

Tracking N-terminal protein processing from the Golgi to the chromatophore of a rhizarian amoeba

Thierry Meinnel

▶ **To cite this version:**

Thierry Meinnel. Tracking N-terminal protein processing from the Golgi to the chromatophore of a rhizarian amoeba. Plant Physiology, 2022, 189 (3), pp.1226-1231. 10.1093/plphys/kiac173 . hal-03754961

HAL Id: hal-03754961
https://hal.science/hal-03754961

Submitted on 20 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

https://doi.org/10.1093/plphys/kiac173 PLANT PHYSIOLOGY 2022: 189: 1226–1231

Plant Physiology®

Letter

Tracking N-terminal protein processing from the Golgi to the chromatophore of a rhizarian amoeba

Thierry Meinnel [1,*]

1 Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198 Gif-sur-Yvette, France

*Author for correspondence: thierry.meinnel@i2bc.paris-saclay.fr

T.M. wrote the paper and performed the analysis.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plphys/pages/General-Instructions) is: T. Meinnel (thierry.meinnel@i2bc.paris-saclay.fr).

Dear Editor,

Sample analysis from various ecosystems has substantially enriched recent knowledge about the diversity of life (Karsenti et al., 2011). One major aim of these efforts—facilitated by rapid advances in single and meta-genomic DNA sequencing—has been to explore the diversity of "invisible" or "uncultivable" organisms. While prokaryotes and viruses are by far the most abundant and biodiverse organisms on Earth (Lynch et al., 2014), the breadth of eukaryotic diversity is now better appreciated by complete genome sequences of model organisms in each class, facilitating and reinforcing classification efforts (Supplemental Figure S1). Several complete and well-annotated nuclear and organelle genomes contributing to eukaryotic diversity are now available through unique portals (Grigoriev et al., 2021).

The current representation of the eukaryotic tree of life is based on the dichotomy between Unikonta and Bikonta, both groups containing unicellular and multicellular organisms (Supplemental Figure S1). Unlike Unikonta, which are basically heterotrophic organisms, Bikonta often photosynthesize via a unique organelle, the plastid. There are three major phylogenetic branches of Bikonta: Archaeplastida, Chromalveolata, and Excavata (Supplemental Figure S1). Of these, the Archaeplastida include plants, green algae, and red algae, all of which are characterized by a common ancient primary endosymbiosis of a photosynthetic β-cyanobacterium ∼1.6 billion years ago. This ancestral inclusion led to primitive photosynthetic organisms. Red and green algae (Supplemental Figure S1) are among the closest representative fossils of this primordial event, since most other photosynthetic subgroups—Euglenozoa, Stramenopiles (diatoms), and Apicomplexa (Supplemental Figure S1)—acquired a secondary inclusion of a photosynthetic eukaryote (green or red algae) via phagocytosis into another non-photosynthetic eukaryote much more recently, ∼190 million years ago (Falciatore et al., 2020). Other organelles, like in dinoflagellates, have a tertiary origin (Yoon et al., 2005).

Primary endosymbiosis of photosynthetic bacteria represents not only a remarkably ancient event but also a unique event creating sophisticated interdependence between the plastid and nucleus: genes migrated from the plastid to the nucleus and proteins were retargeted from the nuclear genome to the plastid. Exactly how proteins traffic to the plastid has been studied for decades in Archaeplastida. Most proteins display an N-terminal transit peptide (TP) that allows them to cross the plastid membranes, and this leader peptide is usually cleaved off once the protein enters the plastid (Christian et al., 2020). In the stroma, the proteins usually undergo further modifications including N-terminal acetylation (NTA) of the new N-terminus (Giglione and Meinnel, 2021) and some are retargeted to the thylakoid for further processing. Secondary endosymbiosis also gave rise to complex import mechanisms (Agrawal and Striepen, 2010; Hopkins et al., 2012).

