Converting antimicrobial into targeting peptides reveals key features governing protein import into mitochondria and chloroplasts

Oliver D Caspari[1*†], Clotilde Garrido[1], Chris O Law[2], Yves Choquet[1], Francis-André Wollman[1], Ingrid Lafontaine[1*]

[1]UMR7141 (CNRS/Sorbonne Université), Institut de Biologie Physico-Chimique, 13 Rue Pierre et Marie Curie, 75005 Paris, France

[2]Centre for Microscopy and Cellular Imaging, Biology Dept. Loyola Campus of Concordia University, 7141 Sherbrooke W., Montréal H4B 1R6, QC, Canada

*corresponding authors:

ODC: **odc20@cantab.net**

IL: **ingrid.lafontaine@ibpc.fr**

current addresses:

[†]Department of Microbiology, Institut Pasteur, 28 Rue Docteur Roux, 75015 Paris, France

[¶]UMR7238 (CNRS/Sorbonne Université), Institut de Biologie Paris-Seine, Sorbonne Jussieu 7-9 quai Saint Bernard, 75005 Paris, France

Summary

We asked what peptide features govern targeting to the mitochondria versus the chloroplast using antimicrobial peptides as a starting point. This approach was inspired by the endosymbiotic hypothesis that organelle-targeting peptides derive from antimicrobial amphipathic peptides delivered by the host cell, to which organelle progenitors became resistant. To explore the molecular changes required to convert antimicrobial into targeting peptides, we expressed a set of 13 antimicrobial peptides in *Chlamydomonas reinhardtii*. Peptides were systematically modified to test distinctive features of mitochondrial and chloroplast targeting peptides, and we assessed their targeting potential by following the intracellular localization and maturation of a Venus fluorescent reporter used as cargo protein. Mitochondrial targeting can be achieved by some unmodified antimicrobial peptide sequences. Targeting to both organelles is improved by replacing Lysines with Arginines. Chloroplast targeting is enabled by the presence of flanking unstructured sequences, additional constraints consistent with chloroplast endosymbiosis having occurred in a cell that already contained mitochondria. If indeed targeting peptides evolved from antimicrobial peptides, required modifications imply a temporal evolutionary scenario with an early exchange of cationic residues, and a late acquisition of chloroplast specific motifs.

Keywords:

1. Antimicrobial peptide
2. *Chlamydomonas reinhardtii*
3. Chloroplast
4. Endosymbiotic organellogenesis
5. Mitochondria
6. Presequence
7. Targeting peptide
8. Transit peptide

Introduction

Mitochondria and chloroplasts arose through endosymbiosis and retain their own genomes, yet the vast majority of organellar proteins is encoded in the nucleus, translated in the cytoplasm and imported into the organelles (Wiedemann and Pfanner, 2017; Chotewutmontri et al., 2017). N-terminal targeting peptides (TPs) that are cleaved off upon import provide the information on targeting, although their primary structures are very diverse (Bruce, 2001). Yet chloroplast transit peptides (cTP) and mitochondrial presequences (mTP) have very similar physico-chemical properties making reliable differentiation often challenging. A multitude of studies has found sequence elements contributing to specificity determination, and prediction programs have been improving (Tardif et al., 2012; Armenteros et al., 2019), yet a mechanistic understanding of how targeting information is encoded has remained elusive.

Here, we use antimicrobial peptides (AMP) as an original chassis, to test the contribution of key TP features towards targeting efficiency and specificity. Part of the innate immune system, AMPs are produced by virtually all types of cells in a bid to kill or control microbial adversaries (Joo et al., 2016; Lazzaro et al., 2020). AMPs have recently been proposed to be at the evolutionary origin of TPs (Wollman, 2016; Caspari and Lafontaine, 2021). The proposed evolutionary scenario posits that early during endosymbiotic organellogenesis of first the mitochondrion and later the chloroplast, the host cell used AMPs to attack the bacterial proto-endosymbiont. A bacterial resistance mechanism, whereby the AMP is imported into the bacterial cytoplasm, would have generated a pathway for host proteins to reach the bacterial cytosol – a plausible first step in the evolution of a protein import machinery. Cationic, Helical-Amphipathic Ribosomally-produced AMPs (HA-RAMPs) and TPs share key physico-chemical properties and they were shown, in some instances, to retain cross-functionality (Garrido et al., 2020): several TPs showed antimicrobial activity and selected HA-RAMPs, fused to a cleavage-site containing TP element, were shown to promote the import of a Venus fluorescent protein into either the mitochondria or the chloroplast of the model green alga *Chlamydomonas reinhardtii*.

The main part of mTPs and the central element of cTPs most closely resemble HA-RAMPs on account of a shared cationic, amphipathic helical structure that often encompasses the entire length of HA-RAMPs (Caspari and Lafontaine, 2021) and mTPs (von Heijne, 1986; von Heijne et al., 1989). Although plant cTPs have been described as unstructured (von Heijne and Nishikawa, 1991), algal cTPs more closely

85 resemble mTPs in being helical (Franzén et al., 1990). Helices were observed by NMR
86 in membrane-mimetic environments in both algal and plant cTPs (Lancelin et al., 1996;
87 Bruce, 1998; Krimm et al., 1999; Wienk et al., 2000; Bruce, 2001), and signatures of
88 amphipathic helices can be detected in a majority of cTPs (Garrido et al., 2020).
89 In addition to the helices, both mTPs and cTPs contain recognition sites at the C-
90 terminus where processing peptidases in the mitochondrial matrix (MPP) and the
91 chloroplast stroma (SPP) cleave off the cargo protein (Teixeira and Glaser, 2013).
92 These recognition sites encompass some 10 residues upstream of the cleavage site and
93 are structurally distinct from the rest of the TPs, even showing a weak sequence
94 conservation (von Heijne et al., 1989; Tardif et al., 2012; Köhler et al., 2015). While
95 targeting information is usually contained within mTP sequences upstream of the
96 cleavage site, targeting by cTPs shorter than ca. 60 amino acids often requires
97 downstream unstructured sequence stretches in the N-terminal domain of the mature
98 protein (Bionda et al., 2010; Caspari, 2022). Besides the amphipathic helical module
99 and the C-terminal cleavage module shared between mTPs and cTPs, it has been argued
100 that distinct features at their N-termini conferred organelle specificity to each set of
101 targeting peptides (von Heijne et al., 1989; Ivey and Bruce, 2000; Ivey et al., 2000;
102 Bhushan et al., 2006; Chotewutmontri et al., 2012; Chotewutmontri and Bruce, 2015;
103 Köhler et al., 2015; Chotewutmontri et al., 2017; Lee et al., 2019).
104 In this study, we systematically introduced modifications into a diversity of HA-
105 RAMPs in a bid to generate targeting to the mitochondria or chloroplast in
106 Chlamydomonas. This dataset provides new insights into how different TP elements
107 contribute to differential targeting and to the efficiency of protein import. Being similar
108 in physico-chemical properties to TPs (Garrido et al., 2020), HA-RAMPs provide a
109 privileged vantage point from which to study sequence elements that govern targeting.
110 In our choice of HA-RAMPs, we aimed to reflect the diversity of available sequences
111 by choosing representatives of different HA-RAMP families based on similarity with
112 TPs. We show that some of our 13 HA-RAMPs natively contain TP-like sequence
113 elements, with some HA-RAMPs being prone to chloroplast targeting while others
114 show a preference for the mitochondria. Furthermore, we provide evidence for a critical
115 functional difference in cationic residues with Lysine (K) being used in HA-RAMPs
116 and Arginine (R) in TPs.

117 **Results**

118 ***HA-RAMPs display varying degrees of similarity to TPs***

119 Figure 1 shows the major sequence features of the 13 HA-RAMPs used in the present
120 study (Fig. 1a) together with those of a typical cTP and a typical mTP (Fig. 1b). On the
121 right-hand side, these peptides are represented according to their proportion of α-helical
122 amphipathic structure. The HA-RAMPs Brevinin-2ISb (B2I), Magainin 2 (MII),
123 Ranatuerin-2G (R2G), Dermaseptin S4 (DS4), Dermadistinctin-M (DDM), Brevinin-
124 1E (B1E), Cecropin-P3 (CP3), Sarcotoxin-1D (S1D), Esculentin-1SEA (E1S),
125 Leucocin A (LCA), SI Moricin (SIM), Bacillocin 1580 (B15), and Enterocin HF (EHF)
126 were chosen such that different AMP families would be represented with varying
127 proximity to TPs (Table S1) and thus a range of physico-chemical properties be
128 explored (Fig. S1). As negative controls, two peptides that lack predicted amphipathic
129 helices were chosen from among randomly generated sequences (Fig. 1c).

130 To explore targeting, the 13 HA-RAMPs were systematically modified (Fig. 2), notably
131 by adding TP elements from the N- and/or C-terminal domains of Chlamydomonas
132 Rubisco activase (RCA1) cTP, or a C-terminal domain of similar length from
133 mitochondrial γ-carbonic anhydrase 2 (CAG2) mTP (Fig. 1b, Fig. 2). In a bid to keep
134 peptidase recognition sites intact, C-terminal elements were designed to start 10
135 residues upstream of the cleavage site, even though this slightly truncates TP helices.
136 Peptides carrying the 15 amino-acid long cTP N-terminus (cTP-N) will be denoted ^{c}P.
137 Similarly, peptides with C-terminal elements from RCA1-cTP (cTP-C) or CAG2-mTP
138 (mTP-C) will be denoted P^{c} or P^{m} respectively. Peptide variants are used to drive the
139 subcellular localization of a Venus fluorescent reporter protein, which was assessed
140 using fluorescence microscopy; in Fig. 2 an executive overview of observed Venus
141 localization is presented for all constructs. We previously validated fluorescence
142 localization biochemically in a small number of strains by showing that the Venus-
143 FLAG reporter is retained within isolated mitochondria or chloroplasts (Garrido et al.,
144 2020). Here, we use automated image segmentation to validate our subcellular
145 localization assessment (Fig. S2, Supplementary text). Localization was obtained using
146 stable Chlamydomonas expression lines, generated by introducing DNA sequences
147 encoding peptides upstream of the Venus coding sequence in a bicistronic expression
148 vector (Caspari, 2020) using Gibson assembly, with transformation cassettes integrated
149 into the Chlamydomonas nuclear genome at random sites via electroporation.

150 Micrographs for three biological replicates (i.e., independent insertion lines) per
151 construct are displayed in Fig. S4-S21. The reader will be asked to return to Fig. 2
152 throughout, with Fig. 3 and Fig. 4 providing selected micrographs as examples
153 highlighting particular points of interest.

154 ***Some unmodified HA-RAMPs generate mitochondrial targeting***

155 In the absence of any modifications, B2I and MII were capable of organelle-targeting
156 visible in fluorescence microscopy (Fig. 2 row A, Fig. S3-S15). When equipped with
157 B2I, the fluorescent reporter Venus shows mitochondrial localization (Fig. 3A, Fig.
158 S3): the Venus signal is observed as a characteristic network pattern that co-localizes
159 with MitoTracker fluorescence (cf. Fig. S2A, S4). By contrast, in the case of MII only
160 a fraction of the fluorescence signal in the Venus channel shows co-localization with
161 the MitoTracker dye, signifying that targeting is only partial (Fig. 3A, Fig. S5). Note
162 that in these epifluorescence images, some autofluorescence emanating from the
163 chloroplast is present in the Venus channel (Fig. 3B). In all images, display brightness
164 was adjusted such that Venus localization would be clearly visible, and where possible
165 this included removing autofluorescence. Brightness settings were chosen as indicated
166 next to each image for full transparency. In cases where low Venus accumulation
167 necessitated brightness values low enough for autofluorescence to be visible in the
168 Venus channel, a black dotted line is visible in the brightness display, indicating that
169 some autofluorescence signal co-localizing with chlorophyll should be expected in the
170 Venus channel independent of the genuine Venus localization.

171 ***K/R content contributes to functional divergence between HA-RAMP and TP***

172 Extant HA-RAMPs and organellar TPs display very few differences in their amino-acid
173 content (Supplementary text, Fig. S20). As expected, both are poor in acidic residues
174 (D, E) but enriched in basic residues (K or R). However, their complement in basic
175 residues is markedly different (Fig. 3C): HA-RAMPs are rich in Lysines (K) whereas
176 TPs are rich in Arginines (R).
177 In order to see whether these contrasting differences in K/R ratio contributed to the
178 functional divergence between HA-RAMPs and TPs, we substituted all instances of K
179 by R in HA-RAMPs and of R by K in TPs. In the rest of the text, the basic amino acid
180 mostly present in a given peptide P are indicated by a subscript as P_R or P_K. HA-RAMPs
181 with K→R transition showed reduced antimicrobial activity, as illustrated on Fig. 3D
182 by the increased minimum inhibitory concentrations for MII, DS4 or B1E.

