 10 min at 4°C. The RNA pellet was washed two times with 75 % ethanol (v/v in diethyldicarbonate-treated [DEPC] water) by vigorous vortexing for 1 min followed by centrifugation. The pellet was air dried and resuspended in DEPC-water. Approx. 10 µg of RNA could be isolated per individual *P. chromatophora* culture.

Residual DNA was removed by the addition of 1 Unit DNAse I (Thermo Scientific RevertAid Kit) per 2.5 µg RNA as well as the appropriate volume of 10 x reaction buffer and incubation at 37°C for 30 min.

RNA was purified from the enzymatic reaction using the RNeasy MinElute Kit (Qiagen) according to the manufacturer´s instructions.

The RevertAid Kit (Thermo Scientific) was used according to the manufacturer´s instructions for Oligo dT_{18}-primed cDNA first strand synthesis.

4.2.1.3 Polymerase Chain Reaction (PCR)

DNA-fragments used for cloning were PCR-amplified using Phusion DNA Polymerase (New England Biolabs). It possesses 3´→ 5´ proofreading exonuclease activity and generates blunt-ended products. Colony-PCR was used to check microbial transformants for the presence of the respective transgene. In these cases, whole cells were scratched from agar plates and added to the PCR-reaction directly. Cells are lysed during an elongated initial denaturation phase to release DNA that serves as PCR-template. Some templates required the usage of GC buffer instead of the default HF buffer, and the addition of DMSO. Usually, a reaction volume of 20 µl or 50 µl and 30-35 PCR-cycles were used.

Component	Final conc.
H_2O	variable
5x Phusion Buffer[a]	1 x
10 mM dNTPs	200 µM each
10 µM primer fw	0.5 µM
10 µM primer rev	0.5 µM
DMSO (optional)	3 % (v/v)
template DNA	variable[b]
Phusion DNA Polymerase	0.02 U/µl

Cycle step	Protocol	
Initial denaturation	98°C	30 sec-3 min
Denaturation	98°C	10 sec
Annealing	variable[c]	10-30 sec
Extension	72°C	30 sec/kb
Final extension	72°C	5-10 min

[a] HF (high fidelity) or GC (for difficult templates) versions.
[b] Plasmids: 0.05-0.5 ng/µl, gDNA/cDNA: 1-5 ng/µl
[c] Primers < 20 nt: use T_m of lower T_m primer; Primers > 20 nt: use T_m of lower T_m primer + 3°C. The annealing step can be skipped, when the primer T_m is > 69°C.

4.2.1.4 Agarose Gel Electrophoresis

DNA fragments were separated according to their length electrophoretically in 0.8 % (w/v) agarose TAE-gels. 2 % (w/v) gels were used when a more accurate separation of fragments < 3000 bp was desired. Ethidium bromide was added to visualize DNA under UV-light. GeneRuler 1 kb or 100 bp Plus DNA Ladders (Thermo Fisher Scientific) were used as length standards.

TAE-buffer	
Ingredients	**Concentration**
Tris base	40 mM
Glacial acetic acid	20 mM
Na_2-EDTA	1 mM

4.2.1.5 DNA Separation in Agarose Gels and Gel Extraction

PCR-products or restriction fragments were separated from contaminating DNA, enzymes, or other substances via agarose gel electrophoresis. The QIAquick Gel Extraction Kit (Qiagen) was used according to the manufacturer´s instructions to extract pure DNA from agarose gels.

4.2.1.6 Preparation of Plasmid DNA

Plasmid DNA was purified from 5 ml overnight cultures of *E. coli* using the QIAprep Spin Miniprep Kit (Qiagen) according to the manufacturer´s instructions.

4.2.1.7 DNA Sequencing

Sanger sequencing was performed by Eurofins Genomics. Mix2Seq overnight sequencing service was used. Plasmid DNA (approx. 1000 ng) or PCR products (approx. 150 ng) and 20 pmol of the respective sequencing primer were mixed in the provided vial and sent to the sequencing facility.

4.2.1.8 Conventional Molecular Cloning

Conventional restriction/ligation cloning was used for the construction of vectors GPN021 and GPN022. *P. chromatophora PTr1* and *PTr4* ORFs were PCR-amplified (for primers see 4.1.2), purified from 0.8 % (w/v) agarose gels and blunt-end ligated with the pJET1.2/blunt cloning vector using the CloneJET PCR Cloning Kit (Thermo Scientific) according to the manufacturer´s instructions. *E. coli* Top10 cells were transformed and plasmid DNA was isolated from the transformants. Once the correctness of the inserted ORFs was verified by sequencing, 3 µg of Plasmid DNA was restricted with the high-fidelity endonucleases *EcoRI* and either *XhoI* or *HindIII* (New England Biolabs) to excise *PTr1* and *PTr4* ORFs from pJET, respectively. Expression vector pTAC was restricted accordingly. The pTAC vector backbone and both fragments were gel-purified and used in a 1:5 molar ratio for ligation by T4 DNA ligase (New England Biolabs) at RT for 1 h. *E. coli* Top10 cells were transformed with the ligation mixture and ampicillin-resistant transformants were checked for the presence of a correctly assembled vector via colony-PCR and control restriction digestions. Plasmid DNA was isolated as described before, sequenced, and used to transform *E. coli ΔuhpT* strain JW3641-2.

4.2.2 Protein Biochemical Methods

4.2.2.1 Determination of Protein Concentration

The protein concentration in complex mixtures (e.g. cell lysates or fractions) was determined in a Pierce 660 nm assay (Thermo Scientific) according to the manufacturer´s instructions. The absorbance of the protein/reagent mixtures was measured in an Infinite M200 (TECAN) plate reader at 660 nm. A

Neuhoff assay (Neuhoff *et al.*, 1979) was used when the sample contained detergents, as these are washed out after protein staining with amido black during several rounds of de-staining and thereby render the assay more robust for such samples. The absorbance of amido black was measured in a Novaspec II spectrophotometer (Pharmacia Biotech) at 630 nm. BSA (Bovine Serum Albumine) standards were prepared in the same buffers and contained the same additives as the samples to be determined.

4.2.2.2 TCA-precipitation of Protein

Protein was extracted from intact cells or chromatophores or from lysates through precipitation with trichloracetic acid (TCA). 1/10 volume of 100 % (w/v) TCA was added to the samples and the mixture was incubated for 30 min on ice and centrifuged at 21,000 x g for 20 min. Protein pellets were washed twice by the addition of ice-cold acetone, incubation for 10 min on ice, and centrifugation at 21,000 x g for 10 min. Pellets were air dried and resuspended in the desired buffer.

4.2.2.3 SDS-PAGE

SDS-PAGE (Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis) was used to separate denatured proteins according to their molecular weight. Depending on the application, either FastGene Bis-Tris 4-20 % gradient gels (Nippon Genetics) or 10 % Tris-glycine gels were used. Bis-Tris gels were run in MES buffer at 100-140 V and 55-100 mA. Tris-Glycine gels were run in TGS buffer at 40 mA/gel. PageRuler Prestained or Unstained protein ladders (Thermo Scientific) were used.

MES-buffer	
Ingredients	**Concentration**
Tris base	50 mM
MES	50 mM
EDTA	1 mM
SDS	0.1 % (w/v)

Sample buffer (1x)	
Ingredients	**Concentration**
Tris-HCl pH 7.0	35 mM
Glycerol	7.5 % (v/v)
DTT	150 mM
SDS	3 % (w/v)
Bromophenol blue	0.005 % (w/v)

TGS-buffer	
Ingredients	**Concentration**
Tris base	25 mM
Glycine	192 mM
SDS	0.1 % (w/v)

Tris-Glycine Separating gel	
Ingredients	**Concentration**
Tris-HCl pH 8.8	375 mM
30:0.8 acrylamide (30 %)	33 % (v/v)
Glycerol	1.25 %
SDS	0.10 %
APS	0.05 %
TEMED	0.10 %

Tris-Glycine Stacking gel	
Ingredients	**Concentration**
Tris-HCl pH 6.8	125 mM
30:0.8 acrylamide (30 %)	16 % (v/v)
SDS	0.10 %
APS	0.05 %
TEMED	0.10 %

4.2.2.4 Silver Staining of Protein in Polyacrylamide Gels

Proteins separated in SDS-polyacrylamide gels were silver-stained using a slightly modified protocol according to (Blum *et al.*, 1987), by subjecting the gels to a series of incubation steps under gentle agitation in different solutions and rinsing with water after each step. First, gels were incubated in fixing solution for 1 h and subsequently washed in 30 % ethanol for 30 min. The gels were then incubated in 0.02 % (w/v) sodium-thiosulfate for 1 min precisely and stained with silver nitrate for 20 min and protected from light. Subsequently, the gels were incubated in developer solution until the desired intensity of staining was achieved. The reaction was stopped after incubation in stopping solution for 5-10 min.

Fixing solution	
Ingredients	**Concentration**
Acetic acid	10 % (v/v)
Methanol	50 % (v/v)

Staining solution	
Ingredients	**Concentration**
$AgNO_3$	12 mM
37 % formaldehyde	0.015 % (v/v)

Developer	
Ingredients	**Concentration**
Na_2CO_3	565 mM
37 % formaldehyde	0.05 % (v/v)
$Na_2S_2O_3$	25 µM

Stopping solution	
Ingredients	**Concentration**
Acetic acid	18 % (v/v)
Methanol	75 % (v/v)

4.2.2.5 Western Blotting of Proteins

Proteins separated via SDS-PAGE were transferred onto methanol-activated PVDF (Polyvinyliden fluoride) membranes (Amersham Hybond-P, GE Healthcare) according to the following procedure. Transfer was performed electrophoretically using a semi-dry western blot chamber (846-015-200, Biometra). Two filter papers (GE Healthcare, Whatman paper) were soaked in anode buffer 1 and stacked on the chamber´s anode, followed by another filter paper soaked in anode buffer 2. The PVDF membrane was activated in methanol for 1 min, rinsed with water, equilibrated in anode buffer 2, and placed on top of the filter paper stack. On top of that, the polyacrylamide gel containing the proteins was carefully placed, followed by three further filter papers soaked in cathode buffer. Air bubbles between the layers were carefully pushed out. Proteins were transferred for 1 h at 60 mA per stack that was assembled in the chamber.

Anode buffer 1		**Anode buffer 2**		**Cathode buffer**	
Ingredients	**Concentration**	**Ingredients**	**Concentration**	**Ingredients**	**Concentration**
Tris-HCl pH 10.4	300 mM	Tris-HCl pH 10.4	30 mM	Tris-HCl pH 9.4	25 mM
Methanol	15 % (v/v)	Methanol	15 % (v/v)	Methanol	15 % (v/v)
				Ɛ-Aminocapronic acid	40 mM

4.2.2.6 Protein Immunodetection on PVDF Membranes

To detect individual proteins on a PVDF membrane, the membrane was treated with a specific antibody that binds to the protein to be detected, followed by a secondary antibody that specifically binds to the primary antibody and is coupled to an enzyme (horseradish peroxidase, HRP) that produces a chemiluminescent signal when its substrate is added. To block free binding sites on the PVDF membrane, it was first incubated in Blocking solution for 1 h under gentle agitation. The membrane was then sealed in a plastic bag and a small volume of blocking solution supplemented with the appropriate amount of primary antibody was added. The membrane was incubated overnight on a wheel at 4°C to allow binding of the antibody to its antigen. Excess antibody was washed away successively three times with fresh TBST buffer under gentle agitation. A 1:3000 dilution of HRP-coupled anti-rabbit IgG secondary antibody (Thermo Fisher Scientific) in blocking solution was applied equally to the primary antibody, but required only 1-3 h of incubation at RT. Excess antibody was washed away successively as described before. HRP substrate (SuperSignal West Pico PLUS, Thermo Scientific) was added to the membrane and chemiluminescence was detected using the ImageQuant LAS-4000 mini (GE Healthcare).

TBST	
Ingredients	**Concentration**
Tris-HCl pH 7.5	50 mM
NaCl	150 mM
Tween 20	0.05 % (v/v)

Blocking solution	
Ingredients	**Concentration**
Skimmed milk powder	5 % (w/v)[a]

[a] In TBST

4.2.2.7 Immunofluorescence Microscopy of V5-tagged Transporters in *S. cerevisiae*

The subcellular localization of V5-tagged transporters in *S. cerevisiae* was evaluated by immunofluorescence microscopy. To do so, cells were fixed in formaldehyde, the cell wall was enzymatically disrupted to obtain spheroblasts and antigens were detected immunologically according to Baggett *et al.*, 2003.

10 ml overnight cultures were grown in selective SD minimal medium to an OD_{600} of 0.5-1 and supplemented with 2.5 ml of 5 x concentrated fixation cocktail. Cells were incubated for 90 min under continuous shaking at 200 rpm. Cultures were harvested at 800 x g and resuspended in 1 ml of fixation cocktail. Cells were transferred to a 1.5 ml tube and incubated for 23 h in a rotator. Cells were then harvested for 5 min at 800 x g and resuspended in 1 ml of spheroblasting solution. Cells were incubated for 150 min at 30 °C and spheroblasting was checked microscopically after 100 min. If necessary, 5 more units of lyticase enzyme were added. Spheroblasts were then pelleted for 5 min at 300 x g and resuspended in 0.5 ml SHA buffer containing 1 % (v/v) TritonX-100. Spheroblasts were incubated for 10 min, washed twice in 1 ml SHA buffer, and resuspended to an OD_{600} of approx. 24. 15 µl of the

suspension was applied dropwise to poly-L-lysine (0.01 % solution, Sigma Aldrich) coated slides and incubated for 15-30 min in a moisture chamber. Access liquid was removed carefully. For blocking, 15-30 µl of WT buffer was added to the slides and incubated for 30 min. WT buffer was removed and 15 µl of primary antibody (mouse anti-V5 αb27671 [Sigma Aldrich], 1:1000 in WT buffer) was added to the blocked spheroblasts on the slides and slides were incubated overnight at 4°C in a moisture chamber. Samples were washed five times carefully with 15-30 µl of WT buffer. 20 µl of secondary antibody (1 drop of Alexafluor 488 donkey anti-mouse IgG in 500 µl WT buffer) was added to the spheroblasts and slides were incubated for 30 min in the dark. Spheroblasts were then again washed five times carefully with 15-30 µl of WT buffer. All liquid was removed carefully and 1 drop of SlowFade Diamond antifade mounting solution (Molecular Probes) was applied to the slides. Coverslips were placed to the slides. The slides were then covered by a paper towel, and the coverslips were pressed down firmly but carefully by placing a weight for 1-2 min. Finally, the coverslips were sealed with nail polish. Images for localization studies were acquired using an epifluorescence microscope (Zeiss Axio Imager.M1) and image manipulation was performed using the Metamorph software package (version 7, Molecular Devices).

5 x Fixation cocktail	
Ingredients	**Concentration**
Potassium phosphate pH 6.5	500 mM
37 % formaldehyde	20 % (v/v)

SHA buffer	
Ingredients	**Concentration**
HEPES pH 7.5	100 mM
Sorbitol	1 M

Spheroblasting solution[a]	
Ingredients	**Concentration**
β-mercaptoethanol	0.2 % (v/v)
Lyticase[b]	5-10 U/ml

[a] Prepared in SHA buffer

[b] From *Anthrobacter luteus*, (Sigma Aldrich)

W5 buffer	
Ingredients	**Concentration**
HEPES pH 7.5	50 mM
NaCl	150 mM
Tween 20	0.1 % (v/v)
IgG-free BSA	5 % (w/v)

4.2.2.8 Generation and Affinity Purification of Peptide Antibodies

Specific antibodies against the nuclear encoded candidate chromatophore envelope transporter BGATr were generated by Davids Biotechnologie, Regensburg. As the protein could not be produced and purified in sufficient amounts, peptide antibodies were generated. Therefore, two short peptides were synthesized via solid phase synthesis and conjugated to a carrier protein (KLH, Keyhole Limpet Hemocyanin) to enable optimal antigen presentation to the immune system. The peptides correspond to portions of the transporter that are likely not parts of transmembrane helices and thus should be accessible when the antibodies are used on the native protein. Other factors to be considered are the peptides predicted antigenicity (antigenic force for a good antibody response), the predicted solubility

(surface probability for solubility and antigen presentation), and predicted epitopes (epitopes for inducing B-cells for specific antibodies). A rabbit was immunized five times consecutively with both peptides and the antiserum could be harvested after 63 days.

Peptide #1 AWHHIKGTFDDPTAVGEVSH

Peptide #2 AERSERKDRSNERKRTTTES

α-BGATr antibodies were purified from the serum using the synthetic peptides that were used for antibody generation coupled to NHS Mag Sepharose magnetic beads (GE Healthcare). NHS (N-hydroxysuccinimide) is a ligand to which molecules with primary amino groups bind covalently. The manufacturer´s protocol was modified for the purification of intact antibodies as follows. 25 µl of sepharose beads were prepared by removing the storage solution and short incubation in 500 µl ice-cold equilibration buffer. The peptide solution (1 mg peptide in coupling buffer) was added and incubated for at least 30 min with slow end-over-end mixing. Afterwards, residual NHS active groups were blocked by three consecutive alternately treatments with blocking buffer A and blocking buffer B. Beads were then equilibrated for binding by the addition of 500 µl binding buffer. 1 ml of α-BGATr antiserum diluted in 300 µl binding buffer was added to the beads and incubated for 1-2 h with slow end-over-end mixing to allow binding of the antibodies to the bead-bound peptides. Beads were then washed two times with 500 µl wash buffer and once with water. The antibodies were eluted from the beads by a low-pH elution buffer in two consecutive steps using 100 µl and 50 µl, respectively. 25 µl of neutralization buffer were added to the eluate immediately to neutralize the pH. 0.05 % sodium azide and 1 mg/ml BSA were added to avoid microbial contamination and protease cleavage of the antibodies, respectively.

Coupling buffer (1.2 x)		**Blocking buffer A**		**Blocking buffer B**	
Ingredients	**Concentration**	**Ingredients**	**Concentration**	**Ingredients**	**Concentration**
$NaHCO_3$	230 mM	Ethanolamine	500 mM	Na-acetate	100 mM
NaCl	666 mM	NaCl	500 mM	NaCl	500 mM
pH 8.3		pH 8.3		pH 4.0	

Binding/Wash buffer		**Elution buffer**		**Neutralization buffer**	
Ingredients	**Concentration**	**Ingredients**	**Concentration**	**Ingredients**	**Concentration**
Tris-HCl pH 7.5	50 mM	Glycine	100 mM	Tris-HCl pH 9.0	1 M
NaCl	500 mM	pH 2.0			
Tween 20	0.05 % (v/v)				

4.2.3 Microbiological Methods

4.2.3.1 *P. chromatophora*

Cultivation of *P. chromatophora*

P. chromatophora was cultivated in Fernbach-flasks containing 500 ml WARIS-H medium (see 4.1.4.1) at 17°C and 15 µE at a 14 h/10 h light/dark cycle (fluorescent tubes: MASTER TL-D Super 80 840, Philips). Cultures were checked for the absence of contaminating bacteria regularly by plating aliquots on solid medium composed of 1/10 Bacterial Standard medium and 9/10 WARIS-H medium.

4.2.3.2 *Synechocystis* sp. PCC6803

Cultivation of *Synechocystis* sp. PCC6803

Synechocystis was cultivated at 28°C and approx. 30-40 µE constant illumination in liquid BG11 medium (see 4.1.4.2) shaking at 150 rpm or on solidified BG11 medium in tissue culture dishes (TPP Techno Plastic Products AG). The cell density of liquid cultures was determined photometrically by measuring the optical density at 750 nm (OD_{750}) with a Genesys 10S UV-Vis spectrophotometer (Thermo Scientific). Cultures were diluted to an OD_{750} in the range of 0.2 and 0.8 for the measurements to guarantee linearity of cell density and OD measured. The respective culture medium was used for blank measurements.

Transformation of *Synechocystis* sp. PCC6803

A modified version of the protocol from Hagemann & Zuther, 1992 was used. Cultures (5 ml per vector to be introduced) were grown to logarithmic growth phase (OD_{750} of approx. 1-3) in BG11, harvested by centrifugation at 3000 x g for 10 min, and cells were resuspended in BG11 medium (1/50 volume of the original culture). 100 µl of the cell suspension were added approx. 1-2.5 µg of pAIISK plasmid DNA (concentration 150-650 ng/µl) and mixed carefully. Cells were incubated overnight at 28°C under shaking in total darkness. Cells were spread on solid BG11 and again incubated overnight, this time under dim light (30-40 µE light source, protected by a semitransparent fleece). Cells were further incubated for 48 h at full light intensity (30-40 µE). Afterwards, 500 µl of BG11 medium containing 5 µg/ml Spec were spread underneath the BG11 solid medium disc to allow for slow diffusion of the antibiotic to the cells on the agar surface. Transformants formed colonies after 14 d, which were then spread on fresh BG11 medium containing 5 µg/ml Spec. After 7-10 d these transformants were transferred to BG11 medium containing 25 µg/ml Spec. Resistant transformants were checked for the genomic integration of the pAIISK-cassette (containing the ORF to be expressed and spectinomycin resistance gene *aadA*) into the *psbAII* locus via Colony-PCR. The generated PCR-fragments were purified from agarose gels and also used for sequencing of the integrated ORFs. For cryoconservation

of the strains generated, cells were grown in liquid culture and 2 ml were harvested by centrifugation, resuspended in DMSO, and stored at -80°C.

4.2.3.3 *E. coli*

Cultivation of *E. coli*

E. coli was cultivated in liquid LB medium (shaking at 200 rpm or in a rotating wheel) or on solidified LB medium (see 4.1.4.3). Cultures were incubated at 37°C unless otherwise stated. The cell density of liquid cultures was determined photometrically by measuring the optical density at 600 nm (OD_{600}) with a Novaspec II (Pharmacia Biotech) spectrophotometer or a Genesys 10S UV-Vis spectrophotometer (Thermo Scientific). Cultures were diluted to an OD_{600} in the range of 0.2 and 0.8 for the measurements to guarantee linearity of cell density and OD measured. The respective culture medium was used for blank measurements.

Preparation of Frozen Chemically Competent *E. coli* Cells

An *E. coli* overnight culture was used to inoculate a culture of an OD_{600} = 0.1 in pre-warmed LB medium supplemented with 10 mM $MgCl_2$ and $MgSO_4$, respectively. Cells were incubated at 37°C and under shaking at 200 rpm until an OD_{600} of 0.6 was accomplished. The culture was then immediately cooled down on ice and incubated for 30 min at 4°C. Cells were harvested for 8 min at 3000 x g, resuspended in pre-cooled RF1 solution (1/3 volume of the original culture), and incubated for further 30 min on ice. Cells were then pelleted again, resuspended in pre-cooled RF2 solution (1/20 volume of the original culture), and incubated for further 30 min on ice. Aliquots of 50 µl were prepared in sterile tubes, frozen in liquid nitrogen, and stored at -80°C until use.

RF1 solution			**RF2 solution**		
Ingredients	**Concentration**		**Ingredients**	**Concentration**	
Potassium acetate	30 mM	2.95 g/l	MOPS	10 mM	2.1 g/l
RbCl	100 mM	12 g/l	RbCl	10 mM	1.2 g/l
$CaCl_2$ x 2 H_2O	10 mM	15 g/l	$CaCl_2$ x 2 H_2O	75 mM	11 g/l
Glycerol	15 % (w/v)	150 g/l	Glycerol	15 % (w/v)	150 g/l
$MnCl_2$ x 4 H_2O	50 mM	9.90 g/l			

The pH of both solutions was adjusted to 5.8 with glacial acetic acid and the solutions were sterile filtered.