As primary endosymbiosis is such a rare event, knowledge of at least one other independent similar event and how genes and proteins exchange between two genomes is key to a better understanding of eukaryogenesis, its trajectory,

Received February 03, 2022. Accepted March 16, 2022. Advance access publication April 21, 2022

© The Author(s) 2022. Published by Oxford University Press on behalf of American Society of Plant Biologists.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

Downloaded from https://academic.oup.com/plphys/article/189/3/1226/6571825 by CNRS user on 28 June 2022

and evolutionary events. *Paulinellidae*, which include the freshwater *Paulinella chromatophore* and *Paulinella micropora* and the marine *Paulinella longichromatophora* (Lhee et al., 2019), are euglyphid testate amoebae of the Rhizaria supergroup (Figure 1) displaying a siliceous shell and a photosynthetic organelle, like diatoms. Unlike diatoms, however, these organisms are unique as their plastid—the chromatophore—originates from the much more recent primary endosymbiosis of an α-cyanobacterium only ~100 million years ago and independent of Archaeplastida (Stephens et al., 2021). Despite difficulties in analyzing these organisms due to partial gene annotation (see, e.g. the 1 Gbp *P. micropora* http://cyanophora.rutgers.edu/P_micropora/ [Lhee et al., 2021]; the *P. chromatophora* genome is 10-fold larger) and lack of genetic manipulation tools, *Paulinella* spp. are useful models of primary endosymbiosis, a major event in eukaryogenesis that also involves the mitochondrion (Szathmáry, 2015).