183 We used Chlamydomonas Rubisco activase (RCA1) cTP ($^{c}CH_{R}^{c}$) and mitochondrial γ-
184 carbonic anhydrase 2 (CAG2) mTP (MH_{R}^{m}) as positive controls for chloroplast and
185 mitochondrial targeting respectively (Fig. 3E, Fig. S18-S19). Note that TP helical
186 fragments, stripped of their N- and C-terminal domains, are denoted as MH for mTP
187 and CH for cTP (as detailed in Fig. 1B). When equipped with RCA1-cTP ($^{c}CH_{R}^{c}$),
188 Venus shows two chloroplast localization features: a diffuse signal that colocalizes with
189 chlorophyll fluorescence, and a bright spot where there is a drop in chlorophyll
190 fluorescence. Both of these features are genuine markers of chloroplast localization: in
191 Chlamydomonas, the single cup-shaped chloroplast has a reliable morphology, and the
192 dip in chlorophyll fluorescence at the base of the chloroplast is a well-established
193 marker of the pyrenoid, a proteinaceous chloroplast sub-compartment that contains a
194 large majority of Rubisco (Fig. S2B) (Mackinder et al., 2016; Caspari et al., 2017;
195 Mackinder et al., 2017; Caspari, 2022). While small proteins like Venus can enter the
196 pyrenoid, most thylakoid membranes are excluded, leading to the observed decrease in
197 chlorophyll autofluorescence in this spot. Venus channel signal emanating from the
198 pyrenoid is thus a useful visual guide to true chloroplast localization of the fluorescent
199 reporter (Caspari, 2022). Note that due to cTP-C containing a Rubisco-binding motif
200 (Meyer et al., 2020), Venus accumulation in the Rubisco microcompartment, the
201 pyrenoid, is particularly pronounced in constructs carrying this element. When natively
202 R-rich RCA1-cTP and CAG2-mTP sequences were subjected to systematic R→K
203 substitutions ($^{c}CH_{K}^{c}$, MH_{K}^{m}), their ability to target was reduced or abolished (Fig. 3E,
204 Fig. S18-S19). These experiments demonstrate the respective functional contributions
205 of K and R residues to antimicrobial and organelle targeting activity.
206 We then systematically re-examined the organelle targeting ability of the set of 13 HA-
207 RAMP that had undergone K→R substitutions (Fig. 2 row B, Fig. S3-S15). MII_{R} now
208 shows much improved mitochondrial targeting (Fig. 3F), with a large majority of Venus
209 colocalizing with MitoTracker (note that the MitoTracker occasionally appears to stain
210 the cell envelope in addition to the mitochondria, thus not all the MitoTracker signal
211 colocalizes with Venus; cf. Fig. S2A). $R2G_{R}$ as well as $CP3_{R}$ show a gain of partial
212 targeting, as evidenced by significant overlap of Venus and MitoTracker fluorescence
213 (Fig. 3F). Thus, an exchange of K into R increases targeting by HA-RAMP constructs.
214 This effect is not exclusive to mitochondrial targeting (Fig. 3G): Across constructs (i.e.
215 including those containing additional modifications), the K→R switch enabled or
216 improved mitochondrial targeting in 8 cases (Fig. 2, rows = capital letters, columns =

217 lower case letters; gain: ABcg, CDgi, GHf; improve: ABb, CDc, GHc) and chloroplast
218 targeting in 15 cases, including 3 cases of dual targeting (Fig. 2, gain: CDm, EFg,
219 GHgilm, KLg; improve: EFm, GHk, KLafj; dual: IJab, KLc). Lost or decreased
220 targeting was observed in only three cases (Fig. 2, loss: IJgj, decrease: EFh). See Fig.
221 S24 for illustrative examples.

222 ***TP C-termini matter for targeting***

223 Addition of mTP-C, the cleavage-site containing C-terminal element of CAG2-mTP
224 enabled partial mitochondrial targeting in two constructs ($R2G_K$ and $E1S_R$) and
225 improved mitochondrial targeting in a further two constructs ($R2G_R$ and $CP3_R$), in
226 addition to continued targeting by B2I and MII (Fig. 2 rows CD). As an indicative
227 example, gain of targeting in $R2G_K$ is shown in Fig. 4A. The impact of adding cTP-C,
228 the C-terminal element derived from RCA1-cTP, is even more important: cTP-C
229 significantly enabled chloroplast targeting, which could be seen in nine HA-RAMP
230 constructs involving $CP3_R$, S1D, LCA, B15 and EHF (Fig. 2 rows E,F). Note that the
231 addition of cTP-C was also compatible with mitochondrial localization by B2I and MII
232 (Fig. 2 rows E,F). Gain of partial chloroplast targeting in LCA_K and full targeting in
233 EHF_R are shown as examples in Fig. 4A. The addition of mTP-C also enabled
234 chloroplast localization by EHF_R (Fig. 4A, Fig. 2 row D). Low Venus accumulation in
235 EHF_R and EHF_R^m means brightness needed to be set low enough that autofluorescence
236 accounts for at least some of the signal that is colocalized with the chlorophyll channel
237 (Fig. 4A). EHF_R shows Venus signal in the cytoplasm, but not within the pyrenoid,
238 which was interpreted as an absence of targeting. By contrast, Venus signal emanating
239 from within the pyrenoid in EHF_R^m provides unambiguous evidence for chloroplast
240 localization.
241 In TP controls, deletion of mTP-C reduces but does not totally abolish mitochondrial
242 targeting of CAG2-mTP (MH_R), while replacing mTP-C with cTP-C (MH_R^c) partially
243 retargets the construct to the chloroplast (Fig. 4B; Fig. 2 column q, rows BDF). By
244 contrast, deletion of cTP-C abolishes chloroplast targeting by RCA1-cTP (^cCH_R), as
245 does replacing cTP-C with mTP-C ($^cCH_R^m$) (Fig. 4B; Fig. 2 column p, rows HIL).

246 ***cTP N-termini matter for chloroplast targeting***

247 The sole addition of cTP-N generated at least partial chloroplast targeting in 6 HA-
248 RAMP constructs, notably by SIM (Fig. 4C), but also $CP3_R$, $E1S_R$, $B15_R$ and EHF_R
249 (Fig. 2 rows G,H). However, cTP-N also enabled mitochondrial targeting in three HA-

250 RAMPs that had not previously shown mitochondrial localization, B1E (Fig. 4C),
251 DDM_R and S1D (Fig. 2 rows G,H).
252 The importance of cTP-N as a chloroplast determinant becomes more evident when
253 combined with a C-terminal element, as shown in Fig. 4D using the example of $^cB2I_R$:
254 while $^cB2I_R$ (partially) targets the mitochondria, addition of mTP-C ($^cB2I_R^m$) results in
255 dual targeting to the chloroplast in addition to the mitochondria, and addition of cTP-C
256 ($^cB2I_R^c$) results in targeting only to the chloroplast. Across HA-RAMP constructs,
257 combining cTP-N with mTP-C (Fig. 2 rows I,J) resulted in 6 cases of dual targeting
258 ($B2I_R$, MII_R, B1E, S1D) and 10 cases of chloroplast targeting ($CP3_K$, E1S, LCA_K, SIM,
259 B15, EHF). Combining cTP-N with cTP-C (Fig. 2 rows KL) generated at least partial
260 chloroplast targeting in 22 out of 26 HA-RAMP constructs (including 1 instance of dual
261 targeting by $R2G_R$); only DS4 fails to show any chloroplast targeting.
262 Finally, Fig. 4E (also Fig. 2 column q, rows JL) shows that while addition of cTP-N to
263 the CAG2-mTP abolished mitochondrial targeting ($^cMH_R^m$), the cTP-N/cTP-C
264 combination ($^cMH_R^c$) retargets to the chloroplast (see Fig. 4B for controls without cTP-
265 N). Replacing the native amphipathic helix of RCA1-cTP between cTP-N and cTP-C
266 with a poly-Alanine peptide of equal length ($^cAA^c$), or fusing cTP-N and cTP-C directly
267 with no intervening peptide (^{c}-^{c}) both lead to partial chloroplast targeting (Fig. 4E, Fig
268 2 rows KL, columns rs). The two latter experiments demonstrate that there are enough
269 determinants for recognition of the chloroplast translocon dispersed between the Nter
270 and Cter of a cTP, to target Venus to the chloroplast, albeit with a lower efficiency than
271 when an amphipathic helix is present in between. That the nature of the intervening
272 peptide matters can be further seen in constructs that fail to target the chloroplast in the
273 presence of cTP-N and cTP-C, such as RP2 (Fig. 4E), but also RP1, $CP3_K$, DS4 and
274 $R2G_K$ (Fig. 2 rows KL).

275 ***HA-RAMPs dominate targeting specificity***

276 Considering Fig. 2 by columns reveals that organelle specificity is, to a large extent,
277 determined by HA-RAMPs. Only 3 HA-RAMPs (B2I, R2G, MII; referred to hereafter
278 as the mt-set) account for >70% of all constructs where mitochondrial targeting is seen.
279 Similarly, 5 HA-RAMPs (E1S, LCA, SIM, B15, EHF; referred to as the cp-set) account
280 for the majority (~57%) of all chloroplast targeting, and for two thirds when excluding
281 the cTP-N/cTP-C combination that generates chloroplast targeting across most HA-
282 RAMPs. In some instances, a set of peptide modifications may switch targeting from

283 the mitochondria to the chloroplast, yet the major effect of modifications, i.e.
284 exchanging K→R (Fig. 3, Fig. S24) or adding TP elements (Fig. 4), is to enhance the
285 targeting ability to an organelle determined by the HA-RAMP primary sequence
286 properties (Fig. 2).

287 ***Probing cleavage of HA-RAMP-driven reporter constructs by immunoblotting***

288 To characterize the maturation of HA-RAMP–targeted proteins upon organellar import,
289 we performed immunoblotting experiments using whole-cell extracts probed with a
290 FLAG antibody targeting the Venus-FLAG reporter (Fig. 5). Indicative examples were
291 selected for clarity; a more comprehensive overview is provided in Fig. S28.
292 In the absence of a dedicated cleavage site (Fig. 5A), we found some processing did
293 occur but preproteins are also maintained, as evidenced by two bands being present in
294 a given lane. The upper band corresponds to unprocessed Venus, which is a fusion of
295 the reporter with the HA-RAMP construct and migrates at varying positions depending
296 on the length of the presequence. The lower band corresponds to the processed form
297 migrating closer to the "free" Venus position (lane a) depending on the exact site of
298 cleavage. In HA-RAMP constructs, the majority of Venus remained in the unprocessed
299 top band for constructs targeting the mitochondria (lanes d-h). Nonetheless, the
300 presence of a faint processed form as lower bands just above free Venus shows that
301 some processing occurs. Such processing is suggestive of import but not conclusive,
302 since proteolysis can occur unlinked to import. The proportion of processed versus
303 unprocessed Venus preprotein is much higher in partially chloroplast-targeted $^{c}SIM_{K}$
304 (lane d), hinting at more efficient degradation of the unprocessed form in the
305 chloroplast.
306 Addition of a cleavage site improves processing across organelles. In the presence of
307 cTP-C (Fig. 5B), which contains the RCA1-cTP cleavage site, HA-RAMP constructs
308 that show evidence of organellar targeting appear to be processed at a site
309 corresponding to the cTP control (lane h), independently of whether the mitochondria
310 or the chloroplast are targeted (lanes c-g). Constructs that do not target (lanes a,b; see
311 also Fig. S28C lanes a,e,i) also appear to be processed, but at a site different to the one
312 used in organelles, suggesting cTP-C may be recognized by cytoplasmic peptidases.
313 Fig. 5C shows constructs equipped with mTP-C, which contains the CAG2-mTP
314 cleavage site. Here, the mTP control (lane a) shows two bands, consistent with partial
315 targeting observed by microscopy (cf. Fig. 3E, Fig. 4B). The top band thus likely

316 corresponds to the preprotein and the bottom band to the mature form within the
317 mitochondria. HA-RAMP constructs equipped with mTP-C that target either
318 mitochondria or the chloroplast (lanes b-g) show lower bands that migrate at or near
319 the cleaved mTP control. These results indicate that mTP and cTP cleavage sites are
320 recognized in both organelles.
321 A switch from K to R (Fig. 5D), which improved targeting, also increased the amount
322 of processed form relative to that of the unprocessed form in all cases, whether due to
323 a gain of targeting (lanes f-i) or an increase in efficiency (lanes b-e), suggesting that
324 cleavage can serve as a proxy for targeting. Indeed, constructs that show targeting, as
325 judged from microscopy, also show evidence of cleavage in immunoblots: second
326 bands are present for constructs lacking cleavage sites (Fig. S28A,D), while cleavage
327 site-containing constructs migrate at the size expected for processed cTP-C (Fig.
328 S28C,F) or mTP-C (Fig. S28B,E). We note that faint processed bands can be seen for
329 several additional constructs (Fig. S28), notably among unmodified, HA-RAMP (Fig.
330 S28A) suggesting partial targeting may occur in these cases, backed up by high image
331 quantification values e.g. for $CP3_K$ and $R2G_K$ (Fig. S2EF, S5A, S9A). Such very partial
332 targeting is however below the detection limit of our targeting assessment based on
333 fluorescence microscopy.

334 ***Chloroplast targeting involves longer, less helical peptides***

335 To understand what differentiates mt-set from cp-set HA-RAMPs, we examined the
336 sequence characteristics of our 13 HA-RAMPs with those of well-characterized
337 Chlamydomonas targeting peptides (Fig. 6). In Chlamydomonas, cTPs are on average
338 49 residues in length (Fig. 6A), and significantly longer than mTPs (t-test, p=0.0017)
339 that are on average 37 residues long. The difference is even greater than shown here
340 given that many cTPs require a contribution from post-cleavage site residues for
341 successful targeting (Bionda et al., 2010; Caspari, 2022). Consistently, cp-set HA-
342 RAMPs (in green, Fig. 6A) are longer than mt-set HA-RAMPs (in orange; p=0.0184)
343 and require further elongation by addition of TP elements before targeting can be
344 observed (Fig. 2).
345 Most mTPs fold into an amphipathic helix for ~80% of the sequence on average, Fig.
346 6B) starting directly from their N-terminus (Fig. 6C). The fraction dedicated to
347 amphipathic helix formation in cTPs is significantly lower (Fig. 6B, p<0.0001), with a
348 longer stretch of non-helical sequence at the N-terminus (Fig. 6B, p<0.0001). In line

349 with this contrast, 4 out of 5 of the cp-set HA-RAMPs have longer sequence stretches
350 upstream of the main helix than the mt-set HA-RAMPs (Fig. 1, Fig. 6C). (Franzén et
351 al., 1990) (Supplementary text, Fig. S22, Fig. S23). Also note that cTPs are predicted
352 to be more prone to protein interaction than mTPs (Supplementary text, Fig. S26, Fig.
353 S27).
354

355 **Discussion**

356 ***How to make TPs from HA-RAMPs***
357 To gain insight into peptide features that govern targeting, we used a related but
358 separate class of peptides, HA-RAMPs, as chassis and stacked modifications to
359 generate targeting into mitochondria or chloroplasts, or in some cases into both (Fig.
360 7A).
361 ***Arginines target better than Lysines***
362 We showed that a K-to-R-switch increases targeting efficacy to both organelles across
363 many constructs. R is more common than K in cTPs across green algae, vascular plants,
364 red algae, glaucophytes and many secondary plastids (Patron and Waller, 2007), but
365 the functional significance for targeting efficacy had not previously been recognized.
366 Our observation is in line with a previous report that a R-to-K switch at the N-ter of an
367 Arabidopsis mTP abolished targeting (Lee at al., 2019). The underlying mechanism
368 warrants further research. Contributing factors may be differences in bulkiness and pKa
369 (Li et al., 2013; Li et al., 2017), or in trans-acting factors that regulate targeting. For
370 example, R features in consensus sequences of the cleavage sites (Tardif et al., 2012;
371 Calvo et al., 2017); recognition of R by processing peptidases may thus potentially
372 account for increased chloroplast-localization among cTP-N bearing constructs (Fig. 2
373 rows GH). Since K is used for ubiquination (Mattiroli and Sixma, 2014), a preference
374 for R over K in TPs may help to protect preproteins from degradation while they transit
375 in the cytosol (Fig. 7A). We note that the 20S proteasome is already present in archaea,
376 with effectors specifically targeting K, like in the eukaryotic cytosol (Maupin-Furlow
377 et al., 2006; Maupin-Furlow, 2013) which means that K residues were likely targets for
378 protein degradation when primary endosymbiosis led to formation of the proto-
379 mitochondrion.