Transformation of *E. coli*

50 µl of frozen chemically competent *E. coli* cells (see above) were thawed on ice and added 10-20 ng of DNA. The mixture was incubated for 10 min on ice. Cells were heat-shocked at 42°C for 45 sec and incubated another 10 min on ice. 1 ml of LB medium was added and the cells were incubated for 30 min at 37°C under agitation to develop resistance to a given antibiotic by expression of the respective

resistance gene that was introduced with the DNA into the cells. Cells were then pelleted at 8000 x g, collected in approx. 200 µl LB medium and spread on LB plates supplemented with the respective antibiotic. Transformants grew overnight at 37°C.

4.2.3.4 *S. cerevisiae*

Cultivation of *S. cerevisiae*

S. cerevisiae was cultivated in different liquid media (in baffled flasks shaking at 200 rpm or in a rotating wheel) or on solidified media (see 4.1.4.4). Cultures were incubated at 28°C. The cell density of liquid cultures was determined photometrically by measuring the optical density at 600 nm (OD_{600}) with a Novaspec II (Pharmacia Biotech) spectrophotometer or a Genesys 10S UV-Vis spectrophotometer (Thermo Scientific). Cultures were diluted to an OD_{600} in the range of 0.2 and 0.8 for the measurements to guarantee linearity of cell density and OD measured. The respective culture medium was used for blank measurements.

Preparation of Frozen Competent *S. cerevisiae* Cells

The protocol of Gietz & Schiestl, 2007 was employed. A well grown *S. cerevisiae* overnight culture was used to inoculate a culture with an OD_{600} = 0.5. Yeast cells were grown in baffled flasks filled up to only 10-20 % of their max. filling volume with 2 x concentrated YEP medium to enhance growth of the cells. Cells were incubated for 4-6 h at 28°C and 200 rpm until an OD_{600} of 2 was reached. Cells were harvested by centrifugation at 3000 x g for 5 min, resuspended in sterile water (1/2 volume of the original culture), pelleted again, and resuspended again in sterile water (1/100 volume of the original culture). Cells were finally pelleted and resuspended in FCC solution (1/100 volume of the original culture). 50 µl aliquots were then frozen at -80°C. Therefore, tubes were placed into a styrofoam rack, insulated in a larger styrofoam or cardboard box filled with paper or styrofoam chips to achieve slow overnight freezing of the cells. Cells were stored at -80°C until use. 200 ml of culture gave around 60 aliquots.

FCC solution	
Ingredients	**Concentration**
Glycerol	5 % (v/v)
DMSO	10 % (v/v)

Transformation of *S. cerevisiae* and *In Vivo* Cloning via PCR-directed Homologous Recombination

A slightly modified version of the protocol from Gietz & Schiestl, 2007 was employed for transformation of frozen competent yeast cells (see above). A transformation mixture composed of 260 µl PEG 3350, 36 µl lithium acetate, and 10 µl ssDNA per aliquot of cells to be transformed was

prepared; a master mix was prepared when several aliquots had to be transformed. Approx. 50 ng of circular vector DNA in 54 µl of sterile water were added to the mixture. An aliquot of frozen competent yeast cells was thawed and cells were pelleted at 3000 x g for 30 sec. The transformation mixture containing the DNA was added to the cells and mixed thoroughly. Cells were incubated for 60 min at 42°C. Subsequently, cells were pelleted and either directly resuspended in 200 µl water and spread on selective medium (when uracil-auxotrophy was used for selection of transformants) or incubated in 2 ml medium for 4 h at 28°C under agitation (when G-418 was used for selection of transformants) to develop resistance to G-418 prior to spreading on selective medium. Transformants grew after 48 h at 28°C.

Solution	**Concentration**	
PEG 3350	50 % (w/v)	500 g/l
Lithium acetate	1 M	66 g/l
Carrier ssDNA[a]	-	10 mg/ml

[a] Salmon sperm DNA was resuspended in sterile 10 mM Tris-EDTA-buffer (pH8) by stirring at 4°C for several hours and was then stored at -20°C. Before usage the DNA was cooked for 5 min at 96°C and then immediately cooled down on ice, so the DNA remained single stranded.

To achieve the directed integration of DNA fragments into expression vectors, *in vivo* cloning via homologous recombination was utilized (Oldenburg *et al.*, 1997). For this purpose, sequences of 30-50 bp, that are identical to the vector regions upstream and downstream of the desired site of integration, have to be added to the DNA fragment to be integrated. Thus, sequences identical to the promoters phxt7 or pM25 and the tCYC1 terminator were added via PCR to the 5´- and 3´-ends of transporter ORFs cloned in this study, respectively (see primers in section 4.1.2). The vectors were linearized using restriction enzymes that cut close to the site of integration. The PCR fragments as well as the linearized vectors were then gel-purified and used to transform *S. cerevisiae* as described above, following a molar ratio of vector and insert of 1:2-1:5. Integration of inserts was checked via colony-PCR. Selected PCR-products were gel-purified and sequenced.

4.2.4 Methods for the Determination of Transporter Substrates

4.2.4.1 Complementation Experiments

Eight nuclear encoded transporters were tested for their capability in transporting metabolites that are most likely imported to or exported from the chromatophore. For this purpose, the corresponding ORFs were amplified from *P. chromatophora* cDNA and cloned into different expression vectors (see 4.1.1 and 4.1.3). Expression was induced in different transporter knockout strains, each of them unable to import specific metabolites across their plasma membranes. When grown on agar solidified medium supplemented with one of those metabolites as the sole source of either carbon or nitrogen, only such

cells should be able to form colonies, which express a *P. chromatophora* transporter able to facilitate import of the given carbon- or nitrogen compound. In some experiments toxic substances were added to the medium instead. Here the uptake of the substances is displayed by growth inhibition. Thus, complementation of the knockout mutant phenotype (i.e. complete or partial restorage of the wild type phenotype) on a specific substrate by a *P. chromatophora* transporter indicates that the given substance indeed represents a possible natural substrate of that transporter in *P. chromatophora*.

Serial Dilution Experiments with *Synechocystis* sp. PCC6803

20 ml cultures in selective liquid BG11 medium were inoculated with cryoconserved *Synechocystis* cells and grown under standard conditions (see 4.2.3.2) for approx. 1 week to an OD_{750} around 1-2. The OD_{750} was determined carefully and a serial dilution corresponding to OD_{750}s of 0.04 to 0.0004 was prepared in BG11. 2µl of each dilution were dripped under sterile conditions on selective agar-solidified BG11 medium supplemented with putative transporter substrates. Following evaporation of excess liquid from the agar surface, plates were sealed and incubated side by side for 7-20 days at 28°C and approx. 30 µE constant illumination.

Serial Dilution Experiments with *E. coli*

3 ml overnight cultures in selective liquid LB medium were inoculated with *E. coli* cells and grown under standard conditions (see 4.2.3.3). The OD_{600} was determined carefully and a serial dilution corresponding to OD_{600}s of 2.5 to $2.5x10^{-6}$ was prepared in LB medium. 2µl of each dilution were dripped under sterile conditions on selective agar-solidified M9 minimal medium supplemented with putative transporter substrates. Following evaporation of excess liquid from the agar surface, plates were sealed and incubated for 2 days at 28°C.

Serial Dilution Experiments with *S. cerevisiae*

5 ml overnight cultures in selective liquid YEP or SD medium were inoculated with cryoconserved *S. cerevisiae* cells and grown under standard conditions (see 4.2.3.4). The OD_{600} was determined carefully and a serial dilution corresponding to OD_{600}s of 0.2 to 0.0002 was prepared in dextrose-free YEP medium or nitrogen-free SD medium. 2µl of each dilution were dripped under sterile conditions on selective agar-solidified YEP or SD medium supplemented with putative transporter substrates. Following evaporation of excess liquid from the agar surface, plates were sealed and incubated for 2-10 days at 28°C.

4.2.4.2 Extraction of Intracellular Amino Acids from Cyanobacteria and Quantification via HPLC

To compare the uptake of several externally applied amino acids into wildtype and natB-transporter mutants of *Synechocystis*, metabolites were extracted from cultures and the contained amino acids were quantified via HPLC. For this purpose, cultures were inoculated in selective liquid BG11 medium and grown under standard conditions (see 4.2.3.2) to an OD_{750} around 0.8. Each of the cultures was split into two 25 ml cultures and only one of these was supplemented with an amino acid, respectively. The cultures were incubated for another 25 h and the OD_{750} was determined carefully. Cells according to 15 ml of an OD_{750} = 1 were pelleted from each culture at 9000 x g and 4°C for 5 min and washed twice with 7.5 ml of ice-cold BG11 to remove external amino acids. Metabolites were extracted according to Rademacher *et al.*, 2016: Cell pellets were frozen in liquid nitrogen. 1.7 ml of 70 % HPLC-grade ethanol (precooled for 1 h at -20 °C) were added and the cells were vortexed three times for 1 min at full speed. Following each round of vortexing the extracts were cooled on ice. Extracts were cleared by centrifugation at 11,000 x g and 4°C for 2 min and 1.2 ml of each supernatant was transferred to a fresh tube. The liquid metabolite extracts were stored at -80°C.

For HPLC analysis amino acids had to be derivatized to reduce their polarity and thereby increase retention times and improve UV-fluorescence. The AccQ-Tag Ultra derivatization kit (Waters) was used. In a first step extracts were dried in a SpeedVac vacuum centrifuge until all solvent was evaporated. The samples were then properly re-dissolved in 50 µl of HPLC-grade water and kept on ice. 10 µl of each sample was added to a new tube containing 30 µl of borate buffer supplemented with 16.7 µM Norvalin. 10 µl of AccQ-tag Ultra reagent (containing 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate [AQC] in acetonitrile) were added and the mixture was incubated for 10 min protected from light at 55°C and rigorous shaking in a heating block. To remove residual cellular debris samples were centrifuged at 16,000 x g for 2 min, supernatants were loaded on a centrifugal nylon filter (0.2 µm, VWR), and passed through the filter at 14,000 x g for 2 min. Samples were then transferred to HPLC capillaries and stored at 4 °C until analysis.

1 µl of each sample was analyzed in a 1290 Infinity UHPLC (Agilent) using an AccQ-Tag Ultra Column (2.1 x 100 mm, Waters). Elution of derivates was performed with a gradient of 99.9 % → 5 % AccQ-Tag Ultra Eluent A (10 % in water, Waters) and 0.1 % → 95 % Eluent B (100 % acetonitrile) at 0,7 ml/min and 60°C. Derivatized compounds were detected at 260 nm by a diode array detector (Agilent). The following dilutions of 1 nmol/µl amino acid standards (Agilent) were co-derivatized and used for calibration: H_2O, 5 pmol, 10 pmol, 50 pmol, 100 pmol, 200 pmol. For the quantification of tryptophan a mixture of glutamine, asparagine, and tryptophan (dilutions 5-200 pmol) was additionally used as a standard.

4.2.4.3 Uptake Assay for Radiolabeled Glycine

To determine potential uptake of glycine by *P. chromatophora* putative Glycine/Alanine transporter, the uptake of ^{14}C-glycine into natB-mutant *Synechocystis* cells (impaired in glycine import) expressing the transporter was determined over time and compared to wildtype cells. For this purpose, cultures were inoculated in selective liquid BG11 medium with *Synechocystis* cells and grown under standard conditions (see 4.2.3.2). Cells according to 750 µl of an OD_{750} = 1.4 were pelleted from each culture at 5000 x g for 5 min and washed in cold Tricin buffer (25 mM Tricin-NaOH pH 8.1). Cells were resuspended in 750 µl Tricin buffer and 750 µl of Transport mix (Tricin buffer supplemented with 0.4 mM glycine and 10 µl/ml ^{14}C-glycine [stock concentration: 0.1 mCi/ml, specific activity: 115 mCi/mmol]) was added and mixed properly. Thus, the assay was carried out with an OD_{750} = 0.7 of cells and a concentration of 0.2 mM glycine. Cells were incubated at 28°C and ambient light and samples of 200 µl were applied to plain 0.45 µm filters (Supor-450, PALL Life Sciences) after 1, 2, 8, 16, and 32 min. All liquid was removed immediately by a vacuum pump and filters were washed with 10 ml of Tricin buffer under continuous vacuum to remove extracellular glycine from the cells on the filter. The filters were transferred to 10 ml of Rotiszint eco plus scintillation cocktail (Roth) and intracellular ^{14}C radioactivity was quantified in a liquid scintillation spectrometer (Beckman LS 6000 SC). Triplicates of 5 µl of Transport mix were co-analyzed as a standard.

4.2.5 Epon Embedding and Transmission Electron Microscopy (TEM)

P. chromatophora cells or isolated chromatophores were fixed in WARIS-H medium or isolation buffer, respectively, containing 1.25 % glutaraldehyde for 45 min on ice. Post-fixation was done for 30 min at room temperature in 1 % OsO_4 in WARIS-H medium or isolation buffer. Fixed cells and chromatophores were washed, mixed with 14.5 % (w/v) BSA, pelleted, and the pellet fixed with 2.5 % glutaraldehyde for 20 min at room temperature. The fixed pellet was dehydrated in rising concentrations of ethanol (from 60 % to 100 % at -20°C) and then infiltrated with Epon using propylene oxide as a transition solvent. Epon was polymerized at 60°C for 24 h. 70 nm ultrathin sections were prepared, fixed on nickel grids, and contrasted with uranyl acetate and lead citrate according to Reynolds, 1963. A Hitachi H7100 transmission electron microscope (Hitachi, Tokyo, Japan) with Morada camera (EMSIS GmbH, Münster, Germany) operated at 100 kV was used for electron microscopic analysis.

4.2.6 Isolation of Chromatophores

P. chromatophora cells were pelleted at 200 x g and 4°C for 10 min and washed three times with cold isolation buffer. The resulting pellet was resuspended in 30 ml of isolation buffer and cells were broken in a cell disruptor (Constant Systems) at 0.5 kbar and 5°C. To completely recover cellular material from the cell disruptor, the system was washed with 20 ml isolation buffer. Free chromatophores and cell debris were pelleted at 2000 x g and 4°C for 10 min. The pellet was resuspended in 1-3 ml of cold 20 %

Percoll in isolation buffer plus protease inhibitor and loaded on a discontinuous 20-80 % Percoll gradient (material from 1-2 cultures might be loaded per gradient). After centrifugation at 750 x g and 4°C for 15 min, intact chromatophores formed a band at the 80/60 % interphase and were collected with a pipette. To increase yield, the pellet at the bottom of the tube was washed in isolation buffer at 2000 x g and 4°C for 5 min and subjected to another round of cell disruption and chromatophore isolation as described before. Chromatophores from both isolations were pooled, diluted 1:5 in isolation buffer, and pelleted at 5000 x g and 4°C for 10 min; the supernatant was carefully removed with a pipette. To increase purity, isolated chromatophores were re-isolated from another, fresh Percoll gradient and subsequently washed several times in isolation buffer at 5000 x g to remove Percoll. Recovered chromatophores were resuspended in 100-200 µl of isolation buffer. To asses yield and purity, a sample of chromatophores was diluted 1:100 and chromatophores were counted in a hemocytometer (Neubauer improved). Chromatophores were frozen in liquid nitrogen and stored at - 80°C until further use. All steps were carried out at 4°C or on ice and all solutions were pre-cooled to 4°C. One well grown 500 ml *P. chromatophora* culture might yield 1-3 x10^6 chromatophores.

Isolation buffer[a]	
Ingredients	**Concentration**
HEPES pH 7.5	50 mM
sucrose	250 mM
NaCl	125 mM
EGTA	2 mM
$MgCl_2$	2 mM

[a] The buffer was sterile filtered and stored at 4°C until usage. For preparation of Percoll gradients, cell lysis, and all further steps involving free chromatophores, isolation buffer was supplemented with protease inhibitor cocktail (Roche cOmplete).

4.2.7 Analysis of the Chromatophore N-Terminome

The HUNTER method allows for the mass spectrometric determination of protein N-termini, including physiologically relevant proteoforms generated by posttranslational proteolytic processing, in microscale proteome samples (Weng *et al.*, 2019). It is based on whole proteome chemical modification of protein N-termini, followed by tryptic digest and depletion of non-modified tryptic (internal) peptides. Thus, mainly N-terminal peptides are identified in subsequent mass spectrometric analysis and enable mapping of protein processing sites. The method was applied here to determine possible specific cleavage sites in the chromatophore transit peptide (crTP), since plastidial and mitochondrial transit peptides are usually cleaved off following protein translocation across organellar envelopes.

4.2.7.1 Preparation of Chromatophore Lysate

Chromatophore lysate was prepared in triplicates from approx. 6 x 10^6 chromatophores, respectively, isolated from axenic *P. chromatophora* cultures (see 4.2.6). Immediately following isolation, chromatophores were resuspended in 500 µl wash buffer and pelleted by centrifugation at 5000 x g for 5 min in a 1.5 ml tube. 300 µl lysis buffer and approx. 200 µl acid-washed glass beads (0.4-0.6 mm diameter) were added and the mixture was vortexed at full speed for 5 min to lyse chromatophores. Additionally, the mixture was cooked for 10 min at 95°C and again vortexed at full speed for 5 min. The lysate was transferred to a new tube. Glass beads were washed with 100 µl lysis buffer and the fraction was combined with the rest of the lysate. Cell debris and residual glass beads were removed by centrifugation at 20,000 x g for 10 min and the clear lysate was transferred to a new tube. The protein concentration of a 1:3 dilution was determined in a 660 nm assay and the samples were frozen in liquid nitrogen and stored at -80°C until further processing. 200-250 µg of chromatophore lysate protein could be isolated per triplicate. Protein LoBind tubes (Eppendorf) and epT.I.P.S. LoRetention tips (Eppendorf) were used during all steps. Protease inhibitor cocktail (Roche cOmplete) was added to all buffers used.

Wash buffer		**Lysis buffer**	
Ingredients	**Concentration**	**Ingredients**	**Concentration**
HEPES pH 7.5	100 mM	HEPES pH 7.5	100 mM
EDTA	5 mM	EDTA	5 mM
		Guanidine-HCl	6 M

4.2.7.2 High-efficiency Undecanal-based N-Termini Enrichment (HUNTER)

Further processing and mass spectrometric analysis of the samples was carried out by Andreas Perrar at the Central Institute for Engineering, Electronics and Analytics, Analytics department (ZEA-3) of the Forschungszentrum Jülich GmbH according to the following protocol.

In a first step, protein disulfide bonds were reduced by the addition of DTT to a concentration of 10 mM and incubation of the samples at 37°C for 30 min. Reduced cysteine residues were carbamidomethylated by the addition of chloroacetamide to a concentration of 50 mM and incubation at RT for 30 min in the dark to prevent reformation of disulfide bonds. Chloroacetamide was quenched by further addition of DTT to 50 mM final concentration and incubation at RT for 20 min. Reduced proteins were now purified from the lysate samples using SP3 magnetic beads (1:1 mixture of SpeedBead Magnetic Carboxylate Modified Particles 65152105050250 and 45152105050250, GE Healthcare). The samples were added ethanol to 80 % (v/v) to enable protein binding to the beads. 1 µl bead suspension was added per 20 µg of protein and the suspension was shortly mixed in a sonication bath and incubated on a rotary shaker at RT for 20 min. Afterwards, the samples were

placed in a magnetic rack to enable the removal of all liquid while beads stay attached to the tube's wall. The beads were washed twice consecutively with 400 µl of 90 % (v/v) acetonitrile under short sonication and magnetic separation for liquid removal. After the final washing step, beads were air dried for max. 3 min. Proteins were detached from the beads by the addition of 30 µl 100 mM HEPES buffer (pH 7.4) but the beads remained in the samples during the next steps.

N-terminal alpha-amines were now labeled via reductive dimethylation by the addition of formaldehyde isotope CD_2O to a final concentration of 30 mM and cyanoborohydride ($NaBH_3CN$) to 15 mM. The samples were incubated at 37°C for 1 h in a thermomixer with heated lid at 1200 rpm. Another 30 mM formaldehyde and 15 mM cyanoborohydride were added and incubated as before to ensure complete labeling. 500 mM Tris buffer (pH 7.4) was added subsequently to quench the reaction at 37°C and 1200 rpm for 30 min. Proteins were again bound to the SP3 magnetic beads still present in the samples and purified as described earlier.

Proteins were now subjected to tryptic digest to generate peptides that can be identified via mass spectrometry. 30 µl of digest buffer (100 mM HEPES pH7.4, 5 mM $CaCl_2$) and 1 µg of MS approved porcine pancreas trypsin (Serva) per 100 µg of protein were added and the mixture was incubated at 37°C and 1200 rpm overnight inside a thermomixer with heated lid. Notably, cleavages can only occur C-terminal to arginine as cleavages at lysine residues are blocked due to dimethylation of lysine side chains.

The non-dimethylated alpha-amines on the newly generated N-termini of internal tryptic peptides were now hydrophobically tagged using undecanal. Ethanol and undecanal were added to final concentrations of 40 % (v/v) and 50 µg per 1 µg protein, respectively, and mixed gently by inverting. The tagging reaction was started by the addition of cyanoborohydride to 30 mM and the samples were incubated at 37°C and 1200 rpm for 1 h. HR-X spin columns (Macherey-Nagel) containing hydrophobic polystyrene-divinylbenzene copolymer were used to deplete undecanal-labeled peptides from the samples. Sample volume was filled up to 400 µl with 40 % (v/v) ethanol and magnetic beads were rinsed thrice while standing on the magnetic stand. The samples were loaded on HR-X columns (activated beforehand by two additions of 400 µl methanol and equilibrated 2 x with 400 µl 40 % (v/v) ethanol), centrifuged for 1 min at 50 x g, and the eluate was collected. Another 400 µl 40 % (v/v) ethanol were loaded on the columns to elute remaining peptides by centrifugation as described before. Both fractions were combined and the solvent was evaporated in a vacuum centrifuge at 60°C to complete dryness.

Enriched N-terminal peptides were resuspended in 40 µl of 0.1 % (v/v) formic acid ($1 < pH < 3$) and loaded on self-packed double layer C18 stage tips (activated with 20 µl of a mixture of 50 % acetonitrile

and 0.1 % formic acid and equilibrated with 40 µl 0.1 % formic acid). Bound peptides were washed with 50 µl 0.1 % formic acid and eluted with 20 µl of a mixture of 50 % acetonitrile and 0.1 % formic acid. The solvent was evaporated in a vacuum centrifuge at 60°C. Peptides were resuspended in 15 µl 0.1 % formic acid and the peptide concentration was determined photometrically with a NanoDrop 2000c spectral photometer (Thermo Fischer Scientific) at 280 nm against a series of peptide standards.