A recent publication in *Plant Physiology* by Oberleitner et al. (2022) expands the current knowledge in this field. The authors extracted chromatophores from *P. chromatophora*. Using state-of-the-art mass spectrometry called "High-efficiency Undecanal-based N-Termini EnRichment" (HUNTER) (Weng et al., 2019), they assessed N-terminal protein processing in the organelle. HUNTER involves chemical labeling of free amino groups (i.e. N-termini and lysine chains) with dimethylation before trypsin-mediated protein hydrolysis into small fragments compatible with mass spectrometry. In this case, trypsin only cleaves after arginines because lysines are modified. After enrichment of N-terminal peptides, this approach allows easy recognition of protein N-termini, which are N-tagged due to chemical dimethylation or natural NTA, the most frequent N-terminal modification of eukaryotic proteins (Aksnes et al., 2019). Searches were carried out using >60,000 open-reading frames (ORFs) translated from a *P. chromatophora* transcriptome and the

```
g77960.t1-P. micropora                                                         MGVNVDTIILIVHTIYLILVKMSMPQLPSHEIQHRFATWAHRARITVDPAQVCKRISYLGEYSRDTRNQVL  71

g77960.t1-P. micropora           VSFHSVESGKESSFMSSKRCGVVNVHLSAGSVGPPWVCLATSPARDAIGVFRFFSFGTLASHLLVTMAAAAPGGPGGDPVVIQIRCVGGVAAGGATLAPR 171
m.87086-P. chromatophora                                                   MDPTVMIDTVIAVPSMTASVISFIIPPARLRFMANPAEIVIIQIRCVGGVAAGGATLAPR  60
m.108443- P. chromatophora                                                                                MANPAEIVIIQIRCVGGVAAGGATLAPR  28
A0A0G4IN79-P. brassicae                                                                                  MPPKFDAGATVVIYLRVFGGEPAPASTLAPK  31
P30050-H. sapiens                                                                                        MPPKFDPNEIKVVYLRCTGGEVGATSALAPK  31
A0A125YT49-T. gondii                                                                                     MAPKFDPSEVKYIYLRQVGGEVGASSVLAPK  31
P508832-A. thaliana                                                                                      MPPKLDPSQIVDVYVRVTGGEVGAASSLAPK  31
A0A813EJD3-P. glacialis                                                                                  MAPKFDANEVKVVFLRQYGGEQAPSSVLAPK  31
P0CX53-S. cerevisiae                                                                                     MPPKFDPNEVKYLYLRAVGGEVGASAALAPK  31
XP_011775943.1-T. brucei                                                                                 MPPKFDPNQEITVVVRAVGGEVPATASLAPK  31
QLA09605.1-E. gracilis                                                                                   MPPKFDPNETKIIYVRATGGEMGATSTLAPK  31
Q54J50-D. discoideum                                                                                     MPPKVDPSEKVEVFLRVCGGEAGAMSTLAPK  31
                                                                                                           **:*    * : :*  **  . .::***:

g77960.t1-P. micropora           LGPLGLSAKKVGDDIAKETM-PYKGLKVTVKLTIQNRQAAIEVVPTAAMLLIQSLNEPERDRKKEKNIKHNGNISMEQVYKIARQLRPRSRAVFFKGTVC 270
m.87086-P. chromatophora         LGPLGLSAKKVGDDIAKETM-PYKGLKVTVKLTIQNRQAAIEVVPTAAMLLIQSLNEPERDRKKEKNIKHNGNISMEQVYKIARQLRPRSRAVFFKGTVC 159
m.108443- P. chromatophora       LGPLGLSAKKVGDDIAKETM-PYKGLKVTVKLIIQNRQAAITVVPTASMLLIQALNEPERDRKKEKNIKHSGNLSMEQIYKVARTLRPRSRAKEFSGTVR 127
A0A0G4IN79-P. brassicae          VGPHGLSPKKIGEDIAKATT-QYKGLRVTCKLTVQNRQAQVEVVPSASSLVIGALKEPIRDRKKEKNIKHDGNLSFEAILDIARTMRPRSQAKELKGTVK 130
P30050-H. sapiens                IGPLGLSPKKVGDDIAKATG-DWKGLRITVKLTIQNRQAQIEVVPSASALIIKALKEPPRDRKKQKNIKHSGNITFDEIVNIARQMRHRSLARELSGTIK 130
A0A125YT49-T. gondii             LGPLGMSPKKVGDDIAKATV-AWKGLKVTVKLAVQNRQATVEVVPSASSLIIKELKEPPRDRKKVKNIKHSGNLSLDQVFSIARTMRAKSQAWEFSGTVK 130
P50883-A. thaliana               IGPLGLAPKKIGEDIAKETAKEWKGLRVTVKLTVQNRQAKVTVVPSAAALVIKALKEPERDRKKVKNIKHNGNISFDDVTEIARIMRPRSIAKELSGTVR 131
A0A813EJD3-P. glacialis          VGPLGMSPKKVGDDIVKGTS-AWKGIRVTVKLTIQNRAAKVDVEPNATSLLIKALKEPMRDRKKTKNIKHSGNLTKEVVFDICRQMRKKSLAKEFSGTVK 130
P0CX53-S. cerevisiae             IGPLGLSPKKVGEDIAKATK-EFKGIKVTVQLKIQNRQAAASVVPSASSLVITALKEPPRDRKKDKNVKHSGNIQLDEIIEIARQMRDKSFGRTLASVTK 131
XP_011775943.1-T. brucei         VGPLGLNAKKIGEDIAKSTK-DWKGLKVTCQLRVKNRVATVVVTPSVASRLIRALKEPPRDRKKVKNIKHDGNIAFSEILKIAKESAPNSMGASMKSVVM 131
QLA09605.1-E. gracilis           VGPLGLNAKKVGEDIMKASK-EWKGLRVTVKLTIKNRQATAEIVPSSTSLVLRALKEPPRDRKKVKHIKHDGNITLEEIIDIARIMRPRSMARELKGTVK 131
Q54J50-D. discoideum             LGPLGVSPKKVGDDIAKATQ-PWKGMKVSVKLTIQNRIAVPEVLPSASALVIKALKEPPRDRKKEKNIKHNGNIPLEEICKIAKTMRFKSLAVDFKGSVL 131
                                 :** **:.**:*:**** *   :***:** ** :***** : ***:*: *:* :*:** ***** *****.**:*:: : .:** :**** *   :.***

g77960.t1-P. micropora           EMLGTCVSIGCTVNDTDPKTLIENIQSGEL-VVPEE 305
m.87086-P. chromatophora         EMLGTCVSIGCTVNDTDPKTLIENIQSGEL-VVPEE 194
m.108443- P. chromatophora       EMLGTCVSIGCTVNDTNPQDLIESIKSGEL-VVPEE 162
A0A0G4IN79-P. brassicae          EILGTCFSVGCTVDGQAPRDVQTAIDNGEIE-VPAA 165
P30050-H. sapiens                EILGTAQSVGCNVDGRHPHDIIDDINSGAVE-CPAS 165
A0A125YT49-T. gondii             EILGTCNSVGCTVDGKSPKQVQRMLDDGEID-CPEA 165
P50883-A. thaliana               EILGTCVSVGCTVDGKDPKDIQQEIQDGEVE-IPEN 166
A0A813EJD3-P. glacialis          EILGTCRAIGCTVDGQSPIDLTEAIDNDEWE-TPKE 165
P0CX53-S. cerevisiae             EILGTAQSVGCRVDFKNPHDIIEGINAGEIE-IPEN 166
XP_011775943.1-T. brucei         EVLGTAVSIGCTIDGEHPRAIQEKVQEGKLK-VP-N 165
QLA09605.1-E. gracilis           EILGTAFSVGCTVEGRNPKDVCDDVGAGIIE-IPEK 166
Q54J50-D. discoideum             EILGTAHSVGCKVNGKSPRDIQAGIQSGEIEVVEPK 167
                                 *:****.*:****:.  *: :   *..**: :*
```