380 ***Unstructured sequences contribute to chloroplast determination***

381 Differences between cTP and mTP N-termini have previously been recognized to
382 contribute to differential targeting (Bhushan et al., 2006), but the sequence features
383 underlying this differentiation have been a matter of debate. In vascular plants, the
384 presence of a N-terminal multi-Arginine motif was shown to prevent chloroplast
385 import, leading to a proposal that this feature was solely responsible for differential
386 targeting (Lee et al., 2019; Lee et al., 2020). A different research effort has focused on
387 the presence of Hsp70 binding sites within cTP N-termini as crucial for enabling import
388 (Chotewutmontri et al., 2012; Chotewutmontri and Bruce, 2015; Chotewutmontri et al.,
389 2017). Here, we find that multiple Arginines are not uncommon in Chlamydomonas
390 cTP N-termini (Fig. S23) and that Hsp70 binding sites are equally present in mTP N-
391 termini (Fig. S26). We found that HA-RAMPs with intrinsically unstructured N-termini
392 proved able to support chloroplast targeting, even in the absence of a cTP-N addition.
393 This finding makes it extremely unlikely that differential import relies on specific
394 peptide-receptor interactions mediated by co-evolved sequence motifs. Instead, we
395 argue that the feature that differentiates cTPs from mTPs is the presence of an
396 unstructured sequence upstream of the amphipathic helix in cTP N-termini. This view
397 re-interprets the tripartite structure of cTPs (von Heijne et al., 1989), which is conserved
398 all the way from glaucocystophyte algae to vascular plants (Köhler et al., 2015), by
399 describing cTPs as composed of a central amphipathic helix flanked by N- and C-
400 terminal unstructured sequence elements. This central helix has a weaker signal than
401 the helix in mTPs (von Heijne and Nishikawa, 1991) and may only form upon
402 membrane contact (Bruce, 2000; Garrido et al., 2020).

403 We also found that chloroplast targeting was further improved by unstructured C-
404 termini. That TP C-termini are important to enable import was previously noted, with
405 cTP and mTP C-termini thought to be functionally interchangeable (Lee et al., 2019).
406 Here, we find that the nature of the C-terminus does matter: the more unstructured cTP-
407 C enabled chloroplast targeting more often and more effectively than mTP-C, which
408 contains a long predicted amphipathic helix. The need to contain unstructured
409 sequences (von Heijne and Nishikawa, 1991) as well as the increased length of cTPs,
410 even extending beyond the cleavage site in many cases (Bionda et al., 2010; Caspari,
411 2022), could be mechanistically related to the import system (Fig. 7A): whereas
412 mitochondrial import makes use of the proton gradient to power uptake of positively
413 charged presequences (Martin et al., 1991; Garg and Gould, 2016), energized

414 chloroplast import requires the cTP to stretch across both TOC and TIC and contact the
415 translocation motor (Chotewutmontri et al., 2017; Richardson et al., 2018; Nakai,
416 2018). Structured sequence elements can thus impede import, including the helix at the
417 Venus N-terminus (Rekas et al., 2002). Consistent with this view, HA-RAMPs with
418 unstructured C-termini, such as SIM, generated chloroplast targeting in the absence of
419 an added TP-C.
420 Unstructured elements also likely provide protein-protein interaction motifs. For
421 instance, TOC-interaction has been attributed to semi-conserved 'FGLK' motifs (Pilon
422 et al., 1995; Lee et al., 2009; Chotewutmontri et al., 2012; Holbrook et al., 2016),
423 although the requirement for F appears to be relaxed in Chlamydomonas (Fig. S26E)
424 (Razzak et al., 2017). The presence of GLK sites with high predicted interactivity in
425 cTP-N and cTP-C should contribute to the high chloroplast targeting potential of
426 constructs equipped with both elements (Fig. S26).

427 ***Amphipathic helices contribute to differential targeting***

428 While N- and C-terminal elements aid specificity determination, particularly for
429 chloroplast import, their influence fails to explain why some HA-RAMPs target
430 preferentially the mitochondria or exclusively the chloroplast. This observation
431 suggests that targeting specificity also is determined by some sequence properties of
432 the amphipathic helical elements, which in the case of HA-RAMPs serve to mediate
433 insertion into specific target membranes. It is of note that individual mTPs and cTPs
434 also interact with membrane bilayers (von Heijne et al., 1989; Bruce, 1998). For
435 instance, the Rubisco small subunit cTP interacts with chloroplast-mimetic membranes
436 only in presence of the chloroplast-specific galactolipids (Pinnaduwage and Bruce,
437 1996). Taken together these observations suggest that amphipathic helices may interact
438 specifically with the membranes of the targeted organelle (Patron and Waller, 2007;
439 Lazzaro et al., 2020). A direct interaction with the membrane bilayer, before that with
440 proteins of the translocons, would provide a basic mechanism for a first step in
441 differential organelle targeting (Fig. 7a). It could explain how TPs can be functionally
442 specific whilst diverse in sequence.

443 ***A possible series of events for the evolution of targeting peptides***

444 A small subset of randomly chosen sequences is able to inefficiently deliver proteins
445 into extant mitochondria, relying on amphipathic helices (Baker and Schatz, 1987;
446 Lemire et al., 1989). Thus, the existing protein import machinery recognizes

447 amphipathic helical peptides. However, an origin of TPs from random sequences does
448 not explain how protein import into organelles would have spontaneously occurred in
449 absence of the extant translocons.
450 Here, our use of antimicrobial peptides that harbour an amphipathic helix to further
451 understand organelle targeting specificity, was inspired by the hypothesis that HA-
452 RAMPs may have given rise to TPs during endosymbiotic organellogenesis because of
453 an "import and destroy" mechanism from HA-RAMP-resistant bacteria (Wollman,
454 2016; Caspari and Lafontaine, 2021). In this context, combining the organelle-targeting
455 behaviour of all those constructs that convert antimicrobial into targeting peptides can
456 be translated into a temporal evolutionary scenario (Fig. 7B) that adds to previously
457 proposed models (Lemire et al., 1989; Garg and Gould, 2016; Lee and Hwang, 2021).
458 First, we found that several unaltered HA-RAMPs can deliver cargo, suggesting the
459 evolution of mTPs from HA-RAMPs would have been straightforward. However,
460 K→R was found to reduce toxicity and increase targeting, a dual effect that would have
461 produced a strong selection pressure favouring this exchange as an early step. Since
462 mitochondria appear to tolerate the presence of preproteins containing unaltered HA-
463 RAMPs, the addition of cleavage sites that allow presequences to be degraded
464 separately from the cargo protein (Kmiec et al., 2014) would have come as a second
465 step.
466 That chloroplast targeting requires additional discriminating elements is consistent with
467 cTPs evolving in a cell that already had mTPs. There are two possible scenarios for the
468 origin of cTPs. Firstly, cTPs might have evolved directly from HA-RAMPs that already
469 contained unstructured sequence elements (Fig. 7B), as seen for our cp-set HA-RAMPs.
470 Secondly, as has been suggested earlier (Lee and Hwang, 2021), cTPs might have co-
471 opted existing mTPs. According to the present scenario, these mTPs already contained
472 R and a cleavage site, yet would still have been recognized by cyanobacterial HA-
473 RAMP importers. In this case, the key innovation that generated cTPs may simply have
474 been the addition of unstructured, possibly TOC-interacting elements to mTPs, which
475 we showed here to be sufficient to retarget the mt-set HA-RAMPs equipped with TP
476 C-termini at least partially to the chloroplast. The cTP N-terminus likely originally
477 started with an F, given that in glaucophytes and rhodophytes (and many derived
478 secondary plastids), a conserved N-terminal F plays a role in chloroplast import (Patron
479 and Waller, 2007; Wunder et al., 2007; Köhler et al., 2015). This observation led to the
480 idea that cTPs may have originated from a re-use of a C-terminal F-based motif

481 involved in secretion via OMP85 beta-barrel proteins (Robert et al., 2006; Knopp et al.,
482 2020) in (cyano)bacteria, from which TOC75 evolved. To this end, periplasmic
483 POTRA domains responsible for substrate-recognition were proposed to have flipped
484 orientation and now point into the host cytoplasm (Bullmann et al., 2010; Sommer et
485 al., 2011). However, several studies of plant TOC75 have since consistently found
486 POTRA-domains to be in the inter-membrane space, not the cytoplasm (Chen et al.,
487 2016; Paila et al., 2016; Gross et al., 2020). Without this cytoplasmic receptor, there is
488 no mechanism for how an N-terminal F could have acted as an import-enabling proto-
489 cTP (Knopp et al., 2020).
490 Note that our proposed scenario makes no prediction about whether host/proto-
491 organelle interactions were mutualistic or antagonistic. Either way, antimicrobial
492 peptides are likely to have been part of the suite of tools used in these host/proto-
493 symbiont interactions. In particular, the role of antimicrobial peptides in mutualistic
494 symbioses (Mergaert, 2018) includes one of the best documented cases for defensive
495 antimicrobial peptide import into bacteria (Guefrachi et al., 2015).

496 ***Conclusion***

497 Investigating the steps required to generate TPs from HA-RAMPs has allowed us to
498 uncover a number of novel mechanistic insights. Notably, we discovered a role for
499 Arginines for targeting efficacy, and delineated the contributions of N- and C-terminal
500 elements in targeting specificity. Our work also suggests that whether due to common
501 descent or convergence, the similarities between TPs and HA-RAMPs point to TPs
502 interacting with membrane lipids as an early targeting step. As any evolutionary
503 scenario, our plausible pathway from HA-RAMPs to TPs must be considered as a
504 working hypothesis that will need to be assessed further by a series of bioinformatics
505 and laboratory-controlled evolutionary experiments. Further understanding of peptide-
506 lipid interactions and the phylogeny of import machinery components should shed new
507 light on the evolution and functioning of organelle TPs.

508 **Materials and Methods**

509 ***Construct generation***

510 Venus expression constructs were designed in SnapGene (v4.3.11) and generated by
511 integrating PCR-amplified (Q5 hot start high-fidelity, #M0515, New England Biolabs)

512 DNA fragments into plasmid pODC53 (Caspari, 2020) upstream of Venus using
513 Gibson assembly (NEBuilder HiFi DNA assembly, #E5520S, New England Biolabs).
514 Chlamydomonas TP sequences were amplified from genomic DNA extracted from
515 strain T222+ (CC-5101). Templates for codon optimized HA-RAMP, RP and R→K
516 modified TP sequences were obtained by gene synthesis (Eurofins Genomics). Correct
517 assembly was verified by sequencing (Eurofins Genomics). Linear transformation
518 cassettes were generated through restriction digest with *Eco*RV (New England
519 Biolabs).

520 ***Transformation and fluorescence screen conditions***

521 Constructs were transformed into wild-type strain T222+ (CC-5101) using the protocol
522 described previously (Onishi and Pringle, 2016), except using 4µl of DNA at 2µg/µl.
523 Transformants (≥24 per construct) selected for Paromomycin resistance were grown in
524 200µl TAP in 96-well plates under 50µmol photons m^{-2} s^{-2} for 3-5 days and then
525 screened for Venus expression in a fluorescence plate reader (CLARIOstar, BMG
526 Labtech) as described previously (Caspari, 2020).

527 ***Microscopy***

528 Cells were grown in 96-well plates as described before (Garrido et al., 2020). Strains
529 with suspected mitochondrial targeting were subjected to 0.1 µM MitoTracker Red
530 CMXRos (ThermoFisher) in growth medium for 30 min in the dark and washed with
531 TAP prior to taking images. Epifluorescence microscopy was performed with cells
532 added into 200µl TAP medium, or SEM (250 mM Sucrose, 1 mM EDTA, in 10 mM
533 MOPS-KOH pH 7.2) to stop swimming, in poly-l-lysine (Sigma Aldrich) coated 8-well
534 µ-slides (Ibidi), using the following setup: microscope – Axio Observer.Z1 inverted
535 microscope (Zeiss); objective – α Plan-Apochromat 100x/1.46 Oil DIC M27 (Zeiss);
536 oil – immersol 518 F (Zeiss); camera – ORCA-flash4.0 digital camera (Hamamatsu);
537 LEDs – 470 nm (chlorophyll), 505 nm (Venus) and white light (MitoTracker, filtered
538 to >535nm using Zeiss beam splitter 423052-0104-000) in Colibri.2 LED system
539 (Zeiss); filter cubes: filter 46HE YFP shift free (Zeiss, 520–550 nm emission) for
540 Venus, custom-made filter set (559/34 BrightLine HC, Beamsplitter T 585 LP, 607/36
541 BrightLine HC; AHF Analysentechnik AG) for MitoTracker, and filter set 50 (Zeiss,
542 665–715 nm emission) for chlorophyll. Images were adjusted in Fiji (Schindelin et al.,
543 2012) (ImageJ version 2.0.0) as described before (Garrido et al., 2020), and final figures
544 assembled in PowerPoint (Microsoft PowerPoint for Mac 2011, Version 14.6.3).

545 ***Automated Image Analysis***

546 A custom ImageJ Macro was written to enable automated image segmentation of
547 epifluorescence micrographs in Fiji. Briefly, in a given micrograph, fluorescence
548 intensities were normalized and individual cells were detected using marker-controlled
549 watershed from the MorphoLibJ library (Legland et al., 2016). For each cell, auto-
550 thresholding the chlorophyll channel using the Huang method was used to generate a
551 chlorophyll mask. Separately, the chlorophyll channel was subjected to a series of
552 morphological filters and gaussian blurring followed by auto-thresholding, to finally
553 detect round holes between 0.1-5µm within the ensuing binary image, to be saved as a
554 pyrenoid mask. In MitoTracker images, the MitoTracker channel was subjected to
555 morphological filtering followed by auto-thresholding with Otsu to generate a
556 mitochondrial mask. Venus channel intensities were recorded for each compartment.
557 For each cell, rotated images containing binary masks as extra channels were saved.
558 Data was compiled, analysed and plotted in R, v3.6.1 (https://www.r-project.org/),
559 using RStudio, 2022.07.2+576 (https://www.rstudio.com/).