4.2.7.3 Mass Spectrometric Analysis and Protein Identification

Analysis of samples was performed on a two-column nano-HPLC setup consisting of an Ultimate 3000 nano-RSLC system with a reverse-phase trap column (2 cm µPAC trapping column, PharmaFluidics) and a reverse-phase analytical column (50 cm µPAC column, PharmaFluidics) with a gradient from 2 to 30 % of solution B for 90 min (A: H_2O + 0.1 % formaldehyde, B: acetonitrile + 0.1 % formaldehyde). One complete run lasted for 2 h per sample. Separated samples were introduced into a high-resolution Q-TOF mass spectrometer (Impact II, Bruker) using a nano-spray ion source (CaptiveSpray, Bruker). Data was acquired in line-mode in a mass range from 100 to 1400 m/z at an acquisition rate of 10 Hz using the Bruker HyStar Software (v5.1, Bruker Daltonics). The top 14 most intense ions were selected for fragmentation. Fragmentation spectra were dynamically acquired with a target TIC (= total ion current) of 25k and a minimal frequency of 5 Hz and a maximal frequency of 20 Hz. Fragment spectra were acquired with stepped parameters, each with half of the acquisition time dedicated for each precursor: 80 µs transfer time, 7.5 eV collision energy, and a collision radio frequency (RF) of 1500 Vpp or 120 µs transfer time, 10 eV collision energy, and a collision RF of 1700 Vpp.

Obtained mass spectrometric data was queried in a database search using MaxQuant v1.6.8.0 (MPI for Biochemistry, Planegg, Germany) applying the standard settings for Bruker Q-TOF instruments. Searches were carried out using 60,108 sequences translated from a *P. chromatophora* transcriptome and 867 sequences derived from translated chromatophore ORFs (Singer *et al.*, 2017). A database containing common contaminants, embedded in MaxQuant, was also included in the query. A decoy database was created by enabling the “revert” option. For HUNTER queries dimethylation of lysines and protein N-termini was set as fixed label (+34 Da due to CD_2O). Digestion mode was changed to semispecific (free N-terminus) ArgC and Oxidation (M), acetylation (peptide N-term) and Glu/Gln -> pyro-Glu were set as variable modifications, while carbamidomethylation of cysteines was set as fixed modification. Requantify option was enabled and maximal peptide length for unspecific searches was set to 40 aa. The minimal number of ratio count was set to 1.

4.2.8 Mass Spectrometric Analysis of the Insoluble Chromatophore Proteome

4.2.8.1 Protein Fractionation

CM (chromatophore membrane) and PM (*Paulinella* whole cell membrane) samples

Isolated chromatophores or *P. chromatophora* cells were washed with Buffer I for 5 min at 20,000 x g or three times for 10 min at 200 x g, respectively. Pellets were resuspended in 10 ml Buffer I and broken by two passages in a cell disrupter at 2.4 kbar and 5°C. To completely recover cellular material from the cell disruptor, the system was washed with 10 ml Buffer I. Lysates were supplemented with 500 mM NaCl (final concentration) and passed five times through a 0.6 mm cannula. Cell debris was removed by two successive centrifugations at 15,500 x g and 4°C for 10 min. The supernatant was subjected to ultracentrifugation for 1 h at 150,000 x g and 4°C (Beckmann L-80XL optima ultracentrifuge, Rotor 70.1 Ti at 50,000 rpm). Pellets were rinsed and resuspended in 100 mM Na_2CO_3 pH > 11 and incubated for 1 h, intermitted by 15 passes through a 0.6 mm cannula. Then, insoluble proteins were collected by ultracentrifugation and subsequently rinsed and resuspended in Buffer II by passage through a cannula. Finally, the insoluble fraction was pelleted by ultracentrifugation and solubilized at 36°C in 100 µl or 150 µl Buffer II supplemented with 1 % TritonX-100, 1 % Na-deoxycholate, and 0.1 % SDS. Aliquots for determination of protein concentration, SDS-PAGE analysis, and MS-analysis (the latter two were supplemented with SDS sample buffer) were frozen in liquid nitrogen and stored at -80°C until usage. $50x10^6$ chromatophores isolated from 26 cultures delivered 10-20 µg of CM protein per triplicate. 60-70 µg of PM protein could be isolated per triplicate from three *P. chromatophora* cultures.

CL (chromatophore lysate) samples

Intact isolated chromatophores were washed in Buffer I at 20,000 x g and 4°C for 5 min and resuspended in 270 µl Buffer I. Protein was then extracted by precipitation with 10 % trichloracetic acid, i.e. 30 µl of cold 100 % (w/v) trichloracetic acid was added to the sample and vortexed, followed by incubation for 30 min on ice. Protein was pelleted at 21,000 x g and 4°C for 20 min and pellets were washed twice with ice cold acetone by incubation on ice for 10 min followed by centrifugation for 10 min at 21,000 x g and 4°C, respectively. Pellets were air dried and finally resuspended in 100 µl Buffer II plus detergents at 36°C. $9x10^6$ chromatophores isolated from three cultures delivered 13-20 µg of CL protein per triplicate.

All steps were performed at 4°C or on ice, protease inhibitor cocktail (Roche cOmplete) was added to all buffers used. Protein LoBind tubes (Eppendorf) and epT.I.P.S. LoRetention tips (Eppendorf) were used.

Buffer I		Buffer II		SDS sample buffer	
Ingredients	**Concentration**	**Ingredients**	**Concentration**	**Ingredients**	**Final concentration**
HEPES pH 7.5	50 mM	Tris-HCl pH 7.5	10 mM	Tris-HCl pH 7.0	35 mM
NaCl	125 mM	NaCl	150 mM	Glycerol	7.5 %
EDTA	0.5 mM	EDTA	0.5 mM	SDS	3 %
				DTT	150 mM
				Bromophenol blue	0.005 % (w/v)

4.2.8.2 Mass Spectrometric Analysis and Protein Identification

Sample preparation and subsequent MS/MS analysis of three independent preparations of CM, PM, and CL samples was essentially carried out as described (Singer *et al.*, 2017). Briefly, proteins were in-gel digested in (per sample) 0.1 µg trypsin in 10 mM ammonium hydrogen carbonate overnight at 37°C and resulting peptides resuspended in 0.1 % trifluoroacetic acid. Two independent MS analyses were performed. In MS experiment 1, 500 ng protein per sample, and in MS experiment 2, 500 ng protein per lysate and 1,500 ng protein per membrane sample was analyzed. Peptides were separated on C18 material by liquid chromatography, injected into a QExactive plus mass spectrometer, and the mass spectrometer was operated as described (Singer *et al.*, 2017). Raw files were further processed with MaxQuant (MPI for Biochemistry, Planegg, Germany) for protein identification and quantification using standard parameters. MaxQuant 1.6.2.10 was used for the MS experiment 1 analysis and MaxQuant 1.6.3.4 for MS experiment 2. Searches were carried out using 60,108 sequences translated from a *P. chromatophora* transcriptome and the 867 translated genes predicted on the chromatophore genome (Singer *et al.*, 2017). Peptides and proteins were accepted at a false discovery rate of 1 %. Proteomic data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol *et al.*, 2019) partner repository with the dataset identifier PXD021087.

4.2.8.3 Protein Enrichment Analysis

Intensities of individual proteins were normalized by division of individual intensities in each replicate by the sum of intensities of all proteins identified with ≥ 2 peptides in the same replicate. Each protein was assigned an intensity level representing its log10 transformed mean normalized intensity from three replicates in either fraction added 7 ($log10\ (\overline{normInt}) + 7$), enabling a simple ranking of intensities in a logarithmic range from 0 to 6.

The enrichment factor for each protein in CM as compared to PM or CL samples ($E_{CM/PM}$ or $E_{CM/CL}$, respectively) was calculated as $E_{CM/PM} = \overline{normInt_{CM}} / \overline{normInt_{PM}}$ or $E_{CM/CL} = \overline{normInt_{CM}} / \overline{normInt_{CL}}$ (Tab. S3; missing values [intensity = 0] were excluded from the calculation of means). Proteins with ≥ 3 spectral counts in the chromatophore (i.e. CM+CL fractions) and either $E_{CM/PM}$ >1.5 in at least one of two MS experiments or 0.5< $E_{CM/PM}$ <1.5 in both MS experiments were considered as enriched in

chromatophores (see Fig. A6.3-4). Correspondingly, $E_{CM/CL}$ >1 indicate protein enrichment, $E_{CM/CL}$ <1 depletion in CM samples.

Furthermore, a statistic approach was applied to visualize differences between proteins enriched or exclusively found in a certain fraction. In pairwise comparisons, only proteins were considered showing valid *normInt* values in all three replicates of at least one of the samples being compared. *NormInt* values were log2 transformed and missing values imputed by values from a down shifted normal distribution (width 0.3 SD, down shift 1.8 SD) followed by a pairwise sample comparison based on Student's t-tests and the significance analysis of microarrays algorithm (S_0 = 0.8, FDR 5 %)(Tusher *et al.*, 2001). Differences between individual proteins in CM vs. PM or CM vs. CL samples were calculated as $\overline{log2(normInt_{CM})} - \overline{log2(normInt_{PM})}$ or $\overline{log2(normInt_{CM})} - \overline{log2(normInt_{CL})}$, respectively.

4.2.9 Bioinformatic Analyses

4.2.9.1 Identification and Classification of Transporters

Transporters were classified according to the Transporter Classification Database (Saier *et al.*, 2016). Genes encoding for transporter components in the genomes of the chromatophore (GenBank accession CP000815.1), *Synechococcus* sp. WH5701 (GenBank assembly accession: GCA_000153045.1), and *Synechocystis* sp. PCC6803 (GCA_000340785.1) were predicted using TransAAP, or downloaded from TransportDB 2.0 (Elbourne *et al.*, 2017) for *Arabidopsis thaliana,* and curated manually to estimate the number of individual transport systems and their putative substrates (see Tab. 2.2-1 and S2). Proteomic data from fractionated *Synechocystis* cells was used to identify transport systems localizing to the cyanobacterial plasma membrane (Baers *et al.*, 2019; Liberton *et al.*, 2016; Pisareva *et al.*, 2011). To identify transport systems localizing to the chloroplast envelope a combination of proteomic data from cell/plastid fractionation experiments in *Arabidopsis* (Bouchnak *et al.*, 2019; Bräutigam *et al.*, 2008; Simm *et al.*, 2013) and the SUBAcon database (Hooper *et al.*, 2014; subset "envelope localization experimentally inferred", status January 2020) were used. Genes encoding for β–barrel outer membrane porins in *Arabidopsis* are listed as reported (Breuers *et al.*, 2011; Goetze *et al.*, 2015; Harsman *et al.*, 2016; Wang *et al.*, 2013) or, for *Synechocystis,* inferred by biochemical/cell biological experiments, proteomics or bioinformatic predictions (Huang *et al.*, 2004; Nicolaisen *et al.*, 2009; Qiu *et al.*, 2018). *Synechococcus*-orthologs of *Synechocystis*-transporters were identified using BLAST (cutoff e^{-05}). No outer membrane porins could be identified in the chromatophore genome based on sequence similarity or topology predictions using MCMBB (Bagos *et al.*, 2004).

4.2.9.2 Identification of Sequence Motifs

Sequence motifs were discovered using MEME 5.0.5 algorithm (Bailey and Elkan, 1994) in classic mode. The *P. chromatophora* transcriptome was screened for (i) conserved motifs shown in Fig. 2.1-10 group 2 and 3 and (ii) the degenerate 38 aa motif shown in Fig. 2.2-2 using FIMO 5.0.5 with default settings (Grant *et al.*, 2011). Proteins that contain at least five repeats of the 38 aa motif with a p-value < e^{-10} and / or at least one repeat with a p-value < e^{-20} were considered candidate OPR proteins. Translated transcripts that show significant similarity ($p < e^{-08}$ or $p < 2e^{-07}$) to the conserved motifs in MS-identified group 2 or group 3 proteins were regarded as putative group 2 and 3 proteins, respectively. Occurrences of the group 2 motif in other organisms were identified with PatternSearch (Zimmermann *et al.*, 2018).

5. References

Adams KL and Palmer JD (2003) Evolution of mitochondrial gene content: Gene loss and transfer to the nucleus. *Mol. Phylogenet. Evol.* 29(3): 380–395.

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. *J. Mol. Biol.* 215(3): 403–410.

Ambroggio EE, Separovic F, Bowie JH, Fidelio GD and Bagatolli LA (2005) Direct visualization of membrane leakage induced by the antibiotic peptides: Maculatin, Citropin, and Aurein. *Biophys. J.* 89(3): 1874–1881.

Antelo-Varela M, Bartel J, Quesada-Ganuza A, Appel K, Bernal-Cabas M, Sura T, Otto A, Rasmussen M, Van Dijl JM, Nielsen A, Maaß S and Becher D (2019) Ariadne's Thread in the analytical labyrinth of membrane proteins: Integration of targeted and shotgun proteomics for global absolute quantification of membrane proteins. *Anal. Chem.* 91(18): 11972–11980.

Armbruster U, Hertle A, Makarenko E, Zühlke J, Pribil M, Dietzmann A, Schliebner I, Aseeva E, Fenino E, Scharfenberg M, Voigt C and Leister D (2009) Chloroplast proteins without cleavable transit peptides: Rare exceptions or a major constituent of the chloroplast proteome? *Mol. Plant* 2(6): 1325–1335.

Armenteros JJA, Salvatore M, Emanuelsson O, Winther O, Von Heijne G, Elofsson A and Nielsen H (2019) Detecting sequence signals in targeting peptides using deep learning. *Life Sci. Alliance* 2(5): e201900429.

Atanassova A, Sugita M, Sugiura M, Pajpanova T and Ivanov I (2003) Molecular cloning, expression and characterization of three distinctive genes encoding methionine aminopeptidases in cyanobacterium Synechocystis sp. strain PCC6803. *Arch. Microbiol.* 180(3): 185–193.

Avci FG, Akbulut BS and Ozkirimli E (2018) Membrane active peptides and their biophysical characterization. *Biomolecules* 8(3): 77.

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL and Mori H (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol. Syst. Biol.* 2: 2006.0008.

Baers LL, Breckels LM, Mills LA, Gatto L, Deery MJ, Stevens TJ, Howe CJ, Lilley KS and Lea-Smith DJ (2019) Proteome mapping of a cyanobacterium reveals distinct compartment organization and cell-dispersed metabolism. *Plant Physiol.* 181(4): 1721–1738.

Baggett JJ, Shaw JD, Sciambi CJ, Watson HA and Wendland B (2003) Fluorescent labeling of yeast. *Curr. Protoc. Cell Biol.* 20(1): 4.13.1-4.13.28.

Bagos PG, Liakopoulos TD and Hamodrakas SJ (2004) Finding beta-barrel outer membrane proteins with a markov chain model. *WSEAS Trans. Biol. Biomed.* 2(1): 186–189.

Bailey TL and Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2: 28–36.

Bailey TL and Gribskov M (1998) Combining evidence using p-values: Application to sequence homology searches. *Bioinformatics* 14(1): 48–54.

Barkan A (2011) Expression of plastid genes: Organelle-specific elaborations on a prokaryotic scaffold. *Plant Physiol.* 155(4): 1520–1532.

Barkan A and Small I (2014) Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.* 65(1): 415–442.

Baslam M, Oikawa K, Kitajima-Koga A, Kaneko K and Mitsui T (2016) Golgi-to-plastid trafficking of proteins through secretory pathway: Insights into vesicle-mediated import toward the plastids. *Plant Signal. Behav.* 11(9): 1–5.

Berezhkovskii AM and Bezrukov SM (2018) Stochastic gating as a novel mechanism for channel selectivity. *Biophys. J.* 114(5): 1026–1029.

Berry IJ, Jarocki VM, Tacchi JL, Raymond BBA, Widjaja M, Padula MP and Djordjevic SP (2017) N-terminomics identifies widespread endoproteolysis and novel methionine excision in a genome-reduced bacterial pathogen. *Sci. Rep.* 7: 11063.

Berry IJ, Steele JR and Matthew P (2016) The application of terminomics for the identification of protein start sites and proteoforms in bacteria. *Proteomics* 16: 257–272.

Besnard J, Pratelli R, Zhao C, Sonawala U, Collakova E, Pilot G and Okumoto S (2016) UMAMIT14 is an amino acid exporter involved in phloem unloading in Arabidopsis roots. *J. Exp. Bot.* 67(22): 6385–6397.

Bhattacharya D, Helmchen T and Melkonian M (1995) Molecular evolutionary analyses of nuclear-encoded small subunit ribosomal RNA identify an independent Rhizopod lineage containing the Euglyphina and the Chlorarachniophyta. *J. Eukaryot. Microbiol.* 42(1): 65–69.

Bhushan S, Kuhn C, Berglund AK, Roth C and Glaser E (2006) The role of the N-terminal domain of chloroplast targeting peptides in organellar protein import and miss-sorting. *FEBS Lett.* 580(16): 3966–3972.

Blum H, Beier H and Gross HJ (1987) Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis* 8(2): 93–99.

Bodył A, MacKiewicz P and Gagat P (2012) Organelle evolution: Paulinella breaks a paradigm. *Curr. Biol.* 22(9): 304–306.

Bodył A, Mackiewicz P and Stiller JW (2009) Early steps in plastid evolution: Current ideas and controversies. *BioEssays* 31(11): 1219–1232.

Bodył A, Mackiewicz P and Stiller JW (2010) Comparative genomic studies suggest that the cyanobacterial endosymbionts of the amoeba Paulinella chromatophora possess an import apparatus for nuclear-encoded proteins. *Plant Biol.* 12(4): 639–649.

Boehm E, Zornoza M, Jourdain AA, Magdalena AD, García-Consuegra I, Merino RT, Orduña A, Martín MA, Martinou JC, De La Fuente MA and Simarro M (2016) Role of FAST kinase domains 3 (FASTKD3) in post-transcriptional regulation of mitochondrial gene expression. *J. Biol. Chem.* 291(50): 25877–25887.

Bölter B and Soll J (2016) Once upon a time – Chloroplast protein import research from infancy to future challenges. *Mol. Plant* 9(6): 798–812.

Bonfil DJ, Ronen-Tarazi M, Sültemeyer D, Lieman-Hurwitz J, Schatz D and Kaplan A (1998) A putative HCO3/- transporter in the cyanobacterium Synechococcus sp. strain PCC 7942. *FEBS Lett.* 430(3): 236–240.

Bonissone S, Gupta N, Romine M, Bradshaw RA and Pevzner PA (2013) N-terminal protein processing: A comparative proteogenomic analysis. *Mol. Cell. Proteomics* 12(1): 14–28.

Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ and Madden TL (2012) Domain enhanced lookup time accelerated BLAST. *Biol. Direct* 7: 12.

Bouchnak I, Brugière S, Moyet L, Le Gall S, Salvi D, Kuntz M, Tardif M and Rolland N (2019) Unraveling hidden components of the chloroplast envelope proteome: Opportunities and limits of better MS sensitivity. *Mol. Cell. Proteomics* 18(7): 1285–1306.

Braakman R, Follows MJ and Chisholm SW (2017) Metabolic evolution and the self-organization of ecosystems. *Proc. Natl. Acad. Sci. U. S. A.* 114(15): E3091–E3100.

Bräutigam A, Hoffmann-Benning S and Weber APM (2008) Comparative proteomics of chloroplast envelopes from C3 and C4 plants reveals specific adaptations of the plastid envelope to C4 photosynthesis and candidate proteins required for maintaining C4 metabolite fluxes. *Plant Physiol.* 148(1): 568–579.

Breuers FKH, Bräutigam A and Weber APM (2011) The plastid outer envelope - A highly dynamic interface between plastid and cytoplasm. *Front. Plant Sci.* 2: 97.

Brogden KA (2005) Antimicrobial peptides: Pore formers or metabolic inhibitors in bacteria? *Nat. Rev. Microbiol.* 3(3): 238–250.

Bruce BD (2001) The paradox of plastid transit peptides: Conservation of function despite divergence in primary

structure. *Biochim. Biophys. Acta - Mol. Cell Res.* 1541: 2–21.

Bualuang A, Kageyama H, Tanaka Y, Incharoensakdi A and Takabe T (2015) Functional characterization of a member of alanine or glycine: Cation symporter family in halotolerant cyanobacterium Aphanothece halophytica. *Biosci. Biotechnol. Biochem.* 79(2): 230–235.

Bublitz DAC, Chadwick GL, Magyar JS, Sandoz KM, Brooks DM, Mesnage S, Ladinsky MS, Garber AI, Bjorkman PJ, Orphan VJ and McCutcheon JP (2019) Peptidoglycan production by an insect-bacterial mosaic. *Cell* 179(3): 703-712.e7.

Burdukiewicz M, Sidorczuk K, Rafacz D, Pietluch F, Chilimoniuk J, Rödiger S and Gagat P (2020) Proteomic screening for prediction and design of antimicrobial peptides with ampgram. *Int. J. Mol. Sci.* 21(12): 1–13.

Carro L, Pujic P, Alloisio N, Fournier P, Boubakri H, Hay AE, Poly F, François P, Hocher V, Mergaert P, Balmand S, Rey M, Heddi A and Normand P (2015) Alnus peptides modify membrane porosity and induce the release of nitrogen-rich metabolites from nitrogen-fixing Frankia. *ISME J.* 9(8): 1723–1733.

Cavalier-Smith T (2000) Membrane heredity and early chloroplast evolution. *Trends Plant Sci.* 5(4): 174–182.

Cenci U, Bhattacharya D, Weber APM, Colleoni C, Subtil A and Ball SG (2017) Biotic Host–Pathogen Interactions As Major Drivers of Plastid Endosymbiosis. *Trends Plant Sci.* 22(4): 316–328.

Cenci U, Nitschke F, Steup M, Minassian BA, Colleoni C and Ball SG (2014) Transition from glycogen to starch metabolism in archaeplastida. *Trends Plant Sci.* 19(1): 18–28.

Chen Y, Zhou B, Li J, Tang H, Tang J and Yang Z (2018) Formation and change of chloroplast-located plant metabolites in response to light conditions. *Int. J. Mol. Sci.* 19(3): 654.

Choi JS, Park YH, Oh JH, Kim S, Kwon J and Choi YE (2019) Efficient profiling of detergent-assisted membrane proteome in cyanobacteria. *J. Appl. Phycol.* 32: 1177–1184.

Cline SG, Laughbaum IA and Hamel PP (2017) CCS2, an octatricopeptide-repeat protein, is required for plastid cytochrome c assembly in the green Alga Chlamydomonas reinhardtii. *Front. Plant Sci.* 8: 1306.

Colcombet J, Lopez-Obando M, Heurtevin L, Bernard C, Martin K, Berthomé R and Lurin C (2013) Systematic study of subcellular localization of Arabidopsis PPR proteins confirms a massive targeting to organelles. *RNA Biol.* 10(9): 1557–1575.

Colleoni C, Linka M, Deschamps P, Handford MG, Dupree P, Weber APM and Ball SG (2010) Phylogenetic and biochemical evidence supports the recruitment of an ADP-glucose translocator for the export of photosynthate during plastid endosymbiosis. *Mol. Biol. Evol.* 27(12): 2691–2701.

Crooks G, Hon G, Chandonia J and Brenner S (2004) WebLogo: a sequence logo generator. *Genome Res* 14: 1188–1190.

Dautin N and Bernstein HD (2007) Protein secretion in gram-negative bacteria via the autotransporter pathway. *Annu. Rev. Microbiol.* 61: 89–112.

del Campo EM (2009) Post-transcriptional control of chloroplast gene expression. *Gene Regul. Syst. Bio.* 2009(3): 31–47.

Delaye L, Valadez-Cano C and Pérez-Zamorano B (2016) How really ancient is Paulinella chromatophora? *PLoS Curr.* 8: ecurrents.tol.e68a099364bb1a1e129a17b4e06b0c6b.

