Retracing the evolution from antimicrobial to targeting peptides

Caspari, O.D.[1,2]*, Garrido, C.[1], Choquet, Y.[1], Wollman F.-A.[1], Lafontaine, I[1]*

Affiliations :

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

Current address :

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

***Correspondance :**

ODC : odc20@cantab.net

IL : ingrid.lafontaine@ibpc.fr

Keywords : Evolution, endosymbiosis, organellogenesis, chloroplast, mitochondria, subcellular targeting, protein import, transit peptide, presequence, *Chlamydomonas reinhardtii*

Abstract

We experimentally challenged 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 such antimicrobial peptides into *bona fide* organelle-targeting peptides, we expressed a set of 13 antimicrobial peptides of various origins in the green alga *Chlamydomonas reinhardtii* that serves as a model for both mitochondrial and chloroplast import. The peptides were modified to match 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. We present a temporal evolutionary scenario that emphasizes the early contribution of exchanging Lysines with Arginines in the sequence of the antimicrobial peptide, the evolution of a processing site followed by the addition of unstructured sequence and protein interaction sites that allow the selective targeting to the chloroplast.

Introduction

Antimicrobial peptides (AMPs) have recently been proposed to be at the evolutionary origin of N-terminal targeting peptides (TPs) that direct proteins into mitochondria and chloroplasts[1,2]. Part of the innate immune system, AMPs are produced by virtually all types of cells in a bid to kill or control microbial adversaries[3,4]. The proposed evolutionary scenario posits that early during endosymbiosis, the host used AMPs to attack the bacterial proto-endosymbiont.

44 A bacterial resistance mechanism, whereby the AMP is imported into the
45 bacterial cytoplasm, would have generated a pathway for host proteins to reach
46 the bacterial cytosol – a plausible first step in the evolution of a protein import
47 machinery. Cationic, Helical-Amphipathic Ribosomally-produced AMPs (HA-
48 RAMPs) and TPs share key physico-chemical properties and they were shown, in
49 some instances, to retain cross-functionality[5]: several TPs showed antimicrobial
50 activity and five HA-RAMPs, fused to a cleavage-site containing TP element, were
51 shown to promote the import of a Venus fluorescent protein into either the
52 mitochondria or the chloroplast of the model green alga *Chlamydomonas*
53 *reinhardtii*.
54 The main part of mTPs and the central element of cTPs appear derived from HA-
55 RAMPs on account of a shared cationic, amphipathic helical structure that often
56 encompasses the entire length of the HA-RAMPs[2] and mTPs[6,7]. Although plant
57 cTPs have been described as unstructured[8], algal cTPs more closely resemble
58 mTPs in being helical[9]. Helices were observed by NMR in membrane-mimetic
59 environments in both algal and plant cTPs[10–14], and signatures of amphipathic
60 helices can be detected in a majority of cTPs[5].
61 In addition to the helices, both mTPs and cTPs contain recognition sites at the C-
62 terminus where processing peptidases in the mitochondrial matrix (MPP) and
63 the chloroplast stroma (SPP) cleave off the cargo protein[15]. These recognition
64 sites encompass some 10 residues upstream of the cleavage site and are
65 structurally distinct from the rest of the TPs, even showing a weak sequence
66 conservation[6,16,17]**.** While targeting information is usually contained within mTP
67 sequences upstream of the cleavage site, targeting by cTPs shorter than ca. 60
68 amino acids often requires downstream unstructured sequence stretches in the

69 N-terminal domain of the mature protein[18,19]. Besides the amphipathic helical
70 module and the C-terminal cleavage module shared between mTPs et cTPs, it has
71 been argued that distinct features at their N-termini conferred organelle
72 specificity to each set of targeting peptides [6,16,20–26].
73 In this study, we systematically introduced modifications into candidate HA-
74 RAMPs to retrace likely steps along the evolution of TPs that improved
75 targeting[1,2]. We thus retrace the evolutionary switches from HA-RAMPs to TPs
76 by adding N- and C-terminal TP elements to 13 HA-RAMP sequences, chosen for
77 their similarity to TPs, in a bid to generate targeting to the mitochondria or
78 chloroplast in the model green alga *Chlamydomonas reinhardtii*. Besides
79 providing a guide to the most likely evolutionary sequence of events, this dataset
80 also provides new insights into how different TP elements contribute to
81 differential targeting and to the efficiency of protein import. We show that some
82 HA-RAMPs natively contain such sequence elements, with some HA-RAMPs
83 being prone to chloroplast targeting while others show a preference for the
84 mitochondria. Furthermore, we provide evidence for a critical evolutionary
85 contribution of a substitution of cationic residues from Lysine (K) in HA-RAMPs
86 to Arginine (R) in TPs.

87 Results

88 Figure 1 shows the major sequence features of the 13 HA-RAMPs used in the
89 present study (Fig. 1A) together with those of a typical cTP and a typical mTP
90 (Fig. 1B). On the right-hand side, these peptides are represented according to
91 their proportion of α-helical amphipathic structure. Three HA-RAMPs (Magainin
92 II, Brevinin 1B and Dermaseptin S4) were chosen based on a manual comparison

93 with TP sequences, and a further ten were chosen based on scoring as highly
94 similar to TPs in a recent analysis[5], such that different AMP families would be
95 represented.
96 To explore targeting, the 13 HA-RAMPs were systematically modified (Fig. 2),
97 notably by adding non-helical TP elements from the N- and/or C-terminal
98 domains of Chlamydomonas Rubisco activase (RBCA) cTP, or a C-terminal
99 domain of similar length from mitochondrial γ-carbonic anhydrase 2 (CAG2)
100 mTP (Fig. 1B, Fig. 2). Peptides carrying the 15 amino-acid long cTP N-terminus
101 (cTP-N) will be denoted ^{c}P. Similarly, peptides with C-terminal elements from
102 RBCA-cTP (cTP-C) or CAG2-mTP (mTP-C) will be denoted P^{c} or P^{m} respectively.
103 Micrographs for three biological replicates per construct are displayed on Sup.
104 Figs. 1-18.
105 In the absence of any modifications, only Brevinin-2I (B2I) and Magainin 2 (MII)
106 out of the set 13 HA-RAMPs were capable of organelle-targeting visible in
107 fluorescence microscopy (Fig. 2A). When equipped with B2I, the fluorescent
108 reporter Venus shows mitochondrial localization: the Venus signal is observed as
109 a characteristic network pattern that co-localizes with MitoTracker fluorescence
110 (Fig. 3). By contrast, in the case of MII only a small fraction of fluorescence signal
111 in the Venus channel shows co-localisation with the mitotracker dye, signifying
112 that targeting is only partial (Fig. 3).

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

114 We looked for differences in primary sequence that might impede the ability of
115 the remaining 11 HA-RAMPs to target. As illustrated on Supp. Fig. 19, extant HA-
116 RAMPs and organellar TPs display a few differences in their amino-acid content.

117 HA-RAMPs are richer in helix-breaking Glycines (G) but poorer in Prolines (P),
118 whereas TPs are enriched in Serines (S) compared to HA-RAMPs. Most of the
119 other amino acid frequencies are broadly similar between the two sets of
120 peptides and within range of the uniprot average. As expected, both HA-RAMPs
121 and TPs are poor in acidic residues (D, E) but enriched in basic residues (K or R).
122 However, their complement in basic residues is markedly different (Fig. 4A):
123 HA-RAMPs are rich in Lysines (K) whereas TPs are rich in Arginines (R).
124 In order to see whether these contrasting differences in K/R ratio contributed to
125 the functional divergence between HA-RAMPs and TPs, we substituted all
126 instances of R by K in TPs and of K by R in HA-RAMPs. In the rest of the text, the
127 basic amino acid mostly present in a given peptide P will be indicated by a
128 subscript as P_R or P_K. HA-RAMPs with K→R transition showed reduced
129 antimicrobial activity, as illustrated on Fig. 4B by the increased minimum
130 inhibitory concentrations for MII, Dermaseptin S4 (DS4) or Brevinin 1E (B1E).
131 Conversely, when the natively R-rich Chlamydomonas Rubisco activase (RBCA)
132 cTP ($^{c}CH_{R}{}^{c}$) and mitochondrial γ-carbonic anhydrase 2 (CAG2) mTP ($MH_{R}{}^{m}$)
133 underwent systematic R→K substitutions ($^{c}CH_{K}{}^{c}$, $MH_{K}{}^{m}$), their ability to target
134 the chloroplast and the mitochondria were reduced or abolished (Fig. 4C). Note
135 that TP helical fragments, devoid of N- and C-terminal domains, are denoted as
136 MH for mTP and CH for cTP (Fig. 1B). These experiments demonstrate the
137 respective functional contributions of K and R residues to antimicrobial and
138 organelle targeting activity. We then systematically re-examined the organelle
139 targeting ability of the set of 13 HA-RAMP that had undergone K→R
140 substitutions (Fig. 2B). MII_R now shows full mitochondrial targeting, and $R2G_R$

141 (Ranatuerin 2G) as well as $CP3_R$ (Cecropin P3) show partial targeting, as
142 evidenced by significant overlap of Venus and mitotracker fluorescence (Fig. 4D).

143 Targeting peptide elements contribute to targeting

144 *C-termini matter for targeting*

145 Addition of mTP-C, the cleavage-site containing C-terminal element of CAG2-
146 mTP (Fig. 1B) enabled mitochondrial targeting for $ES1_R$ (Esculentin 1S) and
147 $R2G_K$ and improved the targeting already observed for $R2G_R$ and $CP3_R$, in
148 addition to continued targeting by B2I and MII (Fig. 2CD). The impact of adding
149 cTP-C, the C-terminal element derived from RBCA-cTP (Fig. 1B) is even more
150 important: cTP-C significantly enabled chloroplast targeting (Fig. 2E,F, Fig. 5B),
151 which could be seen in 9 HA-RAMP constructs involving EHF (Enterocin HF),
152 B15 (Bacillocin 1580), LCA (Leucocin A), S1D (Sarcotoxin 1D) and $CP3_R$. The
153 addition of cTP-C to the helical part of the CAG2-mTP ($MH_R{}^c$) even partially
154 retargets the construct to the chloroplast (Fig. 5C), and deletion of cTP-C
155 abolishes chloroplast targeting by RBCA-cTP (^cCH_R). By comparison, deletion of
156 mTP-C reduces but does not totally abolish mitochondrial targeting of CAG2-
157 mTP (MH_R), and addition of mTP-C fails to rescue ^cCH_R ($^cCH_R{}^m$). Note that the
158 addition of cTP-C was also compatible with mitochondrial localization by B2I
159 and MII and the addition of mTP-C also enabled chloroplast localization by EHF_R
160 (Fig. 2D-F, Fig. 5A,B).

161 *N-termini matter for chloroplast targeting*

162 The sole addition of cTP-N generated at least partial chloroplast targeting (Fig.
163 2G,H, Fig. 5D) in 6 HA-RAMP constructs, notably by SI Moricin (SIM) as well as
164 EHF_R, $B15_R$, $E1S_R$ and $CP3_R$. However, cTP-N also enabled mitochondrial

165 targeting (Fig. 2G,H, Fig. 5E) in three HA-RAMPs that had not previously shown
166 mitochondrial localization, B1E, S1D and DDM_R (Dermadistinctin M).
167 The importance of cTP-N as a chloroplast determinant becomes most evident
168 when combined with a C-terminal element: addition of cTP-N to constructs
169 equipped with mTP-C (Fig. 2I,J) resulted in 10 HA-RAMP constructs targeting the
170 chloroplast (EHF, B15, SIM, E1S, LCA_K, $CP3_K$) as well as 6 constructs showing
171 dual targeting to the chloroplast and the mitochondria (S1D, B1E, MII_R, $B2I_R$).
172 Combining cTP-N with cTP-C (Fig. 2KL) generates at least partial chloroplast
173 targeting in 22 out of 26 HA-RAMP constructs (including 1 instance of dual
174 targeting by $R2G_R$); only Dermaseptin S4 (DS4) fails to show any chloroplast
175 targeting. Consider B2I with cTP-N (^cB2I, Fig. 2G-La) as an example of the effect
176 of combining cTP-N with C-termini, shown in Fig. 5F. Equipped with mTP-C,
177 $^cB2I_K^m$ targets the mitochondria whereas addition of cTP-C ($^cB2I_K^c$) generates
178 partial targeting to the chloroplast. Upon K→R, the presence of mTP-C ($^cB2I_R^m$)
179 now supports dual targeting and full chloroplast targeting is observed in the
180 presence of cTP-C ($^cB2I_R^c$).
181 Finally, Fig. 5G shows that while addition of cTP-N to the CAG2-mTP abolished
182 mitochondrial targeting ($^cMH_R^m$), the cTP-N/cTP-C combination also retargets
183 MH to the chloroplast, and enables partial chloroplast targeting of a control in
184 which the central element is composed of a poly-Alanine tract ($^cAA^c$) devoid of
185 amphipathic properties and of a no peptide control (^{c}-^{c}). The two latter
186 experiments demonstrate that there are enough determinants for recognition of
187 the chloroplast translocon dispersed between the Nter and Cter of a cTP, to
188 target Venus to the chloroplast, albeit with a lower efficiency than when an
189 amphipathic helix is present in between.

190 To further explore the role of the N-terminal cTP domain, we performed some *in*
191 *silico* comparisons of the studied peptides. In Principal Component Analyses (on
192 ACC Z-scales[5]) comparing TP N- and C-termini with our 13 HA-RAMPs (Sup. Fig.
193 20), cTPs and mTPs could be differentiated much better by N- compared to C-
194 termini, supporting N-termini as important specificity determinants.
195 Furthermore, even though algal cTP N-termini are more charged, less
196 hydrophobic and more disordered than higher plant cTPs (Sup. Fig. 21) (in line
197 with algal cTPs being more mTP-like[9]), Chlamydomonas cTP N-termini are
198 recognized as chloroplast-targeting in a model trained on Arabidopsis TP and
199 vice versa (Sup. Fig. 21), confirming that essential physico-chemical features of
200 chloroplast determination in cTP N-termini are maintained in both lineages.

201 **HA-RAMPs dominate targeting specificity**

202 Considering Fig. 2 by columns reveals that organelle specificity is, to a large
203 extent, determined by HA-RAMPs. Only 3 HA-RAMPs (B2I, R2G, MII; referred to
204 hereafter as the mt-set) account for >70% of all constructs where mitochondrial
205 targeting is seen. Similarly, 5 HA-RAMPs (E1S, LCA, SIM, B15, EHF; referred to as
206 the cp-set) account for the majority (~57%) of all chloroplast targeting, and for
207 two thirds when excluding the cTP-N/cTP-C combination that generates
208 chloroplast targeting across most HA-RAMPs. In some instances, a set of peptide
209 modifications may switch targeting from the mitochondria to the chloroplast, yet
210 the major effect of modifications, i.e. exchanging K→R (Fig. 4, Sup. Fig. 22) or
211 adding TP elements (Fig. 5, 6), is to enhance the targeting ability to an organelle
212 determined by the HA-RAMP primary sequence properties (Fig. 2).

213 ## Chloroplast targeting involves longer, less helical peptides

214 To understand what differentiates mt-set from cp-set HA-RAMPs, we examined
215 the sequence characteristics of our 13 HA-RAMPs with those of well-
216 characterized Chlamydomonas targeting peptides (Fig. 6). In Chlamydomonas,
217 mTPs are *ca.* 35 residues in length (Fig. 6A). By contrast, most cTPs are
218 significantly longer (p=0.0017). The difference is even greater than shown here
219 given that many cTPs require a contribution from post-cleavage site residues for
220 successful targeting[5,18,19]. Consistently, cp-set HA-RAMPs (in green, Fig. 6A) are
221 longer than mt-set HA-RAMPs (in orange; p=0.0184) and require further
222 elongation by addition of TP elements before targeting can be observed (Fig. 2).
223 Most mTPs fold into an amphipathic helix for most of the sequence (~80% on
224 average, Fig. 6B) starting directly from their N-terminus (Fig. 6C). The fraction
225 dedicated to amphipathic helix formation in cTPs is significantly lower (Fig. 6B,
226 p<0.0001), with a longer stretch of non-helical sequence at the N-terminus (Fig.
227 6C, p<0.0001). A PCA analysis based on these properties, plus the fraction of R
228 and K, recapitulates the combined properties within each peptide (Sup. Fig. 23).
229 The PCA shows that cp-set HA-RAMPs are localized with cTPs in the upper left
230 area of the graph, consistent with a lower helix fraction (Fig. 1, Fig. 6B) due to
231 longer sequence stretches upstream of the main helix than in the mt-set HA-
232 RAMPs (Fig. 1, Fig. 6C, p=0.0199). Consistent with the importance of these
233 helical features, LCA, the most helical among cp-set HA-RAMPs also shows the
234 highest fraction of only partial chloroplast targeting (3 out of 5 targeting
235 constructs, Fig. 2j rows E,F,K). Furthermore, SIM and E1S, the two cp-set HA-
236 RAMPs with the shortest pre-helix segments (Fig7C), both require addition of
237 cTP-N for chloroplast targeting (Fig. 2i,k).

238 *cTPs are predicted to be more prone to protein interaction than mTPs*

239 cTPs appear more likely to interact with proteins than mTPs (Sup. Fig. 24), just
240 like cp-set HA-RAMPs. 'FGLK' motifs that are reportedly TOC interaction sites in
241 plant cTPs[20] appear shortened to 'GLK' motifs in Chlamydomonas but they are
242 associated with an increased protein interaction potential in cTPs compared to
243 mTPs ($p<0.0001$), including within cTP-N and cTP-C elements. However, Hsp70-
244 interaction sites[23], which are commonly found at cTP N-termini[20], also occur at a
245 high frequency in mTP sequences and HA-RAMPs (Sup. Fig. 24, 25).

246 **Probing cleavage of HA-RAMP-driven reporter constructs by immunobloting**

247 To characterize the maturation of HA-RAMP–targeted proteins upon organellar
248 import, we performed immunoblotting experiments (Fig. 7). Indicative examples
249 were selected for clarity; a more comprehensive overview is provided in Sup.
250 Fig. 26.

251 In the absence of a dedicated cleavage site (Fig. 7A), we found some cleavage did
252 occur but preproteins are also maintained: two bands are observed in constructs
253 that were localized in an organelle by fluorescence imaging (lanes d,f-j). The
254 upper band corresponds to unprocessed Venus, which is a fusion of the reporter
255 with the HA-RAMP construct and migrates at varying positions depending on the
256 length of the presequence. The lower band corresponds to the processed form
257 migrating closer to the "free" Venus position (lane c) depending on the exact site
258 of cleavage. The cTP and mTP controls (lanes a,b) contain 23 residues between
259 cleavage site and Venus (Venus+23), and maintain very little preprotein
260 migrating above. In HA-RAMP constructs by contrast, the majority of Venus
261 remained in the unprocessed top band for constructs targeting the mitochondria
262 (lanes f-j). Nonetheless, the presence of a faint processed form in the lower band

263 just above free Venus shows that some cleavage occurs. The proportion of
264 processed versus unprocessed Venus preprotein is much higher in partially
265 chloroplast-targeted $^{c}SIM_{K}$ (lane d), hinting at more efficient degradation of the
266 unprocessed form in the chloroplast (see also Sup. Fig. 26).
267 Addition of cleavage sites improves cleavage across organelles: In the presence
268 of either cTP-C (Fig. 7B, Blot 1) or mTP-C (Fig. 7B, Blot 2), which contain the
269 RBCA-cTP and CAG2-mTP cleavage sites respectively, targeting constructs
270 appear to be cleaved at a site corresponding to TP controls (Venus+23).
271 Constructs equipped with either of the two cleavage sites deliver a processed
272 Venus of similar size, independently of whether the mitochondria or the
273 chloroplast are targeted (Fig. 7B). These results indicate that mTP and cTP
274 cleavage sites are recognized in both organelles. We note that in the case of HA-
275 RAMP constructs equipped with mTP-C, cleavage appears to occur not at the
276 cleavage site previously identified for CAG2 (Fig. 7B Blot 2 lane a, top band,
277 corresponding to Venus+23), but instead at or near a second site closer to "free"
278 Venus (compare Fig. 7B Blot 2 lane a, bottom band to lanes b-g), implying
279 processing within ten residues upstream of Venus. Unlike mTP-C (Fig. 7B Blot 2
280 lane h), cTP-C appears to be processed in the absence of targeting (Fig. 7B Blot 1
281 lanes a,b) at a site different to the one used in organelles (Fig. 7B Blot 1 lanes c-
282 h).
283 A switch from K to R (Fig. 7C), which improved targeting, also increased the
284 amount of processed form relative to that of the unprocessed form in all cases,
285 whether due to a gain of targeting (lanes f-i) or an increase in efficiency (lanes b-
286 e), suggesting that cleavage can serve as a proxy for targeting. Indeed, there is a
287 remarkable correlation between which constructs show cleavage, as observed in

288 immunoblots, and which constructs show targeting, as judged from microscopy
289 (Sup. Fig. 26). We note that faint processed bands can be seen for several
290 unmodified HA-RAMPs (Sup. Fig. 26), suggesting further peptides in addition to
291 $B2I_K$ and MII_K (Fig. 2) may be able to support some amount of targeting that is,
292 however, too partial to be seen in fluorescence microscopy.