560 ***Western Blots***

561 Cells were grown up in liquid culture (10ml TAP, 30µmol photons m^{-2} s^{-1}, 160RPM)
562 until late mid-log phase. 2ml Aliquots were resuspended in 30µl storage buffer (1×
563 Roche cOmplete™ Mini proteinase inhibitor cocktail, 10 mM NaF, 0.2 M DTT, 0.2 M
564 $NaCO_3$) and stored at -20°C. 20µl boiling buffer (1× Roche cOmplete™ Mini
565 proteinase inhibitor cocktail, 10mM NaF, 50 g/L SDS, 200 g/L sucrose) was added,
566 then aliquots were boiled (50 seconds). Cell debris was removed (tabletop centrifuge,
567 maximum speed, 15 minutes, 4°C). Chlorophyll content was estimated
568 spectrophotometrically: 1 µg chlorophyll $\mu l^{-1} = 0.11 \times (OD_{680} - OD_{770})$ for 5 µl diluted
569 in 1 ml water. Samples (10 µg chlorophyll, equal volumes) were run (overnight, room
570 temperature, 18A) on large gels (35×27cm; resolving gel: 12% Acrylamide, 0.32%
571 Bisacrylamide, 0.25 µl/ml Tetramethyl ethylenediamine, 250 mg/L Ammonium
572 Persulfate, 375 mM Tris/HCl pH 8.8; stacking gel: 5% Acrylamide, 0.133%
573 Bisacrylamide, 0.666 µl/ml Tetramethyl ethylenediamine, 666 mg/L Ammonium
574 Persulfate, 125 mM Tris/HCl pH 6.8). Proteins (<25kDa, >75kDa) were transferred
575 (0.1 µm nitrocellulose membranes, 1h, 0.8 A cm^{-2}) as follows: cathode – 5 filter papers
576 (FP, 3 mm, Whatman) soaked in transfer buffer 1 (40 mM Aminocaproic acid, 20%
577 Isopropanol, 25 mM Tris/HCl pH 9.4) – gel – membrane – 2 FPs soaked in transfer

578 buffer 2 (20% Isopropanol, 25 mM Tris/HCl pH 10.4) – 3 FPs soaked in transfer buffer
579 3 (20% Isopropanol, 300 mM Tris/HCl pH 10.4) – anode. Membranes were fixed
580 (Ponceau Red), incubated (1h, room temperature) in block (30 g/L skimmed milk
581 powder, 0.1% tween-20, 1x PBS: 140 mM NaCl, 30 mM KCl, 100 mM Na_2HPO_4, 15
582 mM KH_2PO_4), immunolabelled (overnight, 4°C) using α-FLAG 1° (Sigma Aldrich
583 F1804, diluted 1:10 000 in block), washed (0.1% tween-20, 1x PBS), treated (1h, room
584 temperature) with horseradish peroxidase-conjugated α-mouse 2° (diluted 1:10 000 in
585 block), washed, revealed (ECL; ChemiDoc, Biorad). Blots were processed using
586 ImageLab (Version 6.0.0 build 26, BioRad), final figures were assembled in
587 PowerPoint (Microsoft).

588 ***Antimicrobial activity assay***

589 Minimum inhibitory concentration assays were performed as described before (Garrido
590 et al., 2020).

591 ***Sequence Data Set***

592 TPs with experimentally-confirmed cleavage sites were recovered from proteomic
593 studies: *Chlamydomonas reinhardtii* cTP (Terashima 2011), mTP (Tardif 2012);
594 *Arabidopsis thaliana* cTP (Ge 2008, Rowland 2015), mTP (Huang 2009). For each
595 peptide, we recovered the full-length protein sequence from NCBI and UNIPROT.
596 Cytoplasmic control sequences were generated by taking N-terminal sequence stretches
597 of random length, matching the distribution of peptide lengths observed in our
598 Chlamydomonas TP dataset, from a random subset of Chlamydomonas proteins with a
599 validated cytoplasmic location in UNIPROT. For PCAs and calculating amino acid
600 frequencies, the same HA-RAMP, signal peptide and non-Chlamydomonas TP
601 sequences were used as before (Garrido et al., 2020).

602 ***Amphipathic helix prediction***

603 Amphipatic α-helices are predicted as described previously (Garrido et al., 2020),
604 following the principle of the Heliquest algorithm (Gautier et al., 2008). Briefly, the
605 approach aims to establish the longest sequence stretch that contains identifiable
606 hydrophilic and hydrophobic faces. The algorithm is iterated such that multiple non-
607 overlapping helices can be found within a given peptide (Fig. 1). Consequently, helix
608 fractions are calculated as number of residues within all predicted helices divided by

609 the total number of residues in the peptide; for evaluating the number of upstream
610 residues, only the longest helix was considered (Fig. 6, Fig. S25).

611 ***ACC terms***

612 In order to evaluate the physicochemical properties of our peptides we used the
613 approach we previously described (Garrido et al., 2020). Briefly, each amino acid is
614 described in terms of 3 'Z-scale' values (Hellberg et al., 1987) that can be interpreted
615 as representing a residue's hydrophobic, steric and electronic properties. Auto-cross-
616 covariances (ACC) between nearby residues are calculated up to a distance of four
617 amino acids, generating a quantitative representation of a given peptide in terms of 36
618 ACC terms. Euclidian distances between HA-RAMP ACC term vectors and the
619 barycentre of Chlamydomonas TPs were used as a measure of similarity (Table S1).

620 ***Visualization***

621 We performed principal component analysis (PCA) to visualize the relationships
622 between peptides as described by their 36 ACC terms (Fig. S1, S22) or by their 5 salient
623 TP properties (Fig. S25) using python package sklearn v0.22.1 (Pedregosa et al., 2011).

624 ***Analysis of TP N-termini***

625 TP N-termini, defined as the N-terminal 15 amino acids, were analysed (Fig. S23) as
626 follows: Charge profiles were generated as described in the literature (Chotewutmontri
627 et al., 2012). Hydrophobicity of TP N-termini was estimated using the Heliquest
628 standalone application (Gautier et al., 2008). To evaluate disorder, we used IUpred2A,
629 a software that calculates the probability for each residue of being in a disordered region
630 (Erdős and Dosztányi, 2020). We used the default 'Long' setting, which has been
631 reported to be more accurate than the alternative 'Short' setting (Nielsen and Mulder,
632 2019). The disorder of a given sequence was taken as the mean of the probability values
633 for each residue (average residue disorder probability).

634 ***Statistical prediction***

635 To evaluate the predictive power of ACC terms (Garrido et al., 2020) obtained for 15
636 residue-long TP N-termini with regards to localization (Fig. S23) we used a binomial
637 logistic regression classifier. We performed one hundred 5-fold cross validation runs.
638 In each set of one hundred runs, we randomly selected sequences such that the same
639 number of mTPs and cTPs were used. For C. reinhardtii we used 33 mTP and 33 cTP
640 and for A. thaliana we use 29 mTP and 29 cTP sequences. We used an elastic net

641 penalty in a saga solver with an l1-ratio (a-parameter) of 0, which is equivalent to using
642 only ridge penalty, where all features take part in the model, and a C parameter (1/λ) of
643 0.1. The a and 1/λ parameters were optimised with 10-fold cross validation. Firstly,
644 when 1/λ = 1, the best accuracy (0.82) was obtained with a between 0 and 0.09.
645 Secondly, with a = 0, the best accuracy (0.82) was obtained with 1/λ = 0.1. A Logistic
646 Regression model with elastic net penalty ratio of 0.15 (scikit-learn Python package
647 v0.22.1) trained on class I HA-RAMPs (Garrido et al., 2020) and Chlamydomonas TPs
648 was used to evaluate how similar potential HA-RAMP candidates are to TPs (Table
649 S1). Custom scripts were written in python (v3.7.6).

650 ***Interaction site prediction***

651 Values for the Boman index, a quantitative proxy for whether a peptide is more likely
652 to interact with proteins (high values) or lipids (low values), were calculated as
653 described in the literature (Boman, 2003). Anchor2 interactivity values, a second proxy
654 for protein interaction potential developed for disordered sequences (Mészáros et al.,
655 2009), were calculated using the IUPred2A standalone application and webserver
656 (Mészáros et al., 2018; Erdős and Dosztányi, 2020). Interaction sites for Hsp70 were
657 predicted as described by Ivey and co-workers (Ivey et al., 2000) based on experimental
658 affinity values for individual amino acids. Putatively TOC-interacting 'FGLK' sites
659 were established by searching for presence of F and [P or G] and [K or R] and [A or L
660 or V], and absence of [D and E], within each 8-residue window of a sequence,
661 corresponding to rule 22 by Chotewutmontri and co-workers (Chotewutmontri et al.,
662 2012) that was recommended by the authors in personal communication. 'FGLK-1'
663 sites were established the same way, but requiring the presence of only three of the four
664 positive determinants. Custom scripts were implemented in R using RStudio.

665 ***Statistical analysis***

666 Chlamydomonas TP distributions (Fig. 6, Fig. S26) were compared using two-sided t-
667 tests (n=34 for mTPs and n=85 for cTPs), and associated mt-set and cp-set HA-RAMPs
668 were compared using one-sided t-tests based on the trends set by TPs (n=3 for mt-set
669 and n=5 for cp-set HA-RAMPs) in R using RStudio. Multiple Kruskal statistical tests
670 are performed (same Chlamydomonas mTPs and cTPs as above, plus n=382 HA-
671 RAMPs) to evaluate the distribution of amino acids (Fig. 3c, Fig. S21) in the different
672 groups, followed by Dunn post-hoc tests (scipy v1.4.1).

673 **Acknowledgements**
674 We are very grateful to Florian Schober and Katherine Madden for their effort and
675 enthusiasm in helping us generate a quantitative image quantification dataset. Thank
676 you to Tiffina Benhamou and Gabriel Chemin, who worked on aspects of this study
677 with ODC during their summer internships. We also thank Prakitchai Chotewutmontri
678 and Barry Bruce for their help in correctly replicating their Hsp70-binding and FGLK-
679 motif finding algorithms.
680
681 The following financial support is gratefully acknowledged: the Centre National de la
682 Recherche Scientifique and Sorbonne University for annual funding to UMR7141; the
683 Agence National de la Recherche for (a) "ChloroMitoRAMP" ANR grant (ANR-19-
684 CE13-0009) and (b) "LabEx Dynamo" (ANR-LABX-011), both of which provided
685 postdoctoral support to ODC, and (c) "MATHTEST" grant (ANR-18-CE13-0027)
686 which provided doctoral support to CG; finally, the Fondation Edmond Rothschild
687 which provided complementary financial support to ODC and CG. The funders had no
688 role in the design of the study; in the collection, analyses, or interpretation of data; in
689 the writing of the manuscript, or in the decision to publish the results.

690 **Author contributions**
691 FAW, IL and YC conceptualized the project and acquired funds. ODC devised,
692 conducted and visualized the wet-lab experimental part of the investigation, and was
693 chiefly responsible for assigning subcellular localization interpretations based on
694 microscopic evidence. CG carried out the dry-lab/bioinformatic part of the
695 investigation and associated visualization under the supervision of IL, writing custom
696 software to do so (ODC contributed to the analysis of FGLK and Hsp70-binding sites).
697 COL developed the custom imageJ macros used for automated image segmentation.
698 ODC wrote the original draft of the manuscript. FAW, IL and ODC reviewed and edited
699 the manuscript. All authors approved the final manuscript. The authors declare no
700 competing interests.

701 **Data availability**
702 All data are available in the main text or the supplementary materials. Custom code
703 generated in the course of this project will be made available without restrictions upon

704 request to the authors. All plasmids, and one independent insertion line per construct
705 (except RP constructs), are available through the Chlamydomonas resource centre
706 (https://www.chlamycollection.org/).

707 **References**

708 **Armenteros, J. J. A., Salvatore, M., Emanuelsson, O., Winther, O., Von Heijne,**
709 **G., Elofsson, A., and Nielsen, H.** (2019). Detecting sequence signals in targeting
710 peptides using deep learning. *Life Sci Alliance* **2**:1–14
711 https://doi.org/10.26508/lsa.201900429.

712 **Baker, A., and Schatz, G.** (1987). Sequences from a prokaryotic genome or the mouse
713 dihydrofolate reductase gene can restore the import of a truncated precursor
714 protein into yeast mitochondria. *Proceedings of the National Academy of Sciences*
715 **84**:3117–3121 https://doi.org/10.1073/pnas.84.10.3117.

716 **Bhushan, S., Kuhn, C., Berglund, A., Roth, C., and Glaser, E.** (2006). The role of
717 the N-terminal domain of chloroplast targeting peptides in organellar protein
718 import and miss-sorting. *FEBS* **580**:3966–3972
719 https://doi.org/10.1016/j.febslet.2006.06.018.

720 **Bionda, T., Tillmann, B., Simm, S., Beilstein, K., Ruprecht, M., and Schleiff, E.**
721 (2010). Chloroplast import signals: The length requirement for translocation in
722 vitro and in vivo. *J Mol Biol* **402**:510–523
723 https://doi.org/10.1016/j.jmb.2010.07.052.

724 **Boman, H. G.** (2003). Antibacterial peptides: basic facts and emerging concepts. *J*
725 *Intern Med* **254**:197–215 https://doi.org/doi:10.1046/j.1365-2796.2003.01228.x.

726 **Bruce, B. D.** (1998). The role of lipids in plastid protein transport. *Plant Mol Biol*
727 **38**:223–46 https://doi.org/10.1023/A:1006094308805.

728 **Bruce, B. D.** (2000). Chloroplast transit peptides: Structure, function and evolution.
729 *Trends Cell Biol* **10**:440–447 https://doi.org/10.1016/S0962-8924(00)01833-X.

730 **Bruce, B. D.** (2001). The paradox of plastid transit peptides: Conservation of function
731 despite divergence in primary structure. *Biochim Biophys Acta Mol Cell Res*
732 **1541**:2–21 https://doi.org/10.1016/S0167-4889(01)00149-5.