Figure 1 Alignment of the full-length ORFs encoding the 60S ribosomal protein L12 (RPL12A) from various organisms. The ORFs encoding ribosomal protein L12 were aligned with ClustalX2.1. The ORF sequences derived from gene models from the Viridaeplanta Arabidopsis (*Arabidopsis thaliana*) (At2g37190, P508832), the rhizarian Cercozoa causative agent of cabbage clubroot, *Plasmodiophora* brassicae (P508832) as the closest relative of *Paulinellidae*, *Toxoplasma gondii* (A0A125YT49), *Homo sapiens* (P30050), *Saccharomyces cerevisiae* (P0CX53), *Trypanosoma bruce*i (XP_011775943.1), *Euglena gracilis* (QLA09605.1), *Dictyostelium discoideum* (Q54J50), the dinoflagellate *Polarella glacialis* (A0A813EJD3), and *Paulinella micropora* (g77960.t1). Both ORFs from *P. chromatophora* (m.87086 and m.108443) are derived from the transcriptome. None of the two corresponding L12 transcripts display the trans-splicing tag (CTTTTCTG) and is therefore expected to correspond to a full-length transcript (nucleotide sequences of all *P. chromatophora* transcripts are available at https://www.ncbi.nlm.nih.gov/bioproject/PRJNA311736). The various in-frame start codons are colored in red. The numbering of the amino acid sequence is indicated on the right side in italics. The zone colored in yellow in *Paulinella* sequences shows significant divergence from the other otherwise very conserved sequences below. The N-terminal acetylated peptide sequence retrieved in this study is shown in green. NB PSI-BLAST of the N-terminal extension (137 residues) of *P. micropora* shows no homology with any known eukaryotic protein.

Downloaded from https://academic.oup.com/plphys/article/189/3/1226/6571825 by CNRS user on 28 June 2022

867 ORFs derived from the translated chromatophore genome (Singer et al., 2017), all available at https://www.ebi.ac.uk/pride/archive/projects/PXD006531 and UniProtKB. The 28 ORFs from the mitochondrial genome—only available in *P. micropora* (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA587507)—were not considered.

In their previous classical proteomic analysis, the same group had already suggested that chromatophores import two classes of nuclear-encoded proteins: one featuring N-terminal targeting signals (crTP) also functional in plants and the other devoid of these signals (Singer et al., 2017). The short imported proteins without crTP were long thought to traffic via the Golgi (Nowack and Grossman, 2012). Thanks to the cutting-edge mass spectrometry and sequence analysis performed in the current study, the authors make a decisive step forward in our understanding of proteins with a crTP. They show that this ∼200-residue pre-sequence—which is longer than a chloroplast TP (<100 residues)—is bipartite and only partially removed once in the chromatophore. The first part of the crTP contains a motif (Yxxφ, where φ is a hydrophobic residue) believed to mediate protein trafficking via the Golgi through adaptor protein 1 complexes that bind to ADP-ribosylation factors (Park and Guo, 2014). The second domain is involved in translocation across the chromatophore inner membrane. Finally, the authors show that in addition to proteolytic cleavage of the crTP upon chromatophore import, many new N-termini also undergo NTA. This modification is also frequently found on cytosolic nuclear-encoded proteins and chromatophore-encoded proteins. The authors suggest that there is a dedicated cytosolic acetylation machinery similar to that seen in other eukaryotes (NatA/B/C/D/E). In addition, the authors suggest the existence of a dedicated NTA machinery in the chromatophore, but they failed to identify homologs with NatGs seen in all Archaeplastida (Giglione and Meinnel, 2021).

The study suggests that the recent acquisition of a novel organelle such as the plastid involves the Golgi as the initial route for protein targeting. This mechanism has already been suggested in *Oryza sativa* (rice, a land plant) for a couple of proteins both featuring cleavage of a 20–30 amino acid signal peptide and N-terminal glycosylation (Kitajima et al., 2009; Kaneko et al., 2016). Whether other Bikonta display such a mechanism is unknown and how Archaeplastida have progressively gained direct transfer from the cytosol to the plastid for a majority of the imported ORFs is a conundrum. Identifying other organisms with intermediary endosymbiosis is critical to resolving this problem.