293 Discussion

294 How TPs may have evolved from HA-RAMPs

295 In an assessment of the hypothesis that extant TPs are derived from a subclass of
296 antimicrobial peptides that harbor an amphipathic helix, the HA-RAMPs[1,2], we
297 studied a systematic series of HA-RAMP constructs. This dataset allows us to
298 propose an evolutionary scenario (Fig. 8) for the conversion of such peptides
299 from antimicrobial to *bona fide* targeting peptides whose properties allow
300 differential targeting between mitochondria and chloroplasts. We showed that a
301 K to R switch reduces cytotoxicity of HA-RAMPs, while increasing targeting. This
302 dual effect must produce a strong selection pressure favoring this exchange at an
303 early step of the endosymbiosis process (Fig. 8A). The underlying mechanism
304 warrants further research, but the finding is in line with the fact that Arginines
305 provide a better uptake efficiency than other cationic residues because of the
306 guanidinium group forming bidentate hydrogen bonds with membrane
307 phosphate groups[27,28]. Differences in bulkiness or in pKa[29,30], or trans-acting
308 factors such as regulation of targeting through ubiquination of K[31], may also play
309 a role. Mitochondria appear able to tolerate the presence of HA-RAMP-
310 containing preproteins, suggesting mTPs could have evolved relatively easily

311 from HA-RAMPs. Even unaltered HA-RAMPs can in some cases deliver cargo,
312 providing a plausible starting point for protein import. Thus addition of cleavage
313 sites would likely have come as a second step (Fig. 8B), allowing presequences to
314 be degraded separately from the cargo protein[32]. Interestingly, Arginines feature
315 prominently in Chlamydomonas cleavage site consensus sequences[17], thus
316 cleavage sites may be generated by the K→R switch, potentially accounting for
317 increased chloroplast-localization among cTP-N bearing constructs (Fig. 2GH).
318 Since cTPs evolved in a cell that already had mTPs, additional discriminating
319 elements were needed to avoid large scale mistargeting. We found that N-
320 terminal elements interact with suitable C-termini to determine chloroplast
321 targeting. The ~15 N-terminal residues have been suggested as specificity
322 determinants in the literature before[20–22], however, the contribution of this
323 element to chloroplast targeting has remained under debate. Our data suggests
324 that rather than providing an absence of Arginines[21,33] (Sup. Fig. 21) or a
325 difference in Hsp70 binding sites[20,25,26] (Sup. Fig. 24), cTP N-termini appear to
326 contribute unstructured sequence upstream of the amphipathic helix. Thus HA-
327 RAMPs that intrinsically contained relatively unstructured N-termini proved
328 most able to support chloroplast targeting even in the absence of cTP-N. The
329 need to contain unstructured sequence[8] as well as the increased length of cTPs,
330 even extending beyond the cleavage site in many cases[5,18,19], appear
331 mechanistically related to the import system. Whereas mitochondrial import
332 makes use of the proton gradient across the inner membrane to power uptake of
333 positively charged presequences[34,35], energized chloroplast import requires the
334 cTP to be elongated enough to reach across both TOC and TIC and contact the
335 translocation motor[20,36]. Highly structured sequence elements such as a helix at

336 the Venus N-terminus[37] can thus impede chloroplast targeting in the absence of a
337 sufficiently long presequence. HA-RAMPs with suitably unstructured C-termini,
338 such as SIM in this study, were thus most likely to generate chloroplast targeting
339 even in the absence of an added TP C-terminus. Similarly, higher chloroplast
340 targeting among cTP-C constructs may be due to mTP-C being more structured,
341 containing a long predicted amphipathic helix. Unstructured elements at N- and
342 C- termini may also contribute to chloroplast targeting by enabling interactions
343 with protein receptors associated with TOC. In basal algal lineages, a conserved F
344 at cTP N-termini appears to play a role in contacting TOC through an interaction
345 with cytoplasmic POTRA domains of TOC75[38–40], mimicking Omp85-interactions
346 in the bacterial cytoplasm mediated by C-terminal sequence motifs[41], which may
347 make a TOC-interacting N-terminus an ancient feature of cTPs. In the green
348 lineage, TOC-interaction has instead been attributed to semi-conserved 'FGLK'
349 motifs[25,42–44], although the requirement for F appears to be relaxed in
350 Chlamydomonas (Fig. 5G)[45]. The presence of GLK sites with high predicted
351 interactivity in cTP-N and cTP-C may explain the high chloroplast targeting
352 potential of constructs equipped with both elements. Thus the evolution of the
353 chloroplast import machinery and chloroplast-specific TPs likely went hand in
354 hand. To generate unstructured, protein interactive N- and C-termini, cTPs may
355 have evolved from HA-RAMPs that already contained unstructured sequence
356 elements (Fig. 8C), as seen for cp-set HA-RAMPs. Alternatively, cTPs may have
357 co-opted existing mTPs that already contained R and a cleavage site, yet were
358 still recognized by cyanobacterial HA-RAMP importers which were part of their
359 resistance mechanism to antimicrobial attack. In this case, the key innovation
360 that generated cTPs may simply have been the addition of an unstructured,

361 possibly TOC-interacting N-terminus to mTPs (Fig. 8D), which we showed here
362 to be sufficient to retarget mt-set HA-RAMPs equipped with TP C-termini at least
363 partially to the chloroplast.

364 ## The legacy of HA-RAMPs: Amphipathic helices and membrane interaction

365 A direct interaction with the membrane bilayer, rather than protein receptors,
366 would provide a basic mechanism for a first step in organelle targeting and could
367 explain how TPs can be functionally specific whilst diverse in sequence. Core
368 components of the outer membrane translocons, such as the translocation pores
369 TOM40 and TOC75 are widely conserved, and both yeast TOM40 and plant
370 TOC75 retain the ability to interact with TPs in the absence of dedicated receptor
371 proteins[46–48]. In contrast to the core components, dedicated TP receptor proteins
372 are not conserved between the key eukaryotic lineages and should be regarded
373 as later addition to translocation machineries[49,50]. Where receptor interactions
374 are required for targeting, binding sites need to be restored when using
375 heterologously expressed proteins, e.g. a leading F needs to be added to plant
376 cTPs to generate cyanelle-targeting in glaucophytes[51] (see previous paragraph).
377 Yet for the most part, heterologous expression of TPs across even large
378 phylogenetic distances generates reliable targeting in both plastids and
379 mitochondria[52–54], such as ophistokont mTPs in trypanosomes and cyanelle cTPs
380 in plant chloroplasts. Most strikingly, chromatophore targeting peptides from
381 *Paulinella chromatophora*, an amoeba that has undergone a primary
382 endosymbiosis with an α-cyanobacterium ca. 1 Mya, target tobacco
383 chloroplasts[55]. TP-enabled targeting must therefore use a universally shared
384 mechanism. In HA-RAMPs, amphipathic helices serve to insert into target

385 membranes. Individual mTPs and cTPs have also been shown to interact with
386 membrane bilayers[6,14]. Our finding that targeting specificity is largely
387 determined by the sequence of the HA-RAMPs points to amphipathic helical
388 elements being directly involved in determining target specificity. While N- and
389 C-terminal elements aid specificity determination, particularly for chloroplast
390 import, their influence fails to explain why some HA-RAMPs target preferentially
391 the mitochondria or exclusively the chloroplast. In analogy to Rubisco small
392 subunit cTP interacting with chloroplast-mimetic membranes only in presence of
393 the chloroplast-specific galactolipid MGDG[56], we suggest amphipathic helices
394 may interact specifically with the membranes of the targeted organelle[3]. Further
395 understanding of lipid interactions, the relationship between extant translocons
396 and bacterial resistance importers, as well as of the peptidases involved in the
397 degradation of TPs, should shed new light on the evolutionary process that
398 allowed the stable integration of proto-organelles through endosymbiosis.

399 Materials and Methods

400 Construct generation

401 Venus expression constructs were designed in SnapGene (v4.3.11) and
402 generated by integrating PCR-amplified (Q5 hot start high-fidelity, #M0515, New
403 England Biolabs) DNA fragments into plasmid pODC53[57] upstream of Venus
404 using Gibson assembly (NEBuilder HiFi DNA assembly, #E5520S, New England
405 Biolabs). Chlamydomonas TP sequences were amplified from genomic DNA
406 extracted from strain T222+ (CC-5101). Templates for codon optimized HA-
407 RAMP, RP and R→K modified TP sequences were obtained by gene synthesis

408 (Eurofins Genomics). Correct assembly was verified by sequencing (Eurofins
409 Genomics). Linear transformation cassettes were generated through restriction
410 digest with *Eco*RV (New England Biolabs).

411 ## Transformation and fluorescence screen conditions

412 Constructs were transformed into wild-type strain T222+ (CC-5101) using the
413 protocol described previously[58], except using 4µl of DNA at 2µg/µl.
414 Transformants selected for Paromomycin resistance were grown in 200µl TAP in
415 96-well plates under 50µmol photons m^{-2} s^{-2} for 3-5 days and then screened for
416 Venus expression in a fluorescence plate reader (CLARIOstar, BMG Labtech) as
417 described previously[57].

418 ## Microscopy

419 Cells were grown in 96-well plates and epifluorescence microscopy was carried
420 out as described before[5]. Strains with suspected mitochondrial targeting were
421 subjected to 0.1 µM MitoTracker Red CMXRos (ThermoFisher) in growth
422 medium for 30 min in the dark and washed with TAP prior to taking
423 epifluorescence images with the same settings as before[5], using a white light LED
424 and a beam splitter at 535nm (423052-0104-000, Zeiss) in combination with a
425 custom-made filter set (559/34 BrightLine HC, Beamsplitter T 585 LP, 607/36
426 BrightLine HC; AHF Analysentechnik AG) for detecting the MitoTracker signal.
427 Images were adjusted in Fiji[59] (ImgageJ version 2.0.0) as described before[5], and
428 final figures assembled in PowerPoint (Microsoft PowerPoint for Mac 2011,
429 Version 14.6.3).

430 ## Western Blots

431 Cells were grown up in 10ml TAP Erlenmeyer Flasks under 30µmol photons m^{-2}
432 s^{-1}, shaken at 160RPM until late mid-log phase. 2ml Aliquots were harvested by
433 centrifugation at room temperature in a tabletop centrifuge, resuspended in 30µl
434 storage buffer (1× Roche cOmplete™ Mini proteinase inhibitor cocktail, 10 mM
435 NaF, 0.2 M DTT, 0.2 M $NaCO_3$) and kept at -20°C. 20µl boiling buffer (1× Roche
436 cOmplete™ Mini proteinase inhibitor cocktail, 10mM NaF, 50 g/L SDS, 200 g/L
437 sucrose) was added to thawed aliquots, which were then boiled for 50 seconds.
438 Cell debris was removed by centrifugation (tabletop centrifuge, maximum speed,
439 15 minutes, 4°C). Chlorophyll content of 5 µl supernatant in 1 ml water was
440 estimated as 1 µg chlorophyll μl^{-1} = 0.11 × (OD_{680} – OD_{770}), with Absorbances at
441 680 nm (OD_{680}) and 770nm (OD_{770}) recorded in a spectrophotometer. Samples
442 were diluted to load identical volumes, and a constant 10 µg chlorophyll were
443 loaded onto large gels (35×27cm; resolving gel: 12% Acrylamide, 0.32%
444 Bisacrylamide, 0.25 µl/ml Tetramethylethylenediamine, 250 mg/L Ammonium
445 Persulfate, 375 mM Tris/HCl pH 8.8; stacking gel: 5% Acrylamide, 0.133%
446 Bisacrylamide, 0.666 µl/ml Tetramethylethylenediamine, 666 mg/L Ammonium
447 Persulfate, 125 mM Tris/HCl pH 6.8) alongside a molecular size marker
448 (Precision Plus Dual Extra Prestained Protein Standards, BioRad), and gels run
449 overnight at room temperature at 18 A. Proteins in the size range <25 kDA to
450 >75 kDa were transferred onto 0.1 µm nitrocellulose membranes for 1h at 0.8 A
451 cm^{-2}, using 5 layers of 3 mm filter paper (Whatman) soaked in transfer buffer 1
452 (40 mM Aminocaproic acid, 20% Isopropanol, 25 mM Tris/HCl pH 9.4) between
453 at the cathode, followed by gel, membrane and 2 filter papers soaked in transfer
454 buffer 2 (20% Isopropanol, 25 mM Tris/HCl pH 10.4), and finally 3 filter papers

455 soaked in transfer buffer 3 (20% Isopropanol, 300 mM Tris/HCl pH 10.4) at the
456 anode. Proteins were fixed and visualized by Ponceau Red staining, incubated for
457 1h at room temperature in block (30 g/L skimmed milk powder, 0.1% tween-20,
458 1x PBS: 140 mM NaCl, 30 mM KCl, 100 mM Na_2HPO_4, 15 mM KH_2PO_4) and
459 immunolabelled overnight at 4°C using an α-FLAG 1° antibody (Sigma Aldrich
460 F1804, diluted 1:10 000 in block). Membranes were thoroughly washed (0.1%
461 tween-20, 1x PBS), treated for 1h at room temperature with horseradish
462 peroxidase-conjugated α-mouse 2° antibody (diluted 1:10 000 in block), washed
463 again and revealed using ECL recorded on a ChemiDoc (Biorad). Blots were
464 analyzed and converted into TIFF format using the Image Lab software (Version
465 6.0.0 build 26, BioRad), final figures were assembled in PowerPoint (Microsoft).

466 ## Antimicrobial activity assay

467 Minimum inhibitory concentration assays were performed as described before[5].

468 ## Sequence Data Set

469 TPs with experimentally-confirmed cleavage sites were recovered from
470 proteomic studies: *Chlamydomonas reinhardtii* cTP (Terashima 2011), mTP
471 (Tardif 2012); *Arabidopsis thaliana* cTP (Ge 2008, Rowland 2015), mTP (Huang
472 2009). For each peptide, we recovered the full-length protein sequence from
473 NCBI and UNIPROT. Cytoplasmic control sequences were generated by taking N-
474 terminal sequence stretches of random length, matching the distribution of
475 peptide lengths observed in our Chlamydomonas TP dataset, from a random
476 subset of Chlamydomonas proteins with a validated cytoplasmic location in
477 UNIPROT. For calculating amino acid frequencies, the same HA-RAMP sequences
478 were used as before[5].

479 ## Amphipathic helix prediction

480 Amphipatic α-helices are predicted as described previously[5], following the
481 principle of the Heliquest algorithm[60]. Briefly, the approach aims to establish the
482 longest sequence stretch that contains identifiable hydrophilic and hydrophobic
483 faces. The algorithm is iterated such that multiple non-overlapping helices can be
484 found within a given peptide (Fig. 1). Consequently, helix fractions are calculated
485 as number of residues within all predicted helices divided by the total number of
486 residues in the peptide; for evaluating the number of upstream residues, only the
487 longest helix was considered (Fig. 7, Sup. Fig. 23).

488 ## ACC terms

489 In order to evaluate the physicochemical properties of our peptides we used the
490 approach we previously described[5]. Briefly, each amino acid is described in
491 terms of 3 'Z-scale' values[61] that can be interpreted as representing a residue's
492 hydrophobic, steric and electronic properties. Auto-cross-covariances (ACC)
493 between nearby residues are calculated up to a distance of four amino acids,
494 generating a quantitative representation of a given peptide in terms of 36 ACC
495 terms.

496 ## Visualization

497 We performed a principal component analysis (PCA) to visualize the
498 relationships between peptides as described by their 36 ACC terms (Sup. Fig. 20)
499 or by their 5 salient TP properties (Sup. Fig. 23) using python package sklearn
500 v0.22.1[62].

501 ## Analysis of TP N-termini

502 TP N-termini, defined as the N-terminal 15 amino acids, were analysed (Sup. Fig.
503 21) as follows: Charge profiles were generated as described in the literature[25].
504 Hydrophobicity of TP N-termini was estimated using the Heliquest standalone
505 application[60]. To evaluate disorder, we used IUpred2A, a software that calculates
506 the probability for each residue of being in a disordered region[63]. We used the
507 default 'Long' setting, which has been reported to be more accurate than the
508 alternative 'Short' setting[64]. The disorder of a given sequence was taken as the
509 mean of the probability values for each residue (average residue disorder
510 probability).

511 **Statistical prediction**

512 To evaluate the predictive power of ACC terms[5] obtained for 15 residue-long TP
513 N-termini with regards to localization (Sup. Fig. 21) we used a binomial logistic
514 regression classifier. We performed one hundred 5-fold cross validation runs. In
515 each set of one hundred runs, we randomly selected sequences such that the
516 same number of mTPs and cTPs were used. For *C. reinhardtii* we used 33 mTP
517 and 33 cTP and for *A. thaliana* we use 29 mTP and 29 cTP sequences. We used an
518 elastic net penalty in a saga solver with an l1-ratio (a-parameter) of 0, which is
519 equivalent to using only ridge penalty, where all features take part in the model,
520 and a C parameter ($1/\lambda$) of 0.1. The a and $1/\lambda$ parameters were optimised with
521 10-fold cross validation. Firstly, when $1/\lambda = 1$, the best accuracy (0.82) was
522 obtained with a between 0 and 0.09. Secondly, with a = 0, the best accuracy
523 (0.82) was obtained with $1/\lambda = 0.1$. Custom scripts were written in python
524 (v3.7.6).

525 ## Interaction site prediction

526 Values for the Boman index, a quantitative proxy for whether a peptide is more
527 likely to interact with proteins (high values) or lipids (low values), were
528 calculated as described in the literature[65]. Anchor2 interactivity values, a second
529 proxy for protein interaction potential developed for disordered sequences[66],
530 were calculated using the IUPred2A standalone application and webserver[63,67].
531 Interaction sites for Hsp70 were predicted as described by Ivey and coworkers[23]
532 based on experimental affinity values for individual amino acids. Putatively TOC-
533 interacting 'FGLK' sites were established by searching for presence of F and [P or
534 G] and [K or R] and [A or L or V], and absence of [D and E], within each 8-residue
535 window of a sequence, corresponding to rule 22 by Chotewutmontri and co-
536 workers[25] that was recommended by the authors in personal communication.
537 'FGLK-1' sites were established the same way, but requiring the presence of only
538 three of the four positive determinants. Custom scripts were implemented in R[68]
539 (v3.6.1) using RStudio[69] (v1.2.1335).

540 ## Statistical analysis

541 Chlamydomonas TP distributions (Fig. 6, Sup.Fig. 24) were compared using two-
542 sided t-tests (n=34 for mTPs and n=85 for cTPs), and associated mt-set and cp-
543 set HA-RAMPs were compared using one-sided t-tests based on the trends set by
544 TPs (n=3 for mt-set and n=5 for cp-set HA-RAMPs) in R[68] using RStudio[69].
545 Multiple Kruskal statistical tests are performed (same Chlamydomonas mTPs
546 and cTPs as above, plus n=382 HA-RAMPs) to evaluate the distribution of amino
547 acids (Fig. 4A, Sup. Fig. 19) in the different groups, followed by Dunn post-hoc
548 tests (scipy v1.4.1).

Acknowledgments

Thank you to Tiffina Benhamou and Gabriel Chemin, who worked on aspects of this study with ODC during their summer internships. We also thank Prakitchai Chotwutmontri and Barry Bruce for their help in correctly replicating their Hsp70-binding and FGLK-motif finding algorithms.

Financial support by annual funding from the Centre National de la Recherche Scientifique and Sorbonne University to UMR 7141, by the ChloroMitoRAMP ANR grant (ANR-19-CE13-0009) and by LabEx Dynamo (ANR-LABX-011), as well as by the Rothschild Foundation, is gratefully acknowledged.

Competing interest statement

The authors declare no competing interests. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Data availability statement

The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials.

Code availability statement

Custom code generated in the course of this project will be made available without restrictions upon request to the authors.

568 ## Materials availability statement

569 All constructs and an informative subset of expressing strains will be made
570 available through the Chlamydomonas resource centre.

571 ## Author contributions

572 FAW, IL and YC conceptualized the project and acquired funds. ODC devised,
573 conducted and visualized the wet-lab experimental part of the investigation, and
574 was chiefly responsible for assigning subcellular localization interpretations
575 based on microscopic evidence. CG carried out the dry-lab/bioinformatic part of
576 the investigation and associated visualization under the supervision of IL,
577 writing custom software to do so (ODC contributed to the analysis of FGLK and
578 Hsp70-binding sites). ODC wrote the original draft of the manuscript. FAW, IL,
579 ODC and YC reviewed and edited the manuscript. All authors approved the final
580 manuscript.