733 **Bullmann, L., Haarmann, R., Mirus, O., Bredemeier, R., Hempel, F., Maier, U.**
734 **G., and Schleiff, E.** (2010). Filling the gap, evolutionarily conserved Omp85 in

735 plastids of chromalveolates. *Journal of Biological Chemistry* **285**:6848–6856
736 https://doi.org/10.1074/jbc.M109.074807.
737 **Calvo, S. E., Julien, O., Clauser, K. R., Shen, H., Kamer, K. J., Wells, J. A., and**
738 **Mootha, V. K.** (2017). Comparative analysis of mitochondrial N-termini from
739 mouse, human, and yeast. *Molecular and Cellular Proteomics* **16**:512–523
740 https://doi.org/10.1074/mcp.M116.063818.
741 **Caspari, O. D.** (2020). Introduction of a leaky stop codon as molecular tool in
742 Chlamydomonas reinhardtii. *PLoS One* **15**:e0237405
743 https://doi.org/10.1371/journal.pone.0237405.
744 **Caspari, O. D.** (2022). Chloroplast transit peptides often require downstream
745 unstructured sequence in Chlamydomonas reinhardtii. *Front Plant Sci* **13**:825797
746 https://doi.org/10.3389/fpls.2022.825797.
747 **Caspari, O. D., and Lafontaine, I.** (2021). The role of antimicrobial peptides in the
748 evolution of endosymbiotic protein import. *PLoS Pathog* **17**:e1009466
749 https://doi.org/10.1371/journal.ppat.1009466.
750 **Caspari, O. D., Meyer, M. T., Tolleter, D., Wittkopp, T. M., Cunniffe, N. J.,**
751 **Lawson, T., Grossman, A. R., and Griffiths, H.** (2017). Pyrenoid loss in
752 *Chlamydomonas reinhardtii* causes limitations in CO2 supply, but not thylakoid
753 operating efficiency. *J Exp Bot* **68**:3903–3913 https://doi.org/10.1093/jxb/erx197.
754 **Chen, Y. L., Chen, L. J., and Li, H. M.** (2016). Polypeptide transport-associated
755 domains of the Toc75 channel protein are located in the intermembrane space of
756 chloroplasts. *Plant Physiol* **172**:235–243 https://doi.org/10.1104/pp.16.00919.
757 **Chotewutmontri, P., and Bruce, B. D.** (2015). Non-native, N-terminal Hsp70
758 molecular motor recognition elements in transit peptides support plastid protein
759 translocation. *Journal of Biological Chemistry* **290**:7602–7621
760 https://doi.org/10.1074/jbc.M114.633586.
761 **Chotewutmontri, P., Reddick, L. E., McWilliams, D. R., Campbell, I. M., and**
762 **Bruce, B. D.** (2012). Differential Transit Peptide Recognition during Preprotein
763 Binding and Translocation into Flowering Plant Plastids. *Plant Cell* **24**:3040–3059
764 https://doi.org/10.1105/tpc.112.098327.
765 **Chotewutmontri, P., Holbrook, K., and Bruce, B. D.** (2017). Plastid Protein
766 Targeting: Preprotein Recognition and Translocation. *Int Rev Cell Mol Biol*
767 **330**:227–294 https://doi.org/10.1016/bs.ircmb.2016.09.006.

768 **Erdős, G., and Dosztányi, Z.** (2020). Analyzing Protein Disorder with IUPred2A.
769 *Curr Protoc Bioinformatics* **70**:1–15 https://doi.org/10.1002/cpbi.99.
770 **Franzén, L.-G. G., Rochaix, J.-D. D., and von Heijne, G.** (1990). Chloroplast transit
771 peptides from the green alga Chlamydomonas reinhardtii share features with both
772 mitochondrial and higher plant chloroplast presequences. *FEBS Lett* **260**:169–172
773 https://doi.org/10.1016/0014-5793(90)80094-Y.
774 **Garg, S. G., and Gould, S. B.** (2016). The Role of Charge in Protein Targeting
775 Evolution. *Trends Cell Biol* **26**:894–905
776 https://doi.org/10.1016/j.tcb.2016.07.001.
777 **Garrido, C., Caspari, O. D., Choquet, Y., Wollman, F. A., and Lafontaine, I.**
778 (2020). Evidence Supporting an Antimicrobial Origin of Targeting Peptides to
779 Endosymbiotic Organelles. *Cells* **9**:1795 https://doi.org/10.3390/cells9081795.
780 **Gautier, R., Douguet, D., Antonny, B., and Drin, G.** (2008). HELIQUEST: a web
781 server to screen sequences with specific alpha-helical properties. *Bioinformatics*
782 **24**:2101–2102 https://doi.org/10.1093/bioinformatics/btn392.
783 **Gross, L. E., Spies, N., Simm, S., and Schleiff, E.** (2020). Toc75-V/OEP80 is
784 processed during translocation into chloroplasts, and the membrane-embedded
785 form exposes its POTRA domain to the intermembrane space. *FEBS Open Bio*
786 **10**:444–454 https://doi.org/10.1002/2211-5463.12791.
787 **Guefrachi, I., Pierre, O., Timchenko, T., Alunni, B., Barrière, Q., Czernic, P.,**
788 **Villaécija-Aguilar, J.-A., Verly, C., Bourge, M., Fardoux, J., et al.** (2015).
789 Bradyrhizobium BclA Is a Peptide Transporter Required for Bacterial
790 Differentiation in Symbiosis with Aeschynomene Legumes. *Molecular Plant-*
791 *Microbe Interactions* **28**:1155–1166 https://doi.org/10.1094/MPMI-04-15-0094-
792 R.
793 **Hellberg, S., Sjostrom, M., Skagerberg, B., and Wold, S.** (1987). Peptide
794 Quantitative Structure-Activity Relationships, a Multivariate Approach. *J Med*
795 *Chem* **30**:1126–1135 https://doi.org/10.1021/jm00390a003.
796 **Holbrook, K., Subramanian, C., Chotewutmontri, P., Reddick, L. E., Wright, S.,**
797 **Zhang, H., Moncrief, L., and Bruce, B. D.** (2016). Functional Analysis of Semi-
798 conserved Transit Peptide Motifs and Mechanistic Implications in Precursor
799 Targeting and Recognition. *Mol Plant* **9**:1286–1301
800 https://doi.org/10.1016/j.molp.2016.06.004.

801 **Ivey, R. a, and Bruce, B. D.** (2000). In vivo and in vitro interaction of DnaK and a
802 chloroplast transit peptide. *Cell Stress Chaperones* **5**:62–71.
803 **Ivey, R. a, Subramanian, C., and Bruce, B. D.** (2000). Identification of a Hsp70
804 recognition domain within the rubisco small subunit transit peptide. *Plant Physiol*
805 **122**:1289–1299 https://doi.org/10.1104/pp.122.4.1289.
806 **Joo, H. S., Fu, C. I., and Otto, M.** (2016). Bacterial strategies of resistance to
807 antimicrobial peptides. *Philosophical Transactions of the Royal Society B:*
808 *Biological Sciences* **371** https://doi.org/10.1098/rstb.2015.0292.
809 **Kmiec, B., Teixeira, P. F., and Glaser, E.** (2014). Shredding the signal: targeting
810 peptide degradation in mitochondria and chloroplasts. *Trends Plant Sci* **19**:771–8
811 https://doi.org/10.1016/j.tplants.2014.09.004.
812 **Knopp, M., Garg, S. G., Handrich, M., and Gould, S. B.** (2020). Major Changes in
813 Plastid Protein Import and the Origin of the Chloroplastida. *iScience* **23**:100896
814 https://doi.org/10.1016/j.isci.2020.100896.
815 **Köhler, D., Dobritzsch, D., Hoehenwarter, W., Helm, S., Steiner, J. M., and**
816 **Baginsky, S.** (2015). Identification of protein N-termini in Cyanophora paradoxa
817 cyanelles: transit peptide composition and sequence determinants for precursor
818 maturation. *Front Plant Sci* **6**:1–11 https://doi.org/10.3389/fpls.2015.00559.
819 **Krimm, I., Gans, P., Hernandez, J. F., Arlaud, G. J., and Lancelin, J. M.** (1999).
820 A coil-helix instead of a helix-coil motif can be induced in a chloroplast transit
821 peptide from Chlamydomonas reinhardtii. *Eur J Biochem* **265**:171–180
822 https://doi.org/10.1046/j.1432-1327.1999.00701.x.
823 **Lancelin, J. M., Gans, P., Bouchayer, E., Bally, I., Arlaud, G. J., and Jacquot, J.**
824 **P.** (1996). NMR structures of a mitochondrial transit peptide from the green alga
825 Chlamydomonas reinhardtii. *FEBS Lett* **391**:203–208
826 https://doi.org/10.1016/0014-5793(96)00734-X.
827 **Lazzaro, B. P., Zasloff, M., and Rolff, J.** (2020). Antimicrobial peptides: Application
828 informed by evolution. *Science (1979)* **368**
829 https://doi.org/10.1126/science.aau5480.
830 **Lee, D. W., and Hwang, I.** (2021). Understanding the evolution of endosymbiotic
831 organelles based on the targeting sequences of organellar proteins. *New*
832 *Phytologist* **230**:924–930 https://doi.org/10.1111/nph.17167.
833 **Lee, D. W., Lee, S., Oh, Y. J., and Hwang, I.** (2009). Multiple sequence motifs in the
834 rubisco small subunit transit peptide independently contribute to Toc159-

835 dependent import of proteins into chloroplasts. *Plant Physiol* **151**:129–141
836 https://doi.org/10.1104/pp.109.140673.
837 **Lee, D. W., Lee, S., Lee, J., Woo, S., Razzak, M. A., Vitale, A., and Hwang, I.**
838 (2019). Molecular Mechanism of the Specificity of Protein Import into
839 Chloroplasts and Mitochondria in Plant Cells. *Mol Plant* **12**:951–966
840 https://doi.org/10.1016/j.molp.2019.03.003.
841 **Lee, D. W., Lee, S., Min, C.-K., Park, C., Kim, J.-M., Hwang, C.-S., Park, S. K.,**
842 **Cho, N.-H., and Hwang, I.** (2020). Cross-Species Functional Conservation and
843 Possible Origin of the N-Terminal Specificity Domain of Mitochondrial
844 Presequences. *Front Plant Sci* **11**:1–7 https://doi.org/10.3389/fpls.2020.00064.
845 **Legland, D., Arganda-Carreras, I., and Andrey, P.** (2016). MorphoLibJ: Integrated
846 library and plugins for mathematical morphology with ImageJ. *Bioinformatics*
847 **32**:3532–3534 https://doi.org/10.1093/BIOINFORMATICS/BTW413.
848 **Lemire, B. D., Fankhauser, C., Baker, A., and Schatz, G.** (1989). The mitochondrial
849 targeting function of randomly generated peptide sequences correlates with
850 predicted helical amphiphilicity. *Journal of Biological Chemistry* **264**:20206–
851 20215 https://doi.org/10.1016/S0021-9258(19)47048-8.
852 **Li, L., Vorobyov, I., and Allen, T. W.** (2013). The different interactions of lysine and
853 arginine side chains with lipid membranes. *Journal of Physical Chemistry B*
854 **117**:11906–11920 https://doi.org/10.1021/jp405418y.
855 **Li, J., Koh, J.-J., Liu, S., Lakshminarayanan, R., Verma, C. S., and Beuerman, R.**
856 **W.** (2017). Membrane Active Antimicrobial Peptides: Translating Mechanistic
857 Insights to Design. *Front Neurosci* **11**:1–18
858 https://doi.org/10.3389/fnins.2017.00073.
859 **Mackinder, L. C. M., Meyer, M. T., Mettler-Altmann, T., Chen, V. K., Mitchell,**
860 **M. C., Caspari, O. D., Freeman Rosenzweig, E. S., Pallesen, L., Reeves, G.,**
861 **Itakura, A., et al.** (2016). A repeat protein links Rubisco to form the eukaryotic
862 carbon concentrating organelle. *Proceedings of the National Academy of Sciences*
863 **113**:5958–5963 https://doi.org/10.1073/pnas.1522866113.
864 **Mackinder, L. C. M., Chen, C., Leib, R. D., Patena, W., Blum, S. R., Rodman, M.,**
865 **Ramundo, S., Adams, C. M., and Jonikas, M. C.** (2017). A Spatial Interactome
866 Reveals the Protein Organization of the Algal CO2-Concentrating Mechanism.
867 *Cell* **171**:133–147 https://doi.org/10.1016/j.cell.2017.08.044.

868 **Martin, J., Mahlke, K., and Pfanner, N.** (1991). Role of an energized inner membrane
869 in mitochondrial protein import: ΔΨ drives the movement of presequences.
870 *Journal of Biological Chemistry* **266**:18051–18057 https://doi.org/10.1016/s0021-
871 9258(18)55235-2.
872 **Mattiroli, F., and Sixma, T. K.** (2014). Lysine-targeting specificity in ubiquitin and
873 ubiquitin-like modification pathways. *Nat Struct Mol Biol* **21**:308–316
874 https://doi.org/10.1038/nsmb.2792.
875 **Maupin-Furlow, J. A.** (2013). Ubiquitin-like proteins and their roles in archaea.
876 *Trends Microbiol* **21**:31–38 https://doi.org/10.1016/j.tim.2012.09.006.
877 **Maupin-Furlow, J. A., Humbard, M. A., Kirkland, P. A., Li, W., Reuter, C. J.,**
878 **Wright, A. J., and Zhou, G.** (2006). Proteasomes from Structure to Function:
879 Perspectives from Archaea. *Curr Top Dev Biol* **75**:125–169
880 https://doi.org/10.1016/S0070-2153(06)75005-0.
881 **Mergaert, P.** (2018). Role of antimicrobial peptides in controlling symbiotic bacterial
882 populations. *Nat Prod Rep* **35**:336–356 https://doi.org/10.1039/c7np00056a.
883 **Mészáros, B., Simon, I., and Dosztányi, Z.** (2009). Prediction of Protein Binding
884 Regions in Disordered Proteins. *PLoS One* **5**:e1000376
885 https://doi.org/10.1371/Citation.
886 **Mészáros, B., Erdös, G., and Dosztányi, Z.** (2018). IUPred2A: Context-dependent
887 prediction of protein disorder as a function of redox state and protein binding.
888 *Nucleic Acids Res* **46**:W329–W337 https://doi.org/10.1093/nar/gky384.
889 **Meyer, M., Itakura, A., Patena, W., Wang, L., He, S., Emrich-Mills, T., Lau, C.**
890 **S., Yates, G., Mackinder, L. C. M., and Jonikas, M. C.** (2020). Assembly of the
891 algal CO2-fixing organelle, the pyrenoid, is guided by a Rubisco-binding motif.
892 *Sci Adv* **6**:eabd2408 https://doi.org/10.1126/sciadv.abd2408.
893 **Nakai, M.** (2018). New Perspectives on Chloroplast Protein Import. *Plant Cell Physiol*
894 **59**:1111–1119 https://doi.org/10.1093/pcp/pcy083.
895 **Nielsen, J. T., and Mulder, F. A. A.** (2019). Quality and bias of protein disorder
896 predictors. *Sci Rep* **9**:1–11 https://doi.org/10.1038/s41598-019-41644-w.
897 **Onishi, M., and Pringle, J. R.** (2016). Robust Transgene Expression from Bicistronic
898 mRNA in the Green Alga *Chlamydomonas reinhardtii*. *G3 (Bethesda)* **6**:4115–
899 4125 https://doi.org/10.1534/g3.116.033035.
900 **Paila, Y. D., Richardson, L. G. L., Inoue, H., Parks, E. S., McMahon, J., Inoue, K.,**
901 **and Schnell, D. J.** (2016). Multi-functional roles for the polypeptide transport

associated domains of Toc75 in chloroplast protein import. *Elife* **5**:1–29 https://doi.org/10.7554/eLife.12631.