Dilworth M V., Piel MS, Bettaney KE, Ma P, Luo J, Sharples D, Poyner DR, Gross SR, Moncoq K, Henderson PJF, Miroux B and Bill RM (2018) Microbial expression systems for membrane proteins. *Methods* 147: 3–39.

Dinh T V., Bienvenut W V., Linster E, Feldman-Salit A, Jung VA, Meinnel T, Hell R, Giglione C and Wirtz M (2015) Molecular identification and functional characterization of the first Nα-acetyltransferase in plastids by global acetylome profiling. *Proteomics* 15(14): 2426–2435.

Dissmeyer N, Rivas S and Graciet E (2018) Life and death of proteins after protease cleavage: protein degradation by the N-end rule pathway. *New Phytol.* 218(3): 929–935.

Dobson L, Reményi I and Tusnády GE (2015) CCTOP: A Consensus Constrained TOPology prediction web server. *Nucleic Acids Res.* 43(W1): W408–W412.

Domenech P, Kobayashi H, Levier K, Walker GC and Barry CE (2009) BacA, an ABC transporter involved in maintenance of chronic murine infections with mycobacterium tuberculosis. *J. Bacteriol.* 191(2): 477–485.

Drath M, Baier K and Forchhammer K (2009) An alternative methionine aminopeptidase, MAP-A, is required for nitrogen starvation and high-light acclimation in the cyanobacterium Synechocystis sp. PCC 6803. *Microbiology* 155(5): 1427–1439.

Drozdetskiy A, Cole C, Procter J and Barton GJ (2015) JPred4: A protein secondary structure prediction server. *Nucleic Acids Res.* 43(W1): W389–W394.

Duncan RP, Husnik F, Leuven JTVAN and Gilbert DG (2014) Dynamic recruitment of amino acid transporters to the insect / symbiont interface. *Mol. Ecol.* 23: 1608–1623.

Duy D, Wanner G, Meda AR, Von Wirén N, Soll J and Philippar K (2007) PIC1, an ancient permease in Arabidopsis chloroplasts, mediates iron transport. *Plant Cell* 19(3): 986–1006.

Eberhard S, Loiselay C, Drapier D, Bujaldon S, Girard-Bascou J, Kuras R, Choquet Y and Wollman FA (2011) Dual functions of the nucleus-encoded factor TDA1 in trapping and translation activation of atpA transcripts in Chlamydomonas reinhardtii chloroplasts. *Plant J.* 67(6): 1055–1066.

Eisenhut M, Bauwe H and Hagemann M (2007) Glycine accumulation is toxic for the cyanobacterium Synechocystis sp . strain PCC 6803 , but can be compensated by supplementation with magnesium ions. *FEMS Microbiol. Lett.* 277: 232–237.

Eitinger T, Rodionov DA, Grote M and Schneider E (2011) Canonical and ECF-type ATP-binding cassette importers in prokaryotes: Diversity in modular organization and cellular functions. *FEMS Microbiol. Rev.* 35(1): 3–67.

Elbourne LDH, Tetu SG, Hassan KA and Paulsen IT (2017) TransportDB 2.0: A database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Res.* 45(D1): D320–D324.

Embley TM and Martin W (2006) Eukaryotic evolution, changes and challenges. *Nature* 440(7084): 623–630.

Eom JS, Chen LQ, Sosso D, Julius BT, Lin IW, Qu XQ, Braun DM and Frommer WB (2015) SWEETs, transporters for intracellular and intercellular sugar translocation. *Curr. Opin. Plant Biol.* 25: 53–62.

Facchinelli F, Colleoni C, Ball SG and Weber APM (2013) (a) Chlamydia, cyanobiont, or host: Who was on top in the ménage à trois? *Trends Plant Sci.* 18(12): 673–679.

Facchinelli F, Pribil M, Oster U, Ebert NJ, Bhattacharya D, Leister D and Weber APM (2013) (b) Proteomic analysis of the Cyanophora paradoxa muroplast provides clues on early events in plastid endosymbiosis. *Planta* 237(2): 637–651.

Facchinelli F and Weber APM (2011) The metabolite transporters of the plastid envelope: An update. *Front. Plant Sci.* 2: 50.

Farkas A, Maróti G, Dürgo H, Györgypál Z, Lima RM, Medzihradszky KF, Kereszt A, Mergaert P and Kondorosi É (2014) Medicago truncatula symbiotic peptide NCR247 contributes to bacteroid differentiation through multiple mechanisms. *PNAS* 111(14): 5183–5188.

Ferramosca A and Zara V (2013) Biogenesis of mitochondrial carrier proteins: Molecular mechanisms of import into mitochondria. *Biochim. Biophys. Acta - Mol. Cell Res.* 1833(3): 494–502.

Finazzi G, Petroutsos D, Tomizioli M, Flori S, Sautron E, Villanova V, Rolland N and Seigneurin-Berny D (2015) Ions channels/transporters and chloroplast regulation. *Cell Calcium* 58(1): 86–97.

Finkenwirth F, Kirsch F and Eitinger T (2013) Solitary bio Y proteins mediate biotin transport into recombinant Escherichia coli. *J. Bacteriol.* 195(18): 4105–4111.

Fischer K (2011) The import and export business in plastids: Transport processes across the inner envelope membrane. *Plant Physiol.* 155(4): 1511–1519.

Fischer WN, Loo DDF, Koch W, Ludewig U, Boorer KJ, Tegeder M, Rentsch D, Wright EM and Frommer WB (2002) Low and high affinity amino acid H+-cotransporters for cellular import of neutral and charged amino acids. *Plant J.* 29(6): 717–731.

Fischer WW, Hemp J and Johnson JE (2016) Evolution of oxygenic photosynthesis. *Annu. Rev. Earth Planet. Sci.* 44(1): 647–683.

Gabr A, Grossman AR and Bhattacharya D (2020) Paulinella, a model for understanding plastid primary endosymbiosis. *J. Phycol.* 56(4): 837–843.

Gabruk M and Mysliwa-Kurdziel B (2015) Light-dependent protochlorophyllide oxidoreductase: Phylogeny, regulation, and catalytic properties. *Biochemistry* 54(34): 5255–5262.

Gagat P and Mackiewicz P (2014) Protein translocons in photosynthetic organelles of Paulinella chromatophora. *Acta Soc. Bot. Pol.* 83(4): 399–407.

Garg SG and Gould SB (2016) The role of charge in protein targeting evolution. *Trends Cell Biol.* 26(12): 894–905.

Gautier R, Douguet D, Antonny B and Drin G (2008) HELIQUEST: A web server to screen sequences with specific α-helical properties. *Bioinformatics* 24(18): 2101–2102.

Gietz RD and Schiestl RH (2007) Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* 2(1): 1–4.

Giglione C, Boularot A and Meinnel T (2004) Protein N-terminal methionine excision. *Cell. Mol. Life Sci.* 61(12): 1455–1474.

Giglione C, Fieulaine S and Meinnel T (2015) N-terminal protein modifications: Bringing back into play the ribosome. *Biochimie* 114: 134–146.

Giglione C and Meinnel T (2001) Organellar peptide deformylases: Universality of the N-terminal methionine cleavage mechanism. *Trends Plant Sci.* 6(12): 566–572.

Giménez-Mascarell P, González-Recio I, Fernández-Rodríguez C, Oyenarte I, Müller D, Martínez-Chantar ML and Martínez-Cruz LA (2019) Current structural knowledge on the CNNM family of magnesium transport mediators. *Int. J. Mol. Sci.* 20(5): 1135.

Gisin J, Müller A, Pfänder Y, Leimkühler S, Narberhaus F and Masepohl B (2010) A Rhodobacter capsulatus member of a universal permease family imports molybdate and other oxyanions. *J. Bacteriol.* 192(22): 5943–5952.

Goetze TA, Patil M, Jeshen I, Bölter B, Grahl S and Soll J (2015) Oep23 forms an ion channel in the chloroplast outer envelope. *BMC Plant Biol.* 15: 47.

Gohlke U, Pullan L, McDevitt CA, Porcelli I, De Leeuw E, Palmer T, Saibil HR and Berks BC (2005) The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter. *Proc. Natl. Acad. Sci. U. S. A.* 102(30): 10482–10486.

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J and Conesa A (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36(10): 3420–3435.

Götze M, Iacobucci C, Ihling CH and Sinz A (2019) A simple cross-linking/mass spectrometry work flow for studying system-wide protein interactions. *Anal. Chem.* 91: 10236−10244.

Gould SB, Waller RF and McFadden GI (2008) Plastid Evolution. *Annu. Rev. Plant Biol.* 59(1): 491–517.

Grant CE, Bailey TL and Noble WS (2011) FIMO: Scanning for occurrences of a given motif. *Bioinformatics* 27(7): 1017–1018.

Gray MW (2017) Lynn Margulis and the endosymbiont hypothesis: 50 years later. *Mol. Biol. Cell* 28(10): 1285–1287.

Grice AL, Kerr ID and Sansom MSP (1997) Ion channels formed by HIV-1 Vpu : a modelling and simulation study.

FEBS Lett. 405(3): 299–304.

Gross J and Bhattacharya D (2009) (a) Mitochondrial and plastid evolution in eukaryotes: An outsiders' perspective. *Nat. Rev. Genet.* 10(7): 495–505.

Gross J and Bhattacharya D (2009) (b) Revaluating the evolution of the Toc and Tic protein translocons. *Trends Plant Sci.* 14(1): 13–20.

Guefrachi I, Pierre O, Timchenko T, Alunni B, Barrière Q, Czernic P, Villaécija-Aguilar JA, Verly C, Bourge M, Fardoux J, Mars M, Kondorosi E, Giraud E and Mergaert P (2015) Bradyrhizobium BclA is a peptide transporter required for bacterial differentiation in symbiosis with aeschynomene legumes. *Mol. Plant-Microbe Interact.* 28(11): 1155–1166.

Guha S, Ghimire J, Wu E and Wimley WC (2019) Mechanistic landscape of membrane-permeabilizing peptides. *Chem. Rev.* 119(9): 6040–6085.

Hagemann M and Zuther E (1992) Selection and characterization of mutants of the cyanobacterium Synechocystis sp. PCC 6803 unable to tolerate high salt concentrations. *Arch. Microbiol.* 158(6): 429–434.

Hall TA (1999) *BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT*. *Nucleic Acids Symp. Ser.* 41p.: 95–98.

Hammani K, Bonnard G, Bouchoucha A, Gobert A, Pinker F, Salinas T and Giegé P (2014) Helical repeats modular proteins are major players for organelle gene expression. *Biochimie* 100(1): 141–150.

Hamsanathan S and Musser SM (2018) The Tat protein transport system: Intriguing questions and conundrums. *FEMS Microbiol. Lett.* 365(12): 1–11.

Harsman A, Schock A, Hemmis B, Wahl V, Jeshen I, Bartsch P, Schlereth A, Pertl-Obermeyer H, Goetze TA, Soll J, Philippar K and Wagner R (2016) OEP40, a regulated glucose-permeable β-barrel solute channel in the chloroplast outer envelope membrane. *J. Biol. Chem.* 291(34): 17848–17860.

Hatanaka T, Huang W, Wang H, Sugawara M, Prasad PD, Leibach FH and Ganapathy V (2000) Primary structure, functional characteristics and tissue expression pattern of human ATA2, a subtype of amino acid transport system A. *Biochim. Biophys. Acta - Biomembr.* 1467(1): 1–6.

Heins L, Mehrle A, Hemmler R, Wagner R, Küchler M, Hörmann F, Sveshnikov D and Soll J (2002) The preprotein conducting channel at the inner envelope membrane of plastids. *EMBO J.* 21(11): 2616–2625.

Heldt HW and Rapley L (1970) Specific transport of inorganic phosphate, 3-phosphoglycerate and triosephosphates across the inner membrane of the envelope in spinach chloroplasts. *FEBS Lett.* 10(3): 143–148.

Hooper CM, Tanz SK, Castleden IR, Vacher MA, Small ID and Millar AH (2014) SUBAcon: A consensus algorithm for unifying the subcellular localization data of the Arabidopsis proteome. *Bioinformatics* 30(23): 3356–3364.

Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ and Nakai K (2007) WoLF PSORT: Protein localization predictor. *Nucleic Acids Res.* 35(SUPPL.2): 585–587.

Hoshiyasu S, Kohzuma K, Yoshida K, Fujiwara M, Fukao Y, Akiho Yokota and Akashi K (2013) Potential Involvement of N-Terminal Acetylation in the Quantitative Regulation of the ε Subunit of Chloroplast ATP Synthase under Drought Stress. *Biosci. Biotechnol. Biochem.* 77(5): 998–1007.

Huang F, Hedman E, Funk C, Kieselbach T, Schröder WP and Norling B (2004) Isolation of outer membrane of Synechocystis sp. PCC 6803 and its proteomic characterization. *Mol. Cell. Proteomics* 3(6): 586–595.

Huertas MJ, López-Maury L, Giner-Lamia J, Sánchez-Riego AM and Florencio FJ (2014) Metals in cyanobacteria: Analysis of the copper, nickel, cobalt and arsenic homeostasis mechanisms. *Life* 4(4): 865–886.

Huesgen PF, Alami M, Lange PF, Foster LJ, Schröder WP, Overall CM and Green BR (2013) Proteomic amino-termini profiling reveals targeting information for protein import into complex plastids. *PLoS One* 8(9): e74483.

Ito T, Uozumi N, Nakamura T, Takayama S, Matsuda N, Aiba H, Hemmi H and Yoshimura T (2009) The implication of yggT of Escherichia coli in osmotic regulation. *Biosci. Biotechnol. Biochem.* 73(12): 2698–2704.

Jack DL, Yang NM and Saier MH (2001) The drug/metabolite transporter superfamily. *Eur. J. Biochem.* 268(13): 3620–3639.

Jackson C, Clayden S and Reyes-Prieto A (2015) The Glaucophyta: The blue-green plants in a nutshell. *Acta Soc. Bot. Pol.* 84(2): 149–165.

Joce C, Wiener A and Yin H (2011) Transmembrane domain oligomerization propensity determined by ToxR assay. *J. Vis. Exp.* 51: e2721.

Kaneko K, Takamatsu T, Inomata T, Oikawa K, Itoh K, Hirose K, Amano M, Nishimura SI, Toyooka K, Matsuoka K, Pozueta-Romero J and Mitsui T (2016) N-glycomic and microscopic subcellular localization analyses of NPP1, 2 and 6 strongly indicate that trans-golgi compartments participate in the golgi to plastid traffic of nucleotide pyrophosphatase/phosphodiesterases in rice. *Plant Cell Physiol.* 57(8): 1610–1628.

Karkar S, Facchinelli F, Price DC, Weber APM and Bhattacharya D (2015) Metabolic connectivity as a driver of host and endosymbiont integration. *Proc. Natl. Acad. Sci. U. S. A.* 112(33): 10208–10215.

Katoh A, Sonoda M, Katoh H and Ogawa T (1996) Absence of light-induced proton extrusion in a cotA-less mutant of Synechocystis sp. Strain PCC6803. *J. Bacteriol.* 178(18): 5452–5455.

Kearse MG and Wilusz JE (2017) Non-AUG translation : a new start for protein synthesis in eukaryotes. *Genes Dev.* 31: 1717–1731.

Keeling PJ (2010) The endosymbiotic origin, diversification and fate of plastids. *Philos. Trans. R. Soc. B Biol. Sci.* 365(1541): 729–748.

Kelley LA, Mezulis S, Yates CM, Wass MN and Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10(6): 845–858.

Khalid S and Sansom MSP (2006) Molecular dynamics simulations of a bacterial autotransporter: NalP from Neisseria meningitidis. *Mol. Membr. Biol.* 23(6): 499–508.

Kies L (1974) Elektronenmikroskopische Untersuchungen an Paulinella chromatophora Lauterborn, einer Thekamöbe mit blau-grünen Endosymbionten (Cyanellen). *Protoplasma* 80(1–3): 69–89.

Kies L and Kremer BP (1979) Function of cyanelles in the thecamoeba Paulinella chromatophora. *Naturwissenschaften* 66(11): 578–579.

Kim J, Na YJ, Park SJ, Baek SH and Kim DH (2019) Biogenesis of chloroplast outer envelope membrane proteins. *Plant Cell Rep.* 38(7): 783–792.

Kim S and Park MG (2016) Paulinella longichromatophora sp. nov., a new marine photosynthetic testate amoeba containing a chromatophore. *Protist* 167(1): 1–12.

Kitajima A, Asatsuma S, Okada H, Hamada Y, Kaneko K, Nanjo Y, Kawagoe Y, Toyooka K, Matsuoka K, Takeuchi M, Nakano A and Mitsui T (2009) The rice α-amylase glycoprotein is targeted from the golgi apparatus through the secretory pathway to the plastids. *Plant Cell* 21(9): 2844–2858.

Kiyota H, Hirai M and Ikeuchi M (2014) NblA1/A2-dependent homeostasis of amino acid pools during nitrogen starvation in Synechocystis sp. PCC 6803. *Metabolites* 4(3): 517–531.

Klausen MS, Jespersen MC, Nielsen H, Jensen KK, Jurtz VI, Sønderby CK, Sommer MOA, Winther O, Nielsen M, Petersen B and Marcatili P (2019) NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning. *Proteins Struct. Funct. Bioinforma.* 87(6): 520–527.

Kleffmann T, Russenberger D, Von Zychlinski A, Christopher W, Sjölander K, Gruissem W and Baginsky S (2004) The Arabidopsis thaliana chloroplast proteome reveals pathway abundance and novel protein functions. *Curr. Biol.* 14(5): 354–362.

Kleine T (2012) Arabidopsis thaliana mTERF proteins: Evolution and functional classification. *Front. Plant Sci.* 3: 233.

Kleinknecht L, Wang F, Stübe R, Philippar K, Nickelsen J and Bohne AV (2014) RAP, the sole octotricopeptide repeat protein in Arabidopsis, is required for chloroplast 16S rRNA maturation. *Plant Cell* 26(2): 777–787.

Klemens PAW, Patzke K, Deitmer J, Spinner L, Le Hir R, Bellini C, Bedu M, Chardon F, Krapp A and Ekkehard Neuhaus H (2013) Overexpression of the vacuolar sugar carrier AtSWEET16 modifies germination, growth, and stress tolerance in Arabidopsis. *Plant Physiol.* 163(3): 1338–1352.

Knopp M, Garg SG, Handrich M and Gould SB (2020) Major changes in plastid protein import and the origin of the Chloroplastida. *iScience* 23(3): 100896.

Köhler D, Dobritzsch D, Hoehenwarter W, Helm S, Steiner JM and Baginsky S (2015) Identification of protein N-termini in Cyanophora paradoxa cyanelles: Transit peptide composition and sequence determinants for precursor maturation. *Front. Plant Sci.* 6: 559.

Kolaj-Robin O, Russell D, Hayes KA, Pembroke JT and Soulimane T (2015) Cation diffusion facilitator family: Structure and function. *FEBS Lett.* 589(12): 1283–1295.

Kouyianou K, Bock P De, Colaert N, Nikolaki A, Aktoudianaki A, Gevaert K and Tsiotis G (2012) Proteome profiling of the green sulfur bacterium Chlorobaculum tepidum by N-terminal proteomics. *Proteomics* 12: 63–67.

Kovács-Bogdán E, Benz JP, Soll J and Bölter B (2011) Tic20 forms a channel independent of Tic110 in chloroplasts. *BMC Plant Biol.* 11(1): 133.

Kozik P, Francis RW, Seaman MNJ and Robinson MS (2010) A screen for endocytic motifs. *Traffic* 11(6): 843–855.

Krogh A, Larsson B, Von Heijne G and Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* 305(3): 567–580.

Kumar S, Rubino FA, Mendoza AG and Ruiz N (2019) The bacterial lipid II flippase MurJ functions by an alternating-access mechanism. *J. Biol. Chem.* 294(3): 981–990.

Kusnetsov V V. (2018) Chloroplasts: Structure and expression of the plastid genome. *Russ. J. Plant Physiol.* 65(4): 465–476.

Labarre J, Thuriaux P and Chauvat F (1987) Genetic analysis of amino acid transport in the facultatively heterotrophic cyanobacterium Synechocystis sp. strain 6803. *J. Bacteriol.* 169(10): 4668–4673.

Lagarde D, Beuf L and Vermaas W (2000) Increased production of Zeaxanthin and other pigments by application of genetic engineering techniques to Synechocystis sp. strain PCC 6803. *Appl. Environ. Microbiol.* 66(1): 64–72.

Lange PF and Overall CM (2013) Protein TAILS: When termini tell tales of proteolysis and function. *Curr. Opin. Chem. Biol.* 17(1): 73–82.

Laudenbach DE and Grossman AR (1991) Characterization and mutagenesis of sulfur-regulated genes in a cyanobacterium: Evidence for function in sulfate transport. *J. Bacteriol.* 173(9): 2739–2750.

Lauterborn R (1895) Protozoenstudien II. Paulinella chromatophora nov. gen., nov. spec., ein beschalter Rhizopode des Süßwassers mit blaugrünen chromatophorenartigen Einschlüssen. *Zeitschrift für wissenschaftliche Zool.* 59: 537–544.

Lee FCY and Ule J (2018) Advances in CLIP technologies for studies of protein-RNA interactions. *Mol. Cell* 69(3): 354–369.

Lee I and Hong W (2004) RAP – a putative RNA-binding domain. *Trends Biochem. Sci.* 29(11): 567–570.

Lee JM, Cho CH, Park SI, Choi JW, Song HS, West JA, Bhattacharya D and Yoon HS (2016) Parallel evolution of highly conserved plastid genome architecture in red seaweeds and seed plants. *BMC Biol.* 14(1): 1–16.

Leister D (2016) Towards understanding the evolution and functional diversification of DNA-containing plant organelles. *F1000Research* 5: 330.

Lhee D, Yang EC, Kim JI, Nakayama T, Zuccarello G, Andersen RA and Yoon HS (2017) Diversity of the photosynthetic Paulinella species, with the description of Paulinella micropora sp. nov. and the

chromatophore genome sequence for strain KR01. *Protist* 168(2): 155–170.

Liberton M, Saha R, Jacobs JM, Nguyen AY, Gritsenko MA, Smith RD, Koppenaal DW and Pakrasi HB (2016) Global proteomic analysis reveals an exclusive role of thylakoid membranes in bioenergetics of a model cyanobacterium. *Mol. Cell. Proteomics* 15(6): 2021–2032.

Linka N and Esser C (2012) Transport proteins regulate the flux of metabolites and cofactors across the membrane of plant peroxisomes. *Front. Plant Sci.* 3: 3.

Linka N, Hurka H, Lang BF, Burger G, Winkler HH, Stamme C, Urbany C, Seil I, Kusch J and Neuhaus HE (2003) Phylogenetic relationships of non-mitochondrial nucleotide transport proteins in bacteria and eukaryotes. *Gene* 306: 27–35.

Linster E and Wirtz M (2018) N-terminal acetylation: An essential protein modification emerges as an important regulator of stress responses. *J. Exp. Bot.* 69(19): 4555–4568.

Lister DL, Bateman JM, Purton S and Howe CJ (2003) DNA transfer from chloroplast to nucleus is much rarer in Chlamydomonas than in tobacco. *Gene* 316: 33–38.