581 ## References

582 1. Wollman, F.-A. An antimicrobial origin of transit peptides accounts for
583 early endosymbiotic events. *Traffic* **17**, 1322–1328 (2016).

584 2. Caspari, O. D. & Lafontaine, I. The role of antimicrobial peptides in the
585 evolution of endosymbiotic protein import. *PLoS Pathog.* **17**, e1009466
586 (2021).

587 3. Lazzaro, B. P., Zasloff, M. & Rolff, J. Antimicrobial peptides: Application
588 informed by evolution. *Science* **368**, (2020).

589 4. Joo, H. S., Fu, C. I. & Otto, M. Bacterial strategies of resistance to
590 antimicrobial peptides. *Philos. Trans. R. Soc. B Biol. Sci.* **371**, (2016).

591 5. Garrido, C., Caspari, O. D., Choquet, Y., Wollman, F. A. & Lafontaine, I.
592 Evidence Supporting an Antimicrobial Origin of Targeting Peptides to
593 Endosymbiotic Organelles. *Cells* **9**, 1795 (2020).

594 6. von Heijne, G., Steppuhn, J. & Herrmann, R. G. Domain structure of
595 mitochondrial and chloroplastic targeting peptides. *Eur. J. Biochem.* **180**,
596 535–545 (1989).

597 7. von Heijne, G. Mitochondrial targeting sequences may form amphiphilic
598 helices. *EMBO J.* **5**, 1335–1342 (1986).

599 8. von Heijne, G. & Nishikawa, K. Chloroplast transit peptides. The perfect
600 random coil? *FEBS Lett.* **278**, 1–3 (1991).

601 9. Franzén, L.-G. G., Rochaix, J.-D. D. & von Heijne, G. Chloroplast transit
602 peptides from the green alga Chlamydomonas reinhardtii share features
603 with both mitochondrial and higher plant chloroplast presequences. *FEBS*
604 *Lett.* **260**, 169–172 (1990).

605 10. Lancelin, J. M. *et al.* NMR structures of a mitochondrial transit peptide from
606 the green alga Chlamydomonas reinhardtii. *FEBS Lett.* **391**, 203–208
607 (1996).

608 11. Krimm, I., Gans, P., Hernandez, J. F., Arlaud, G. J. & Lancelin, J. M. A coil-helix
609 instead of a helix-coil motif can be induced in a chloroplast transit peptide
610 from Chlamydomonas reinhardtii. *Eur. J. Biochem.* **265**, 171–180 (1999).

611 12. Wienk, H. L. J., Wechselberger, R. W., Czisch, M. & Kruijff, B. De. Structure ,
612 Dynamics , and Insertion of a Chloroplast Targeting Peptide in Mixed.
613 8219–8227 (2000).

614 13. Bruce, B. D. The paradox of plastid transit peptides: Conservation of
615 function despite divergence in primary structure. *Biochim. Biophys. Acta -*
616 *Mol. Cell Res.* **1541**, 2–21 (2001).

617 14. Bruce, B. D. The role of lipids in plastid protein transport. *Plant Mol. Biol.*
618 **38**, 223–46 (1998).

619 15. Teixeira, P. F. & Glaser, E. Processing peptidases in mitochondria and
620 chloroplasts. *Biochim. Biophys. Acta - Mol. Cell Res.* **1833**, 360–370 (2013).

621 16. Köhler, D. *et al.* Identification of protein N-termini in Cyanophora
622 paradoxa cyanelles: transit peptide composition and sequence
623 determinants for precursor maturation. *Front. Plant Sci.* **6**, 1–11 (2015).

624 17. Tardif, M. *et al.* Predalgo: A new subcellular localization prediction tool
625 dedicated to green algae. *Mol. Biol. Evol.* **29**, 3625–3639 (2012).

626 18. Bionda, T. *et al.* Chloroplast import signals: The length requirement for
627 translocation in vitro and in vivo. *J. Mol. Biol.* **402**, 510–523 (2010).

628 19. Caspari, O. D. Chloroplast transit peptides often require downstream
629 unstructured sequence in Chlamydomonas reinhardtii. *bioRxiv* (2021).
630 doi:https://www.biorxiv.org/content/10.1101/2021.11.26.470094

631 20. Chotewutmontri, P., Holbrook, K. & Bruce, B. D. Plastid Protein Targeting:
632 Preprotein Recognition and Translocation. *Int. Rev. Cell Mol. Biol.* **330**,
633 227–294 (2017).

634 21. Lee, D. W. *et al.* Molecular Mechanism of the Specificity of Protein Import
635 into Chloroplasts and Mitochondria in Plant Cells. *Mol. Plant* (2019).
636 doi:10.1016/j.molp.2019.03.003

637 22. Bhushan, S., Kuhn, C., Berglund, A., Roth, C. & Glaser, E. The role of the N-
638 terminal domain of chloroplast targeting peptides in organellar protein
639 import and miss-sorting. *FEBS* **580**, 3966–3972 (2006).

640 23. Ivey, R. a, Subramanian, C. & Bruce, B. D. Identification of a Hsp70
641 recognition domain within the rubisco small subunit transit peptide. *Plant*
642 *Physiol.* **122**, 1289–1299 (2000).

643 24. Ivey, R. a & Bruce, B. D. In vivo and in vitro interaction of DnaK and a
644 chloroplast transit peptide. *Cell Stress Chaperones* **5**, 62–71 (2000).

645 25. Chotewutmontri, P., Reddick, L. E., McWilliams, D. R., Campbell, I. M. &
646 Bruce, B. D. Differential Transit Peptide Recognition during Preprotein
647 Binding and Translocation into Flowering Plant Plastids. *Plant Cell Online*
648 **24**, 3040–3059 (2012).

649 26. Chotewutmontri, P. & Bruce, B. D. Non-native, N-terminal Hsp70 molecular
650 motor recognition elements in transit peptides support plastid protein
651 translocation. *J. Biol. Chem.* **290**, 7602–7621 (2015).

652 27. Rothbard, J. B., Jessop, T. C., Lewis, R. S., Murray, B. A. & Wender, P. A. Role
653 of membrane potential and hydrogen bonding in the mechanism of
654 translocation of guanidinium-rich peptides into cells. *J. Am. Chem. Soc.* **126**,
655 9506–9507 (2004).

656 28. Mitchell, D. J., Kim, D. T., Steinman, L., Fathman, C. G. & Rothbard, J. B.
657 Polyarginine enters cells more efficiently than other polycationic
658 homopolymers. *J. Pept. Res.* **56**, 318–25 (2000).

659 29. Li, L., Vorobyov, I. & Allen, T. W. The different interactions of lysine and
660 arginine side chains with lipid membranes. *J. Phys. Chem. B* **117**, 11906–
661 11920 (2013).

662 30. Li, J. *et al.* Membrane Active Antimicrobial Peptides: Translating
663 Mechanistic Insights to Design. *Front. Neurosci.* **11**, 1–18 (2017).

664 31. Mattiroli, F. & Sixma, T. K. Lysine-targeting specificity in ubiquitin and
665 ubiquitin-like modification pathways. *Nat. Struct. Mol. Biol.* **21**, 308–316
666 (2014).

667 32. Kmiec, B., Teixeira, P. F. & Glaser, E. Shredding the signal: targeting peptide
668 degradation in mitochondria and chloroplasts. *Trends Plant Sci.* **19**, 771–8
669 (2014).

670 33. Lee, D. W. *et al.* Cross-Species Functional Conservation and Possible Origin
671 of the N-Terminal Specificity Domain of Mitochondrial Presequences.
672 *Front. Plant Sci.* **11**, 1–7 (2020).

673 34. Garg, S. G. & Gould, S. B. The Role of Charge in Protein Targeting Evolution.
674 *Trends Cell Biol.* **26**, 894–905 (2016).

675 35. Martin, J., Mahlke, K. & Pfanner, N. Role of an energized inner membrane in
676 mitochondrial protein import: ΔΨ drives the movement of presequences. *J.*
677 *Biol. Chem.* **266**, 18051–18057 (1991).

678 36. Nakai, M. New Perspectives on Chloroplast Protein Import. *Plant Cell*
679 *Physiol.* **59**, 1111–1119 (2018).

680 37. Rekas, A., Alattia, J. R., Nagai, T., Miyawaki, A. & Ikura, M. Crystal structure
681 of venus, a yellow fluorescent protein with improved maturation and
682 reduced environmental sensitivity. *J. Biol. Chem.* **277**, 50573–50578
683 (2002).

684 38. Knopp, M., Garg, S. G., Handrich, M. & Gould, S. B. Major Changes in Plastid
685 Protein Import and the Origin of the Chloroplastida. *iScience* **23**, 100896
686 (2020).

687 39. Sommer, M. S. *et al.* Chloroplast Omp85 proteins change orientation during
688 evolution. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 13841–13846 (2011).

689 40. Wunder, T., Martin, R., Löffelhardt, W., Schleiff, E. & Steiner, J. M. The
690 invariant phenylalanine of precursor proteins discloses the importance of
691 Omp85 for protein translocation into cyanelles. *BMC Evol. Biol.* **7**, 236
692 (2007).

693 41. Robert, V. *et al.* Assembly factor Omp85 recognizes its outer membrane
694 protein substrates by a species-specific C-terminal motif. *PLoS Biol.* **4**,
695 1984–1995 (2006).

696 42. Holbrook, K. *et al.* Functional Analysis of Semi-conserved Transit Peptide
697 Motifs and Mechanistic Implications in Precursor Targeting and
698 Recognition. *Mol. Plant* **9**, 1286–1301 (2016).

699 43. Pilon, M. *et al.* Functional domains of the ferredoxin transit sequence
700 involved in chloroplast import. *Journal of Biological Chemistry* **270**, 3882–
701 3893 (1995).

702 44. Lee, D. W., Lee, S., Oh, Y. J. & Hwang, I. Multiple sequence motifs in the
703 rubisco small subunit transit peptide independently contribute to Toc159-
704 dependent import of proteins into chloroplasts. *Plant Physiol.* **151**, 129–
705 141 (2009).

706 45. Razzak, A., Lee, D. W., Yoo, Y. & Hwang, I. Evolution of rubisco complex
707 small subunit transit peptides from algae to plants. *Sci. Rep.* **7**, 9279
708 (2017).

709 46. Hill, K. *et al.* Tom40 forms the hydrophilic channel of the mitochondrial
710 import pore for preproteins. *Nature* **395**, 516–521 (1998).

711 47. Hinnah, S. C., Hill, K., Wagner, R., Schlicher, T. & Soll, J. Reconstitution of a
712 chloroplast protein import channel. *EMBO J.* **16**, 7351–7360 (1997).

713 48. Hinnah, S. C., Wagner, R., Sveshnikova, N., Harrer, R. & Soll, J. The
714 chloroplast protein import channel Toc75: pore properties and interaction
715 with transit peptides. *Biophys. J.* **83**, 899–911 (2002).

716 49. Duncan, O., Murcha, M. W. & Whelan, J. Unique components of the plant
717 mitochondrial protein import apparatus. *Biochim. Biophys. Acta - Mol. Cell*
718 *Res.* **1833**, 304–313 (2013).

719 50. Schneider, A. Evolution of mitochondrial protein import - Lessons from
720 trypanosomes. *Biol. Chem.* **401**, 663–676 (2020).

721 51. Steiner, J. M. & Löffelhardt, W. Protein translocation into and within
722 cyanelles. *Mol. Membr. Biol.* **22**, 123–132 (2005).

723 52. Atkinson, N. *et al.* Introducing an algal carbon-concentrating mechanism
724 into higher plants: Location and incorporation of key components. *Plant*
725 *Biotechnol. J.* **14**, 1302–1315 (2016).

726 53. Steiner, J. M., Yusa, F., Pompe, J. A. & Löffelhardt, W. Homologous protein
727 import machineries in chloroplasts and cyanelles. *Plant J.* **44**, 646–652
728 (2005).

729 54. Mani, J., Meisinger, C. & Schneider, A. Peeping at TOMs - Diverse entry
730 gates to mitochondria provide insights into the evolution of eukaryotes.
731 *Mol. Biol. Evol.* **33**, 337–351 (2016).

732 55. Singer, A. *et al.* Massive Protein Import into the Early-Evolutionary-Stage
733 Photosynthetic Organelle of the Amoeba Paulinella chromatophora. *Curr.*
734 *Biol.* **27**, 2763–2773 (2017).

735 56. Pinnaduwage, P. & Bruce, B. D. In vitro interaction between a chloroplast
736 transit peptide and chloroplast outer envelope lipids is sequence-specific
737 and lipid class-dependent. *J. Biol. Chem.* **271**, 32907–32915 (1996).

738 57. Caspari, O. D. Introduction of a leaky stop codon as molecular tool in
739 Chlamydomonas reinhardtii. *PLoS One* **15**, e0237405 (2020).

740 58. Onishi, M. & Pringle, J. R. Robust Transgene Expression from Bicistronic
741 mRNA in the Green Alga *Chlamydomonas reinhardtii*. *G3 (Bethesda).* **6**,
742 4115–4125 (2016).

743 59. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image
744 analysis. *Nat. Methods* **9**, 676–682 (2012).

745 60. Gautier, R., Douguet, D., Antonny, B. & Drin, G. HELIQUEST: a web server to

746 screen sequences with specific alpha-helical properties. *Bioinformatics* **24**,
747 2101–2102 (2008).

748 61. Hellberg, S., Sjostrom, M., Skagerberg, B. & Wold, S. Peptide Quantitative
749 Structure-Activity Relationships, a Multivariate Approach. *J. Med. Chem.*
750 **30**, 1126–1135 (1987).

751 62. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach.*
752 *Learn. Res.* **12**, 2825–2830 (2011).

753 63. Erdős, G. & Dosztányi, Z. Analyzing Protein Disorder with IUPred2A. *Curr.*
754 *Protoc. Bioinforma.* **70**, 1–15 (2020).

755 64. Nielsen, J. T. & Mulder, F. A. A. Quality and bias of protein disorder
756 predictors. *Sci. Rep.* **9**, 1–11 (2019).

757 65. Boman, H. G. Antibacterial peptides: basic facts and emerging concepts. *J.*
758 *Intern. Med.* **254**, 197–215 (2003).

759 66. Mészáros, B., Simon, I. & Dosztányi, Z. Prediction of Protein Binding
760 Regions in Disordered Proteins. *PLoS One* **5**, e1000376 (2009).

761 67. Mészáros, B., Erdös, G. & Dosztányi, Z. IUPred2A: Context-dependent
762 prediction of protein disorder as a function of redox state and protein
763 binding. *Nucleic Acids Res.* **46**, W329–W337 (2018).

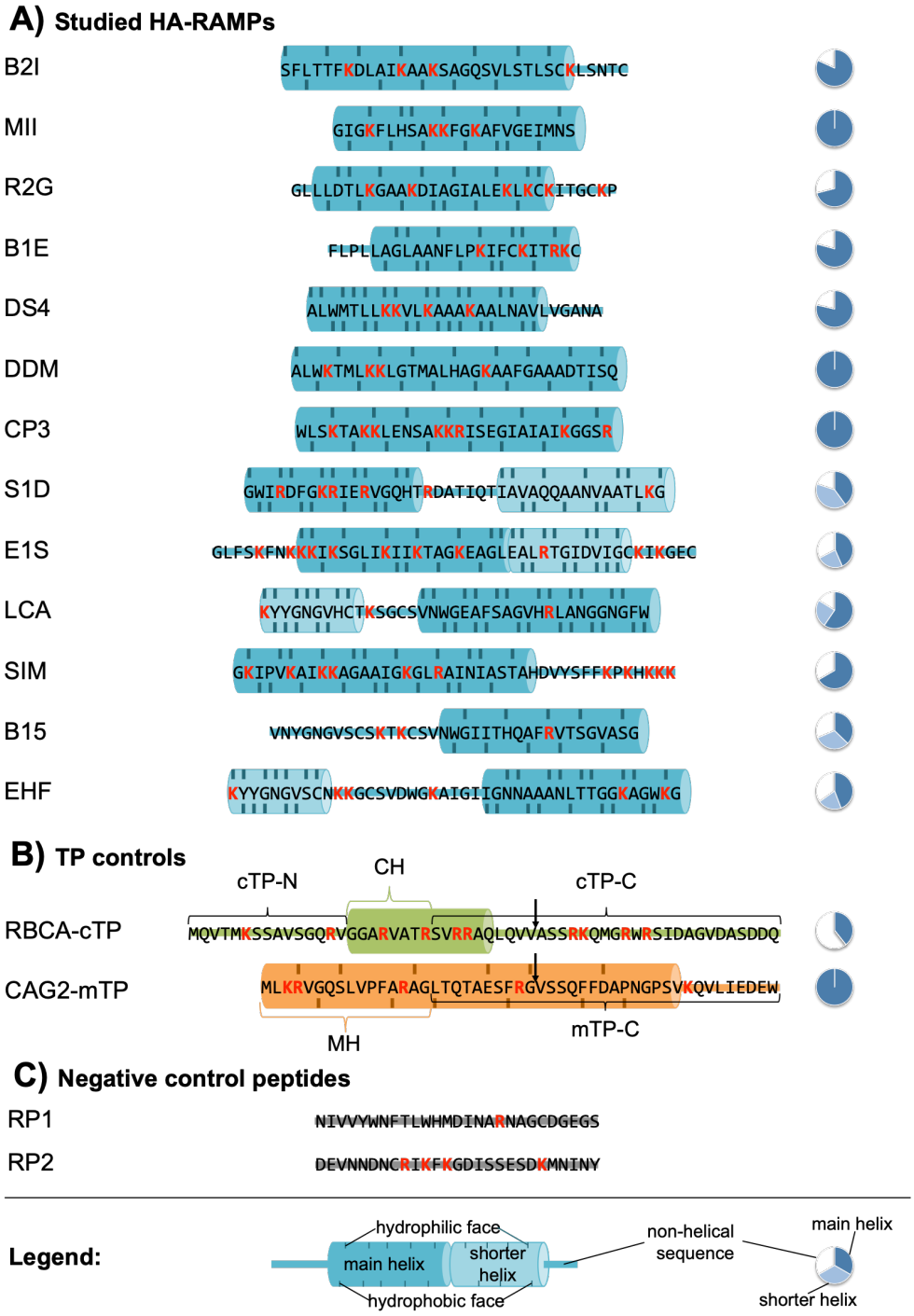
764 68. R Core Team. R: A Language and Environment for Statistical Computing.
765 (2019).

766 69. RStudio Team. RStudio: Integrated Development for R. (2020).

767 70. Meyer, M. *et al.* Assembly of the algal CO2-fixing organelle, the pyrenoid, is
768 guided by a Rubisco-binding motif. *Sci. Adv.* **6**, eabd2408 (2020).