Patron, N. J., and Waller, R. F. (2007). Transit peptide diversity and divergence : a global analysis of plastid targeting signals. *BioEssays* **29**:1048–1058 https://doi.org/10.1002/bies.20638.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**:2825–2830 https://doi.org/10.1289/EHP4713.

Pilon, M., Wienk, H., Sips, W., De Swaaf, M., Talboom, I., Van't Hof, R., De Korte- Kool, G., Demel, R., Weisbeek, P., and De Kruijff, B. (1995). Functional domains of the ferredoxin transit sequence involved in chloroplast import. *Journal of Biological Chemistry* **270**:3882–3893 https://doi.org/10.1074/jbc.270.8.3882.

Pinnaduwage, P., and Bruce, B. D. (1996). In vitro interaction between a chloroplast transit peptide and chloroplast outer envelope lipids is sequence-specific and lipid class-dependent. *Journal of Biological Chemistry* **271**:32907–32915 https://doi.org/10.1074/jbc.271.51.32907.

Razzak, A., Lee, D. W., Yoo, Y., and Hwang, I. (2017). Evolution of rubisco complex small subunit transit peptides from algae to plants. *Sci Rep* **7**:9279 https://doi.org/10.1038/s41598-017-09473-x.

Rekas, A., Alattia, J. R., Nagai, T., Miyawaki, A., and Ikura, M. (2002). Crystal structure of venus, a yellow fluorescent protein with improved maturation and reduced environmental sensitivity. *Journal of Biological Chemistry* **277**:50573–50578 https://doi.org/10.1074/jbc.M209524200.

Richardson, L. G. L., Small, E. L., Inoue, H., and Schnell, D. J. (2018). Molecular topology of the transit peptide during chloroplast protein import. *Plant Cell* **30**:1789–1806 https://doi.org/10.1105/tpc.18.00172.

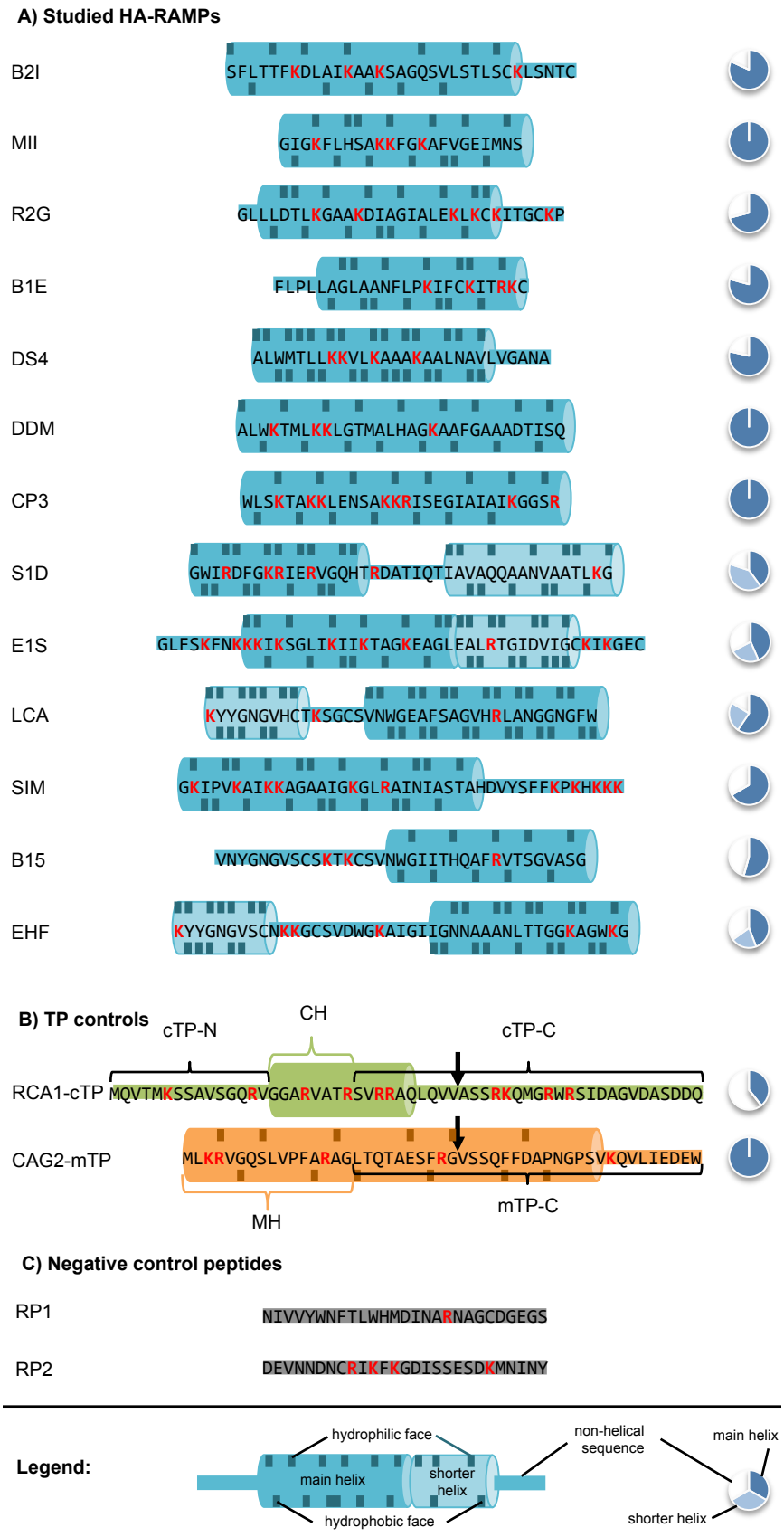
Robert, V., Volokhina, E. B., Senf, F., Bos, M. P., Van Gelder, P., and Tommassen, J. (2006). Assembly factor Omp85 recognizes its outer membrane protein substrates by a species-specific C-terminal motif. *PLoS Biol* **4**:1984–1995 https://doi.org/10.1371/journal.pbio.0040377.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an

935 open-source platform for biological-image analysis. *Nat Methods* **9**:676–682
936 https://doi.org/10.1038/nmeth.2019.
937 **Sommer, M. S., Daum, B., Gross, L. E., Weis, B. L. M., Mirus, O., Abram, L.,**
938 **Maier, U. G., Kühlbrandt, W., and Schleiff, E.** (2011). Chloroplast Omp85
939 proteins change orientation during evolution. *Proc Natl Acad Sci U S A*
940 **108**:13841–13846 https://doi.org/10.1073/pnas.1108626108.
941 **Tardif, M., Atteia, A., Specht, M., Cogne, G., Rolland, N., Brugière, S., Hippler,**
942 **M., Ferro, M., Bruley, C., Peltier, G., et al.** (2012). Predalgo: A new subcellular
943 localization prediction tool dedicated to green algae. *Mol Biol Evol* **29**:3625–3639
944 https://doi.org/10.1093/molbev/mss178.
945 **Teixeira, P. F., and Glaser, E.** (2013). Processing peptidases in mitochondria and
946 chloroplasts. *Biochim Biophys Acta Mol Cell Res* **1833**:360–370
947 https://doi.org/10.1016/j.bbamcr.2012.03.012.
948 **von Heijne, G.** (1986). Mitochondrial targeting sequences may form amphiphilic
949 helices. *EMBO J* **5**:1335–1342 https://doi.org/10.1002/j.1460-
950 2075.1986.tb04364.x.
951 **von Heijne, G., and Nishikawa, K.** (1991). Chloroplast transit peptides. The perfect
952 random coil? *FEBS Lett* **278**:1–3 https://doi.org/10.1016/0014-5793(91)80069-f.
953 **von Heijne, G., Steppuhn, J., and Herrmann, R. G.** (1989). Domain structure of
954 mitochondrial and chloroplastic targeting peptides. *Eur. J. Biochem.* **180**:535–545
955 https://doi.org/10.1111/j.1432-1033.1989.tb14679.x.
956 **Wiedemann, N., and Pfanner, N.** (2017). Mitochondrial Machineries for Protein
957 Import and Assembly. *Annu Rev Biochem* **86**:685–714
958 https://doi.org/10.1146/annurev-biochem-060815-014352.
959 **Wienk, H. L. J., Wechselberger, R. W., Czisch, M., and de Kruijff, B.** (2000).
960 Structure, Dynamics, and Insertion of a Chloroplast Targeting Peptide in Mixed
961 Micelles. *Biochemistry* **39**:8219–8227 https://doi.org/10.1021/bi000110i.
962 **Wollman, F.-A.** (2016). An antimicrobial origin of transit peptides accounts for early
963 endosymbiotic events. *Traffic* **17**:1322–1328 https://doi.org/10.1111/tra.12446.
964 **Wunder, T., Martin, R., Löffelhardt, W., Schleiff, E., and Steiner, J. M.** (2007).
965 The invariant phenylalanine of precursor proteins discloses the importance of
966 Omp85 for protein translocation into cyanelles. *BMC Evol Biol* **7**:236
967 https://doi.org/10.1186/1471-2148-7-236.
968

969 **Figures and captions**

970



971

972 **Fig. 1. Peptide sequences under study.** Amino acid sequences are shown using the
973 one-letter code. Positively charged residues are highlighted in red. The fraction of the
974 sequence predicted to fold into amphipathic helices (see Methods) is provided by a pie-
975 chart to the right of the sequence; for TPs in (**B**) this was calculated up to the cleavage
976 site indicated by a downward arrow. Predicted amphipathic helices are highlighted
977 using a cylinder cartoon, with residues contributing to the hydrophilic/hydrophobic face
978 indicated on the top/bottom. No helix could be predicted within RCA1-cTP, and thus
979 the indicated helix is taken from a published NMR-structure obtained in membrane-
980 mimetic conditions (Krimm et al., 1999). Note that the two helices of E1S are at an
981 angle to each other and therefore cannot form a single continuous amphipathic helix.
982 Abbreviations: (**A**) B2I: Brevinin-2ISb, R2G: Ranatuerin-2G, MII: Magainin 2, B1E:
983 Brevinin-1E, DS4: Dermaseptin S4, DDM: Dermadistinctin-M, CP3: Cecropin-P3,
984 S1D: Sarcotoxin-1D, E1S: Esculentin-1SEA, LCA: Leucocin-A, SIM: SI Moricin,
985 B15: Bacillocin 1580, EHF: Enterocin HF; (**B**) TP – targeting peptide, cTP: chloroplast
986 TP, mTP: mitochondrial TP, RCA1: Rubisco activase, CAG2: γ-carbonic anhydrase,
987 cTP-N: cTP N-terminal element, CH: cTP helix, MH: mTP helix, cTP-C: cTP C-
988 terminal element, mTP-C: mTP C-terminal element; (**C**) RP: random peptide.
989

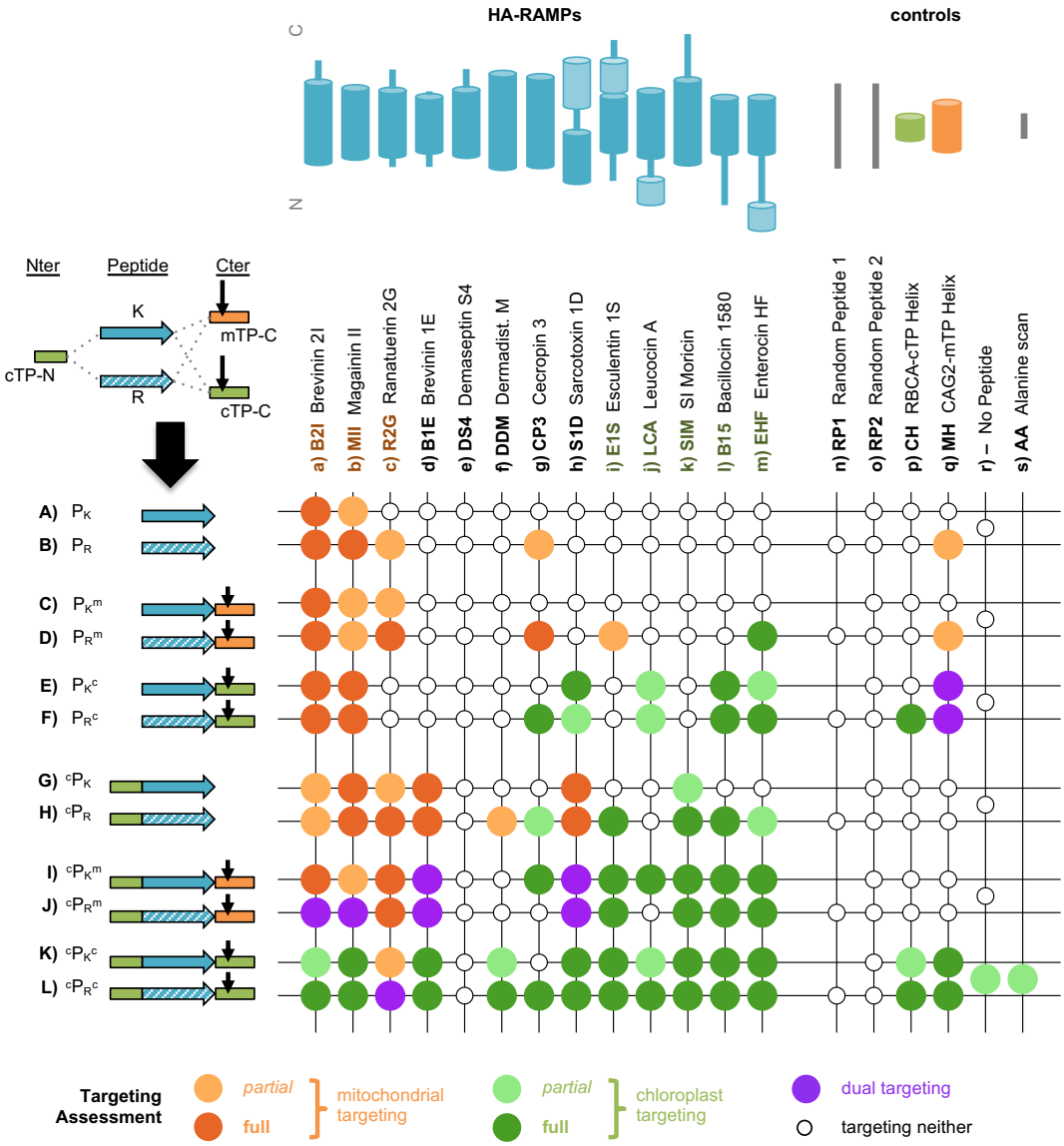


Fig. 2. TP modifications enable HA-RAMP targeting. Chimeric constructs were generated by combining TP elements with HA-RAMPs (see Fig. 1 for sequences; in the $^{c}AA^{c}$ construct in column s, all residues of the helical element "CH" within RCA1-cTP are replaced by Alanines). The overview graph shows in each column (**a-s**) one of the peptides with a cartoon indicating the position of the predicted amphipathic helices within the sequence; and in each row (**A-L**) a combination of peptide, K/R modification, addition of cTP-N (the 15 N-terminal residues upstream of the helix in

RCA1-cTP) and/or a C-terminal TP element (cTP-C or mTP-C, which include -10 residues upstream and +23 residues downstream of the cleavage site for RCA1-cTP and CAG2-mTP respectively) indicated by a cartoon and the following shorthand: P = peptide, $_K$ = contains mostly Lysines, $_R$ = contains mostly Arginines, ^{m} = mTP element, ^{c} = cTP element (^{c}P = cTP-N, P^c = cTP-C). In each case, an overview of observed targeting is provided by a colour code (see figure legend). Images for all constructs are shown in Fig. S3-S20. Note that the present results for B2I$_K^c$, MII$_K^c$, S1D$_K^c$, B15$_K^c$ and EHF$_K^c$ (row E, columns a,b,h,l,m) confirmed our previous report on these strains (Garrido et al., 2020).