Login FH, Balmand S, Vallier A, Vincent-Monégat C, Vigneron A, Weiss-Gayet M, Rochat D and Heddi A (2011) Antimicrobial peptides keep insect endosymbionts under control. *Science* 334: 362–365.

Long W and Cheeseman CI (2015) Structure of, and functional insight into the GLUT family of membrane transporters. *Cell Health Cytoskelet.* 7: 167–183.

Lurin C, Andres C, Aubourg, Sebastien Bellaoui M, Bitton F, Bruyere C, Caboche M, Debast C, Gualberto J, Hoffmann B, Lecharny A, Le Ret M, Martin-Magniette M-L, Mireau H, Peeters N, Renou J-P, Szurek B, Taconnat L and Ian S (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* 16: 2089–2103.

Ma J, Lei HT, Reyes FE, Sanchez-Martinez S, Sarhan MF, Hattne J and Gonen T (2019) Structural basis for substrate binding and specificity of a sodium-alanine symporter AgcS. *Proc. Natl. Acad. Sci. U. S. A.* 116(6): 2086–2090.

Maaß S, Sievers S, Zühlke D, Kuzinski J, Sappa PK, Muntel J, Hessling B, Bernhardt J, Sietmann R, Völker U, Hecker M and Becher D (2011) Efficient, global-scale quantification of absolute protein amounts by integration of targeted mass spectrometry and two-dimensional gel-based proteomics. *Anal. Chem.* 83(7): 2677–2684.

Maaß S, Wachlin G, Bernhardt J, Eymann C, Fromion V, Riedel K, Becher D and Hecker M (2014) Highly precise quantification of protein molecules per cell during stress and starvation responses in bacillus subtilis. *Mol. Cell. Proteomics* 13(9): 2260–2276.

Mackenzie B, Schäfer MKH, Erickson JD, Hedige MA, Weihe E and Varoqui H (2003) Functional properties and cellular distribution of the system A glutamine transporter SNAT1 support specialized roles in central neurons. *J. Biol. Chem.* 278(26): 23720–23730.

Mackiewicz P, Bodył A and Gagat P (2012) (a) Possible import routes of proteins into the cyanobacterial endosymbionts/plastids of Paulinella chromatophora. *Theory Biosci.* 131(1): 1–18.

Mackiewicz P, Bodył A and Gagat P (2012) (b) Protein import into the photosynthetic organelles of Paulinella chromatophora and its implications for primary plastid endosymbiosis. *Symbiosis* 58(1–3): 99–107.

MacLeod PR and MacLeod RA (1986) Cloning in Escherichia coli K-12 of a Na+-dependent transport system from a marine bacterium. *J. Bacteriol.* 165(3): 825–830.

Maloney PC, Ambudkar S V., Anantharam V, Sonna LA and Varadhachary A (1990) Anion-exchange mechanisms in bacteria. *Microbiol. Rev.* 54(1): 1–17.

Mancusso R, Gregorio GG, Liu Q and Wang DN (2012) Structure and mechanism of a bacterial sodium-dependent dicarboxylate transporter. *Nature* 491(7425): 622–626.

Marin B, Nowack ECM, Glöckner G and Melkonian M (2007) The ancestor of the Paulinella chromatophore obtained a carboxysomal operon by horizontal gene transfer from a Nitrococcus-like γ-proteobacterium. *BMC Evol. Biol.* 7: 1–14.

Marin B, Nowack ECM and Melkonian M (2005) A plastid in the making: Evidence for a second primary endosymbiosis. *Protist* 156(4): 425–432.

Markovich D (2012) Sodium-sulfate/carboxylate cotransporters (SLC13). In Bevensee MO (ed.) *Current Topics in Membranes*. Elsevier, 239–256.

Martin WF, Garg S and Zimorski V (2015) Endosymbiotic theories for eukaryote origin. *Philos. Trans. R. Soc. B Biol. Sci.* 370: 20140330.

Matouschek A (2003) Protein unfolding - An important process in vivo? *Curr. Opin. Struct. Biol.* 13(1): 98–109.

Matsuda N, Kobayashi H, Katoh H, Ogawa T, Futatsugi L, Nakamura T, Bakker EP and Uozumi N (2004) Na+-dependent K+ uptake Ktr system from the cyanobacterium Synechocystis sp. PCC 6803 and its role in the early phases of cell adaptation to hyperosmotic shock. *J. Biol. Chem.* 279(52): 54952–54962.

McCutcheon JP (2016) From microbiology to cell biology: when an intracellular bacterium becomes part of its host cell. *Curr. Opin. Cell Biol.* 41: 132–136.

McFadden GI (2014) Origin and evolution of plastids and photosynthesis in eukaryotes. *Cold Spring Harb. Perspect. Biol.* 6(4): a016105.

McFadden GI and Melkonian M (1986) Use of Hepes buffer for microalgal culture media and fixation for electron microscopy. *Phycologia* 25(4): 551–557.

Medeiros JM, Böck D and Pilhofer M (2018) Imaging bacteria inside their host by cryo-focused ion beam milling and electron cryotomography. *Curr. Opin. Microbiol.* 43: 62–68.

Mehrshahi P, Stefano G, Andaloro JM, Brandizzi F, Froehlich JE and DellaPenna D (2013) Transorganellar complementation redefines the biochemical continuity of endoplasmic reticulum and chloroplasts. *Proc. Natl. Acad. Sci. U. S. A.* 110(29): 12126–12131.

Melkonian M and Mollenhauer D (2005) Robert Lauterborn (1869-1952) and his Paulinella chromatophora. *Protist* 156(2): 253–262.

Menéndez N, Braña AF, Salas JA and Méndez C (2007) Involvement of a chromomycin ABC transporter system in secretion of a deacetylated precursor during chromomycin biosynthesis. *Microbiology* 153(9): 3061–3070.

Mereschkowsky K (1905) Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. *Biol. Cent.* 25(18): 593–604.

Mergaert P (2018) Role of antimicrobial peptides in controlling symbiotic bacterial populations. *Nat. Prod. Rep.* 35: 336–356.

Mergaert P, Kikuchi Y, Shigenobu S and Nowack ECM (2017) Metabolic integration of bacterial endosymbionts through antimicrobial peptides. *Trends Microbiol.* 25(9): 703–712.

Mergaert P, Uchiumi T, Alunni B, Evanno G, Cheron A, Catrice O, Mausset AE, Barloy-Hubler F, Galibert F, Kondorosi A and Kondorosi E (2006) Eukaryotic control on bacterial cell cycle and differentiation in the Rhizobium-legume symbiosis. *Proc. Natl. Acad. Sci. U. S. A.* 103(13): 5230–5235.

Micallef L and Rodgers P (2014) euler APE: Drawing area-proportional 3-Venn diagrams using ellipses. *PLoS One* 9(7): e101717.

Montesinos ML, Muro-Pastor AM, Herrero A and Flores E (1998) Ammonium/methylammonium permeases of a cyanobacterium. *J. Biol. Chem.* 273(47): 31463–31470.

Moore BC and Leigh JA (2005) Markerless mutagenesis in Methanococcus maripaludis demonstrates roles for alanine dehydrogenase, alanine racemase, and alanine permease. *J. Bacteriol.* 187(3): 972–979.

Moore DT, Berger BW and Degrado WF (2008) Protein-protein interactions in the membrane: Sequence, structural, and biological motifs. *Structure* 16: 991–1001.

Mukaihara T and Tamura N (2009) Identification of novel Ralstonia solanacearum type III effector proteins through translocation analysis of hrp B-regulated gene products. *Microbiology* 155(7): 2235–2244.

Mulnaes D, Porta N, Clemens R, Apanasenko I, Reiners J, Gremer L, Neudecker P, Smits SHJ and Gohlke H (2020) TopModel: Template-based protein structure prediction at low sequence identity using top-down consensus and deep neural networks. *J. Chem. Theory Comput.* 16: 1953–1967.

Mumberg D, Muller R and Funk M (1994) Regulatable promoters of saccharomyces cerevisiae: Comparison of transcriptional activity and their use for heterologous expression. *Nucleic Acids Res.* 22(25): 5767–5768.

Muntel J, Fromion V, Goelzer A, Maaß S, Mäder U, Büttner K, Hecker M and Becher D (2014) Comprehensive absolute quantification of the cytosolic proteome of bacillus subtilis by data independent, parallel fragmentation in liquid chromatography/mass spectrometry (LC/MSE). *Mol. Cell. Proteomics* 13(4): 1008–1019.

Nakai M (2018) New perspectives on chloroplast protein import. *Plant Cell Physiol.* 59(6): 1111–1119.

Nakamura T, Katoh Y, Shimizu Y, Matsuba Y and Unemoto T (1996) Cloning and sequencing of novel genes from Vibrio alginolyticus that support the growth of K+ uptake-deficient mutant of Escherichia coli. *Biochim. Biophys. Acta - Bioenerg.* 1277(3): 201–208.

Nakanishi T, Kekuda R, Fei YJ, Hatanaka T, Sugawara M, Martindale RG, Leibach FH, Prasad PD and Ganapathy V (2001) Cloning and functional characterization of a new subtype of the amino acid transport system N. *Am. J. Physiol. - Cell Physiol.* 281: 1757–1768.

Nakayama T and Ishida K ichiro (2009) Another acquisition of a primary photosynthetic organelle is underway in Paulinella chromatophora. *Curr. Biol.* 19(7): 284–285.

Nanjo Y, Oka H, Ikarashi N, Kaneko K, Kitajima A, Mitsui T, Muñoz FJ, Rodríguez-López M, Baroja-Fernandez E and Pozueta-Romero J (2006) Rice plastidial N-glycosylated nucleotide pyrophosphatase/phosphodiesterase is transported from the ER-golgi to the chloroplast through the secretory pathway. *Plant Cell* 18(10): 2582–2592.

Neuhoff V, Philipp K, Zimmer H-G and Mesecke S (1979) A simple, versatile, sensitive and volume-independent method for quantitative protein determination which is independent of other external influences. *Hoppe-Seyler´s Zeitschrift für Physiol. Chemie* 360(2): 1657–1670.

Nicolaisen K, Hahn A and Schleiff E (2009) The cell wall in heterocyst formation by Anabaena sp. PCC 7120. *J. Basic Microbiol.* 49(1): 5–24.

Nomura M and Ishida KI (2016) Fine-structural observations on siliceous scale production and shell assembly in the testate amoeba Paulinella chromatophora. *Protist* 167(4): 303–318.

Nomura M, Nakayama T and Ishida KI (2014) Detailed process of shell construction in the photosynthetic testate amoeba Paulinella chromatophora (euglyphid, rhizaria). *J. Eukaryot. Microbiol.* 61(3): 317–321.

Nowack ECM (2014) Paulinella chromatophora - Rethinking the transition from endosymbiont to organelle. *Acta Soc. Bot. Pol.* 83(4): 387–397.

Nowack ECM and Grossman AR (2012) Trafficking of protein into the recently established photosynthetic organelles of Paulinella chromatophora. *Proc. Natl. Acad. Sci. U. S. A.* 109(14): 5340–5345.

Nowack ECM, Melkonian M and Glöckner G (2008) Chromatophore genome sequence of Paulinella sheds light on acquisition of photosynthesis by eukaryotes. *Curr. Biol.* 18(6): 410–418.

Nowack ECM, Price DC, Bhattacharya D, Singer A, Melkonian M and Grossman AR (2016) Gene transfers from diverse bacteria compensate for reductive genome evolution in the chromatophore of Paulinella chromatophora. *Proc. Natl. Acad. Sci. U. S. A.* 113(43): 12214–12219.

Nowack ECM, Vogel H, Groth M, Grossman AR, Melkonian M and Glöckner G (2011) Endosymbiotic gene transfer and transcriptional regulation of transferred genes in Paulinella chromatophora. *Mol. Biol. Evol.* 28(1): 407–422.

Nowack ECM and Weber APM (2018) Genomics-informed insights into endosymbiotic organelle evolution in photosynthetic eukaryotes. *Annu. Rev. Plant Biol.* 69(1): 51–84.

Nugent T and Jones DT (2012) Detecting pore-lining regions in transmembrane protein sequences. *BMC*

Bioinformatics 13: 169.

Oberleitner L, Ehret G and Nowack ECM (2019) Mikrobielle Symbiosen und die Evolution neuer Organellen. *BioSpektrum* 25(3): 268–270.

Oberleitner L, Poschmann G, Macorano L, Schott-Verdugo S, Gohlke H, Stühler K and Nowack ECM (2020) The puzzle of metabolite exchange and identification of putative octotrico peptide repeat expression regulators in the nascent photosynthetic organelles of Paulinella chromatophora. *Front. Microbiol.* 11: 607182.

Oh YJ and Hwang I (2015) Targeting and biogenesis of transporters and channels in chloroplast envelope membranes: Unsolved questions. *Cell Calcium* 58(1): 122–130.

Okazaki S, Okabe S, Higashi M, Shimoda Y, Sato S, Tabata S, Hashiguchi M, Akashi R, Göttfert M and Saeki K (2010) Identification and functional analysis of type III effector proteins in Mesorhizobium Loti. *Mol. Plant-Microbe Interact.* 23(2): 223–234.

Okazaki S, Okabe S, Zehner S, Göttfert M and Saeki K (2008) Symbiotic roles and transcriptional analysis of the type III secretion system in Mesorhizobium loti. In Dakora FD Chimphango SBM Valentine AJ Elmerich C and Newton WE (eds) *Biological Nitrogen Fixation: Towards Poverty Alleviation through Sustainable Agriculture*. Springer, 235.

Oldenburg KR, Vo KT, Michaelis S and Paddon C (1997) Recombination-mediated PCR-directed plasmid construction in vivo in yeast. *Nucleic Acids Res.* 25(2): 451–452.

Omata T, Andriesse X and Hirano A (1993) Identification and characterization of a gene cluster involved in nitrate transport in the cyanobacterium Synechococcus sp. PCC7942. *MGG Mol. Gen. Genet.* 236(2–3): 193–202.

Park SY and Guo X (2014) Adaptor protein complexes and intracellular transport. *Biosci. Rep.* 34(4): 381–390.

Parsons HT, Christiansen K, Knierim B, Carroll A, Ito J, Batth TS, Smith-Moritz AM, Morrison S, Mcinerney P, Hadi MZ, Auer M, Mukhopadhyay A, Petzold CJ, Scheller H V., Loqué D and Heazlewood JL (2012) Isolation and proteomic characterization of the arabidopsis golgi defines functional and novel components involved in plant cell wall biosynthesis. *Plant Physiol.* 159(1): 12–26.

Paulsen IT, Nguyen L, Sliwinski MK, Rabus R and Saier MH (2000) Microbial genome analyses: Comparative transport capabilities in eighteen prokaryotes. *J. Mol. Biol.* 301(1): 75–100.

Peeters N, Carrère S, Anisimova M, Plener L, Cazalé AC and Genin S (2013) Repertoire, unified nomenclature and evolution of the Type III effector gene set in the Ralstonia solanacearum species complex. *BMC Genomics* 14: 859.

Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M, Pérez E, Uszkoreit J, Pfeuffer J, Sachsenberg T, Yilmaz Ş, Tiwary S, Cox J, Audain E, Walzer M, Jarnuczak AF, Ternent T, Brazma A and Vizcaíno JA (2019) The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* 47(D1): D442–D450.

Pisareva T, Kwon J, Oh J, Kim S, Ge C, Wieslander Å, Choi JS and Norling B (2011) Model for membrane organization and protein sorting in the cyanobacterium synechocystis sp. PCC 6803 inferred from proteomics and multivariate sequence analyses. *J. Proteome Res.* 10(8): 3617–3631.

Plohnke N, Seidel T, Kahmann U, Rögner M, Schneider D and Rexroth S (2015) The proteome and lipidome of Synechocystis sp. PCC 6803 cells grown under light-activated heterotrophic conditions. *Mol. Cell. Proteomics* 14(3): 572–584.

Pohland AC and Schneider D (2019) Mg2+ homeostasis and transport in cyanobacteria - at the crossroads of bacterial and chloroplast Mg2+ import. *Biol. Chem.* 400(10): 1289–1301.

Pootakham W, Gonzalez-Ballester D and Grossman AR (2010) Identification and regulation of plasma membrane sulfate transporters in Chlamydomonas. *Plant Physiol.* 153(4): 1653–1668.

Popot JL and De Vitry C (1990) On the microassembly of integral membrane proteins. *Annu. Rev. Biophys. Biophys. Chem.* 19: 369–403.

Preker PJ and Keller W (1998) The HAT helix, a repetitive motif implicated in RNA processing. *Trends Biochem.*

Sci. 23(1): 15–16.

Price DC, Chan CX, Yoon HS, Yang EC, Qiu H, Weber APM, Schwacke R, Gross J, Blouin NA, Lane C, Reyes-Prieto A, Durnford DG, Neilson JAD, Lang BF, Burger G, Steiner JM, Löffelhardt W, Meuser JE, Posewitz MC, Ball S, Arias MC, Henrissat B, Coutinho PM, Rensing SA, Symeonidi A, Doddapaneni H, Green BR, Rajah VD, Boore J and Bhattacharya D (2012) Cyanophora paradoxa genome elucidates origin of photosynthesis in algae and plants. *Science* 335(6070): 843–847.

Price DRG, Duncan RP, Shigenobu S and Wilson ACC (2011) Genome expansion and differential expression of amino acid transporters at the aphid / buchnera symbiotic interface. *Mol. Biol. Evol.* 28(11): 3113–3126.

Price GD, Woodger FJ, Badger MR, Howitt SM and Tucker L (2004) Identification of a SulP-type bicarbonate transporter in marine cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* 101(52): 18228–18233.

Qiu G-W, Lou W-J, Sun C-Y, Yang N, Li Z-K, Li D-L, Zang S-S, Fu F-X, Hutchins DA, Jiang H-B and Qiu B-S (2018) Outer membrane iron uptake pathways in the model cyanobacterium Synechocystis sp. strain PCC 6803. *Appl. Environ. Microbiol.* 84(19): e01512-18.

Qiu H, Price DC, Weber APM, Facchinelli F, Yoon HS and Bhattacharya D (2013) Assessing the bacterial contribution to the plastid proteome. *Trends Plant Sci.* 18(12): 680–687.

Quintero MJ, Montesinos ML, Herrero A and Flores E (2001) Identification of genes encoding amino acid permeases by inactivation of selected ORFs from the Synechocystis genomic sequence. *Genome Res.* 11(12): 2034–2040.

Rademacher N, Kern R, Fujiwara T, Mettler-Altmann T, Miyagishima SY, Hagemann M, Eisenhut M and Weber APM (2016) Photorespiratory glycolate oxidase is essential for the survival of the red alga Cyanidioschyzon merolae under ambient CO2 conditions. *J. Exp. Bot.* 67(10): 3165–3175.

Rahaman A and Lazaridis T (2014) A thermodynamic approach to Alamethicin pore formation. *Biochim. Biophys. Acta* 1838(1): 98–105.

Rahire M, Laroche F, Cerutti L and Rochaix JD (2012) Identification of an OPR protein involved in the translation initiation of the PsaB subunit of photosystem i. *Plant J.* 72(4): 652–661.

Rahnamaeian M, Cytryńska M, Zdybicka-Barabas A, Dobslaff K, Wiesner J, Twyman RM, Zuchner T, Sadd BM, Regoes RR, Schmid-Hempel P and Vilcinskas A (2015) Insect antimicrobial peptides show potentiating functional interactions against Gram-negative bacteria. *Proc. R. Soc. B Biol. Sci.* 282:201502

Raimunda D, Gonzalez-Guerrero M, Leeber III BW and Argüello JM (2011) The transport mechanism of bacterial Cu+ -ATPases: distinct efflux rates adapted to different function. *Bone* 24(3): 467–475.

Rasmussen B, Bekker A and Fletcher IR (2013) Correlation of Paleoproterozoic glaciations based on U-Pb zircon ages for tuff beds in the Transvaal and Huronian Supergroups. *Earth Planet. Sci. Lett.* 382: 173–180.

Rasmussen T (2016) How do mechanosensitive channels sense membrane tension? *Biochem. Soc. Trans.* 44(4): 1019–1025.

Raymond J and Segre D (2006) The effect of oxygen on biochemical networks and the evolution of complex life. *Science* 311(5768): 1764–1767.

Rebsamen M, Pochini L, Stasyk T, De Araújo MEG, Galluccio M, Kandasamy RK, Snijder B, Fauster A, Rudashevskaya EL, Bruckner M, Scorzoni S, Filipek PA, Huber KVM, Bigenzahn JW, Heinz LX, Kraft C, Bennett KL, Indiveri C, Huber LA and Superti-Furga G (2015) SLC38A9 is a component of the lysosomal amino acid sensing machinery that controls mTORC1. *Nature* 519(7544): 477–481.

Reifenberger E, Boles E and Ciriacy M (1997) Kinetic characterization of individual hexose transporters of Saccharomyces cerevisiae and their relation to the triggering mechanisms of glucose repression. *Eur. J. Biochem.* 245(2): 324–333.

Reinders A, Panshyshyn JA and Ward JM (2005) Analysis of transport activity of Arabidopsis sugar alcohol permease homolog AtPLT5. *J. Biol. Chem.* 280(2): 1594–1602.

Ren Q and Paulsen IT (2005) Comparative analyses of fundamental differences in membrane transport

capabilities in prokaryotes and eukaryotes. *PLoS Comput. Biol.* 1(3): e27.

Reyes-Prieto A, Yoon HS, Moustafa A, Yang EC, Andersen RA, Boo SM, Nakayama T, Ishida KI and Bhattacharya D (2010) Differential gene retention in plastids of common recent origin. *Mol. Biol. Evol.* 27(7): 1530–1537.

Reynolds CR, Islam SA and Sternberg MJE (2018) EzMol: A web server wizard for the rapid visualization and image production of protein and nucleic acid structures. *J. Mol. Biol.* 430(15): 2244–2248.

Reynolds ES (1963) The use of lead citrate at high pH as an electron-opaque stain in electron microscopy. *J. Cell Biol.* 17(1): 208–212.

Rhie MN, Yoon HE, Oh HY, Zedler S, Unden G and Kim O Bin (2014) A Na+-coupled C4-dicarboxylate transporter (Asuc_0304) and aerobic growth of Actinobacillus succinogenes on C4-dicarboxylates. *Microbiol. (United Kingdom)* 160: 1533–1544.

Richardson LG and Schnell DJ (2019) Origins, function and regulation of the TOC-TIC general protein import machinery of plastids. *J. Exp. Bot.* 71(4): 1226–1238.

Richter S and Lamppa GK (1998) A chloroplast processing enzyme functions as the general stromal processing peptidase. *Proc. Natl. Acad. Sci. U. S. A.* 95(13): 7463–7468.

Richter S, Zhong R and Lamppa G (2005) Function of the stromal processing peptidase in the chloroplast import pathway. *Physiol. Plant.* 123(4): 362–368.

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

Ris H and Plaut W (1962) Ultrastructure of DNA-containing areas in the chloroplast of Chlamydomonas. *J. Cell Biol.* 13: 383–391.

Rodionov DA, Hebbeln P, Eudes A, Ter Beek J, Rodionova IA, Erkens GB, Slotboom DJ, Gelfand MS, Osterman AL, Hanson AD and Eitinger T (2009) A novel class of modular transporters for vitamins in prokaryotes. *J. Bacteriol.* 91(1): 42–51.

Rodríguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Löffelhardt W, Bohnert HJ, Philippe H and Lang BF (2005) Monophyly of primary photosynthetic eukaryotes: Green plants, red algae, and glaucophytes. *Curr. Biol.* 15(14): 1325–1330.

Rolland N, Dorne A-J, Amoroso G, Sültemeyer DF, Joyard J and Rochaix J-D (1997) Disruption of the plastid ycf10 open reading frame affects uptake of inorganic carbon in the chloroplast of Chlamydomonas. *EMBO J.* 16(22): 6713–6726.

Rouanet C and Nasser W (2001) The PecM protein of the phytopathogenic bacterium Erwinia chrysanthemi, membrane topology and possible involvement in the efflux of the blue pigment indigoidine. *J. Mol. Microbiol. Biotechnol.* 3(2): 309–318.

Rowland E, Kim J, Bhuiyan NH and Van Wijk KJ (2015) The Arabidopsis chloroplast stromal n-terminome: Complexities of amino-terminal protein maturation and stability. *Plant Physiol.* 169(3): 1881–1896.

Sagan/Margulis L (1967) On the origin of mitosing cells. *J. Theor. Biol.* 14(3): 255–274.

Saidijam M, Azizpour S and Patching SG (2018) Comprehensive analysis of the numbers, lengths and amino acid compositions of transmembrane helices in prokaryotic, eukaryotic and viral integral membrane proteins of high-resolution structure. *J. Biomol. Struct. Dyn.* 36(2): 443–464.

Saier MH, Reddy VS, Tsu B V., Ahmed MS, Li C and Moreno-Hagelsieb G (2016) The Transporter Classification Database (TCDB): Recent advances. *Nucleic Acids Res.* 44(D1): D372–D379.

Sasaki Y, Sekiguchi K, Nagano Y and Matsuno R (1993) Chloroplast envelope protein encoded by chloroplast genome. *FEBS Lett.* 316(1): 93–98.

Sato N, Yoshitomi T and Mori-Moriyama N (2020) Characterization and biosynthesis of lipids in Paulinella micropora MYN1. Evidence for efficient integration of chromatophores into cellular lipid metabolism. *Plant Cell Physiol.* 61(5): 869–881.

Sazuka T, Yamaguchi M and Ohara O (1999) Cyano2Dbase updated: Linkage of 234 protein spots to corresponding genes through N-terminal microsequencing. *Electrophoresis* 20(11): 2160–2171.

Schmidt A, Kochanowski K, Vedelaar S, Ahrné E, Volkmer B, Callipo L, Knoops K, Bauer M, Aebersold R and Heinemann M (2016) The quantitative and condition-dependent Escherichia coli proteome. *Nat. Biotechnol.* 34(1): 104–110.

Schweppe DK, Chavez JD, Lee CF, Caudal A, Kruse SE, Stuppard R, Marcinek DJ, Shadel GS, Tian R and Bruce JE (2017) Mitochondrial protein interactome elucidated by chemical cross-linking mass spectrometry. *PNAS* 114(7): 1732–1737.

Seigneurin-Berny D, Gravot A, Auroy P, Mazard C, Kraut A, Finazzi G, Grunwald D, Rappaport F, Vavasseur A, Joyard J, Richaud P and Rolland N (2006) HMA1, a new Cu-ATPase of the chloroplast envelope, is essential for growth under adverse light conditions. *J. Biol. Chem.* 281(5): 2882–2892.

Sestok AE, Linkous RO and Smith AT (2018) Toward a mechanistic understanding of Feo-mediated ferrous iron uptake. *Metallomics* 10(7): 887–898.

Shibata M, Katoh H, Sonoda M, Ohkawa H, Shimoyama M, Fukuzawa H, Kaplan A and Ogawa T (2002) Genes essential to sodium-dependent bicarbonate transport in cyanobacteria: Function and phylogenetic analysis. *J. Biol. Chem.* 277(21): 18658–18664.

Simm S, Papasotiriou DG, Ibrahim M, Leisegang MS, Müller B, Schorge T, Karas M, Mirus O, Sommer MS and Schleiff E (2013) Defining the core proteome of the chloroplast envelope membranes. *Front. Plant Sci.* 4: 11.

Singer A, Poschmann G, Mühlich C, Valadez-Cano C, Hänsch S, Hüren V, Rensing SA, Stühler K and Nowack ECM (2017) Massive protein import into the early-evolutionary-stage photosynthetic organelle of the amoeba Paulinella chromatophora. *Curr. Biol.* 27(18): 2763-2773.e5.

Small I, Peeters N, Legeai F and Lurin C (2004) Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* 4(6): 1581–1590.

Smart OS, Neduvelil JG, Wang X, Wallace BA and Sansomt MSP (1996) HOLE : A program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* 14: 354–360.

Solis-Escalante D, van den Broek M, Kuijpers NGA, Pronk JT, Boles E, Daran JM and Daran-Lapujade P (2015) The genome sequence of the popular hexose-transport-deficient Saccharomyces cerevisiae strain EBY.VW4000 reveals LoxP/Cre-induced translocations and gene loss. *FEMS Yeast Res.* 15(2): fou004.

Sommer MS and Schleiff E (2014) Protein targeting and transport as a necessary consequence of increased cellular complexity. *Cold Spring Harb. Perspect. Biol.* 6(8): a016055.

Song J, Ji C and Zhang JZH (2014) Insights on Na+ binding and conformational dynamics in multidrug and toxic compound extrusion transporter NorM. *Proteins* 82: 240–249.

Soppa J (2010) Protein acetylation in archaea, bacteria, and eukaryotes. *Archaea* 2010: 820681.

Stal LJ and Moezelaar R (1997) Fermentation in cyanobacteria. *FEMS Microbiol. Rev.* 21(2): 179–211.

Stegemann S, Hartmann S, Ruf S and Bock R (2003) High-frequency gene transfer from the chloroplast genome to the nucleus. *Proc. Natl. Acad. Sci. U. S. A.* 100(15): 8828–8833.

Sutherland MC, Tran NL, Tillman DE, Jarodsky JM, Yuan J and Kranz RG (2018) Structure-function analysis of the bifunctional CcsBA heme exporter and cytochrome c synthetase. *MBio* 9(6): e02134-18.

Tardif M, Atteia A, Specht M, Cogne G, Rolland N, Brugière S, Hippler M, Ferro M, Bruley C, Peltier G, Vallon O and Cournac L (2012) Predalgo: A new subcellular localization prediction tool dedicated to green algae. *Mol. Biol. Evol.* 29(12): 3625–3639.

Teese MG and Langosch D (2015) Role of GxxxG motifs in transmembrane domain interactions. *Biochemistry* 54(33): 5125–5135.

Teixeira PF and Glaser E (2013) Processing peptidases in mitochondria and chloroplasts. *Biochim. Biophys. Acta* -

Mol. Cell Res. 1833(2): 360–370.

Teng YS, Su YS, Chen LJ, Lee YJ, Hwang I and Li HM (2006) Tic21 is an essential translocon component for protein translocation across the chloroplast inner envelope membrane. *Plant Cell* 18(9): 2247–2257.

Teper D, Burstein D, Salomon D, Gershovitz M, Pupko T and Sessa G (2016) Identification of novel Xanthomonas euvesicatoria type III effector proteins by a machine-learning approach. *Mol. Plant Pathol.* 17(3): 398–411.

Thompson JD, Higgins DG and Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22(22): 4673–4680.

Tian Q, Taupin J, Elledge S, Robertson M and Anderson P (1995) Fas-activated serine / threonine kinase (FAST) phosphorylates TIA-1 during Fas-mediated apoptosis. *J. exp. Med.* 182: 865–874.

Timmis JN, Ayliff MA, Huang CY and Martin W (2004) Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* 5(2): 123–135.

Tourasse NJ, Choquet Y and Vallon O (2013) PPR proteins of green algae. *RNA Biol.* 10(9): 1526–1542.

Trevisson E, Burlina A, Doimo M, Pertegato V, Casarin A, Cesaro L, Navas P, Basso G, Sartori G and Salviati L (2009) Functional complementation in yeast allows molecular characterization of missense argininosuccinate lyase mutations. *J. Biol. Chem.* 284(42): 28926–28934.

Tripp J, Essl C, Iancu C V., Boles E, Choe JY and Oreb M (2017) Establishing a yeast-based screening system for discovery of human GLUT5 inhibitors and activators. *Sci. Rep.* 7(1): 1–9.

Tsunekawa K, Shijuku T, Hayashimoto M, Kojima Y, Onai K, Morishita M, Ishiura M, Kuroda T, Nakamura T, Kobayashi H, Sato M, Toyooka K, Matsuoka K, Omata T and Uozumi N (2009) Identification and characterization of the Na+/H+ antiporter NhaS3 from the thylakoid membrane of Synechocystis sp. PCC 6803. *J. Biol. Chem.* 284(24): 16513–16521.

Tusher VG, Tibshirani R and Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U. S. A.* 98(9): 5116–5121.

Vaccaro L, Scott KA and Sansom MSP (2008) Gating at both ends and breathing in the middle : conformational dynamics of TolC. *Biophys. J.* 95(12): 5681–5691.

Valadez-Cano C, Olivares-Hernández R, Resendis-Antonio O, DeLuna A and Delaye L (2017) Natural selection drove metabolic specialization of the chromatophore in Paulinella chromatophora. *BMC Evol. Biol.* 17: 99.

Valladares A, Montesinos ML, Herrero A and Flores E (2002) An ABC-type, high-affinity urea permease identified in cyanobacteria. *Mol. Microbiol.* 43(3): 703–715.

Van de Velde W, Zehirov G, Szatmari A, Debreczeny M, Ishihara H, Kevei Z, Farkas A, Mikulass K, Nagy A, Tiricz H, Satiat-Jeunemaître B, Alunni B, Bourge M, Kucho K, Abe M, Kereszt A, Maroti G, Uchiumi T, Kondorosi E and Mergaert P (2010) Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science* 327: 1122–1125.

Varland S, Osberg C and Arnesen T (2015) N-terminal modifications of cellular proteins: The enzymes involved, their substrate specificities and biological effects. *Proteomics* 15(14): 2385–2401.

Villarejo A, Burén S, Larsson S, Déjardin A, Monné M, Rudhe C, Karlsson J, Jansson S, Lerouge P, Rolland N, von Heijne G, Grebe M, Bako L and Samuelsson G (2005) Evidence for a protein transported through the secretory pathway en route to the higher plant chloroplast. *Nat. Cell Biol.* 7(12): 1124–1131.

Waditee R, Hossain GS, Tanaka Y, Nakamura T, Shikata M, Takano J, Takabe T and Takabe T (2004) Isolation and functional characterization of Ca2+/H+ antiporters from cyanobacteria. *J. Biol. Chem.* 279(6): 4330–4338.

Wagner S, Baarst L, Ytterberg AJ, Klussmerer A, Wagner CS, Nord O, Nygren PÅ, Van Wijks KJ and De Gier JW (2007) Consequences of membrane protein overexpression in Escherichia coli. *Mol. Cell. Proteomics* 6(9): 1527–1550.

Wang Y, Chen CH, Hu D, Ulmschneider MB and Ulmschneider JP (2016) Spontaneous formation of structurally

diverse membrane channel architectures from a single antimicrobial peptide. *Nat. Commun.* 7:13535

Wang Z, Anderson NS and Benning C (2013) The phosphatidic acid binding site of the Arabidopsis trigalactosyldiacylglycerol 4 (TGD4) protein required for lipid import into chloroplasts. *J. Biol. Chem.* 288(7): 4763–4771.

Weber A and Flügge UI (2002) Interaction of cytosolic and plastidic nitrogen metabolism in plants. *J. Exp. Bot.* 53(370): 865–874.

Weber A, Servaites JC, Geiger DR, Kofler H, Hille D, Gröner F, Hebbeker U and Flügge UI (2000) Identification, purification, and molecular cloning of a putative plastidic glucose translocator. *Plant Cell* 12(5): 787–801.

Weber APM, Linka M and Bhattacharya D (2006) Single, ancient origin of a plastid metabolite translocator family in Plantae from an endomembrane-derived ancestor. *Eukaryot. Cell* 5(3): 609–612.

Weber APM and Linka N (2011) Connecting the plastid: Transporters of the plastid envelope and their role in linking plastidial with cytosolic metabolism. *Annu. Rev. Plant Biol.* 62(1): 53–77.

Weber APM, Schwacke R and Flügge U-I (2005) Solute transporters of the plastid envelope membrane. *Annu. Rev. Plant Biol.* 56(1): 133–164.

Weng SSH, Demir F, Ergin EK, Dirnberger S, Uzozie A, Tuscher D, Nierves L, Tsui J, Huesgen PF and Lange PF (2019) Sensitive determination of proteolytic proteoforms in limited microscale proteome samples. *Mol. Cell. Proteomics* 18(11): 2335–2347.

Westrich LD, Gotsmann, Vincent Leon Herkt C, Ries F, Kazek T, Trösch R, Ramundo S, Nickelsen J, Armbruster L, Wirtz M, Storchová Z, Raeschle M and Willmund F forthcoming. The proteomic inventory reveals the chloroplast ribosome as nexus within a diverse protein network. *under Revis.*

Wieczorke R, Krampe S, Weierstall T, Freidel K, Hollenberg CP and Boles E (1999) Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in Saccharomyces cerevisiae. *FEBS Lett.* 464(3): 123–128.

Wienk HLJ, Czisch M and De Kruijff B (1999) The structural flexibility of the preferredoxin transit peptide. *FEBS Lett.* 453(3): 318–326.

Wimley WC (2010) Describing the mechanism of antimicrobial peptide action with the interfacial activity model. *ACS Chem. Biol.* 5(10): 905–917.

Woodson JD and Chory J (2008) Coordination of gene expression between organellar and nuclear genomes. *Nat. Rev. Genet.* 9(5): 383–395.

Wyant GA (2017) mTORC1 activator SLC38A9 is required to efflux essential amino acids from lysosomes and use protein as a nutrient. *Cell* 171(3): 642–654.

Yan J, Yao Y, Hong S, Yang Y, Shen C, Zhang Q, Zhang D, Zou T and Yin P (2019) Delineation of pentatricopeptide repeat codes for target RNA prediction. *Nucleic Acids Res.* 47(7): 3728–3738.

Yan Y and Huang S (2019) CHDOCK : a hierarchical docking approach for modeling Cn symmetric homo-oligomeric complexes. *Biophys. Reports* 5(2): 65–72.

Yang MK, Yang YH, Chen Z, Zhang J, Lin Y, Wang Y, Xiong Q, Li T, Ge F, Bryant DA and Zhao JD (2014) Proteogenomic analysis and global discovery of posttranslational modifications in prokaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 111(52): E5633–E5642.

Yin P, Li Q, Yan C, Liu Y, Liu J, Yu F, Wang Z, Long J, He J, Wang HW, Wang J, Zhu JK, Shi Y and Yan N (2013) Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* 504(7478): 168–171.

Yoon HS, Nakayama T, Reyes-Prieto A, Andersen RA, Boo SM, Ishida KI and Bhattacharya D (2009) A single origin of the photosynthetic organelle in different Paulinella lineages. *BMC Evol. Biol.* 9(1): 1–11.

Yoon HS, Reyes-Prieto A, Melkonian M and Bhattacharya D (2006) Minimal plastid genome evolution in the Paulinella endosymbiont. *Curr. Biol.* 16(17): 670–672.

Zakataeva NP, Kutukova EA, Gronskiy S V., Troshin P V., Livshits VA and Aleshin V V. (2006) Export of metabolites by the proteins of the DMT and RhtB families and its possible role in intercellular communication. *Microbiology* 75(4): 438–448.

Zhang R, Nowack ECM, Price DC, Bhattacharya D and Grossman AR (2017) Impact of light intensity and quality on chromatophore and nuclear gene expression in Paulinella chromatophora, an amoeba with nascent photosynthetic organelles. *Plant J.* 90(2): 221–234.

Zhong R, Wan J, Jin R and Lamppa G (2003) A pea antisense gene for the chloroplast stromal processing peptidase yields seedling lethals in Arabidopsis: Survivors show defective GFP import in vivo. *Plant J.* 34(6): 802–812.

Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN and Alva V (2018) A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* 430(15): 2237–2243.

Zulkifli L, Akai M, Yoshikawa A, Shimojima M, Ohta H, Guy HR and Uozumi N (2010) The KtrA and KtrE subunits are required for Na+-dependent K + uptake by KtrB across the plasma membrane in Synechocystis sp. strain PCC 6803. *J. Bacteriol.* 192(19): 5063–5070.

6. Appendix

The enclosed disk contains supplementing Excel spreadsheets S1-S6, as well as an electronic version of this thesis.

6.1 Supplemental Tables

Table S1 N-terminal peptides of chromatophore-encoded and nuclear encoded proteins identified in the chromatophore. N-terminal peptides identified in chromatophore lysate are listed separately for chromatophore-encoded and nuclear encoded proteins. Each peptide is assigned a unique peptide ID. Peptides matching the same unique N-terminus share the same N-terminus ID. The sequence and chemical modification are listed for each peptide. It is also indicated, which modifications have been identified for each unique N-terminus. It is indicated, whether an N-terminus is classified as canonical or non-canonical. The exact position of the identified peptide relative to the protein model is provided, as well as the amino acids present in the P1 position (amino acid preceding the identified peptide) and P1´position (N-terminal amino acid of the identified peptide). For quantification, the summed peptide spectrum matches and the log10 LFQ intensity mean are given. The MaxQuant score (MQ score) is provided, with values > 40 indicating solid peptide identification and values > 100 indicating very solid identification. The protein a peptide matches to is indicated, as well as its annotation (performed with Blast2GO [Götz *et al.,* 2008] with an E value cutoff of e^{-3} and curated manually) and length. The number of unique N-termini and the modifications that were identified for a protein are given. For nuclear encoded proteins, additional information is provided on whether the full-length 5´-sequence of a CDS is available (5´-FL) as indicated by the presence of a spliced leader (SL) sequence at the 5'-end of the transcript or an in frame stop codon upstream of the start ATG. Furthermore, the table shows the class a protein belongs to according to its length (L, long [> 250 aa]; S, short [< 95 aa]; l, likely long), whether a chromatophore transit peptide could be identified on its N-terminus (crTP; see Singer *et al.,* 2017), and whether the protein is regarded an import candidate (MS-identified or predicted, according to Singer *et al.*, 2017 and Oberleitner *et al.*, 2020). For crTP-proteins, the positional range of cleavage sites (see Fig. 2.1-3A and D) is provided. Finally, it is indicated for chromatophore-encoded proteins, whether peptides matching to two or more different positional ranges of a protein (see Fig. 2.1-2A) have been identified, whether canonical N-termini have been identified for a protein, and whether detection of canonical peptides might have been abolished due to lack of trypsin cleavage sites in the N-terminal protein sequence.

Table S2 Transporters and transport systems for ions and metabolites that could be identified in the chromatophore genome, the genome of *Synechocystis* sp. 6803, the genome of *Synechococcus* sp. WH5701, and the nuclear genome of *Arabidopsis thaliana* (only transporters localizing to chloroplasts). Proteins involved in membrane transport were predicted in the chromatophore genome (GenBank accession CP000815.1), the genome of *Synechocystis* sp. 6803 (GenBank assembly accession: GCA_000340785.1), the genome of *Synechococcus* sp. WH5701 (GenBank assembly accession: GCA_000153045.1, AANO00000000.1) with TransAAP (Elbourne *et al.,* 2017), or downloaded from TransportDB 2.0 for *Arabidopsis* nuclear encoded, chloroplast

localized proteins (status January 2020; Elbourne *et al.,* 2017), classified according to the Transporter Classification Database (TCDB; Saier *et al.,* 2016), and curated manually. The TCDB-Number, abbreviated and full family name are provided and for each protein the name of the most similar characterized protein is indicated. For proteins that contribute to multi-subunit transport systems a function as membrane spanning permease component, peripheral component or ABC component is depicted. Subunits that could not be assigned to a transport system are in grey font and were excluded from Figure 2.2-1. OM-pores were added manually to the list and literature providing information on the transported substrate and mode of function is cited. The table also lists the putative substrate and the substrate group as depicted in Figure 2.2-1 for each transporter. The mean normalized ion intensity for each transporter identified in chromatophore membrane samples in this study and the corresponding intensity level depicted in Figure 2.2-7E is provided. For the chromatophore, the putative sub-organellar localization as deduced from the experimentally determined sub-cellular localization of cyanobacterial orthologs or other well-studied members of the respective TCDB-family is indicated. For *Synechocystis*, the putative sub-cellular localization as deduced from proteomics on fractionated cells (Pisareva *et al.,* 2011; Liberton *et al.,* 2016; Baers *et al.,* 2019) is indicated. For *Synechococcus*, all proteins were blasted against the *Synechocystis* sp. PCC6803 genome, followed by reciprocal blast. The most similar *Synechocystis*-protein identified with BlastP (cutoff e^{-05}) was used to deduce the putative sub-cellular localization of individual *Synechococcus*-proteins or whole transport systems. For proteins not producing significant BLAST hits in *Synechocystis* the subcellular localization was deduced from literature on proteins from organisms other than *Synechocystis*. For *Arabidopsis* the putative sub-cellular localization as deduced from the SUBAcon database (Hooper *et al.,* 2014; subset "envelope localization experimentally inferred", status January 2020) or proteomic data from cell/plastid fractionation experiments (Bräutigam *et al.,* 2008; Ferro *et al.,* 2010; Simm *et al.,* 2013; Bouchnak *et al.,* 2019) is indicated. (OM, outer membrane; IM, inner membrane; E, envelope; OE, outer envelope; IE, inner envelope; PM, plasma membrane; TM, thylakoid membrane). Transport systems involved in protein transport were excluded from the table.

Table S3 Overview of proteins identified by protein mass spectrometry in MS/MS experiments 1 and 2. Proteins identified in whole cell membranes (PM) and chromatophore membrane (CM) and lysate (CL) fractions that could be assigned an intensity. It shows how many (unique) peptides were identified that could be assigned to each protein. Annotation was performed with Blast2GO (Götz *et al.,* 2008) with an E value cutoff of e^{-3} and curated manually. The presence of a chromatophore transit peptide (crTP; see Singer *et al.,* 2017), as well as the number of transmembrane helices (TMH) predicted by TMHMM (Krogh *et al.*, 2001) and not being located in the crTP sequence are indicated as well. The table lists the summed MS/MS spectral counts in chromatophore-derived fractions (CM + CL) as well as the average normalized ion intensity from triplicates (see 4.2.8.3). The quotient of average normalized ion intensities in CM and CL represents the enrichment factor of a protein in CM compared to CL. Accordingly the quotient of average normalized ion intensities in CM and PM represents the enrichment factor of a protein in CM compared to PM. It is indicated whether the protein is regarded enriched with high confidence (E(HC)) or low confidence (E(LC)), is depleted in the chromatophore (D) or whether both MS/MS experiments delivered contradictory results (U). It is shown which proteins are classified as import candidates and which ones were already predicted or MS-identified import candidates in Singer *et al.* 2017. A t-

test was performed on proteins found in either three replicates and indicates statistical significance of the enrichment of individual proteins in a fraction. Difference and p-values were visualized in Fig. 2.2-9B and C.