769

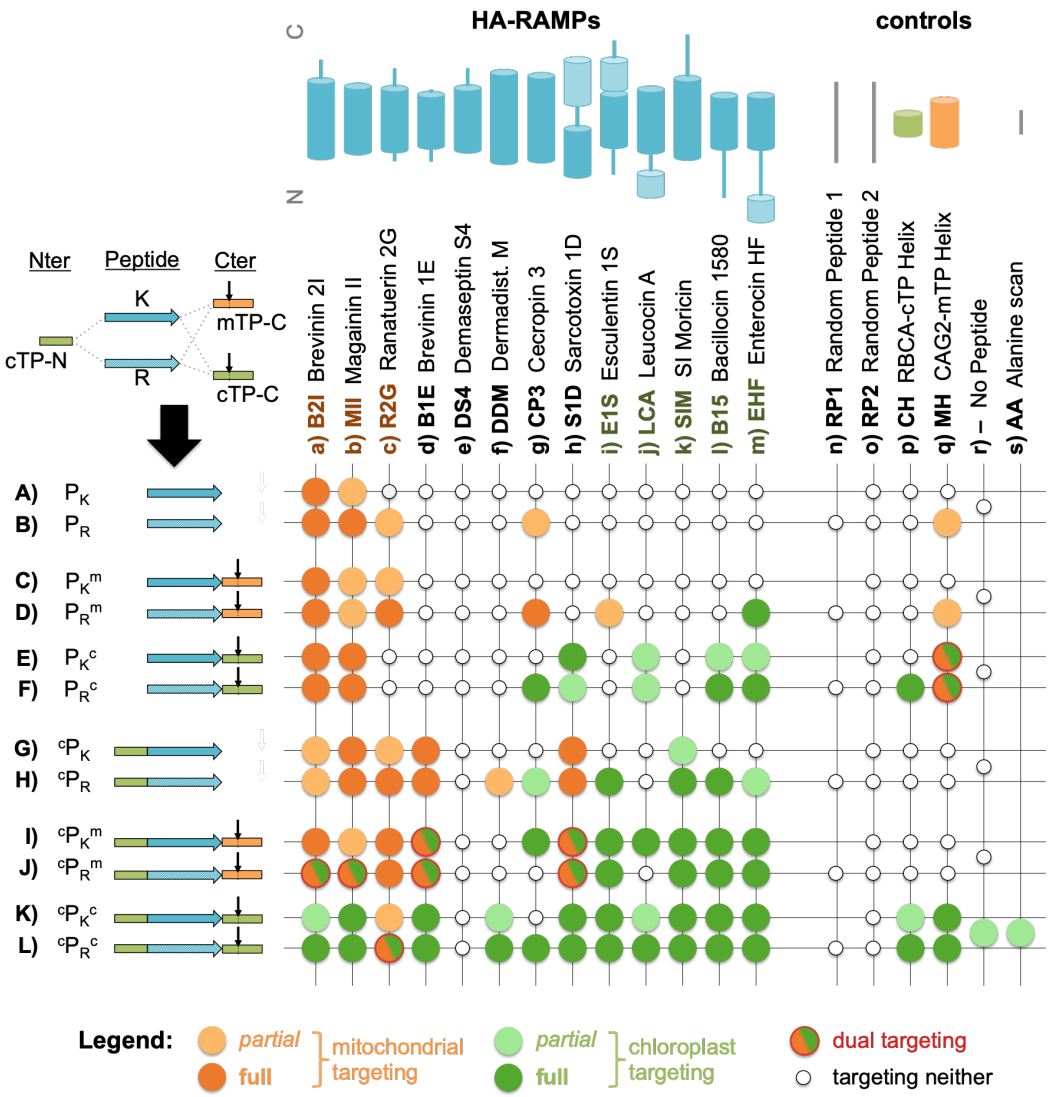
770 ## Figure captions



771

772 **Figure 1. Peptide sequences under study.** Amino acid sequences are shown
773 using the one-letter code. Positively charged residues are highlighted in red. The
774 fraction of the sequence predicted to fold into amphipathic helices (see Methods)
775 is provided by a pie-chart to the right of the sequence; for TPs in (B) this was

776 calculated up to the cleavage site indicated by a downward arrow. Predicted
777 amphipathic helices are highlighted using a cylinder cartoon, with residues
778 contributing to the hydrophilic/hydrophobic face indicated on the top/bottom.
779 No helix could be predicted within RBCA-cTP, and thus the indicated helix is
780 taken from a published NMR-structure obtained in membrane-mimetic
781 conditions[11]. Note that the two helices of E1S are at an angle to each other and
782 therefore cannot form a single continuous amphipathic helix. Abbreviations: **A)**
783 B2I: Brevinin 2ISb, R2G: Ranatuerin 2G, MII: Magainin 2, B1E: Brevinin 1E, DS4:
784 Dermaseptin S4, DDM: Dermadistinctin M, CP3: Cecropin P3, S1D: Sarcotoxin 1D,
785 E1S: Esculentin 1S, LCA: Leucocin A, SIM: SI Moricin, B15: Bacillocin 1580, EHF:
786 Enterocin HF; **B)** TP – targeting peptide, cTP: chloroplast TP, mTP: mitochondrial
787 TP, RBCA: Rubisco activase, CAG2: γ-carbonic anhydrase, cTP-N: cTP N-terminal
788 element, CH: cTP helix, MH: mTP helix, cTP-C: cTP C-terminal element, mTP-C:
789 mTP C-terminal element; **C)** RP: random peptide.



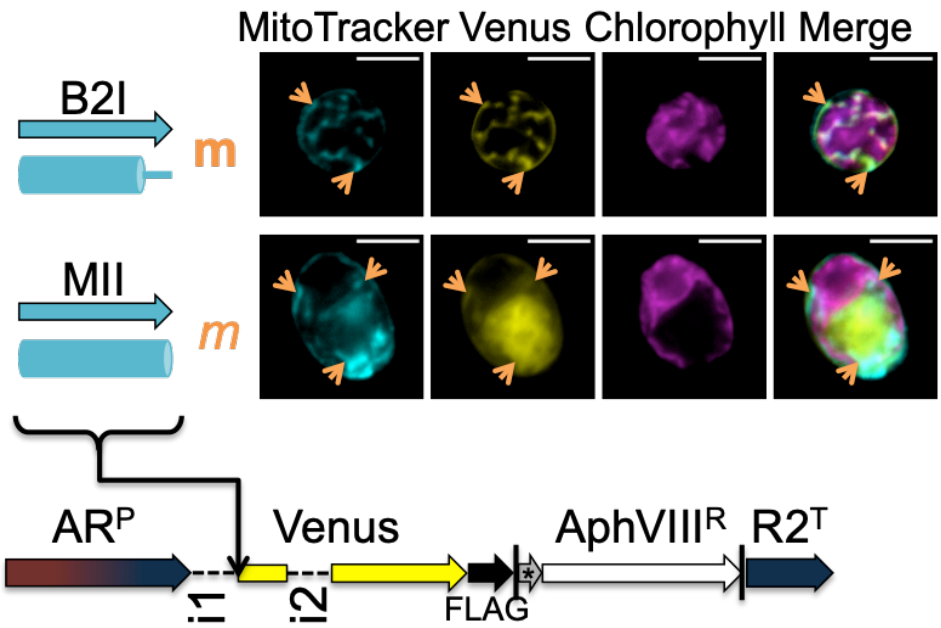
790

791 **Figure 2. TP modifications enable HA-RAMP targeting.** Chimeric constructs
792 were generated by combining TP elements with HA-RAMPs (see Fig. 1 for
793 sequences; in the $^{c}AA^{c}$ construct in column s, all residues of the helical element
794 "CH" within RBCA-cTP are replaced by Alanines). The overview graph shows in
795 each column (a-s) one of the peptides with a cartoon indicating the position of
796 the predicted amphipathic helices within the sequence; and in each row (A-L) a
797 combination of peptide, K/R modification and TP-element indicated by a cartoon
798 and the following shorthand: P = peptide, $_{K}$ = contains mostly Lysines, $_{R}$ =
799 contains mostly Arginines, ^{m} = mTP element, ^{c} = cTP element (^{c}P = cTP-N, P^{c} =
800 cTP-C). In each case, an overview of observed targeting is provided by a colour

801 code (see figure legend). Images for all constructs are shown in Supplementary

802 Fig. 1-18. Note that the present results for B2I$_{K^c}$, MII$_{K^c}$, S1D$_{K^c}$, B15$_{K^c}$ and EHF$_{K^c}$

803 (row E, columns a,b,h,l,m) confirmed our previous report on these strains[5].



804

805 **Figure 3. Unmodified HA-RAMPs can target the mitochondria.** Indicated

806 peptides were inserted upstream of a Venus fluorescent protein reporter in a

807 bicistronic expression system for Chlamydomonas[57]. AR^P: hybrid *HSP70a-RBCS2*

808 promoter, i1: *RBCS2* intron 1, i2: *RBCS2* intron 2, FLAG: FLAG-tag, |: stop codon,

809 *: bicistronic bridge sequence tagcat, AphVIII^R: Paromomycin resistance gene,

810 R2^T: *RBCS2* terminator. Epifluorescence microscopy images of selected examples

811 are shown. False-coloured yellow fluorescence from the Venus channel reports

812 on the subcellular localization of the fluorescent reporter. MitoTracker

813 fluorescence, false-coloured in cyan, indicates the position of mitochondria

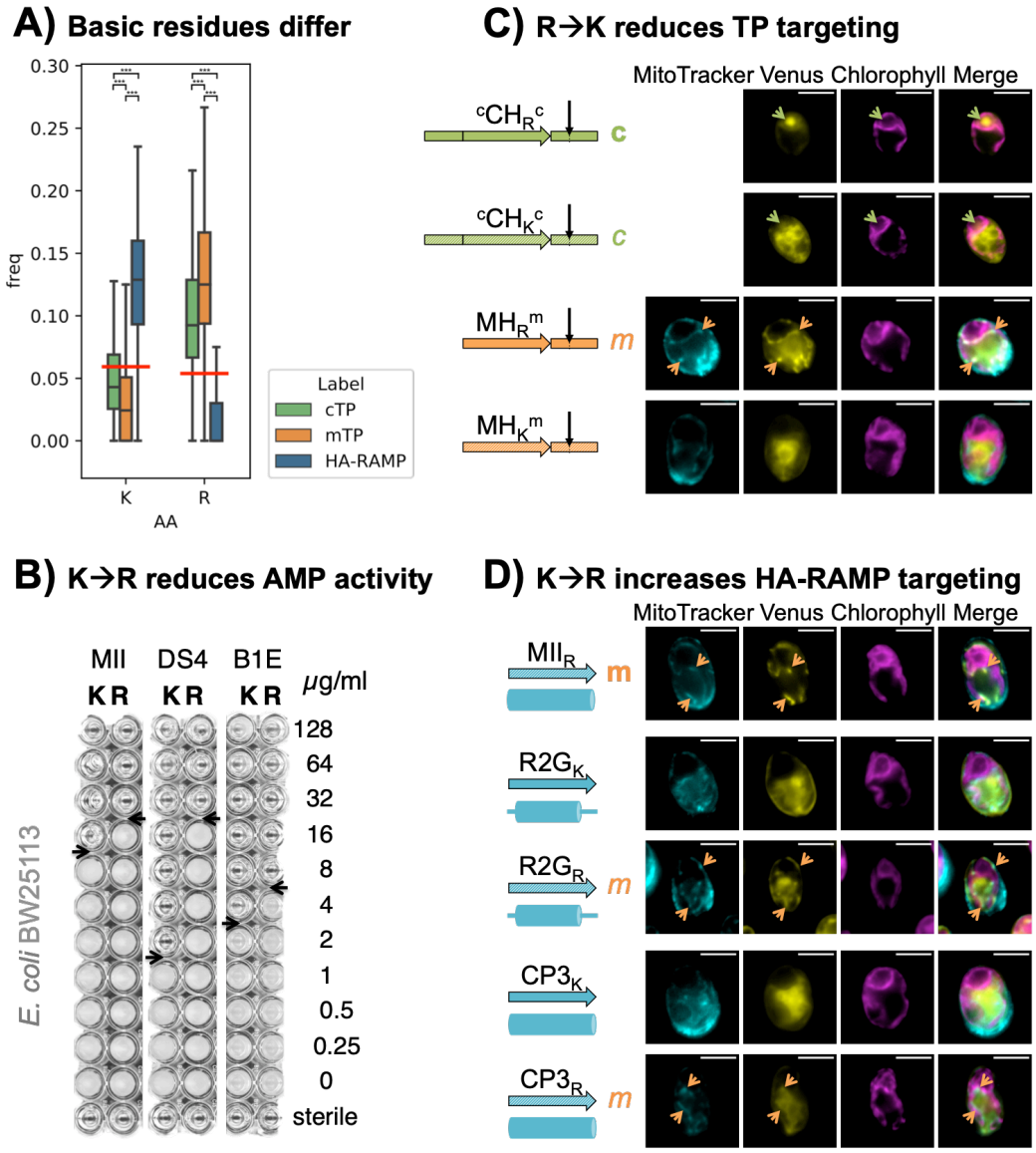
814 (although parts of the cell exterior are sometimes also stained), with salient

815 features highlighted with orange arrows to indicate co-localisation with the

816 Venus channel. Chlorophyll autofluorescence, shown in magenta, indicates the

817 location of the chloroplast. Scale bars are 5µm. Refer to Fig. 1 for sequences and

818 Sup. Fig. 1 and 2 for biological replicates. Where a construct was interpreted as
819 generating reporter localization in mitochondria or chloroplast, this is indicated
820 by an orange 'm' or a green 'c' respectively, in bold for full targeting or in italics
821 for partial targeting.



822

823 **Figure 4. K is for killing, R is for targeting. A)** Lysine (K) and Arginine (R)
824 frequencies for Chlamydomonas TPs and HA-RAMPs are shown as boxplots
825 (center line: median; box limits: upper and lower quartiles; whiskers: min/max
826 values within 1.5x interquartile range). To give a baseline for comparison, their
827 average across the Uniprot database is given as red horizontal line. Statistically

828 significant differences are indicated with stars: (Multiple Kruskal-Wallis tests
829 followed by Dunn post-hoc tests, ***p<0.0001). See Sup. Fig. 19 for all amino
830 acids. **B)** Grayscale photos of antimicrobial activity growth assays in the
831 presence of dilutions of three selected antimicrobial peptides against *E. coli*
832 strain BW25113 show reduced activity when natively K-rich sequences (K, left
833 hand columns) are altered by replacing all Ks by Rs (R, right hand columns). **C**
834 **and D)** Epifluorescence images of selected chimeric TP control (C) or HA-RAMP
835 (D) constructs are shown as in Fig. 3. MH: mTP helical fragment, CH: cTP helical
836 fragment. Orange arrows highlight co-localisation of Venus signal with the
837 MitoTracker dye, demonstrating mitochondrial localisation; green arrows
838 highlight the presence of Venus signal in the pyrenoid, demonstrating
839 chloroplast localization. Where a construct was interpreted as generating
840 reporter localization in mitochondria or chloroplast, this is indicated by an
841 orange 'm' or a green 'c' respectively.

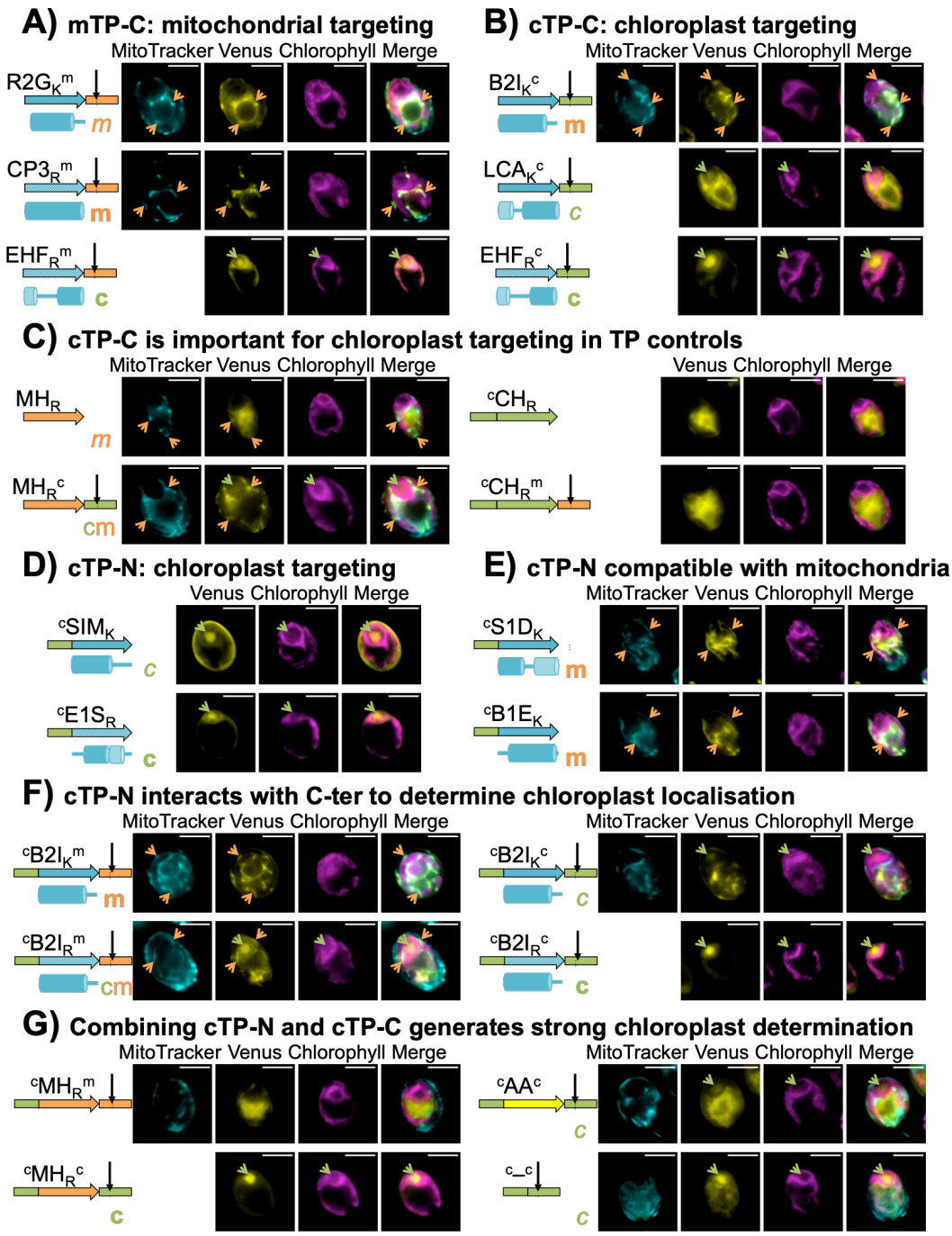


Figure 5. TP N- and C-termini enable chloroplast targeting by HA-RAMPs. Exemplary epifluorescence images are shown as in Fig. 3; same convention as in Fig. 4 for Venus localisation. Note that cTP-C contains a pyrenoid localization motif[70], explaining the particularly pronounced Venus accumulation in the pyrenoid in constructs carrying this element. The result for B2I$_K^c$ in (B) confirms our previous report on this construct[5].

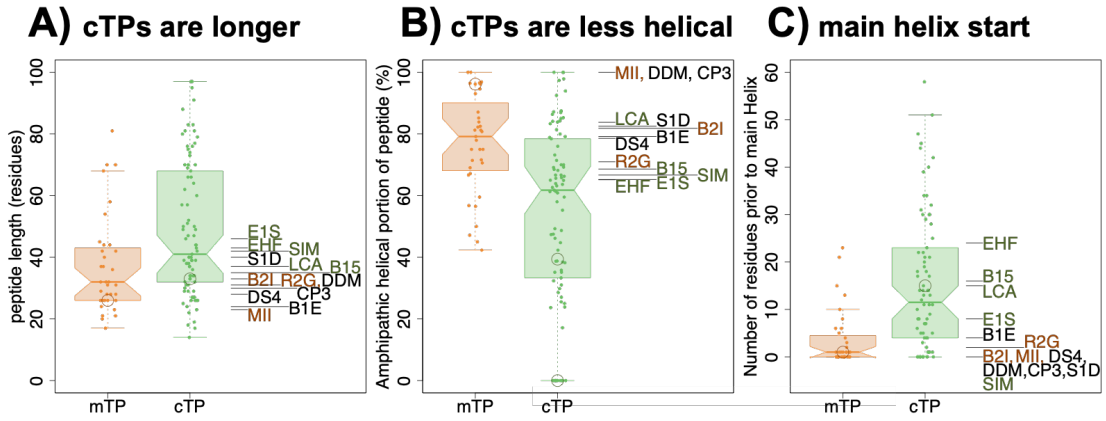


Figure 7. Chloroplast and mitochondrial targeting HA-RAMPs match cTPs and mTPs respectively. For salient properties, Chlamydomonas mTPs and cTPs are compared to our 13 HA-RAMPs. TP distributions are shown as boxplots (center line: median; box limits: upper and lower quartiles; whiskers: min/max values within 1.5x interquartile range), coloured points represent individual peptides. The position of the CAG2-mTP and RBCA-cTP is circled in each graph. The non-zero value for RBCA-cTP in (B) and the single circle in (C) report on the amphipathic helix established by an NMR-study[11], 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.