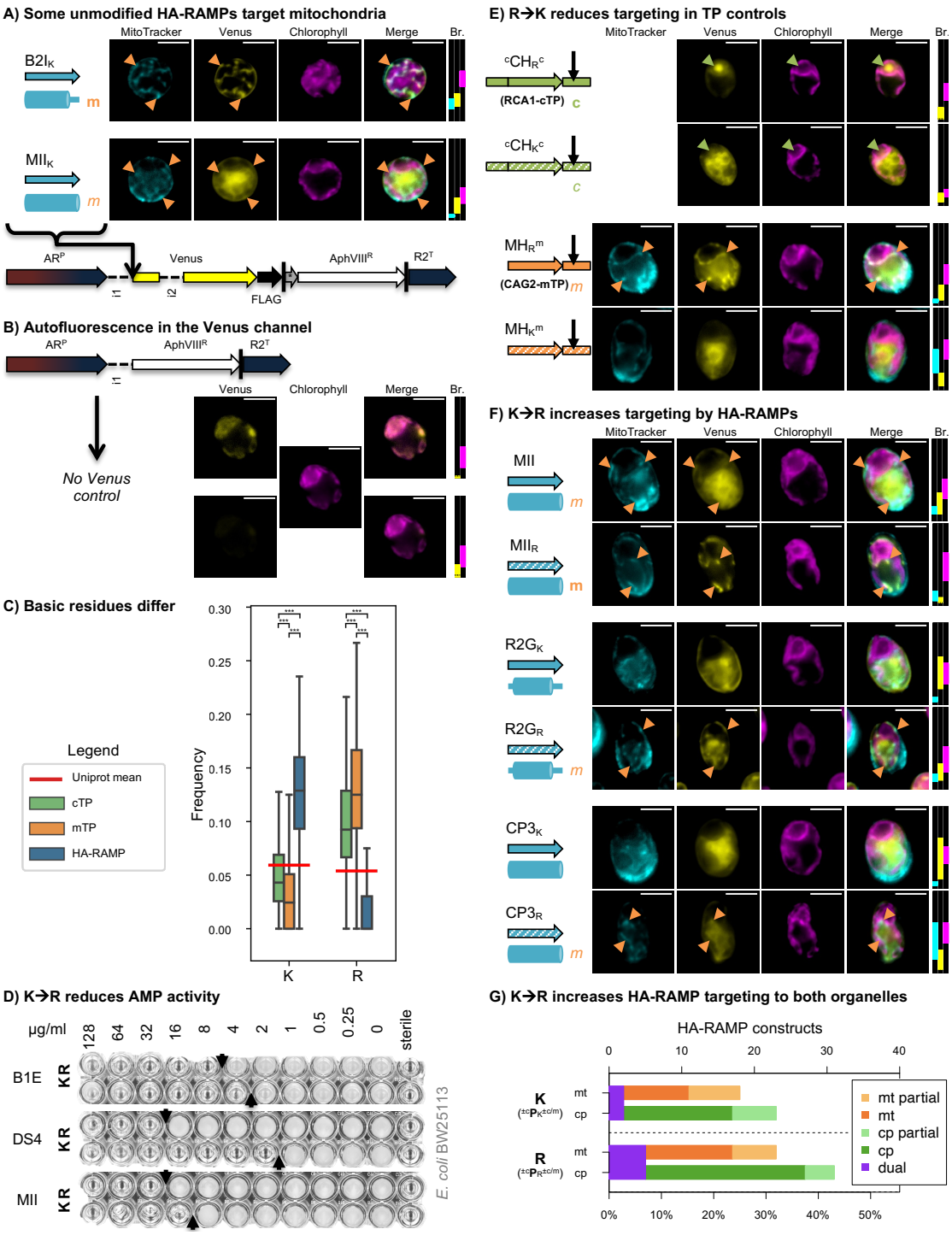
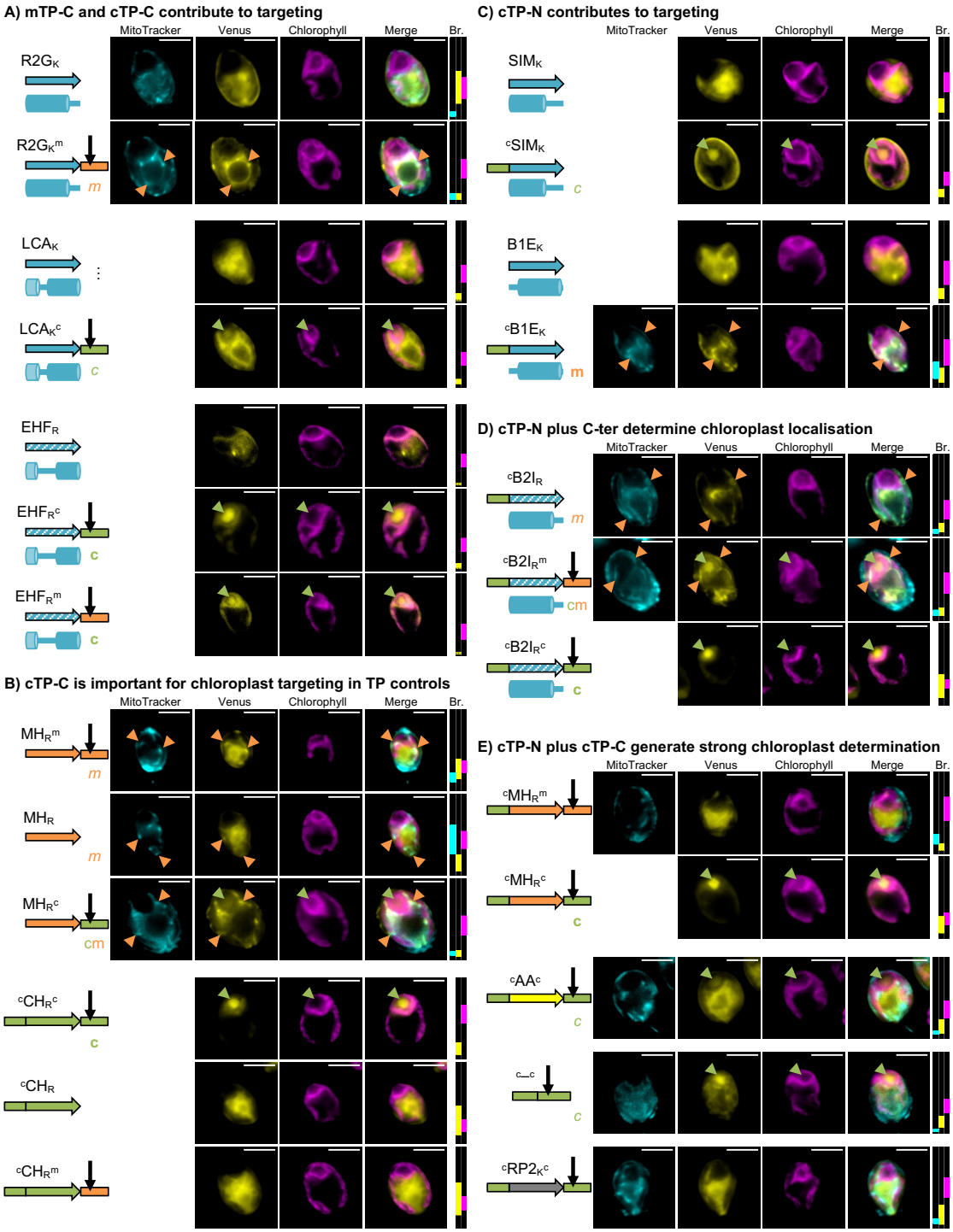


Fig. 3. K is for killing, R is for targeting. (**A**) Indicated peptides were inserted upstream of a Venus fluorescent protein reporter in a bicistronic expression system for Chlamydomonas (Caspari, 2020). AR^P: hybrid *HSP70a-RBCS2* promoter, i1: *RBCS2* intron 1, i2: *RBCS2* intron 2, FLAG: FLAG-tag, |: stop codon, *: bicistronic bridge sequence tagcat, $AphVIII^R$: Paromomycin resistance gene, $R2^T$: *RBCS2* terminator. Epifluorescence microscopy images of selected examples are shown. False-coloured yellow fluorescence from the Venus channel reports on the subcellular localization of

1016 the fluorescent reporter. MitoTracker fluorescence, false-coloured in cyan, indicates the
1017 position of mitochondria (although parts of the cell exterior are sometimes also stained),
1018 with salient features highlighted with orange arrows to indicate co-localization with the
1019 Venus channel. Chlorophyll autofluorescence, shown in magenta, indicates the location
1020 of the chloroplast. Scale bars are 5µm. Refer to Fig. 1 for sequences and Fig. S3 and
1021 S4 for biological replicates. Where a construct was interpreted as generating reporter
1022 localization in mitochondria or chloroplast, this is indicated by an orange 'm' or a green
1023 'c' respectively, in bold for full targeting or in italics for partial targeting. Brightness
1024 (Br.) was adjusted for clarity: fluorescence intensity values were restricted to the range
1025 shown for each channel by matching coloured rectangles. Intensity scales to 0 at the
1026 bottom of the panel, and to 65535 at the top. (**B**) A 'no Venus' control strain, expressing
1027 an empty vector, is shown with two different Venus channel brightness settings to
1028 visualize chloroplast autofluorescence in the Venus channel. Autofluorescence
1029 intensity is typically below 2000, indicated by a black dotted line in Brightness
1030 rectangles. Therefore, if Venus channel brightness is adjusted below 2000,
1031 autofluorescence originating from the chloroplast may be misinterpreted as Venus
1032 located in the chloroplast. **(C)** Lysine (K) and Arginine (R) frequencies for
1033 Chlamydomonas TPs and HA-RAMPs are shown as boxplots (centre line: median; box
1034 limits: upper and lower quartiles; whiskers: min/max values within 1.5x interquartile
1035 range). To give a baseline for comparison, their average across the UNIPROT database
1036 is given as red horizontal line. Statistically significant differences are indicated with
1037 stars: (Multiple Kruskal-Wallis tests followed by Dunn post-hoc tests, ***$p<0.0001$).
1038 See Fig. S21 for all amino acids. (**D**) Grayscale photos of antimicrobial activity growth
1039 assays in the presence of dilutions of three selected antimicrobial peptides against *E.*
1040 *coli* strain BW25113 show reduced activity when natively K-rich sequences (K, left
1041 hand columns) are altered by replacing all Ks by Rs (R, right hand columns). (**E,F**)
1042 Epifluorescence images of selected chimeric (**E**) TP controls (MH: mTP helical
1043 fragment, CH: cTP helical fragment) or (**F**) HA-RAMP constructs are shown as in (A).
1044 Note that images shown for MII_K in (A) and (F) come from independent insertion lines
1045 of the same construct. See Fig. S4, S5, S9, S18, S19 for biological replicates. (**G**) Data
1046 for targeting across all HA-RAMP constructs (i.e. including additional modifications)
1047 is shown, comparing K and R-bearing peptides (cf. Fig. 2 columns a-m).
1048



1049

1050 **Fig. 4. TP N- and C-termini enable chloroplast targeting by HA-RAMPs.**
1051 Exemplary epifluorescence images are shown as in Fig. 3; same convention as in Fig.
1052 3 for Venus localization. Note that images shown for MH_R^m and $^cCH_R^c$ in (B) come
1053 from different independent insertion lines of the same constructs as in Fig. 3E. See Fig.
1054 S3-S20 for biological replicates and further examples of the same trends.