Table S4 Characteristics of proteins identified in the chromatophore insoluble fraction and nuclear encoded chromatophore import candidates. The lists contain all proteins identified in the chromatophore membrane fraction (CM) with ≥3 summed spectral counts in chromatophore derived samples (CM + CL) in at least one of the two MS analyses and all import candidates identified so far (Singer *et al.* 2017 plus this study), either by mass spectrometry of bioinformatic prediction of a crTP. It is indicated for each protein whether the corresponding gene is encoded in the host cell nuclear genome or in the chromatophore genome. Additional information is provided on whether the full-length (FL) CDS is available as indicated by the presence of a spliced leader (SL) sequence at the 5'-end of the transcript or an in frame stop codon upstream of the start ATG (5´-FL) and a stop codon towards the 3'-end of the transcript (FL). Furthermore, the table shows the class a protein belongs to according to its length (L, long; S, short; l, likely long) and whether a chromatophore transit peptide could be identified on its N-terminus (crTP; see Singer *et al.,* 2017), as well as the number of transmembrane helices (TMH) predicted by TMHMM (Krogh *et al.*, 2001) or CCTOP (Dobson *et al.*, 2015) and not being located in the crTP sequence. Annotation was performed with Blast2GO (Götz *et al.,* 2008) with an E value cutoff of e^{-3} and curated manually, assigned KEGG Orthology (KO) and Gene Onthology (GO) terms are provided. Based on this information import candidates were sorted into putative functional categories. The table also lists how many peptides were identified, in how many samples the protein was identified and the average normalized ion intensity from CM-triplicates. The intensity level represents the log10 of the average normalized intensity + 7 and groups intensities from level 1 (low intensity) to 6 (high intensity). The quotient of average normalized ion intensities in CM and PM represents the enrichment factor of a protein in CM compared to PM. Accordingly, the quotient of average normalized ion intensities in CM and CL represents the enrichment factor of a protein in CM compared to CL. It is indicated whether the protein is regarded enriched in CM compared to PM with high confidence (E(HC)) or low confidence (E(LC)), is depleted in the chromatophore (D) or whether both MS/MS experiments delivered contradictory results (U) (Figure 2.2-8A). It is also indicated, whether proteins were already found enriched in the chromatophore in Singer *et al.* 2017. The last columns show which proteins are classified as high confidence (HC) or low confidence (LC) import candidates, and which ones were already predicted or MS-identified import candidates in Singer *et al.* 2017. Proteins identified in both, CM and PM fractions, are visualized in Figure A6.3-4.

Table S5 Single-helix pore-lining predictions, predicted stoichiometry and AMP prediction for group 1 orphan import candidates. MEMSAT-SVM scores are given, where higher values indicate a higher likelihood of having pore-lining residues. As defined by the developer, values above 0.27 are considered indicative of holding a pore-lining region. Stoichiometries are only reported for helices that are predicted to have pore-lining regions. Probabilities of having an AMP pattern are given, as reported by AmpGram, with values above 0.5 indicating a potential AMP region in the sequence. Entries in bold represent MS-identified import candidates, entries in regular fond represent similar proteins identified in *P. chromatophora* transcriptome data.

Table S6 Characteristics of *P. chromatophora* OPR proteins. *P. chromatophora* proteins for which occurrence of the octotrico peptide repeat depicted in Figure 2.3-2C was predicted by FIMO using default settings (Grant *et al.,* 2011). The number of motif occurrences per protein is indicated. Proteins that contain at least 5 repeats of the motif with a p-value < e^{-10} and / or at least 1 repeat with a p-value < e^{-20} were considered candidate OPR proteins. Additional information is provided on whether the full-length CDS is available as indicated by the presence of a spliced leader (SL) sequence at the 5'-end of the transcript or an in frame stop codon upstream of the start ATG (5´-FL) and a stop codon towards the 3'-end of the transcript (FL). Furthermore, the table shows whether a chromatophore transit peptide could be identified on a proteins N-terminus (crTP; see Singer *et al.,* 2017), whether the protein is regarded an import candidate (MS-identified or predicted), and whether it was identified as an import candidate by mass spectrometry in this study or in Singer *et al.* 2017. For 5´-full-length sequences that do not represent chromatophore import candidates subcellular targeting was predicted by different algorithms (Emanuelsson *et al.,* 2000; Armenteros *et al.,* 2019; Small *et al.,* 2004; Horton *et al.,* 2007; mTP, mitochondrial targeting peptide; ER, endoplasmic reticulum; bold letters indicate high confidence of prediction). The presence of *P. chromatophora* OPR conserved regions 1 and 2 (CR1 and CR2), a RAP domain, or a FAST kinase-like domain (FAST_1) was checked via BlastP or DELTA-BLAST (Boratyn *et al.,* 2012) for each of the listed proteins.

6.2 Supplemental Figures: The Chromatophore N-terminome

6.2.1 Principal N-termini

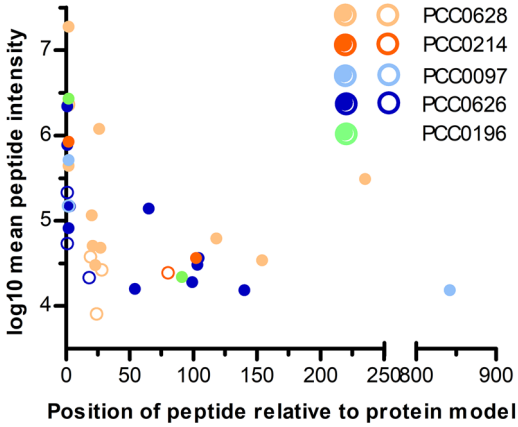


Figure A6.2-1 Principal N-termini might be determined due to high peptide intensities. N-terminal peptides corresponding to up to 12 different unique N-termini were identified by the HUNTER method. Often, peptides of one N-terminus (usually a canonical N-terminus, if present) show much higher intensities than the others and thus likely represent the most prevalent / biologically active proteoform *in vivo*. These N-termini are regarded "principal N-termini". Positions and mean LFQ intensities (from three replicates) are shown for PCC0628, phycobilisome linker polypeptide; PCC0214, ATP synthase subunit B; PCC0097, DNA-directed RNA polymerase beta subunit; PCC0626, phycocyanin alpha subunit; PCC0196, allophycocyanin alpha chain. Dimethylated peptides are represented by ●, acetylated peptides by ○.

6.2.2 Amino Acids in P1 and P1´Positions and Correlation with Peptide Intensity

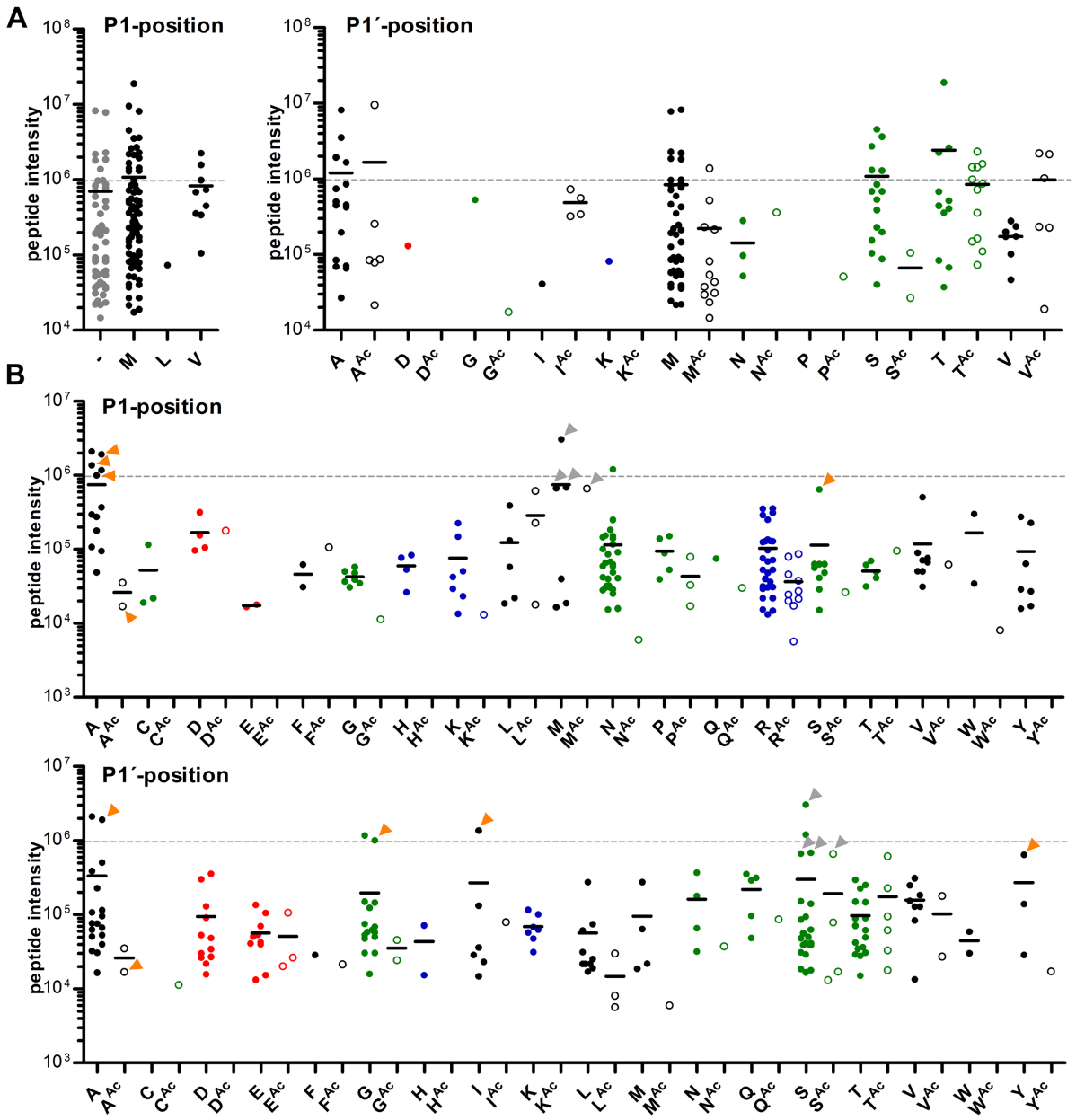


Figure A6.2-2 Amino acids in the P1 and P1´positions of chromatophore-encoded proteins. Intensities of **A** canonical and **B** non-canonical N-terminal peptides identified for chromatophore-encoded proteins, sorted by the identity of the amino acids in P1 positions (amino acids preceding the identified N-terminus) and P1´positions (N-terminal amino acids of the identified N-terminus), respectively. The mean LFQ intensity from three replicates is represented for each peptide. Intensity means for peptides possessing the same amino acid at the P1 or P1´positions are indicated. Peptides for which the intensity could not be calculated were excluded. Dimethylated peptides are represented by ●, acetylated peptides by ○. Black, hydrophobic; green, polar amino acids, cysteine, proline, and glycine; red, negatively charged; blue, positively charged; grey, no P1 position due to unprocessed N-terminus starting with iMet. Valine and leucine in canonical P1 positions are a result of alternative translational start sites (GUG and UUG) and it is nonetheless likely that methionine is incorporated by the ribosome at this position, not valine/leucine (Kearse and Wilusz, 2017). A dashed line marks an intensity of 10^6 in all graphs. Grey

arrowheads indicate putative canonical N-termini that likely have been mis-assigned due to incorrect protein models. Orange arrowheads mark N-termini resulting from SP-cleavage. Notably, the frequencies of individual amino acids shown here do not reflect the overall relative frequencies of amino acids in the predicted chromatophore-encoded proteome.

6.2.3 P1´Amino Acids after crTP Cleavage

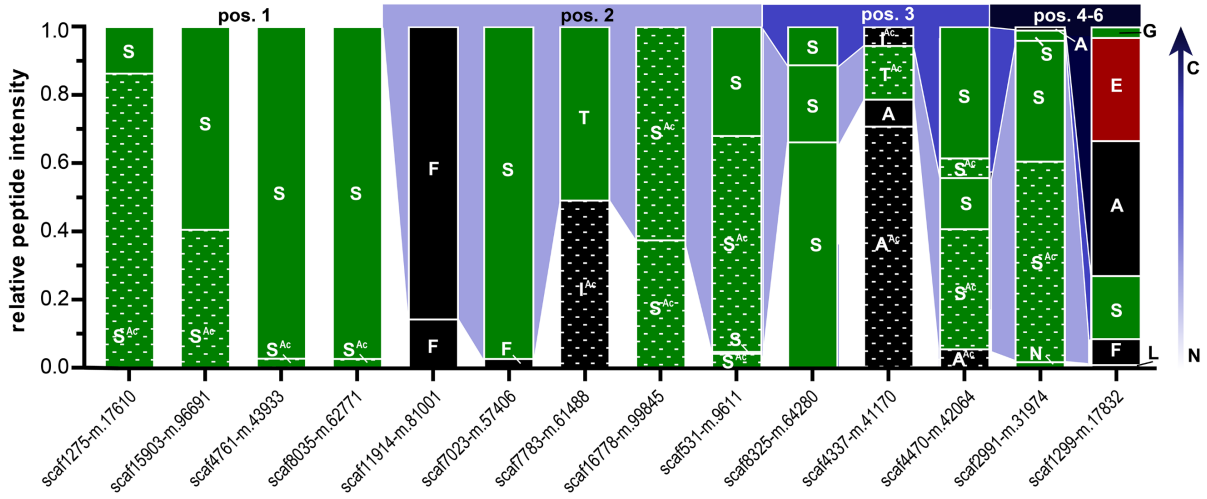


Figure A6.2-3 Amino acids present in the P1´position after crTP cleavage and relative abundance of corresponding N-terminal peptides. Selected crTP processing products are shown (site 1; Fig. 2.1-3D). Amino acids representing N-terminal residues of all site 1 N-terminal peptides identified for each protein are depicted. Bar size corresponds to relative intensity of an N-terminal peptide with respect to all crTP-derived N-terminal peptides identified for a protein (calculated from mean LFQ intensities from three replicates). The relative position of the N-terminus (where pos. 1 is the most N-terminal N-terminus identified for the protein) is indicated by a color code (white to dark blue). Black, hydrophobic; green, polar amino acids and glycine; red, negatively charged; patterned, acetylated.

6.3 Supplemental Figures: Transporters and Metabolite Exchange

6.3.1 Amino Acid Sequences of STr1 and Ptr4

>STr1 (deduced from CCAC0185-comp6040_c0_seq1)

MFSELPWTINTEFVAIWSMVATVFSCFLYFSPINVIREVIRDKSTKQYRGETFVATFANCFLWNNYVL
YKDPKITIYFAMFVCDCIGIFAALSFFCIFTFYCAGRWRYFKMMFYFLAITAILNVALATTGSDAATK
SLVLGLVCSLFNVAMSASPCAVLAEVFRTKSVRFLPLPILTCGLVASLEWLVYAVYIGDIWLTIPITL
GSFLYIMQLVIYWYYGGCHSTRVKNTRRLGVRRSPYISMMPFMVVSGSYFRSMRSGRSVTFASSTSAV
SSAISRTPLDLLEGSSFTTTSTTTNQSGVLSSATSLSVISEPASPGSAPSSQPTAPPSFIEIHPPVVV
AVEPNGHPQLPAPPSFVESELPASSQVEPPSVQLHAPPSFIELPSFSQIEPIESPTLSIPHYVVPSFI
ECKPPSFEPVSETTDSAEENEVQESDFCIPISQ*

>PTr4 (deduced from scaffold19601-size716, scaffold22355-size626, and genome-contig 638419)

MVGVGWWDRWGLAGVYLSVWFLSGLAKSDLNKAALQTFPFPVTVLYVQFLVVSFLTTLLLGAAQRLPY
IPWSRRLLQLVFPYTAFLLLGHIFSQLALESMPVYILKAIKSSEPLVTLCVAWFWLGERCGWRTLLTL
FPICGGVALTTLDKLQTGGQAFWLAGLLAVLSSVSYDVSKVAAKSIFVLGERELPSGQAAAVDSRSDA
LLPPGQAAAVDSRSDAVEEVKREVVSEQVFLTTEQMICEPSVADSTVVAVVRSRLGIRHAPVLLATLQ
AQQEATARLVSTQSFSDSQSESSEHVYLPFTDGPLLQDPVFAPPTELTPPRPISLHPLELGLVSNLLA
LVLLLPVWLWLEEGSLVAPDRQSLWFLVGNGLADFVDSLSGLYFLTLASATSAAVVANLRTVFIIFVS
MIIFHDALTVERICGVALTLVGVFFYQYAIRYEAVQRPGYEAAQSEPPAVAS*

Figure 6.3-1 Deduced amino acid sequences of STr1 and PTr4, for which no full-length protein model exists in NCBI BioProject PRNJA311736. PTr4 contains an insertion of 48 bp that was not present in any Illumina HiSeq read. Presence of the insertion in the *PTr4* transcript was verified experimentally, as the *PTr4* ORF was amplified from two independently synthesized aliquots of *P. chromatophora* cDNA, integrated into pJET cloning vectors, and sequenced.

6.3.2 Application of α-BGATr in Immunogold Electron Microscopy

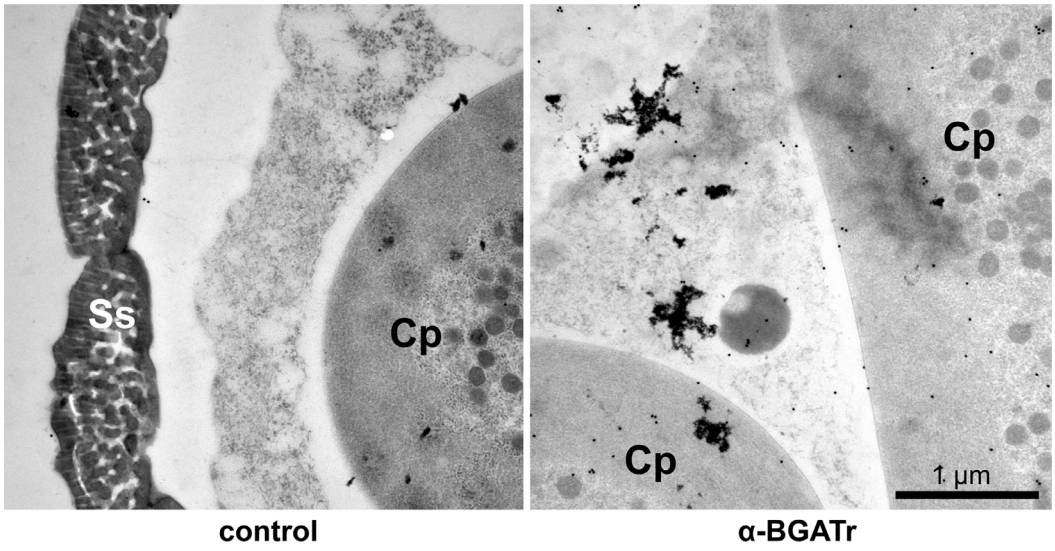


Figure A6.3-2 Affinity purified α-BGATr failed to specifically bind to BGATr in thin sections of LR White embedded *P. chromatophora* cells. Gold particles are distributed uniformly among subcellular structures, with some located at the chromatophore envelope, but others located at thylakoids or carboxysomes and several in host derived structures. Some particles were even located outside of the cells. Control samples (secondary antibody only) are decorated with much fewer particles indicating non-specific binding of the primary but not the secondary antibody. Cp, chromatophore; Ss, silica scales. Procedure: Cells were embedded in LR White resin (medium grade acrylic resin, London resin company) and thin sections were fixed on nickel grids. For immunogold staining thin sections were rehydrated in PBS buffer supplemented with 0.05 % Tween 20 (PBST) and incubated with basic blocking solution (Aurion) for 20 min. Primary antibody (α-BGATr) was diluted 1:1 in blocking solution and allowed to bind to thin sections for 3 h. Unbound antibody was removed by three consecutive 5 min washes in PBST. Control sections were treated identically but incubated for 3 h in blocking solution instead of primary antibody. All sections were incubated for 30 min in a 1:50 dilution of Protein A gold (15 nm, Aurion) in blocking

solution followed by seven short washing steps in PBST. Antibody/antigen complexes were fixed via incubation in 8 % glutaraldehyde for 15 min, washed with filtered dH_2O, and air dried. For contrasting, sections were incubated in uranyl acetate for 15 min, washed with filtered dH_2O, then incubated in lead citrate for 4 min, and again washed with filtered dH_2O (Reynolds, 1963). Images were taken at a magnification of 30,000 x with a Hitachi H-7100 transmission electron microscope (Hitachi, Tokyo, Japan) with Morada camera (EMSIS GmbH, Münster, Germany).

6.3.3 Chromatophore Fractionation Effectively Enriched Membrane Proteins

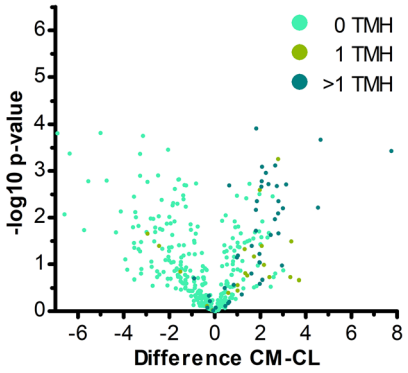


Figure A6.3-3 Membrane proteins are enriched in chromatophore membrane fractions (CM) compared to chromatophore lysate (CL). Enrichment of chromatophore-encoded membrane proteins in MS experiment 2 (for experiment 1 see Fig. 2.1-7C). The difference between the mean log2-transformed normalized intensities of individual proteins in CM and CL samples ($\overline{log2(normInt_{CM})} - \overline{log2(normInt_{CL})}$; Difference) is plotted against significance (-log10 p-values in Student´s t-test) for proteins identified in all three triplicates of either CM or CL or both. A positive difference value indicates protein enrichment in CM, negative values indicate depletion in CM compared to CL samples. Values for proteins detected only in one sample have been imputed and are only shown when their difference is significant. The number of predicted TMHs is indicated by a color code.