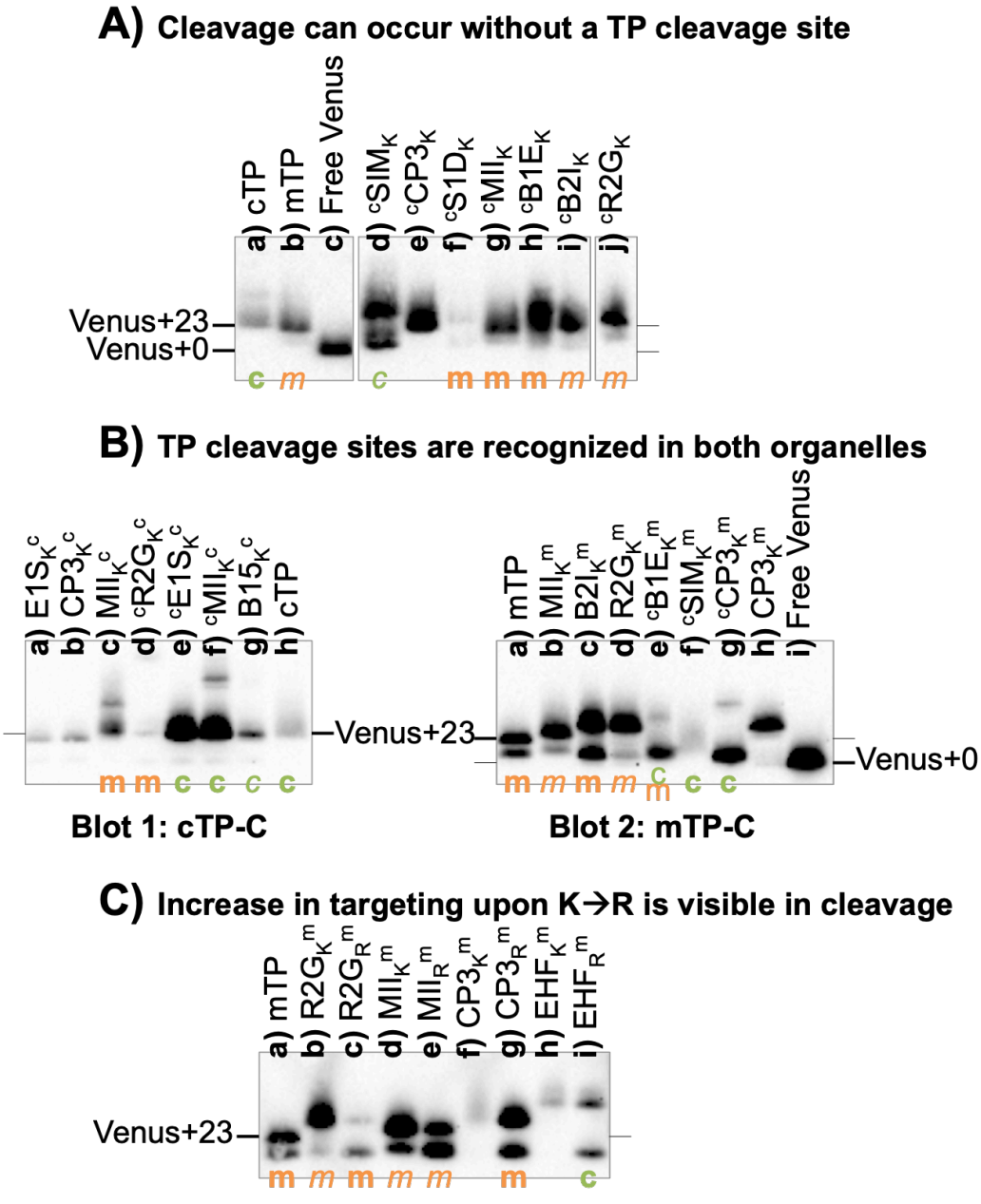
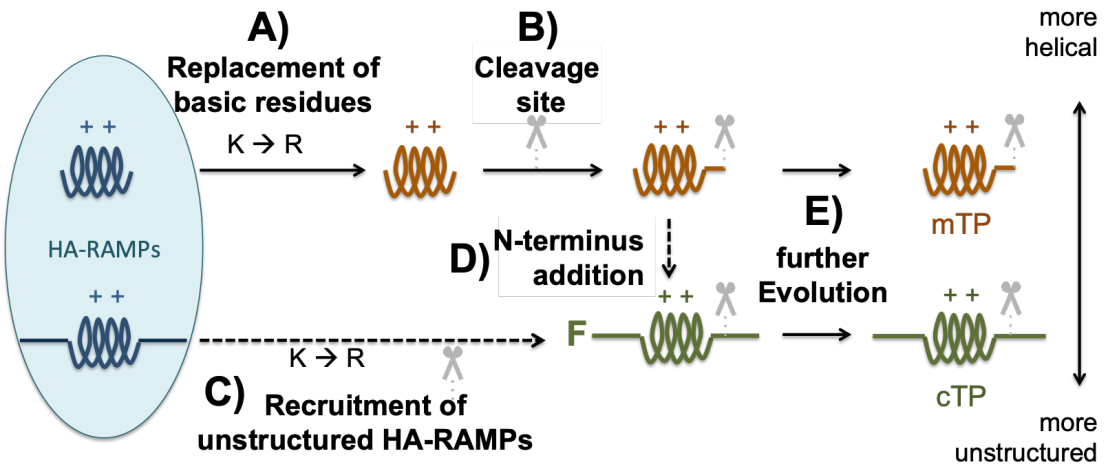


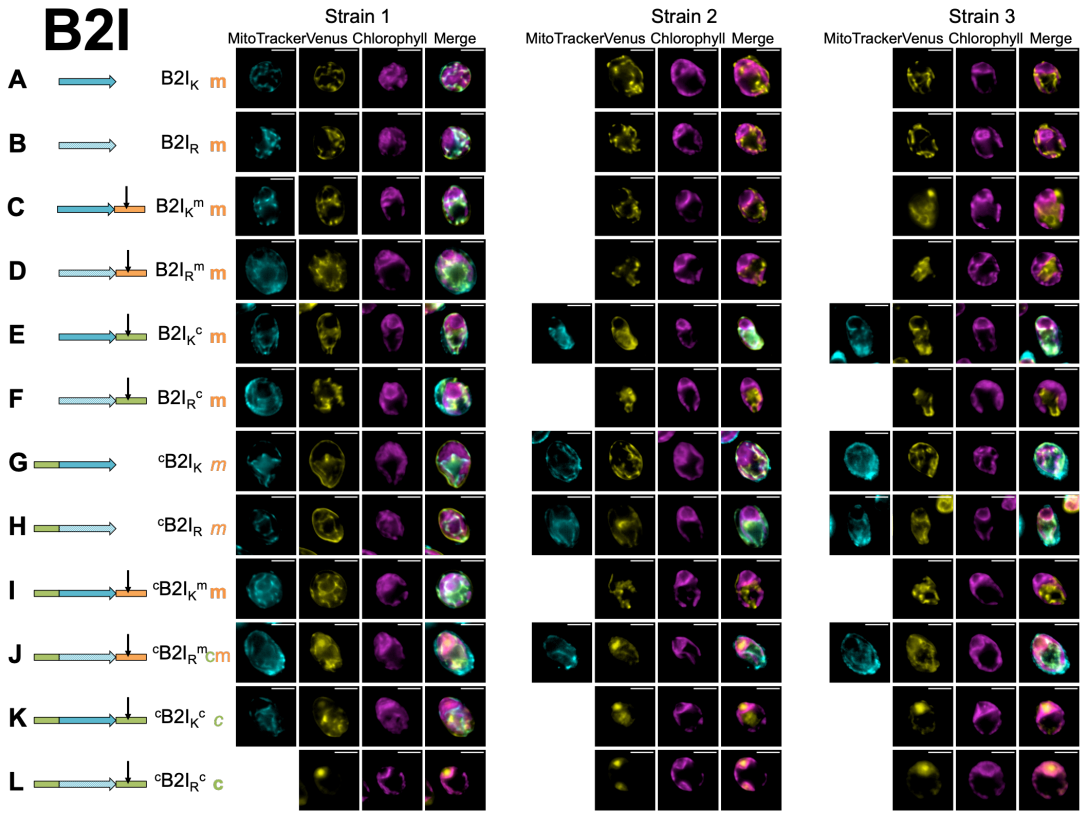
Figure 8. Import is associated with maturation of the preprotein. Western Blots used an α-FLAG antibody on selected constructs as indicated above blots. Where a construct was interpreted as generating reporter localization by fluorescence microscopy (cf. Fig. 2, Sup. Fig. 1-18) in mitochondria or chloroplast, 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 Sup. Fig. 26.



876

877 **Figure 9. A proposed progression for the evolution of targeting peptides**
878 **from antimicrobial origins. A)** HA-RAMPs would have first changed K to R and
879 **B)** acquired a cleavage site to become mTPs. To generate cTPs, either **C)** more
880 unstructured HA-RAMPs were recruited directly to become cTPs by undergoing
881 a K to R shift and acquiring a cleavage site separately, or **D)** mTPs acquired an N-
882 terminal non-helical domain. Early cTP N-termini likely interacted with TOC75
883 POTRA domains via a motif starting with F[39,40]. **E)** Further evolution would have
884 reinforced the differences between cTPs and mTPs to limit mis-targeting.

885



886

887 **Supplementary Fig. 1. Microscopy for biological replicates of Brevinin 2ISb.**

888 Three independent transformant lines (Strains 1-3) are shown for each

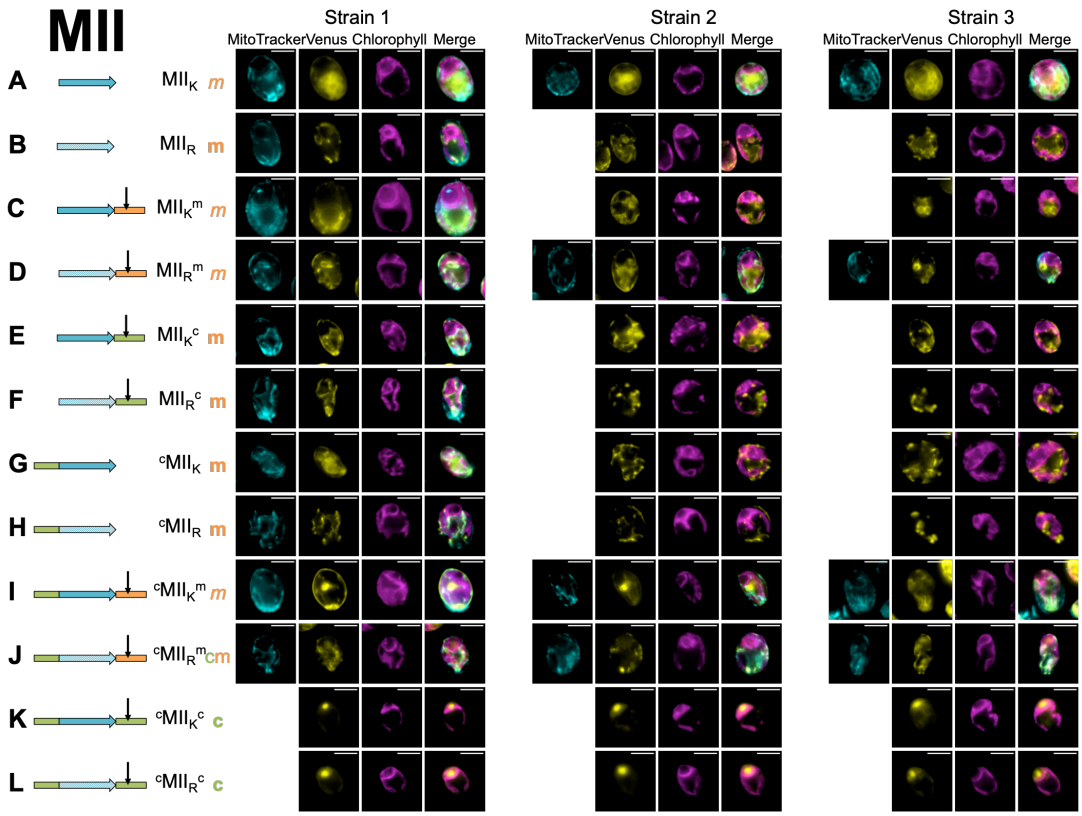
889 combination of modifications (A-L), represented by a cartoon and a shorthand

890 description (cf. Fig. 2). Where a construct was interpreted as generating reporter

891 localization in mitochondria or chloroplast, this is indicated by an orange 'm' or a

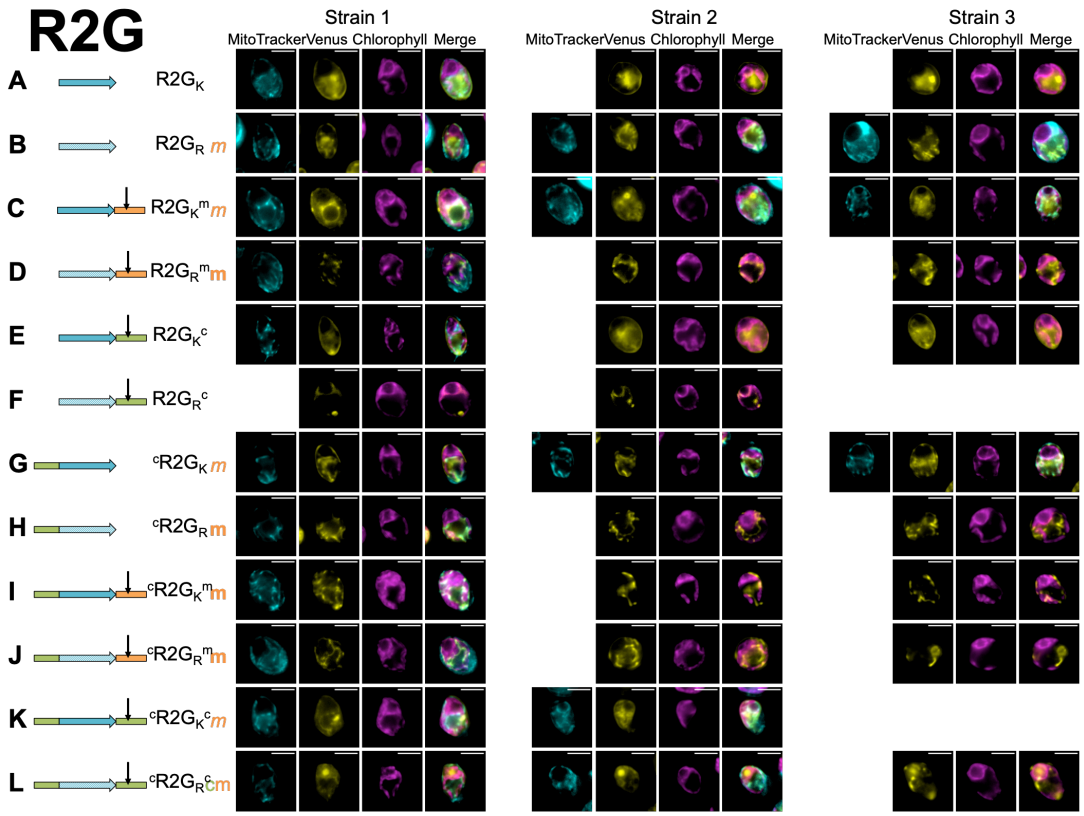
892 green 'c' respectively, in bold for full targeting or in italics for partial targeting.

893 Scale bars are 5μm.

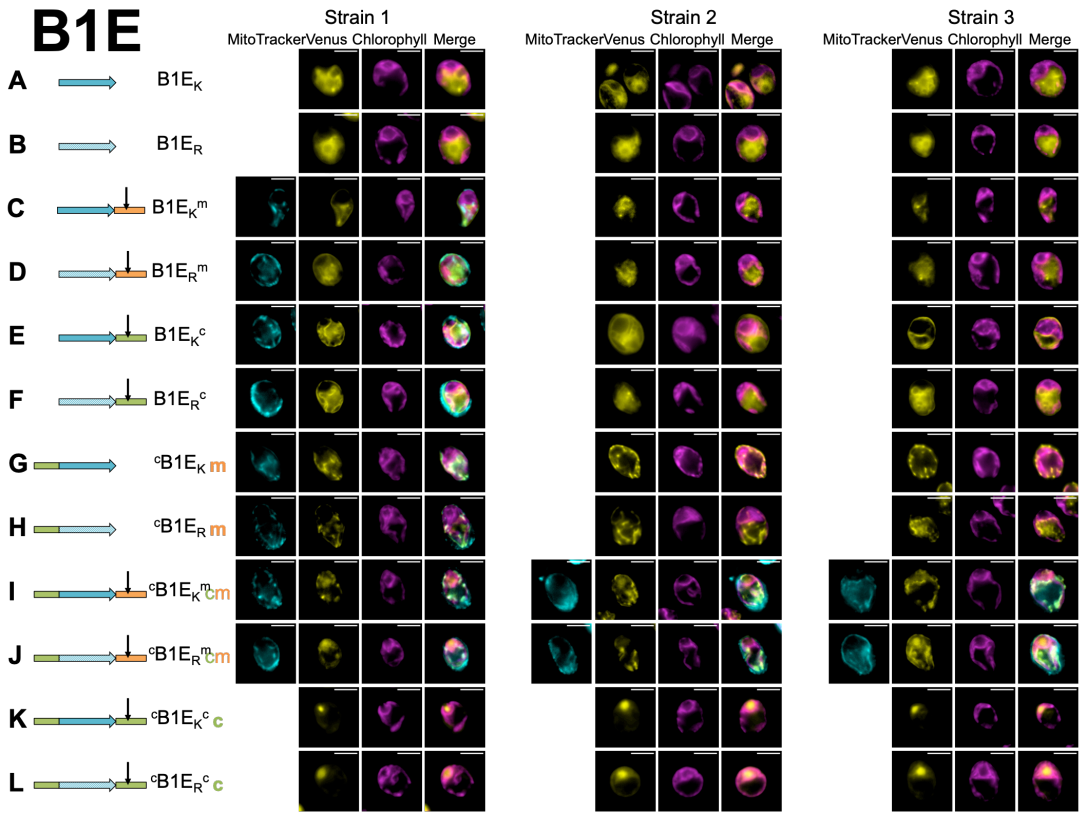


894

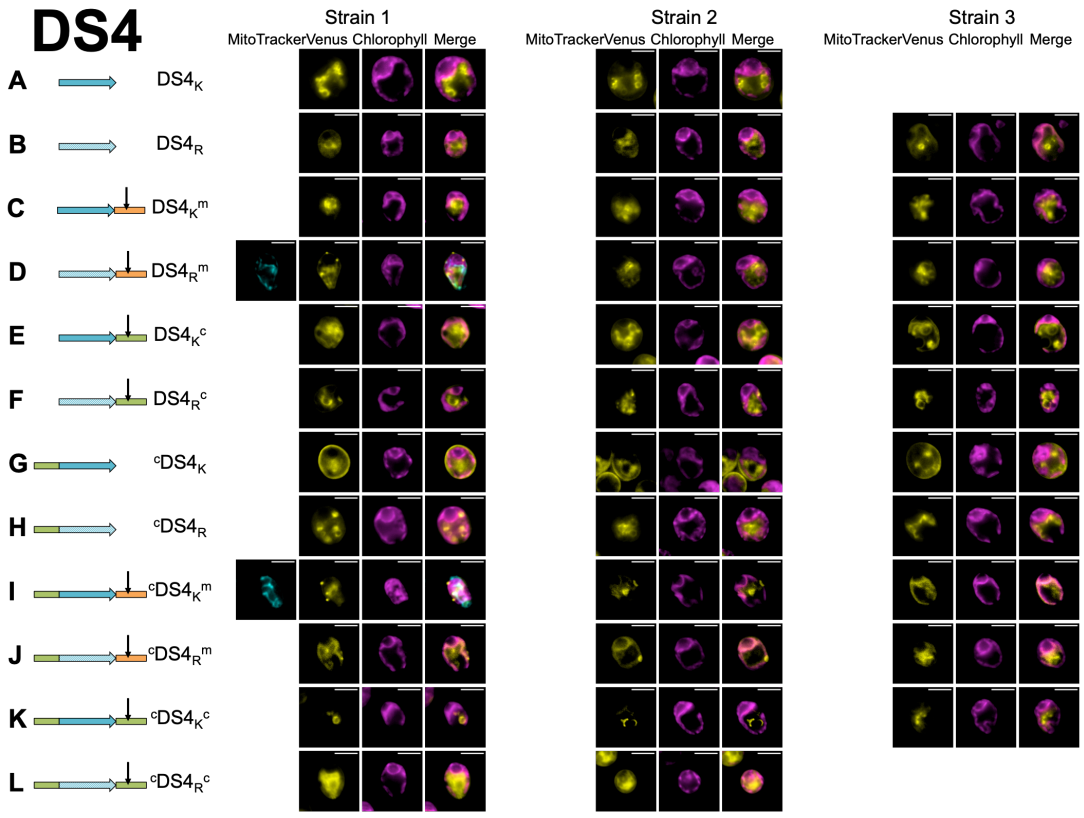
895 **Supplementary Fig. 2. Microscopy for biological replicates of Magainin 2.**
896 Three independent transformant lines (Strains 1-3) are shown for each
897 combination of modifications (A-L), represented by a cartoon and a shorthand
898 description (cf. Fig. 2). Where a construct was interpreted as generating reporter
899 localization in mitochondria or chloroplast, this is indicated by an orange 'm' or a
900 green 'c' respectively, in bold for full targeting or in italics for partial targeting.
901 Scale bars are 5µm.