While the added value of the biological model is very appealing, the HUNTER technology is limited in its capacity to further expand the data, as it relies on the quality of proteome annotation and sample preparation, as with any other mass spectrometry-based approach. As the authors mention, proteins may degrade during sample preparation or as a natural cellular process, and both natural and unnatural proteolytic species can accumulate. The authors indicate that NTA is one marker of natural proteolysis. Due to high instability of the acetyl donor, acetyl-CoA, NTA is unlikely to occur once the organelle is disrupted. For instance, they identified various N-termini of the large subunit of Rubisco encoded in the chromatophore, none of which corresponded to the full length active subunit and several of them showing NTA. Rubisco is highly abundant and it is indeed likely that degradation intermediates may have accumulated. It is striking that these intermediates may be further post-translationally N-acetylated. One cannot exclude that—during fractionation—the chromatophore undergoes proteolytic stress leading in organello to this accumulation. Another bias in the approach is one inherent to subcellular fractionation: in such approaches, there is often contamination with other, closely linked subcellular compartments, such as mitochondria in chloroplast preparations from plants. In this study, the identification of 60S ribosomal (L12) and cytoskeleton (fascin or actin) protein homologs is evidence of such contamination, as mentioned and cross-checked by the authors in their Supplemental Table S1 (column AA, sheet "nucleus-encoded"). It shows cytosolic contamination was <5% and confirms high purity of the chromatophore preparations (see also Supplemental Figure S1 and Supplemental Table S2 in Singer et al., 2017). Among the contaminants, one actin homolog displaying NTA at Glu3 is peculiar, since NTA is a unique processing activity displayed by NAA80 (NatH) on cytosolic actin (Aksnes et al., 2019). Furthermore, this unambiguous characterization reveals that in addition to the aforementioned set of N-terminal acetyl transferases (Nats), there exists a nuclear-encoded NatH in Cercozoa (NatH homologs have not been identified in Archaeplastida to date; Giglione and Meinnel, 2021). This further reinforces the idea that the primary eukaryote engulfing the cyanobacteria at the origin of the chromatophore is distinct. A third limitation arises due to the partial nuclear genomic information, its extreme complexity so far only addressed through 60,000 sequences from the transcriptomic analysis and no gene model, except by comparison with those derived from the *P. micropora* draft genome (Lhee et al., 2021). Many ORFs are only partial and the N-terminus is the most uncertain part, with a majority of ORFs lacking the canonical methionine start codon. As a result, the position of the N-terminus in many identifications is uncertain and it is difficult to conclude whether or not a signal peptide is cleaved or whether the cleavage is a priori canonical (e.g. peptides 631/635). It may also be true that the first Met codon is not the start codon and that another alternative Met (or an unusual non-AUG codon; Kearse and Wilusz, 2017; Na et al., 2018) is used (e.g. peptides 66/693/291 with Met6/32/37 and cleavages at positions 45/46/50). To reinforce their conclusions, the authors filter their search to only transcripts displaying the 5′-trans-splicing tag, demonstrating that the full-length cDNA was retrieved (see their Supplemental Table S1, column W). This corresponds to 26.1% of the nuclear encoded unique transcripts of the transcriptome (Supplemental Figure S6

Downloaded from https://academic.oup.com/plphys/article/189/3/1226/6571825 by CNRS user on 28 June 2022