6.3.4 Determination of Import Candidates

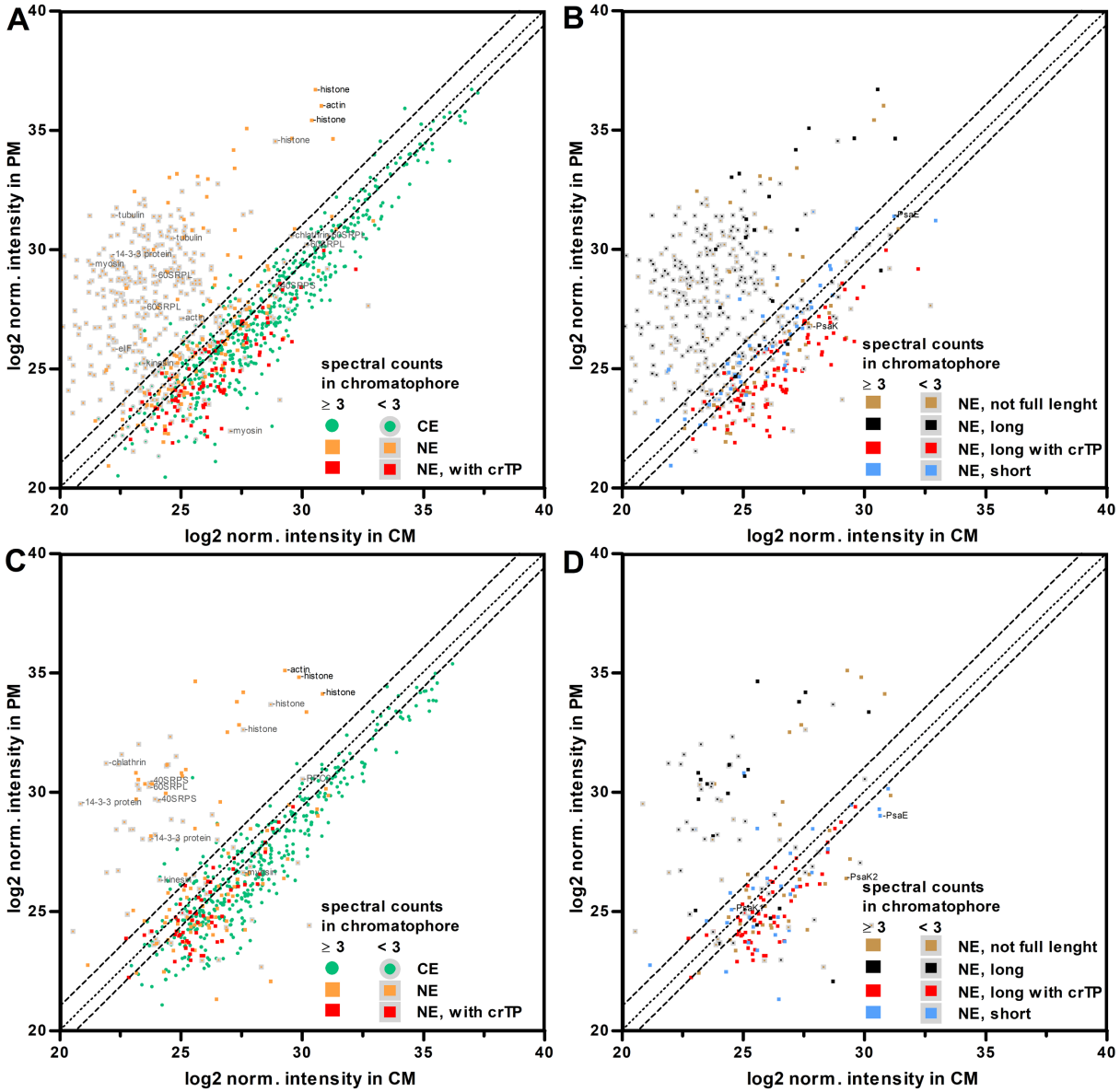
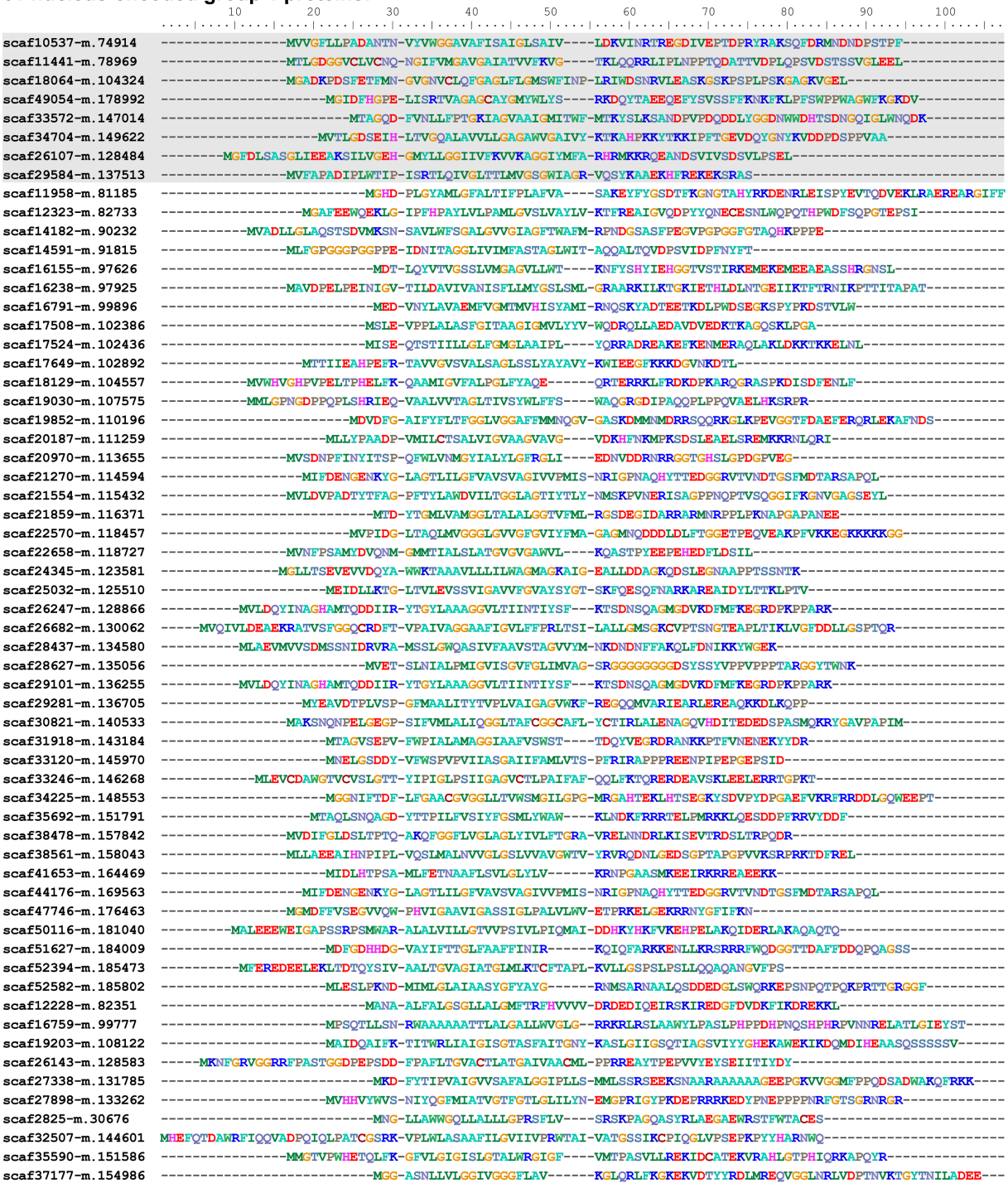


Figure A6.3-4 Calibration of enrichment level thresholds for determination of import candidates. The $log2(\overline{normInt_{CM}})$ (see 4.2.8.3) was plotted against the $log2(\overline{normInt_{PM}})$ for proteins identified in CM and PM samples in **A and B** MS experiment 1 and **C and D** MS experiment 2. CE, chromatophore-encoded; NE, nuclear encoded; NE with crTP, nuclear encoded proteins that contain a crTP; NE long, proteins longer than 125 aa that do not contain a crTP; NE short, proteins shorter than 95 aa; NE not full-length, length and/or presence of crTP cannot be determined due to lacking sequence information (includes also proteins which lack an SL or an in-frame Stop-codon upstream to their putative translation start sites). Proteins identified with < 3 spectral counts in the chromatophore (CM + CL samples) were not considered as import candidates. For the remaining proteins identified with ≥ 3 spectral counts in chromatophore samples, the enrichment factor in CM as compared to PM fraction ($\overline{normInt_{CM}}/\overline{normInt_{PM}}$) was calculated (see Tab. S3 and S4). Proteins enriched in the chromatophore by a fold change of at least 1.5 in at least one out of two MS experiments are regarded "Enriched with high confidence"; proteins depleted in the chromatophore by a fold change < 0.5 in at least one MS experiment are

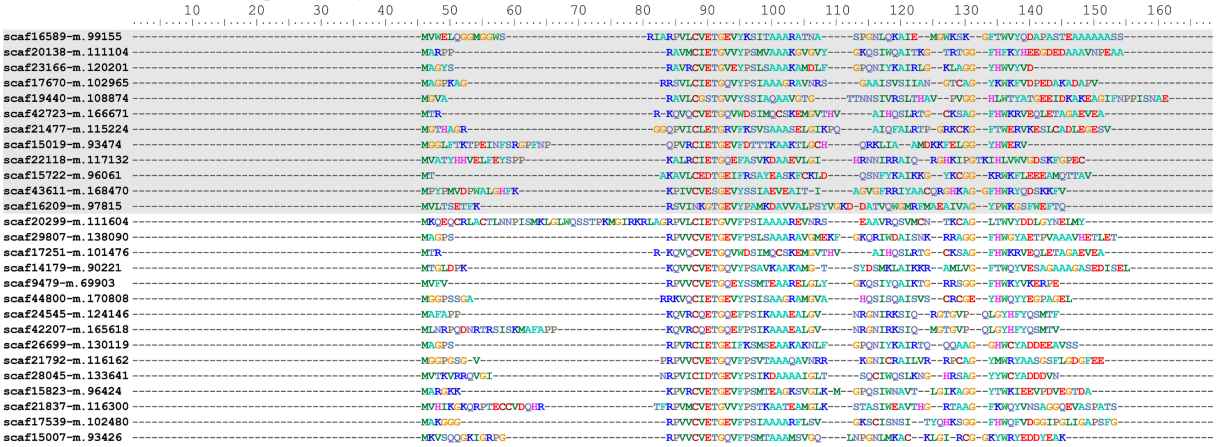
regarded “Depleted”; proteins with weak fold changes ≥ 0.5 but ≤ 1.5 (indicated by dashed lines) in both MS experiments are regarded “Enriched with low confidence”. Proteins enriched in one experiment but depleted in the other are classified as “unclear”. Many of the proteins excluded from the analysis by using the ≥ 3 spectral counts-threshold represent highly expressed contaminating host-proteins like 40S/60S ribosomal proteins, histones, and cytoskeletal components (A and C) and lack a crTP (B and D). The “Enriched with low confidence” range includes several crTP-containing proteins and also chromatophore-encoded proteins (A and C) as well as short import candidates and small proteins involved in photosynthesis (B and D). The range was introduced to calibrate the experimental data based on our knowledge regarding the chromatophore localization of crTP-containing proteins, small proteins involved in photosynthesis, and chromatophore-encoded proteins.

6.3.5 Further Group 1 to 3 Short Proteins Identified in the *P. chromatophora* Transcriptome

61 nucleus-encoded group 1 proteins:

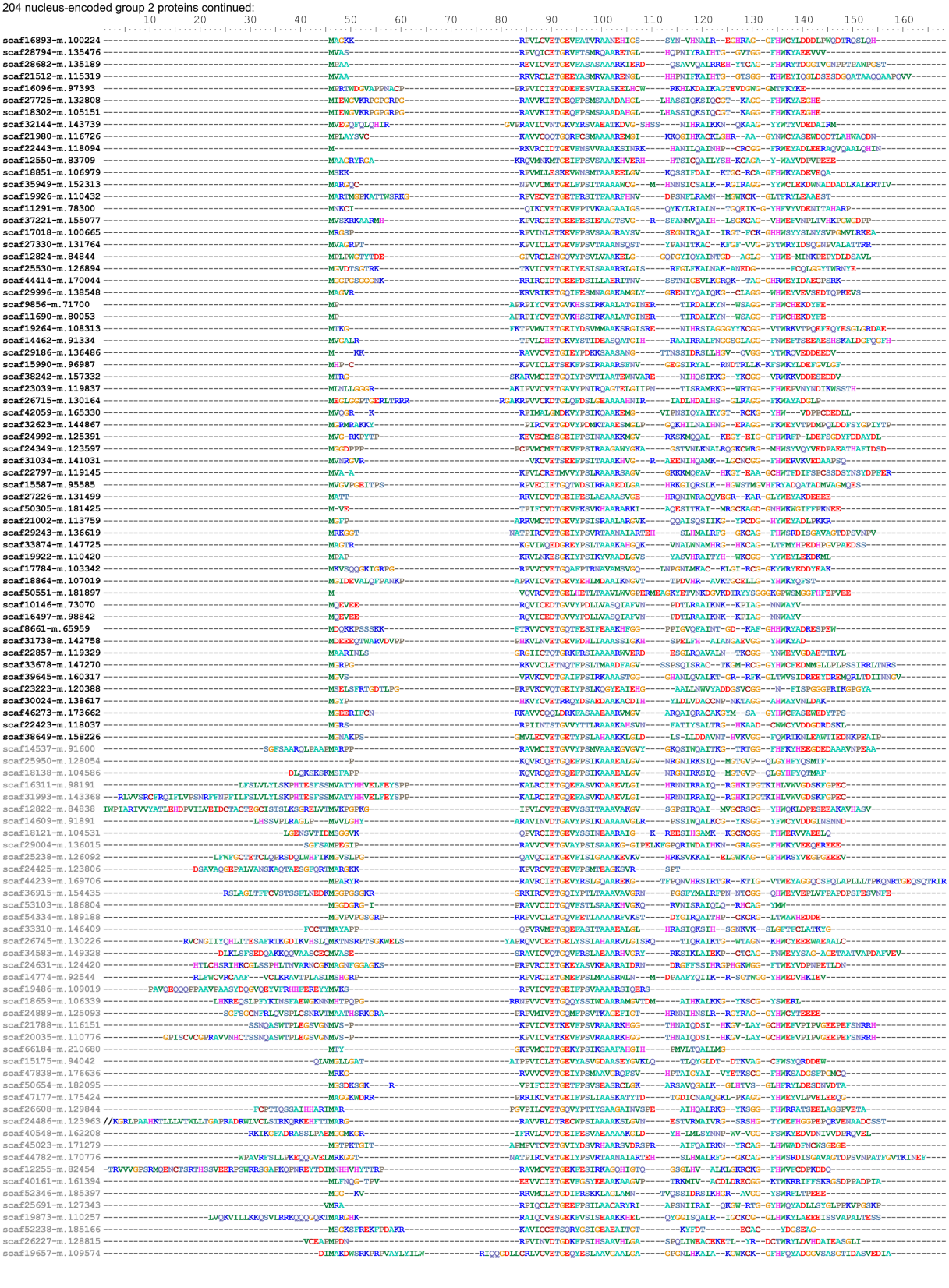


204 nucleus-encoded group 2 proteins:



(continued on next page)

204 nucleus-encoded group 2 proteins continued:



38 nucleus-encoded group 3 proteins:

Figure A6.3-5 Alignments of potential further group 1 to 3 short import candidates identified in the *P. chromatophora* transcriptome. Translated transcripts representing short proteins with a predicted TMH in the N-terminal 2/3 of the sequence that contains > 20 % small amino acids (glycine, alanine, serine) and feature an N-terminus with a net charge ≤ 0 were regarded as putative further group 1 proteins. Translated transcripts that show significant similarity (FIMO, $p < e^{-08}$ or $p < 2e^{-07}$) to the conserved motifs in MS-identified group 2 or group 3 proteins (highlighted in grey; Fig. 2.2-10A) were regarded as putative group 2 and 3 proteins, respectively. 161 and 29 of the proteins are short (< 95 aa), respectively. For the remaining proteins either the start methionine could not be identified unequivocally (as a spliced leader or an in-frame stop codon upstream of the start ATG are missing in the corresponding transcript) and/or a stop codon is missing at the C-terminus (grey sequence titles). Thus, their length cannot be determined. Importantly, none of the identified proteins belonging to either group is unequivocally long (> 250 aa).

6.3.6 Group 1 Short Import Candidates Might Form Oligomeric Pores

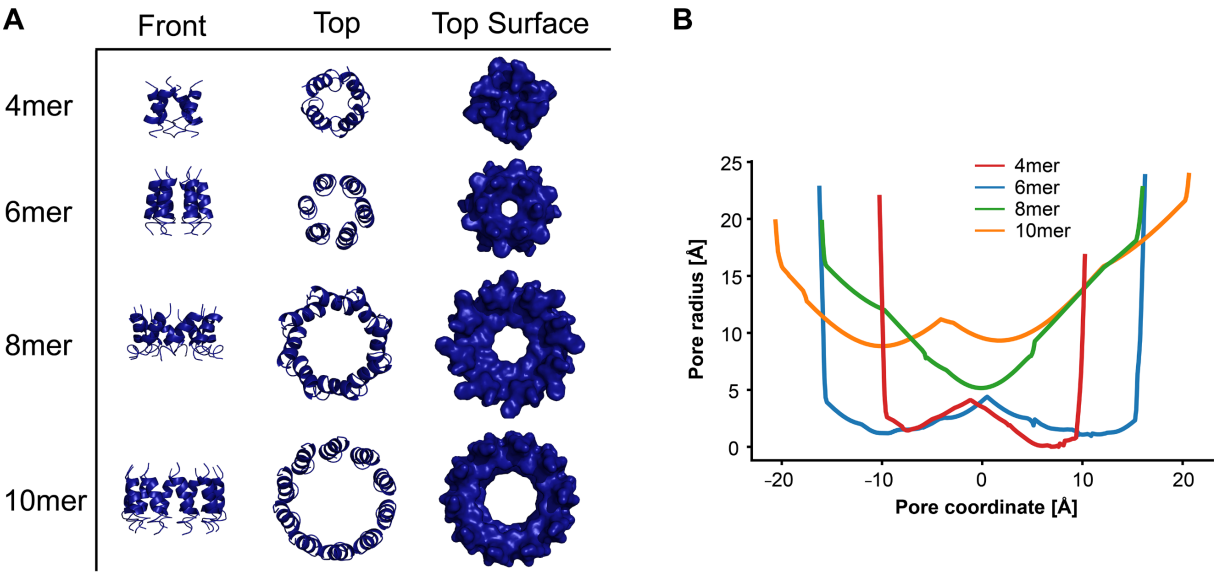


Figure A6.3-6 Hypothetical *n*-mer pore-like assemblies for scaffold29584-m.137513. A The predicted TMH of scaffold29584-m.137513 was extracted from a homology model made with TopModel, primarily based on the cryo-EM structure of the bovine respirasome (PDB ID 5LUF). The structure was arranged in 4, 6, 8, and 10mer assemblies with CHDOCK; the best scoring assembly for each case is shown from the front and top; for the latter, also the molecular surface is shown. According to pore geometry, structures with six or more helices (i.e. of higher order than the predicted stoichiometries [Tab. S5]) might allow permeation of metabolites through the pore, as suggested from previous studies (Grice *et al.*, 1997). **B** Radius along the pore of structures shown in panel A as calculated with HOLE. For reference, the bottleneck radius of common metabolite transporters such as multidrug and Omp channels is close to the radius of a water molecule (1.4 Å)(Khalid and Sansom, 2006; Song *et al.*, 2014; Vaccaro *et al.*, 2008). An increasing pore size begs the question of how such a channel could selectively transport metabolites, while maintaining the electrochemical gradient of the chromatophore. VDAC, in the outer membrane of mitochondria, remains partially selective despite having a pore of 13 Å radius due to a stochastic gating mechanism (Berezhkovskii and Bezrukov, 2018) which might be considered as a model for metabolite transport across the chromatophore inner membrane.

7. Manuscripts

The following manuscript was published on November 27th 2020 in the *Frontiers in Microbiology* journal:

'The puzzle of metabolite exchange and identification of putative octotrico peptide repeat expression regulators in the nascent photosynthetic organelles of *Paulinella chromatophora*'

Authors: Linda Oberleitner, Gereon Poschmann, Luis Macorano, Stephan Schott-Verdugo, Holger Gohlke, Kai Stühler, and Eva C.M. Nowack

Illustrations and tables from the manuscript of which original or modified versions were used in this thesis are listed below:

Figure / table in manuscript	**Figure / table in this thesis**
Figure 1	Figure 2.2-1
Figure 2 A, B	Figure 2.2-6 A, C
Figure 2 C, D	Figure 2.2-7 D, E
Figure 3 A	Figure 2.2-8 A
Figure 3 B, C	Figure 2.2-9 B, C
Figure 4	Figure 2.2-10
Figure 5 A	Figure 2.2-8 B
Figure 5 B	Figure 2.3-1
Figure 6	Figure 2.3-2
Figure S1	Figure A6.3-4
Figure S2 A	Figure 2.2-7 C
Figure S2 B	Figure A6.3-3
Figure S3	Figure A6.3-5
Figure S4	Figure A6.3-6
Table 1	Table 2.2-3
Table 2	Table 2.2-4
Table S1	Table S3
Table S2	Table S2
Table S3	Table S4
Table S4	Table S5
Table S5	Table S6

Eva Nowack and I conceived the study and wrote the manuscript. The material enriched for insoluble proteins of chromatophores and whole cells, as well as chromatophore lysate was prepared by me. Gereon Poschmann and Kai Stühler carried out the mass spectrometric analysis and provided pre-processed data in an Excel spreadsheet containing information on “protein identification parameters”, “peptide spectrum matches”, “ion intensity”, and “statistical significance (t-test)”. The data were further analyzed and evaluated by Eva Nowack and me. I compiled the information on chromatophore-encoded transporters and carried out the comparison with cyanobacterial and plastid transporters. The complete bioinformatic analysis of OPR proteins was carried out by me. Embedding and TEM of *P. chromatophora* cells and isolated chromatophores was done by Luis Macorano. Bioinformatic analysis of AMP-like proteins was done by Eva Nowack, Stephan Schott-Verdugo, Holger Gohlke, and me. Figure A6.3-6 was made by Stephan Schott-Verdugo and Holger Gohlke. All other figures were prepared by me.

I hereby confirm the accuracy of this information.

17.12.2020, Düsseldorf

Acknowledgements

An dieser Stelle möchte ich mich bei allen bedanken, die mich während der letzten Jahre begleitet und bei der Erstellung dieser Arbeit unterstützt haben.

Zuallererst: Danke an meine Chefin und wissenschaftliche Mentorin Eva Nowack für die Möglichkeit, als Teil von Evalution und der kleinen aber feinen *Paulinella*-Gemeinschaft zu promovieren. Vielen Dank für das abwechslungsreiche Projekt und deine ständige Unterstützung und Leitung.
Ich danke außerdem Prof. Dr. Lutz Schmitt für die Annahme des Zweitgutachtens und die hilfreichen Besprechungen als mein SFB-Mentor.
Danke an den Sonderforschungsbereich 1208 (SFB1208) für die Finanzierung meiner Arbeit und an meine vielen SFB-Mitforscher für die zahlreichen konstruktiven und geselligen Seminare, Retreats und Konferenzen.
Ein besonderer Dank gilt all meinen Evalution-Kollegen. Danke an Anna für die hervorragende Vorarbeit und alltägliche Hilfe im Labor. Danke Georg für deine geduldige Unterstützung in Sachen Mikroskopie. Danke auch an Jorge, Sofia, Luis, Jan, Vika und viele weitere aktuelle und ehemalige Kollegen. Von euch habe ich nicht nur eine Menge Methodiken lernen können, sondern dabei auch alltäglich eine lustige Zeit mit euch gehabt. Das Gleiche gilt für die vielen hilfsbereiten Feldis.
Danke außerdem an meine Kollaborationspartner innerhalb und außerhalb des SFB1208, z.B. Andreas Perrar, Gereon Poschmann, Nicole Linka, Marion Eisenhut und Tabea Mettler-Altmann.

Außerdem gilt ein großer Dank natürlich meiner Familie und meinen Freunden. Ihr habt mich geduldig ertragen, wenn die Experimente mal wieder einfach nicht funktionieren wollten. Danke Mama, du bist mein Fels in der Brandung während des gesamten Studiums. Danke auch Papa und Beate für die tatkräftige Hilfe, z.B. beim Renovieren und Umziehen. Danke an Jani, Bine, Sebas, Alina, Alex und Kathrin fürs Mut machen und die regelmäßige Zerstreuung.

Vielen Dank euch allen!