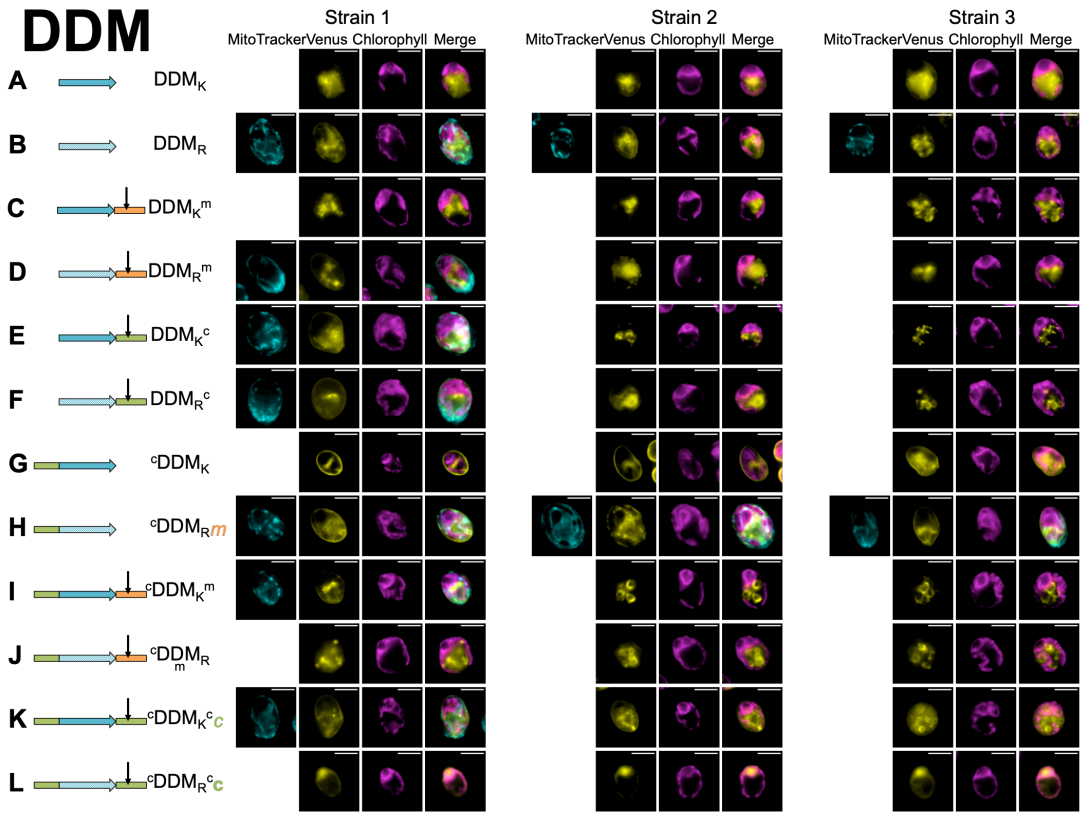
Supplementary Fig. 3. Microscopy for biological replicates of Ranatuerin 2G. Three independent transformant lines (Strains 1-3) are shown for each combination of modifications (A-L), represented by a cartoon and a shorthand description (cf. Fig. 2). Where a construct was interpreted as generating reporter localization 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. Scale bars are 5µm.



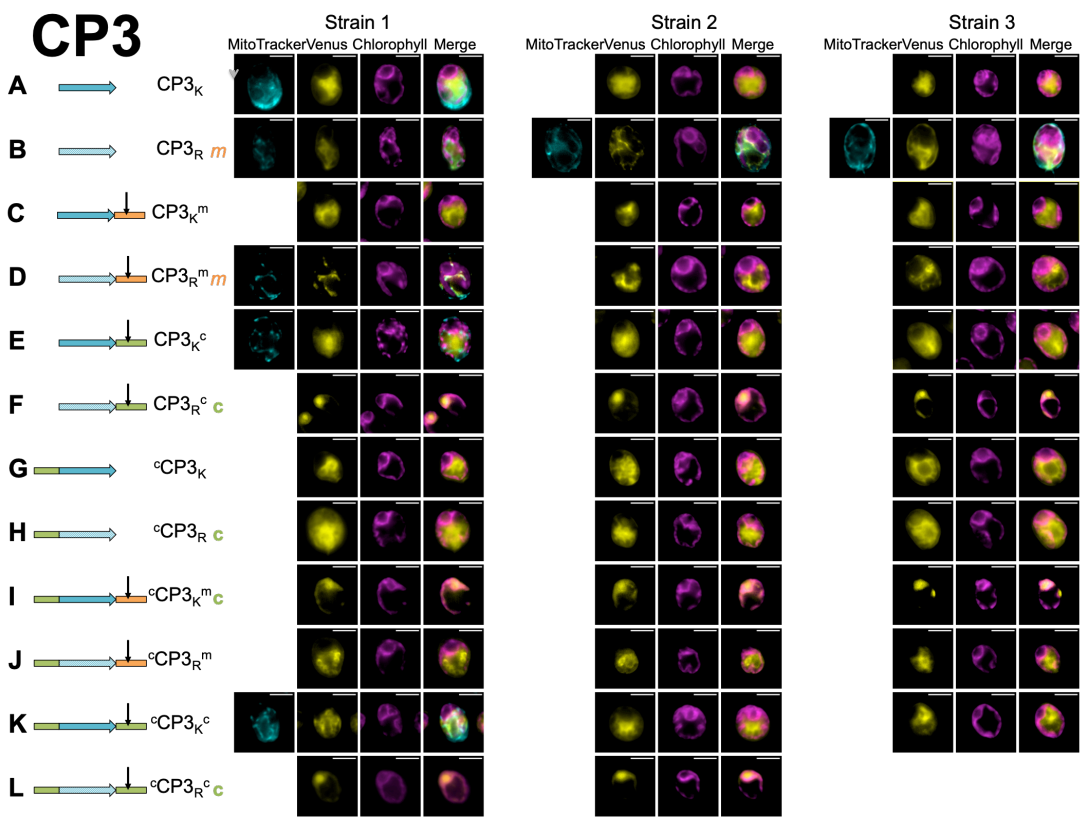
Supplementary Fig. 4. Microscopy for biological replicates of Brevinin 1E. Three independent transformant lines (Strains 1-3) are shown for each combination of modifications (A-L), represented by a cartoon and a shorthand description (cf. Fig. 2). Where a construct was interpreted as generating reporter localization 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. Scale bars are 5µm.



Supplementary Fig. 5. Microscopy for biological replicates of Dermaseptin S4. Three independent transformant lines (Strains 1-3) are shown for each combination of modifications (A-L), represented by a cartoon and a shorthand description (cf. Fig. 2). Where a construct was interpreted as generating reporter localization 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. Scale bars are 5µm.

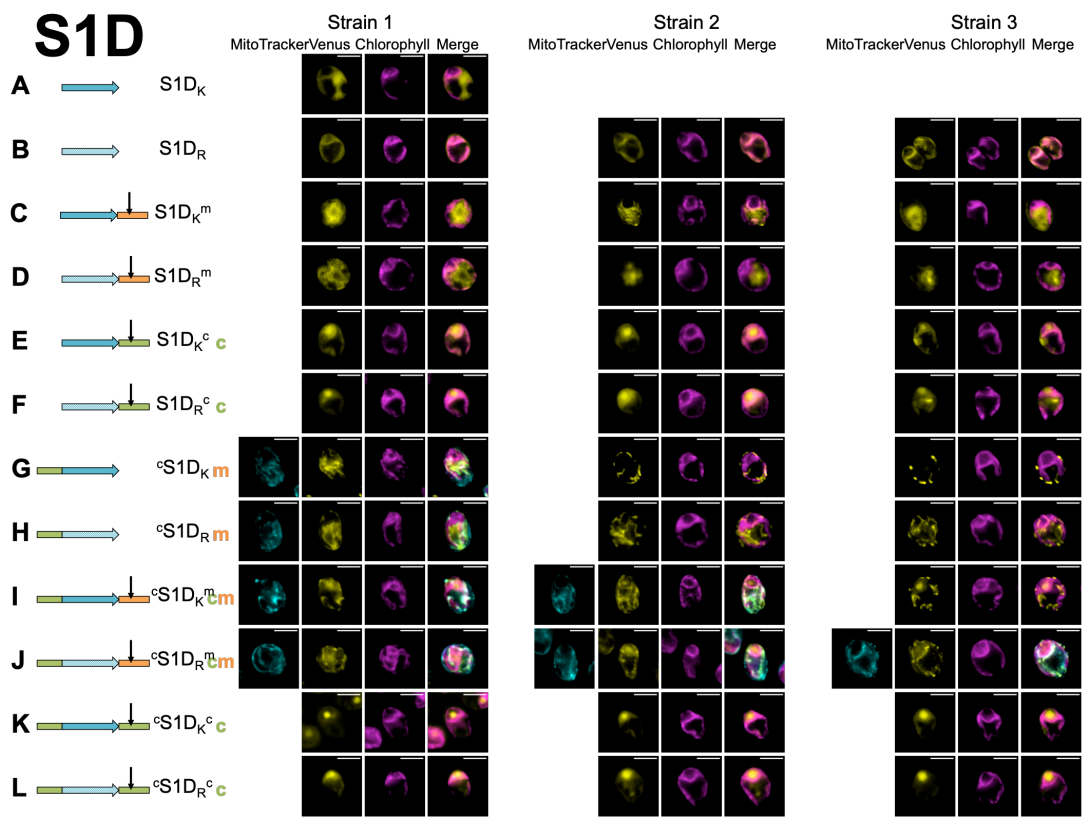


Supplementary Fig. 6. Microscopy for biological replicates of Dermadistinctin M. Three independent transformant lines (Strains 1-3) are shown for each combination of modifications (A-L), represented by a cartoon and a shorthand description (cf. Fig. 2). Where a construct was interpreted as generating reporter localization 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. Scale bars are 5µm.

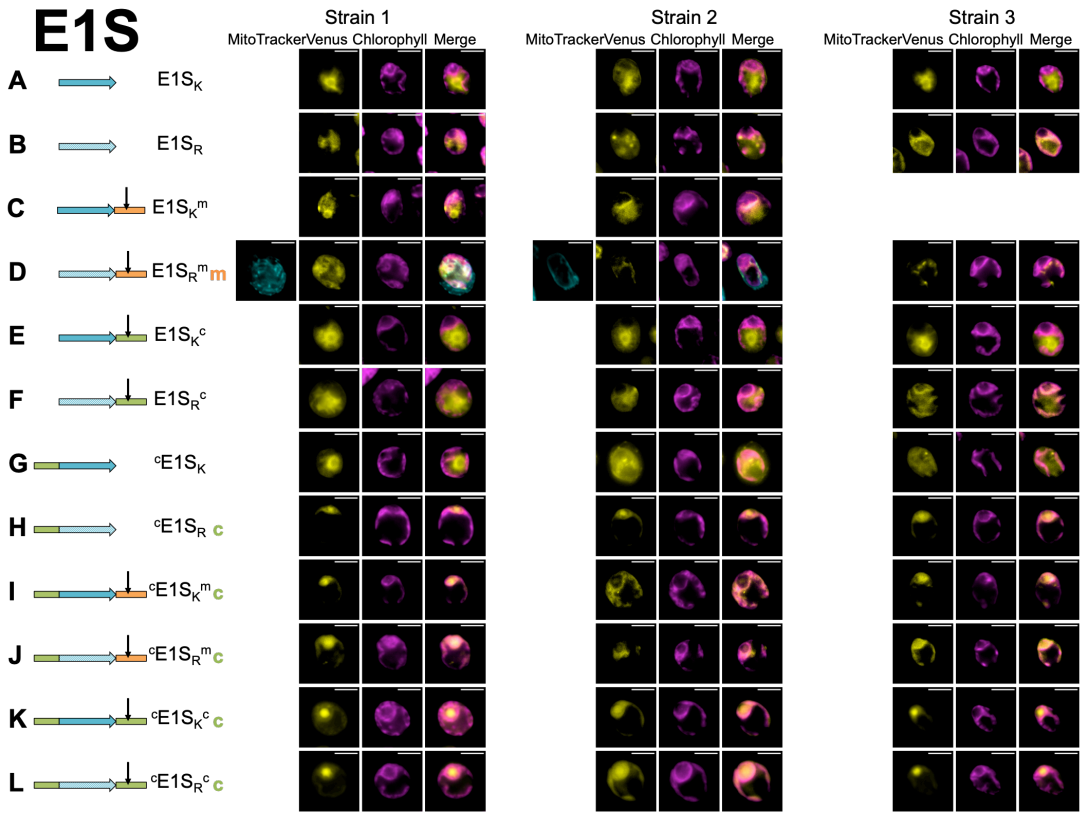


934

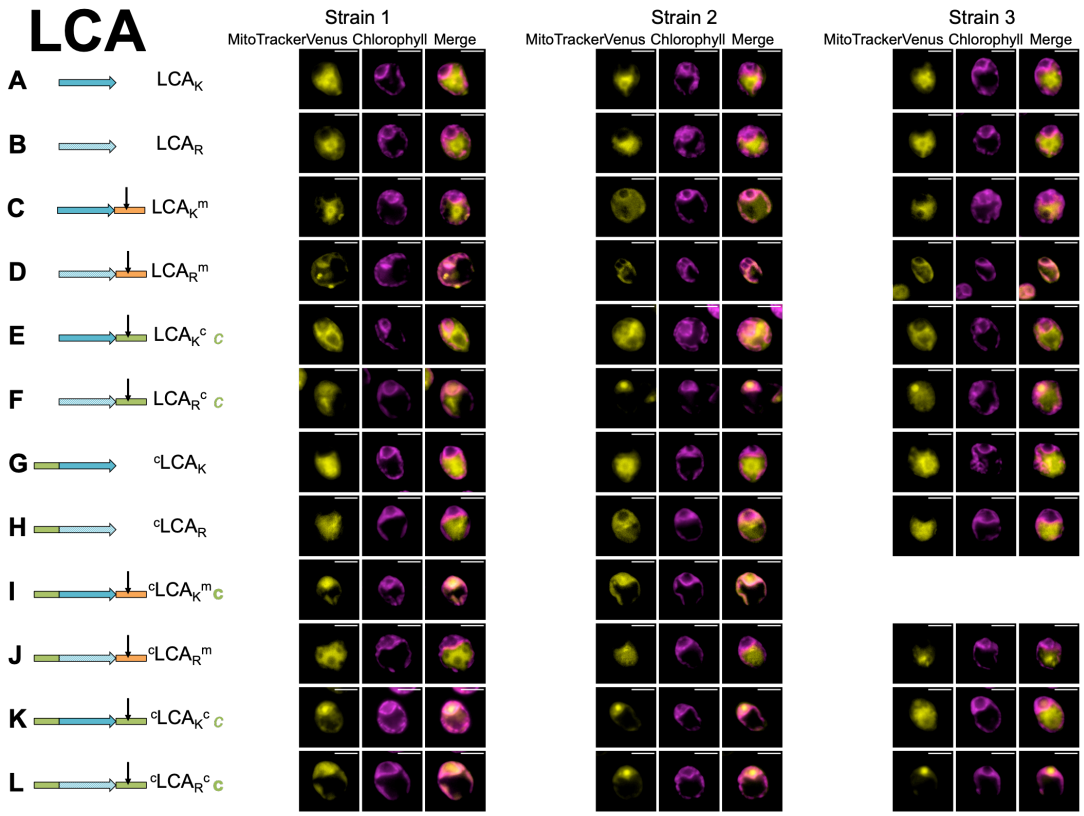
935 **Supplementary Fig. 7. Microscopy for biological replicates of Cecropin P3.**
936 Three independent transformant lines (Strains 1-3) are shown for each
937 combination of modifications (A-L), represented by a cartoon and a shorthand
938 description (cf. Fig. 2). Where a construct was interpreted as generating reporter
939 localization in mitochondria or chloroplast, this is indicated by an orange 'm' or a
940 green 'c' respectively, in bold for full targeting or in italics for partial targeting.
941 Scale bars are 5μm.



Supplementary Fig. 8. Microscopy for biological replicates of Sarcotoxin 1D. Three independent transformant lines (Strains 1-3) are shown for each combination of modifications (A-L), represented by a cartoon and a shorthand description (cf. Fig. 2). Where a construct was interpreted as generating reporter localization 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. Scale bars are 5µm.

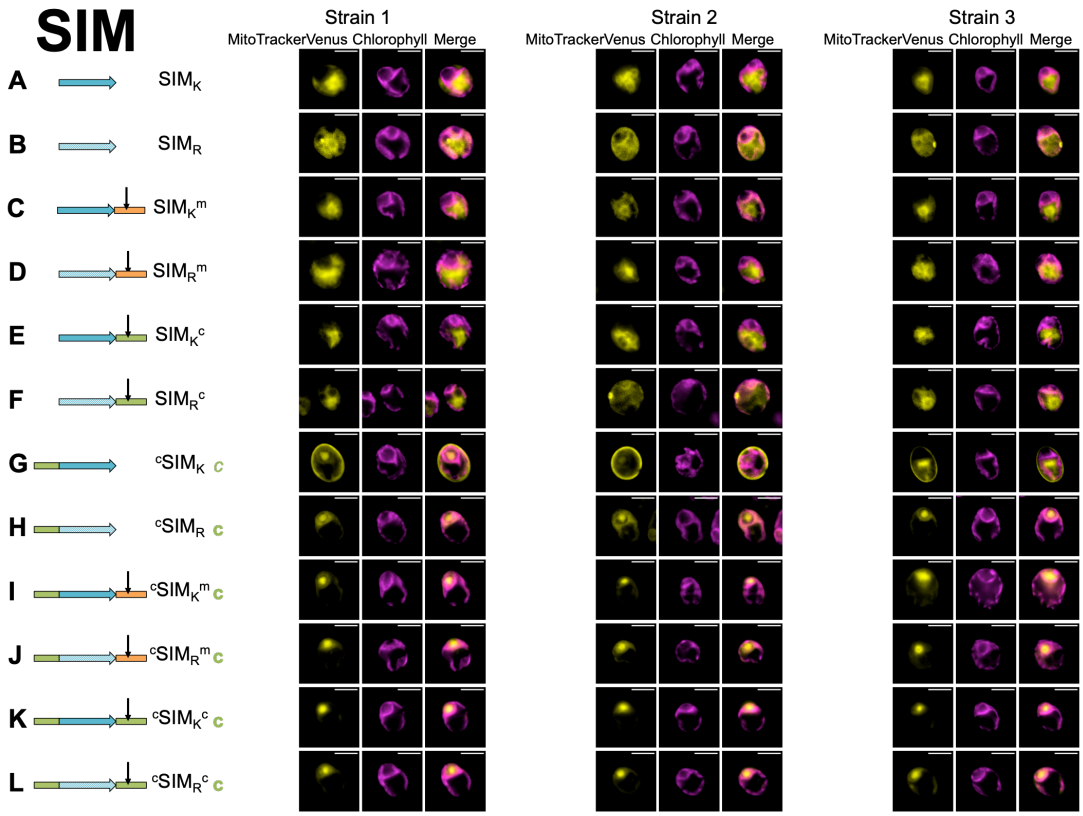


Supplementary Fig. 9. Microscopy for biological replicates of Esculentin 1S. Three independent transformant lines (Strains 1-3) are shown for each combination of modifications (A-L), represented by a cartoon and a shorthand description (cf. Fig. 2). Where a construct was interpreted as generating reporter localization 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. Scale bars are 5µm.