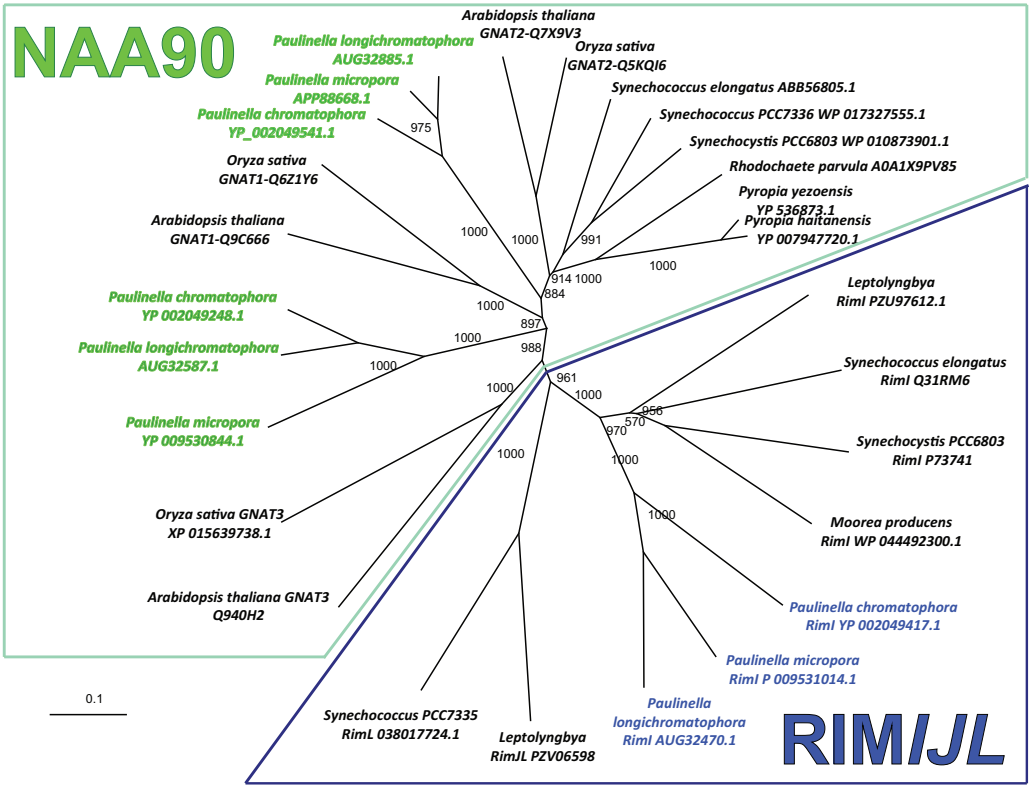


Figure 2 Two gene homologs with strong homologies to cyanobacterial and chloroplast *N*-acetyltransferases NatG occur in the chromatophore genome of *Paulinella* spp. A *Synechococcus elongatus* GNAT2 homolog was used to challenge BLAST *Paulinella* protein diversity. In total, 14 sequences were selected, and the Nats of several cyanobacteria, including RimIJL, were added to NAA90 members from rice (*O. sativa*) and Arabidopsis (*A. thaliana*) and red algae plastids. In total, 27 sequences were aligned and the bootstrap tree was constructed (Grzela et al., 2017). Internal values labeled on each node record the stability of the branch over the 1,000 bootstrap replicates. Together with NAA70, NAA90 is one of two subgroups of NatGs and made of GNAT1/2/3.

in Nowack et al., 2016). An interesting example is ribosomal L12, which corresponds to a strongly conserved protein over the entire length of its sequence—including the N-terminus—in Eukaryotes (Figure 1). In *P. chromatophora*, the L12 protein homolog is encoded by two transcripts (scaffold13378/m.87086 and scaffold19310/m.108443) only differing at their 5′-end, one long and the other much shorter. The retrieved peptide matches codon 35 after the fourth methionine of the long transcript and the other matches codon 2 after the first methionine of the short transcript (Figure 1). While the occurrence of NTA suggests that the peptide is present, the origin of the primary substrate, and therefore the exact N-terminal sequence, therefore needs further analysis. The authors select the short L12 transcript over the long transcript and reject this protein as a cytosolic contaminant. What the long transcript—likely even longer if not containing the trans-splicing tag—corresponds to is unknown. In the *P. micropora* draft genome annotation, the L12 protein can be retrieved (g77960.t1, 88% identity over the 162 residues of the short isoform; Figure 1), which indeed encodes a 140-residue extension. This extension does not match that of the long isoform. While this validates the interpretation of the current analysis, it raises questions about protein synthesis and/or processing-related mechanisms in Chromalveolata (see for instance Gabr et al., 2022).

An important conclusion of this study relates to the occurrence of protein NTA in the chromatophore. The authors mention the similarity with observations in Archaeplastida, namely NTA of proteins of both origins, the "ragging" of proteins devoid of crTP starting on very close residues and undergoing partial trimming after initial cleavage, as observed in plants (Bienvenut et al., 2012; Rowland et al., 2015). A large panel of N-terminal residues may undergo NTA, either through a family of dedicated enzymes like those seen in the cytosol of all eukaryotes or through Nats with relaxed specificity, like in the chloroplast (Bienvenut et al., 2020; Giglione and Meinnel, 2021). It is therefore important to identify the associated Nat(s), but the authors were unsuccessful. In a recent study of Nats in photosynthetic organisms, a family of eight nuclear-encoded and plastid-located Nats (NatG) with relaxed specificity was shown to perform plastid NTA in all Archaeplastida. NatGs are very similar to *N*-acetyltransferases from cyanobacteria like *Synechococcus* or *Leptolyngbya*, all featuring NTA of four photosystem 2 components (Yang et al., 2014; Giglione and Meinnel, 2021). These Nats are specific to cyanobacteria and add to the classical RIMI/J/L-specific set of Nats found in