1055

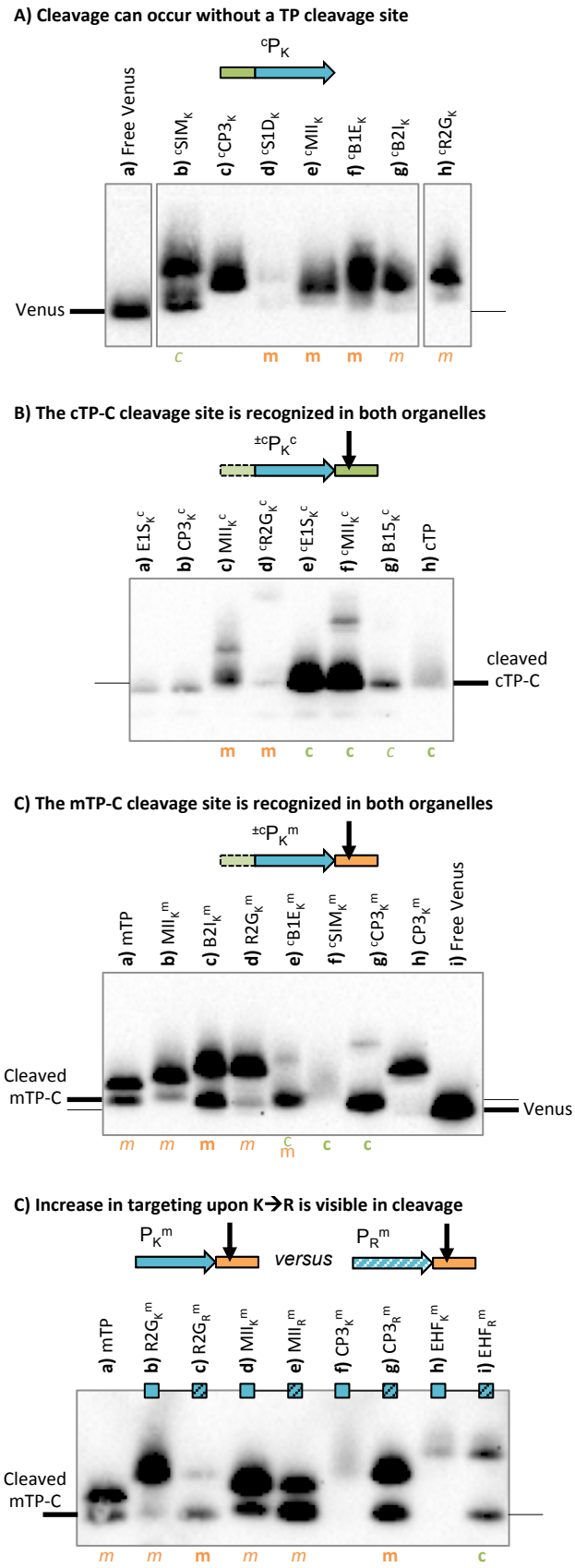


Fig. 5. Import is associated with maturation of the preprotein. Western Blots used an α-FLAG antibody on selected constructs as indicated above blots. Where a construct was interpreted as generating reporter localization by fluorescence microscopy (cf. Fig. 2, Fig. S3-S20) in mitochondria or chloroplast, this is indicated by an orange 'm' or a green 'c' respectively, in bold for full targeting or in italics for partial targeting. In (**A**) some lanes were spliced for clarity; the uncropped blot is provided in Fig. S28. The migration of Venus without any presequence, and Venus with additional amino acids at the N-terminus left over after cleavage of cTP-C and mTP-C, are indicated for reference based on free Venus and cTP/mTP controls respectively. The cTP control is RCA1-cTP (construct $^{c}CH_{R}^{c}$), the mTP control is CAG2-mTP (construct MH_{R}^{m}).

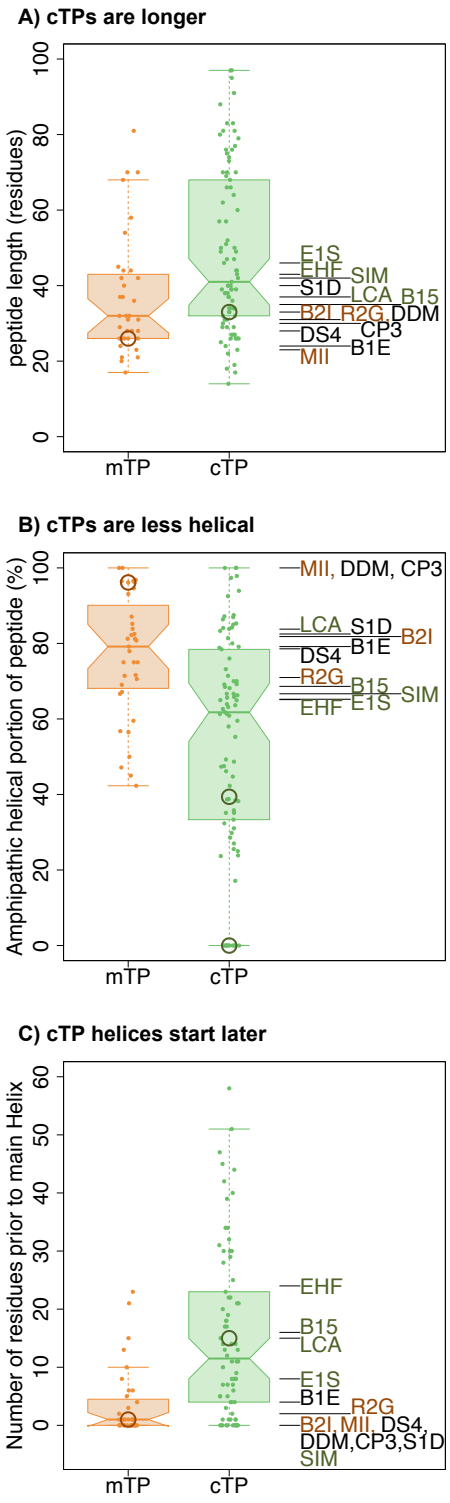


Fig. 6. Chloroplast and mitochondrial targeting HA-RAMPs match cTPs and mTPs respectively. For salient properties, Chlamydomonas mTPs and cTPs are compared to our 13 HA-RAMPs. TP distributions are shown as boxplots (centre line: median; box limits: upper and lower quartiles; whiskers: min/max values within 1.5x interquartile range), coloured points represent individual peptides. The position of the CAG2-mTP and RCA1-cTP is circled in each graph. The non-zero value for RCA1-cTP in (B) and the single circle in (C) report on the amphipathic helix established by an NMR-study (Krimm et al., 1999), since no helix could be predicted by our approach. HA-RAMPs are colour-coded by preferred targeting: orange = mt-set; green = cp-set (cf. Fig. 2). (**A**) cTPs are significantly longer than mTPs (p=0.0017), and cp-set are longer than mt-set HA-RAMPs (p=0.0184). (**B**) Helices make up a significantly smaller fraction of cTPs than mTPs (p<0.0001), and HA-RAMPs show a similar trend (0.1122). (**C**) cTPs contain significantly longer sequence stretches upstream of the longest predicted helix than mTPs (p<0.0001), as do cp-set compared to mt-set HA-RAMPs (p=0.0205). Reported p-values were obtained through two-way t-tests for TPs and one-way t-tests for HA-RAMPs.

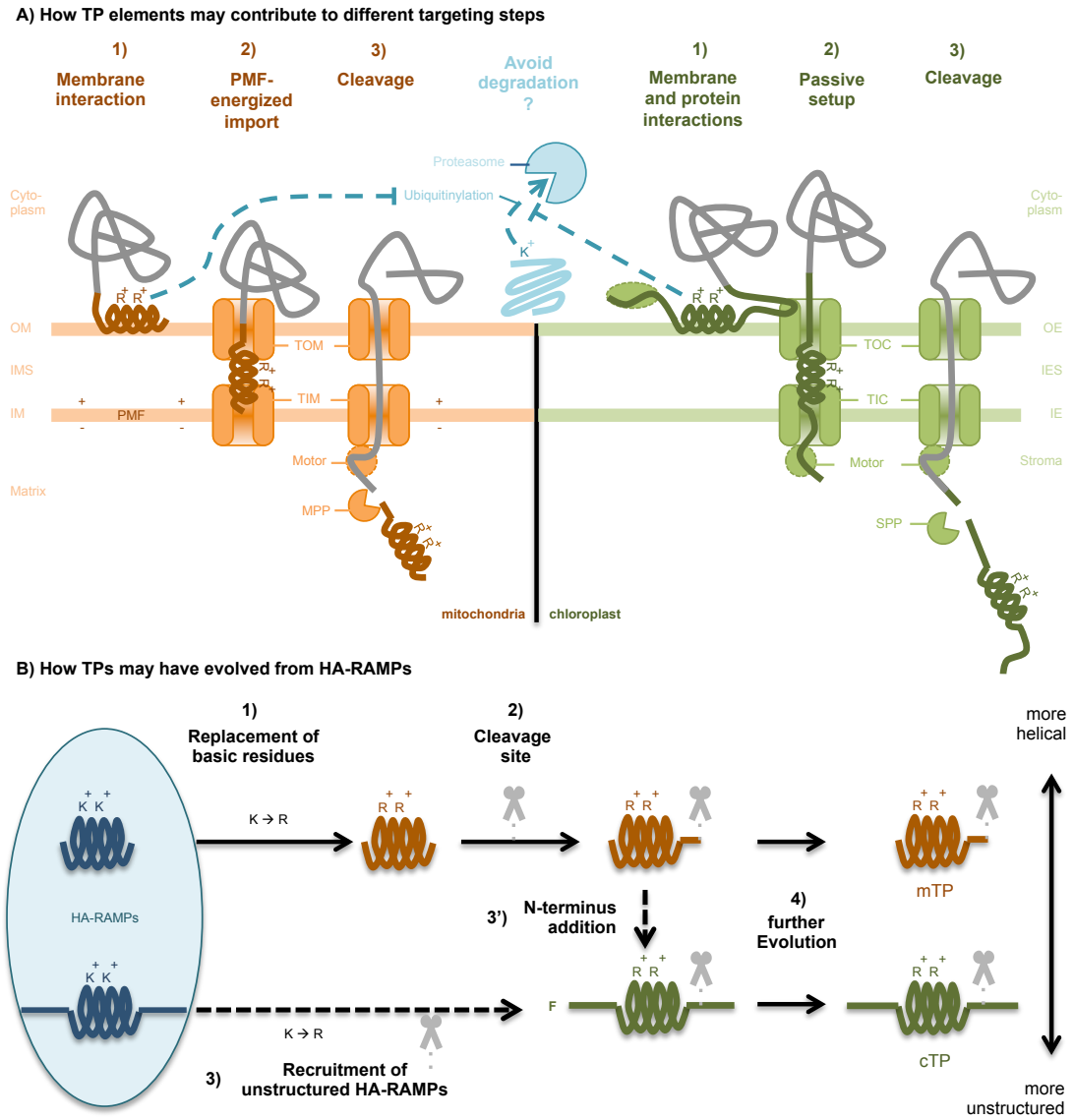


Fig. 7. Proposed functioning and evolution of TPs. Abbreviations: OM – outer membrane, IMS – inter-membrane space, IM – inner membrane, OE – outer envelope, IES – inter-envelope space, IE – inner envelope, PMF – proton motive force, TOM – translocon of the outer mitochondrial membrane, TIM – translocon of the inner mitochondrial membrane, TOC – translocon of the outer chloroplast envelope, TIC – translocon of the inner chloroplast envelope, MPP – matrix processing peptidase, SPP – stromal processing peptidase, R – Arginine, K – Lysine, F – Phenylalanine, mTP – mitochondrial targeting peptide, cTP – chloroplast transit peptide. (**A**) A number of roles for TP elements during protein import are suggested. Since ubiquitination targets K residues, a preference for R in TPs may increase preprotein stability. R may also play a role in lipid interactions. (1) membrane interactions may play a role in enabling differential targeting due to organelle-specific lipid preferences of TP helices. Protein

1132 interactions by cTP N- and C-terminal elements, e.g. with cytosolic factors or TOC-
1133 components (subunits are not shown for simplicity), may also play a role in specific
1134 targeting. (2) Import of positively-charged mTPs across IM is energized by the PMF.
1135 By contrast, a passive setup is required for cTPs to allow N-termini to reach into the
1136 stroma and contact the motor complex, likely contributing to increased length and
1137 relatively unstructured N- and C-termini of cTPs. (3) Sequence elements contributing
1138 to targeting may be present downstream of cleavage sites in cTPs. (**B**) If TPs evolved
1139 from AMPs, 1) HA-RAMPs would have first changed K to R and 2) acquired a cleavage
1140 site to become mTPs. To generate cTPs, either 3) more unstructured HA-RAMPs were
1141 recruited directly to become cTPs by undergoing a K to R shift and acquiring a cleavage
1142 site separately, or 3') mTPs acquired an N-terminal non-helical domain. Early cTP N-
1143 termini likely contained a starting F (Wunder et al., 2007). 4) Further evolution would
1144 have reinforced the differences between cTPs and mTPs to limit mis-targeting.
1145

1146 **Supporting Information:**

1147 **Supplementary Text**

1148 **Fig. S1** HA-RAMP candidates cover a diversity of physico-chemical properties

1149 **Fig. S2** Automated image analysis backs up manual targeting assessment

1150 **Fig. S3** Biological replicates of Brevinin-2ISb

1151 **Fig. S4** Biological replicates of Magainin 2

1152 **Fig. S5** Biological replicates of Ranatuerin-2G

1153 **Fig. S6** Biological replicates of Brevinin-1E

1154 **Fig. S7** Biological replicates of Dermaseptin S4

1155 **Fig. S8** Biological replicates of Dermadistinctin-M

1156 **Fig. S8** Biological replicates of Cecropin-P3

1157 **Fig. S10** Biological replicates of Sarcotoxin-1D

1158 **Fig. S11** Biological replicates of Esculentin-1SEA

1159 **Fig. S12** Biological replicates of Leucocin-A

1160 **Fig. S13** Biological replicates of SI Moricin

1161 **Fig. S14** Biological replicates of Bacillocin 1580

1162 **Fig. S15** Biological replicates of Enterocin HF

1163 **Fig. S16** Biological replicates of negative control Random Peptide 1

1164 **Fig. S17** Biological replicates of negative control Random Peptide 2

1165 **Fig. S18** Biological replicates of Rubisco activase cTP helical element control

1166 **Fig. S19** Biological replicates of γ-carbonic anhydrase 2 mTP helical element control

1167 **Fig. S20** Biological replicates of no-peptide and A-screen controls

1168 **Fig. S21** Comparison of amino acid frequencies reveals K/R shift

1169 **Fig. S22** PCAs reveal that N- but not C-termini differ between cTPs and mTPs

1170 **Fig. S23** Algal and plant cTP N-ter share physicochemical differences against mTP

1171 **Fig. S24** K→R generally improves HA-RAMP targeting

1172 **Fig. S25** HA-RAMP properties determine their targeting propensities

1173 **Fig. S26** Higher protein interactivity predicted for cTPs than mTPs

1174 **Fig. S27** Hsp70 and (F)GLK sites are present in HA-RAMPs

1175 **Fig. S28** Western Blots for K-bearing constructs

1176

1177 **Table S1** Similarity of HA-RAMPs to TPs quantified (Excel file)

1178 **Table S2** Numeric values derived of bioinformatic analyses (Excel file)

1179 **Table S3** Automated image quantification data (Excel file)