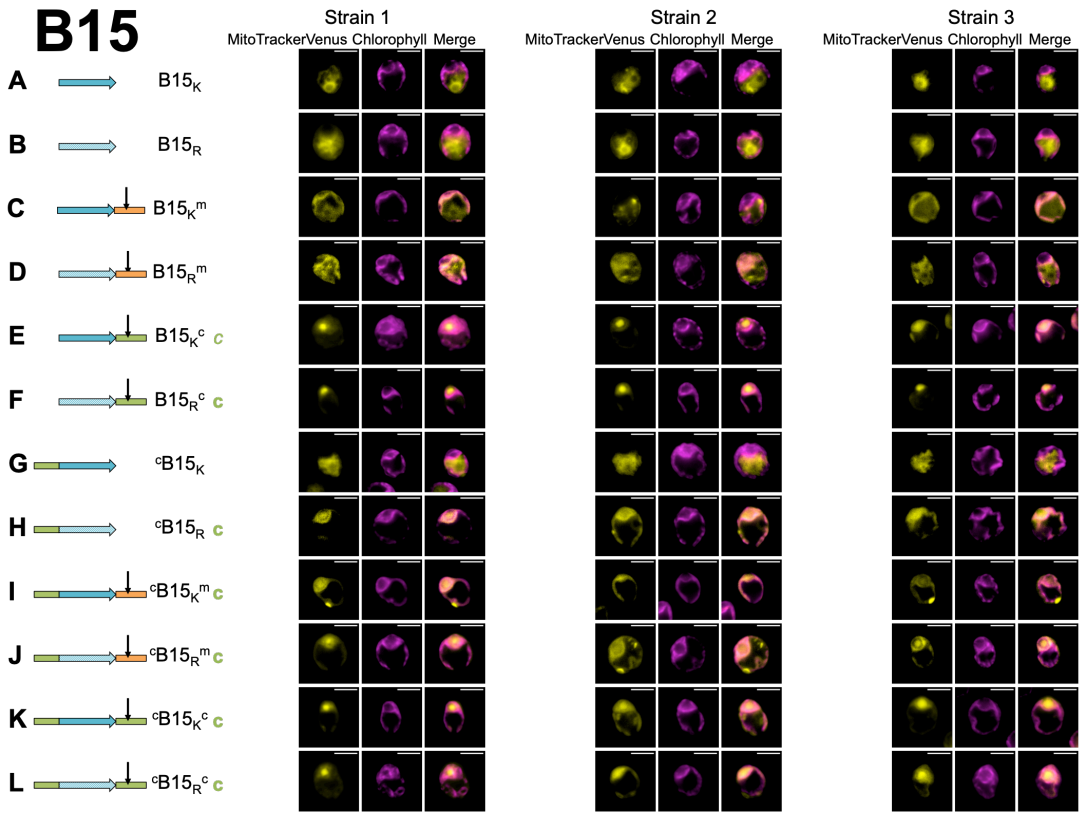


958

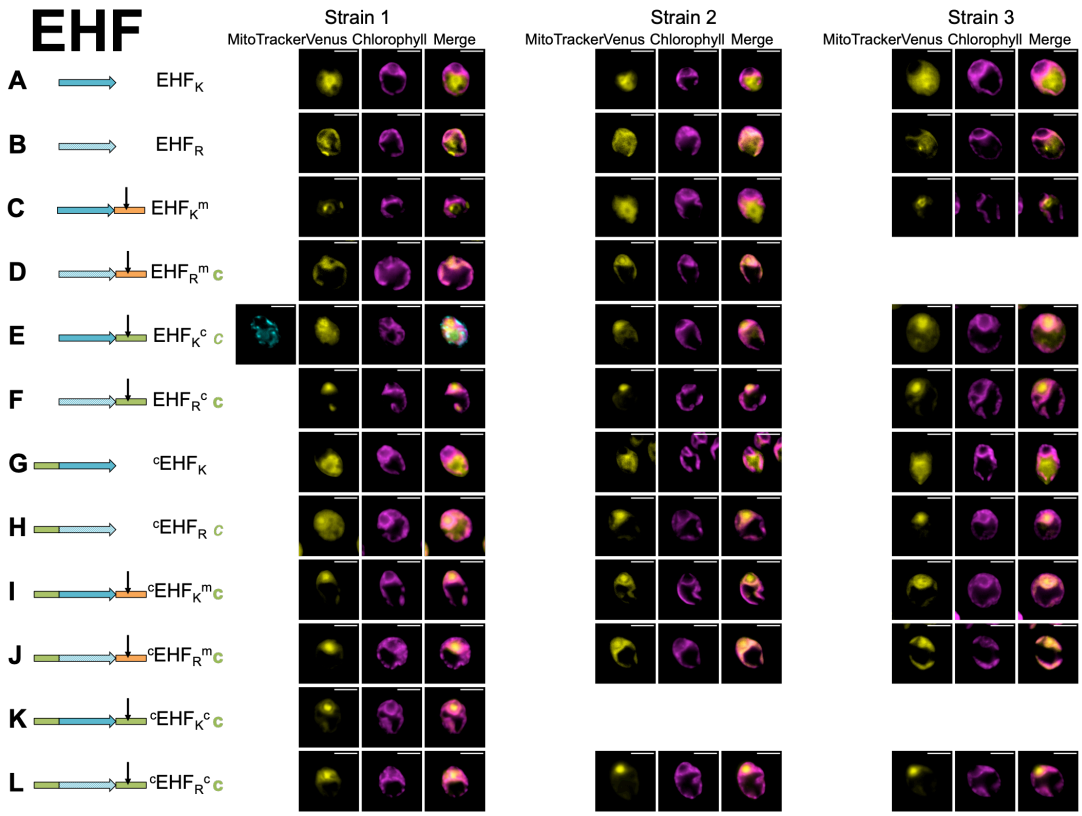
959 **Supplementary Fig. 10. Microscopy for biological replicates of Leucocin A.**
960 Three independent transformant lines (Strains 1-3) are shown for each
961 combination of modifications (A-L), represented by a cartoon and a shorthand
962 description (cf. Fig. 2). Where a construct was interpreted as generating reporter
963 localization in mitochondria or chloroplast, this is indicated by an orange 'm' or a
964 green 'c' respectively, in bold for full targeting or in italics for partial targeting.
965 Scale bars are 5µm.



Supplementary Fig. 11. Microscopy for biological replicates of SI Moricin. Three independent transformant lines (Strains 1-3) are shown for each combination of modifications (A-L), represented by a cartoon and a shorthand description (cf. Fig. 2). Where a construct was interpreted as generating reporter localization 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. Scale bars are 5µm.



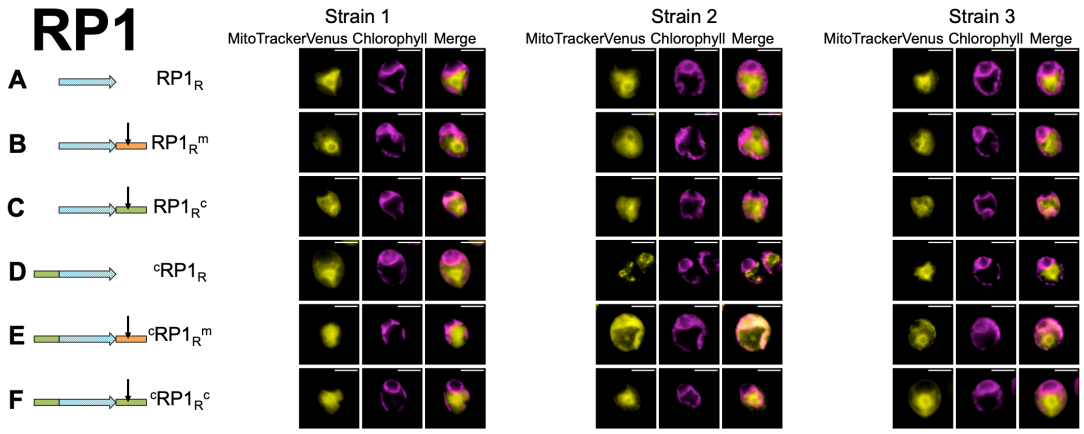
Supplementary Fig. 12. Microscopy for biological replicates of Bacillocin 1580. Three independent transformant lines (Strains 1-3) are shown for each combination of modifications (A-L), represented by a cartoon and a shorthand description (cf. Fig. 2). Where a construct was interpreted as generating reporter localization 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. Scale bars are 5µm.



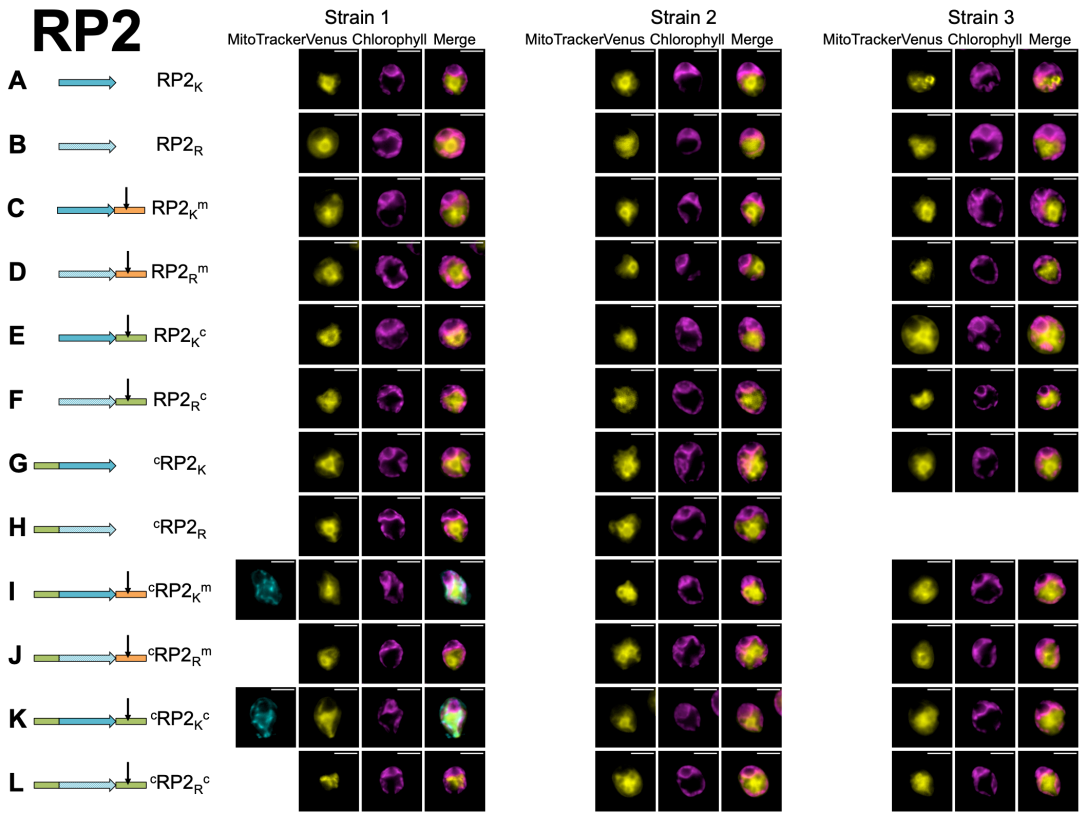
982

983 **Supplementary Fig. 13. Microscopy for biological replicates of Enterocin**
984 **HF.** Three independent transformant lines (Strains 1-3) are shown for each
985 combination of modifications (A-L), represented by a cartoon and a shorthand
986 description (cf. Fig. 2). Where a construct was interpreted as generating reporter
987 localization in mitochondria or chloroplast, this is indicated by an orange 'm' or a
988 green 'c' respectively, in bold for full targeting or in italics for partial targeting.
989 Scale bars are 5µm.

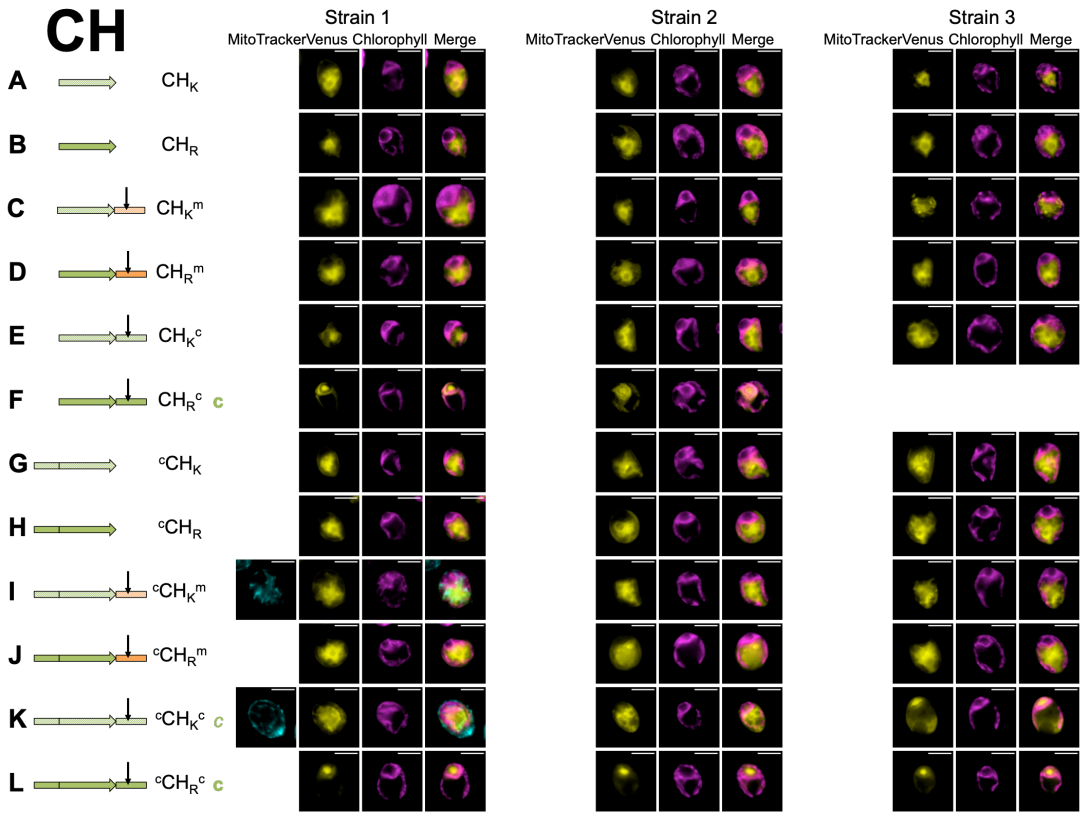
990



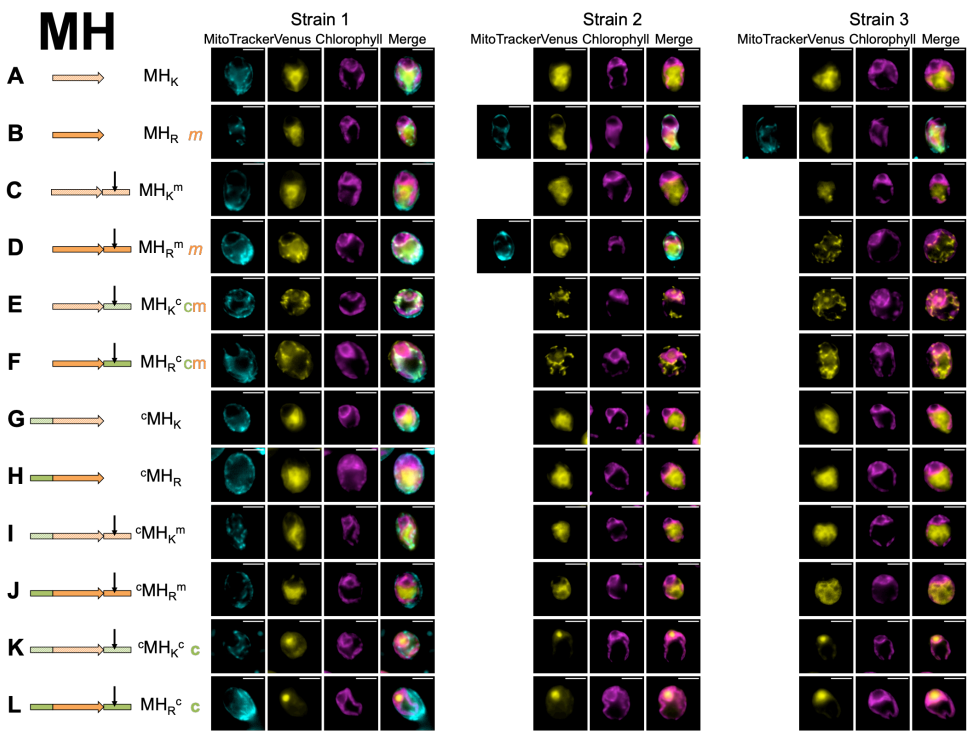
Supplementary Fig. 14. Microscopy for biological replicates of negative control Random Peptide 1. Three independent transformant lines (Strains 1-3) are shown for each combination of modifications (A-F), represented by a cartoon and a shorthand description (cf. Fig. 2). Where a construct was interpreted as generating reporter localization 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. Scale bars are 5µm.



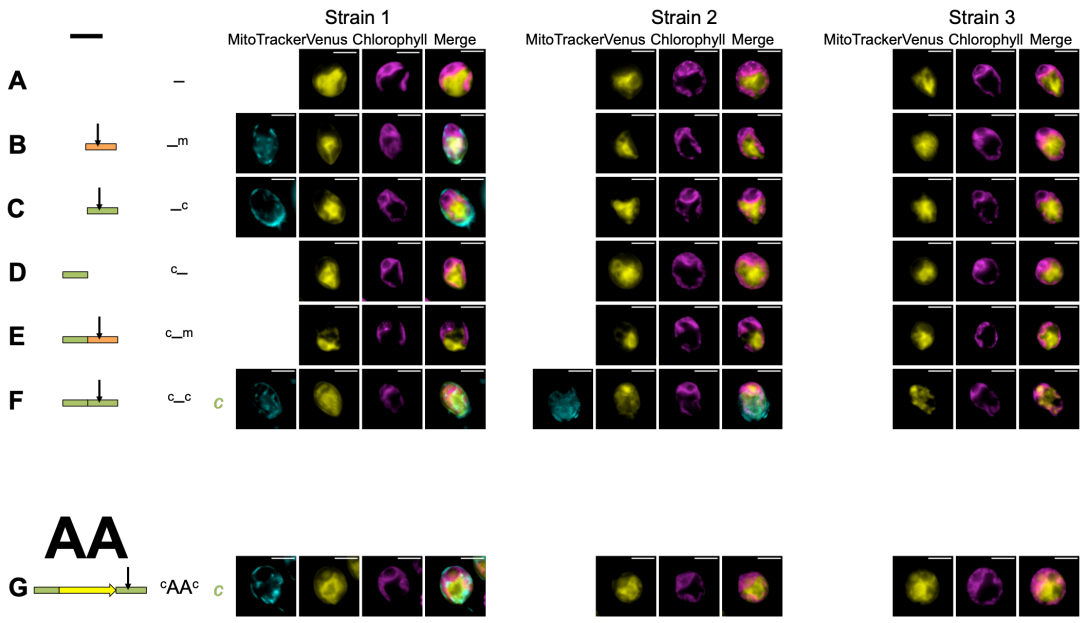
Supplementary Fig. 15. Microscopy for biological replicates of negative control Random Peptide 2. Three independent transformant lines (Strains 1-3) are shown for each combination of modifications (A-L), represented by a cartoon and a shorthand description (cf. Fig. 2). Where a construct was interpreted as generating reporter localization 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. Scale bars are 5µm.



Supplementary Fig. 16. Microscopy for biological replicates of TP control Rubisco activase cTP helical element (CH). Three independent transformant lines (Strains 1-3) are shown for each combination of modifications (A-L), represented by a cartoon and a shorthand description (cf. Fig. 2). Where a construct was interpreted as generating reporter localization 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. Scale bars are 5μm.

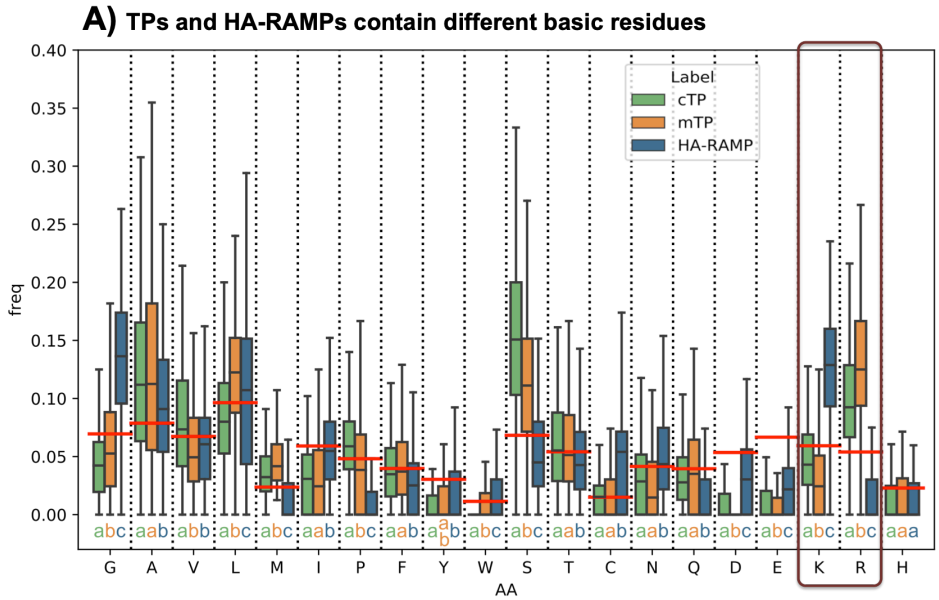


Supplementary Fig. 17. Microscopy for biological replicates of TP control mitochondrial γ-carbonic anhydrase 2 mTP helical element (MH). Three independent transformant lines (Strains 1-3) are shown for each combination of modifications (A-L), represented by a cartoon and a shorthand description (cf. Fig. 2). Where a construct was interpreted as generating reporter localization 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. Scale bars are 5μm.



Supplementary Fig. 18. Microscopy for biological replicates of no-peptide and Alanine-screen controls. Three independent transformant lines (Strains 1-3) are shown for each combination of elements (A-G), represented by a cartoon and a shorthand description (cf. Fig. 2). Where a construct was interpreted as generating reporter localization 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. Scale bars are 5µm.