Downloaded from https://academic.oup.com/plphys/article/189/3/1226/6571825 by CNRS user on 28 June 2022

bacteria. In addition, in the fossil red algae *Pyropia yezoensis* and *Rhodochaete parvula*, a NatG (GNAT2) homolog is "still" encoded in the plastid genome, suggesting that the gene progressively moved to the nucleus. Likewise, using cyanobacteria-specific Nats, one can identify NatG homologs in all three chromatophore genomes (Figure 2). These GNATs are also close homologs to the plant NAA90s, one of the two NatG subgroups (Giglione and Meinnel, 2021). In the absence of reverse genetics tools for Paulinellidae, alternative approaches to characterize and confirm *Paulinella* NatG activities would be interesting to implement (Asensio et al., 2022). One conclusion therefore is that NTA in *Paulinella* spp. arises through the same mechanism seen in Archaeplastida or Cyanobacteria. In this context, it would be interesting to explore the role of NTA in the chromatophore in photosynthesis adaptation, like in plants or other processes involving protein stability (Koskela et al., 2018). Finally, whether GNAT2 displays similar NTA and lysine acetyltransferase activity in *Paulinella* as it does in plants is also unknown.

Supplemental data

The following materials are available in the online version of this article.

Supplemental Figure S1. A schematized tree of life showing the high representation of photosynthetic organisms in biodiversity.

Funding

The work of the team investigating acetylation was supported by the French National Research Agency (ANR) KatNat (ERA-NET, ANR-17-CAPS-0001-01), CanMore (France-Germany PRCI, ANR-20 CE92-0040), and Fondation ARC (ARCPJA32020060002137) grants. The work has benefited from the support of a French State grant (ANR-17-EUR-0007, EUR SPS-GSR) managed by the ANR under an Investments for the Future program (ANR-11-IDEX-0003-02) and from the facilities and expertise of the I2BC proteomic platform SICaPS, supported by IBiSA, Ile de France Region, Plan Cancer, CNRS, and Paris-Saclay University.

Conflict of interest statement. None declared.

References

Agrawal S, Striepen B (2010) More membranes, more proteins: complex protein import mechanisms into secondary plastids. Protist **161**: 672–687

Aksnes H, Ree R, Arnesen T (2019) Co-translational, post-translational, and non-catalytic roles of N-terminal acetyltransferases. Mol Cell **73**: 1097–1114

Asensio T, Dian C, Boyer J-B, Rivière F, Meinnel T, Giglione C (2022) A continuous assay set to screen and characterize novel protein N-acetyltransferases unveils rice general control non-repressible 5-related N-acetyltransferase 2 activity. Front Plant Sci **13**: 832144

Bienvenut WV, Brünje A, Boyer J-B, Mühlenbeck JS, Bernal G, Lassowskat I, Dian C, Linster E, Dinh TV, Koskela MM, et al. (2020) Dual lysine and N-terminal acetyltransferases reveal the complexity underpinning protein acetylation. Mol Syst Biol **16**: e9464

Bienvenut WV, Sumpton D, Martinez A, Lilla S, Espagne C, Meinnel T, Giglione C (2012) Comparative large scale characterization of plant versus mammal proteins reveals similar and idiosyncratic N-alpha-acetylation features. Mol Cell Proteomics **11**: M111 015131

Christian RW, Hewitt SL, Nelson G, Roalson EH, Dhingra A (2020) Plastid transit peptides-where do they come from and where do they all belong? Multi-genome and pan-genomic assessment of chloroplast transit peptide evolution. PeerJ **8**: e9772

Falciatore A, Jaubert M, Bouly J-P, Bailleul B, Mock T (2020) Diatom molecular research comes of age: model species for studying phytoplankton biology and diversity. Plant Cell **32**: 547–572

Gabr A, Stephens TG, Bhattacharya D (2022) Hypothesis: Trans-splicing generates evolutionary novelty in the photosynthetic amoeba *Paulinella*. J Phycol. doi:10.1111/jpy.13247

Giglione C, Meinnel T (2021) Evolution-driven versatility of N terminal acetylation in photoautotrophs. Trends Plant Sci **26**: 375–391

Grigoriev IV, Hayes RD, Calhoun S, Kamel B, Wang A, Ahrendt S, Dusheyko S, Nikitin R, Mondo SJ, Salamov A, et al. (2021) PhycoCosm, a comparative algal genomics resource. Nucleic Acids Res **49**: D1004–D1011

Grzela R, Nusbaum J, Fieulaine S, Lavecchia F, Bienvenut WV, Dian C, Meinnel T, Giglione C (2017) The C-terminal residue of phage Vp16 PDF, the smallest peptide deformylase, acts as an offset element locking the active conformation. Sci Rep **7**: 11041

Hopkins JF, Spencer DF, Laboissiere S, Neilson JA, Eveleigh RJ, Durnford DG, Gray MW, Archibald JM (2012) Proteomics reveals plastid- and periplastid-targeted proteins in the chlorarachniophyte alga *Bigelowiella natans*. Genome Biol Evol **4**: 1391–1406

Kaneko K, Takamatsu T, Inomata T, Oikawa K, Itoh K, Hirose K, Amano M, Nishimura S, Toyooka K, Matsuoka K, et al. (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**: 1610–1628

Karsenti E, Acinas SG, Bork P, Bowler C, De Vargas C, Raes J, Sullivan M, Arendt D, Benzoni F, Claverie JM, et al. (2011) A holistic approach to marine eco-systems biology. PLoS Biol **9**: e1001177

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

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

Koskela MM, Brünje A, Ivanauskaite A, Grabsztunowicz M, Lassowskat I, Neumann U, Dinh TV, Sindlinger J, Schwarzer D, Wirtz M, et al. (2018) Chloroplast acetyltransferase NSI is required for state transitions in *Arabidopsis thaliana*. Plant Cell **30**: 1695–1709

Lhee D, Ha J-S, Kim S, Park MG, Bhattacharya D, Yoon HS (2019) Evolutionary dynamics of the chromatophore genome in three photosynthetic *Paulinella* species. Sci Rep **9**: 2560

Lhee D, Lee J, Ettahi K, Cho CH, Ha JS, Chan YF, Zelzion U, Stephens TG, Price DC, Gabr A, et al. (2021) Amoeba genome reveals dominant host contribution to plastid endosymbiosis. Mol Biol Evol **38**: 344–357

Lynch M, Field MC, Goodson HV, Malik HS, Pereira-Leal JB, Roos DS, Turkewitz AP, Sazer S (2014) Evolutionary cell biology: two origins, one objective. Proc Natl Acad Sci USA **111**: 16990–16994

Downloaded from https://academic.oup.com/plphys/article/189/3/1226/6571825 by CNRS user on 28 June 2022

Na CH, Barbhuiya MA, Kim MS, Verbruggen S, Eacker SM, Pletnikova O, Troncoso JC, Halushka MK, Menschaert G, Overall CM, et al. (2018) Discovery of noncanonical translation initiation sites through mass spectrometric analysis of protein N termini. Genome Res **28**: 25–36

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

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

Oberleitner L, Perrar A, Macorano L, Huesgen PF, Nowack ECM (2022) A bipartite chromatophore transit peptide and N-terminal protein processing in the *Paulinella* chromatophore. Plant Physiol **189**: 152–164

Park SY, Guo X (2014) Adaptor protein complexes and intracellular transport. Biosci Rep **34**: e00123

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

Singer A, Poschmann G, Mühlich C, Valadez-Cano C, Hänsch S, Hüren V, Rensing SA, Stühler K, Nowack ECM (2017) Massive protein import into the early-evolutionary-stage photosynthetic organelle of the amoeba *Paulinella chromatophora*. Curr Biol **27**: 2763–2773.e2765

Stephens TG, Gabr A, Calatrava V, Grossman AR, Bhattacharya D (2021) Why is primary endosymbiosis so rare? New Phytol **231**: 1693–1699

Szathmáry E (2015) Toward major evolutionary transitions theory 2.0. Proc Natl Acad Sci USA **112**: 10104–10111

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

Yang M-K, Yang Y-H, Chen Z, Zhang J, Lin Y, Wang Y, Xiong Q, Li T, Ge F, Bryant DA, et al. (2014) Proteogenomic analysis and global discovery of posttranslational modifications in prokaryotes. Proc Natl Acad Sci USA **111**: E5633–E5642

Yoon HS, Hackett JD, Van Dolah FM, Nosenko T, Lidie KL, Bhattacharya D (2005) Tertiary endosymbiosis driven genome evolution in dinoflagellate algae. Mol Biol Evol **22**: 1299–1308

Downloaded from https://academic.oup.com/plphys/article/189/3/1226/6571825 by CNRS user on 28 June 2022