1034

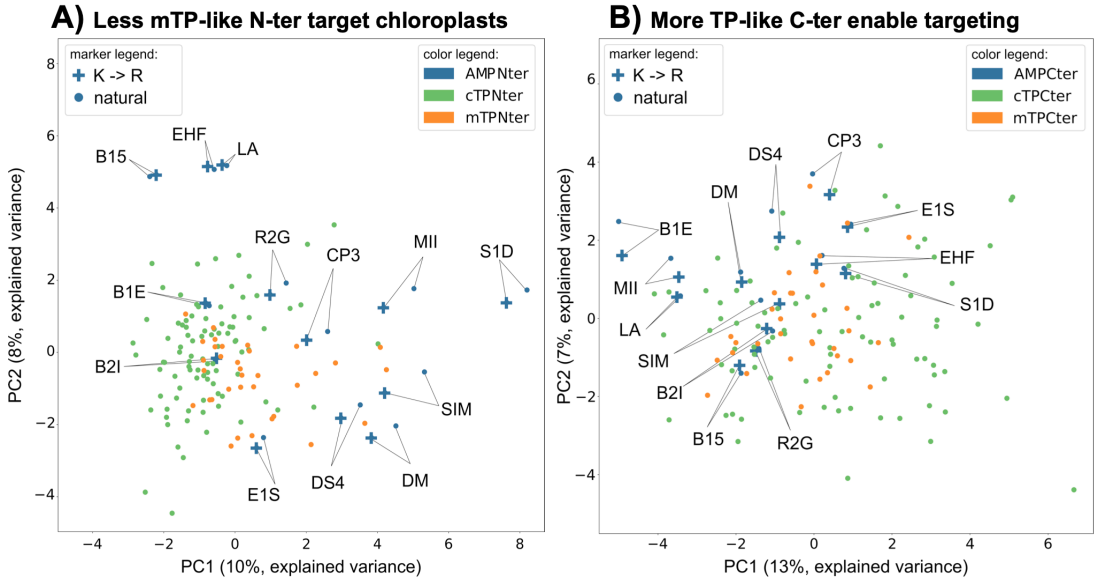


1035

1036 **Supplementary Fig. 19. Comparison of amino acid frequencies reveals K/R**
1037 **shift. A)** Amino acid frequencies are shown as boxplots (center line: median; box
1038 limits: upper and lower quartiles; whiskers: min/max values within 1.5x
1039 interquartile range) for Chlamydomonas cTPs in green, Chlamydomonas mTPs in
1040 orange and HA-RAMPs in blue. To give a baseline for comparison, the average
1041 across uniprot is given for each amino acid as red horizontal line. For each
1042 residue, different letters underneath distributions indicate a significant
1043 difference: groups that share the same letter are significantly different at $p<0.05$
1044 (Multiple Kruskal Wallis tests followed by Dunn post-hoc tests) from groups
1045 attributed a different letter.

1046

1047



1048

1049 **Supplementary Fig. 20. Principal component analyses reveal that N- but not**

1050 **C-termini differ between cTPs and mTPs.** Principal component (PC) analyses

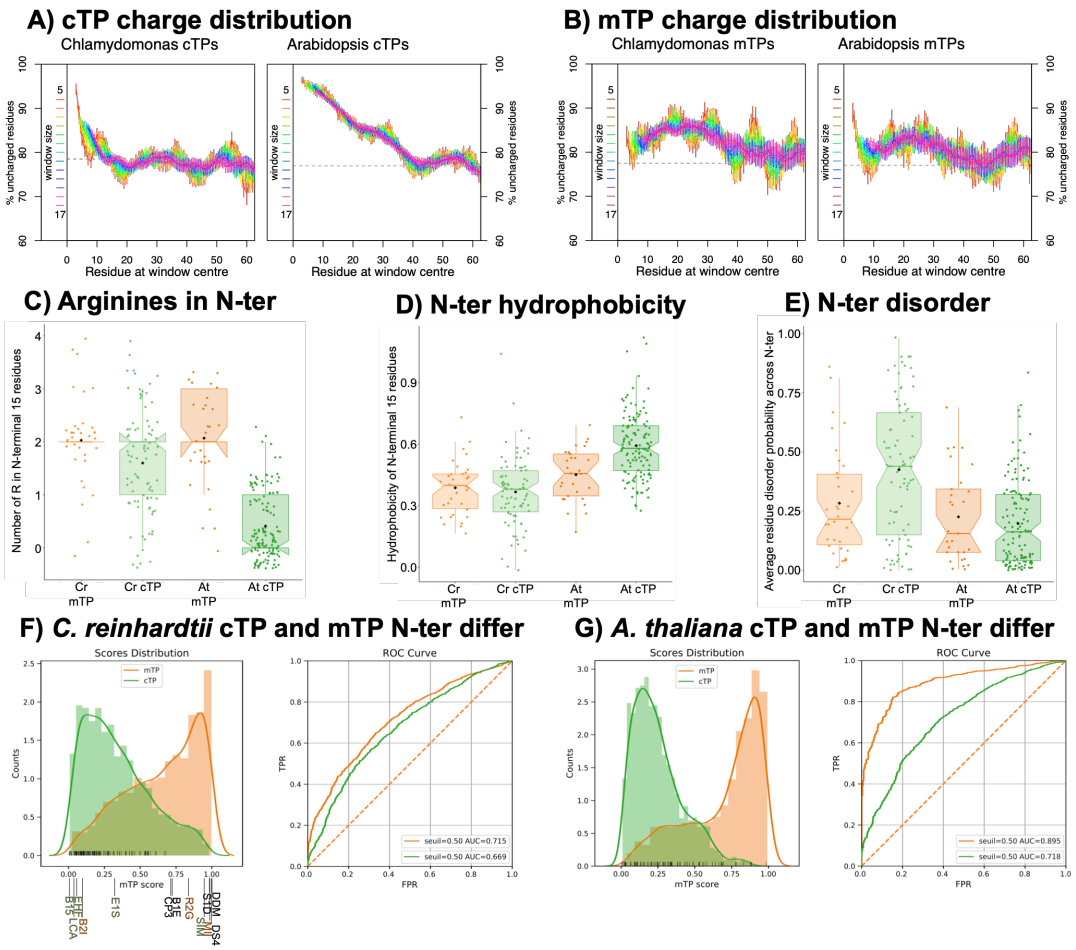
1051 of auto-cross-correlated Z-scale values[5] for **A)** N-termini (15 residues) and **B)** C-

1052 termini (33 residues, encompassing -10 to +23 relative to the cleavage site for

1053 TPs) respectively of Chlamydomonas cTPs (in green) and mTPs (in orange) as

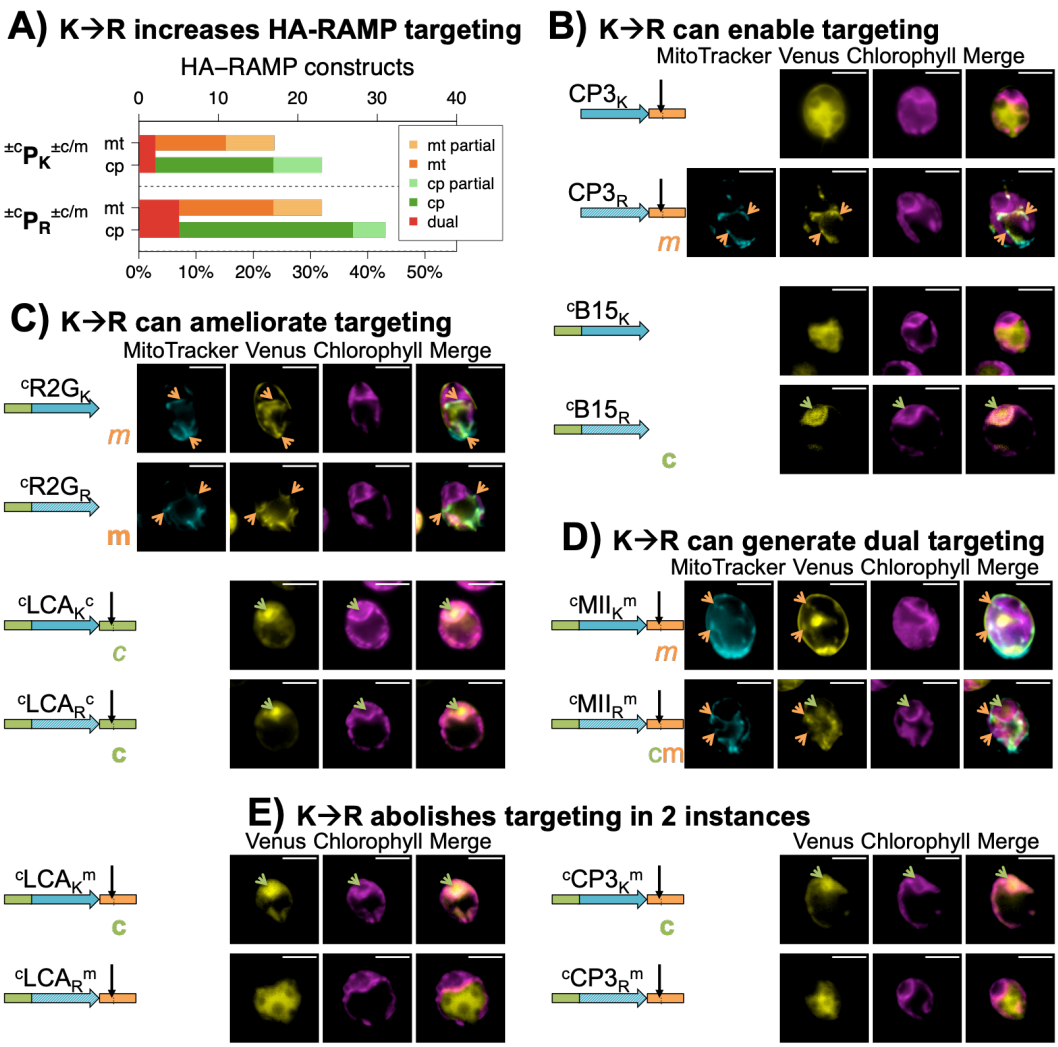
1054 well as the 13 HA-RAMPs under study (in blue; crosses denote HA-RAMPs after

1055 K→R).

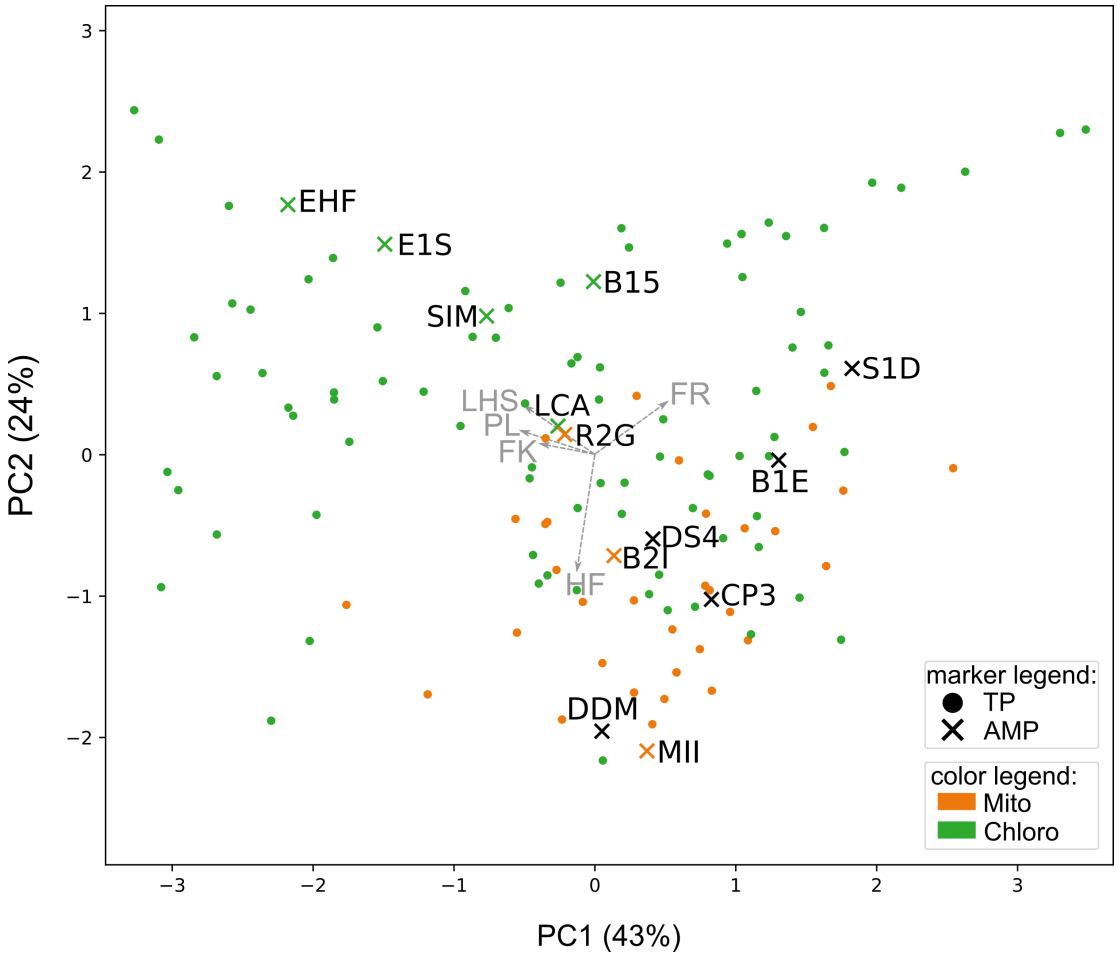


Supplementary Fig. 21. Unusual algal cTP N-termini share physicochemical differences against mTP N-termini with plant counterparts. A) Plots detailing the percentage of uncharged residues[25] (i.e. excluding K/R and D/E) for window sizes ranging from 5 to 17 residues (refer to legend for colour code) show that Chlamydomonas cTPs (on the left) are less uncharged at the N-terminus than Arabidopsis cTPs (on the right). Dashed grey lines indicate the average value for randomized sequences (including mature proteins), to provide an estimate of what would be expected by chance. **B)** Equivalent graphs (as in A) for mTPs show a similar charge profile across Chlamydomonas and Arabidopsis. **C-E)** Distributions of salient values as boxplots (center line: median; box limits: upper and lower quartiles; whiskers: min/max values within 1.5x interquartile

1068 range) for mTPs in orange and cTPs in green with *Chlamydomonas reinhardtii*
1069 (Cr) on the left and *Arabidopsis thaliana* (At) on the right show that: C)
1070 Chlamydomonas cTP N-termini contain R at almost the same frequency as mTP
1071 N-termini, whereas Arabidopsis cTP N-termini contain fewer R; D)
1072 Chlamydomonas TP N-termini are less hydrophobic than Arabidopsis TP N-
1073 termini, with plant cTP N-termini showing the highest hydrophobicity; E)
1074 Chlamydomonas cTP N-termini are more disordered than mTP or Arabidopsis
1075 cTP N-termini. Points represent individual peptides (note that for integer values,
1076 point positions are randomized within ±0.5 in y as well as in x to increase point
1077 separation), population means are shown as black diamonds. **F)** Binomial logistic
1078 regression classifier separating TPs using auto-cross-covariance of Z-scales[5] of
1079 the N-terminal 15 residues for *C. reinhardtii*. The left-hand graph shows the
1080 distribution of cTPs in green and mTPs in orange over the model output 'mTP
1081 score'. Black bars at the bottom of the graph represent scores for cTPs of
1082 Arabidopsis, showing that plant cTPs are recognized as cTPs by the model
1083 trained on algal cTP N-termini (89% recognized as cTPs, with an mTP score
1084 <0.5). Values for our 13 HA-RAMPs are given below the graph. The right-hand
1085 graph shows receiver operating characteristic (ROC) curves, plotting the true
1086 positive rate (TPR) against the false positive rate (FPR), where the area under
1087 the curve (AUC) serves as estimate of the model quality with values above 0.5
1088 indicating that the model is better than random. **G)** Equivalent model (as in F)
1089 trained on *A. thaliana* TPs, with black bars now representing Chlamydomonas
1090 cTPs showing that the majority of algal cTPs are recognized as cTPs by the plant
1091 model (75% recognized as cTPs, with an mTP score <0.5).



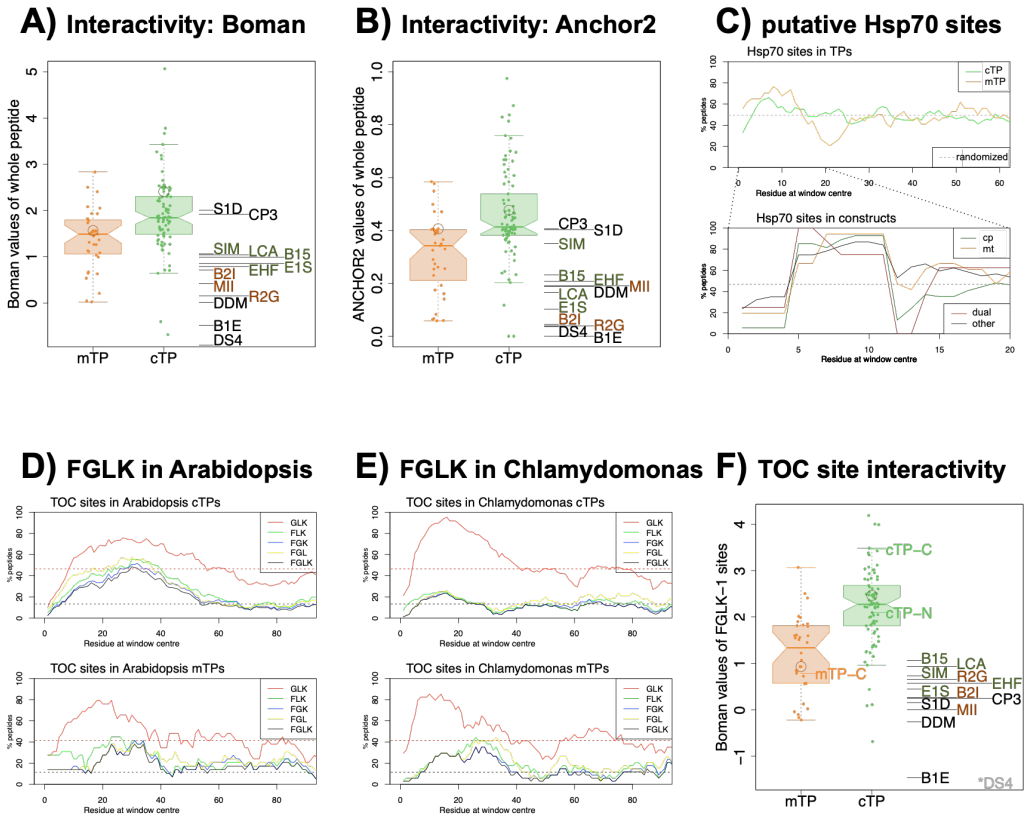
Supplementary Fig. 22. K→R generally improves HA-RAMP targeting. A) An overview graph showing how many, out of 13 HA-RAMPs studied, show targeting to either mitochondria (mt) or chloroplast (cp), or both (dual) before and after K→R represented. **B-E)** Epifluorescence microscopy images of selected examples are shown as in Fig. 3; same convention for Venus localization as in Fig. 4.



Supplementary Fig. 23. HA-RAMP properties determine their targeting propensities. PCA analysis of cTPs (green dots), mTPs (orange dots), cp-set HA-RAMPs (green crosses), mt-set HA-RAMPs (orange crosses) and other HA-RAMPs (dark crosses) based on their length (Peptide Length: PL), number of residues before the longest predicted amphipathic helix (Long Helix Start: LHS), fraction of residues forming the predicted amphipathic helix (Helix Fraction: HF) and fraction of K (Lysine Fraction: KF) and R residues (Arginine Fraction: AF). Arrows on the graph represent the eigenvalues of the individual variables. The first principal component explains 43% of the variance and distinguishes peptides principally according to KF, PL, LHS (pointing left) and RF residues (pointing right). The second principal component explains 24% of the variance and distinguishes peptides principally according to HF. The cp-set HA-RAMPs

1113 group with the most distinguishable cTPs, being longer, with a higher LHS and a

1114 lower HF compared to mt-set HA-RAMPs and mTPs.



1115

1116 **Supplementary Fig. 24. Higher protein interactivity predicted for cTPs than**

1117 **mTPs. A, B and F)** For salient properties, Chlamydomonas mTPs and cTPs are

1118 compared to HA-RAMPs (cp-set in green, mt-set in orange). Distributions are

1119 shown as boxplots (center line: median; box limits: upper and lower quartiles;

1120 whiskers: min/max values within 1.5x interquartile range), coloured points

1121 represent individual peptides **A)** cTPs show increased protein interactivity

1122 (p=0.0003), as do cp-set HA-RAMPs (p= 0.0355), as estimated through Boman

1123 values[65], a proxy developed for AMPs where a value of ca. >2 indicates increased

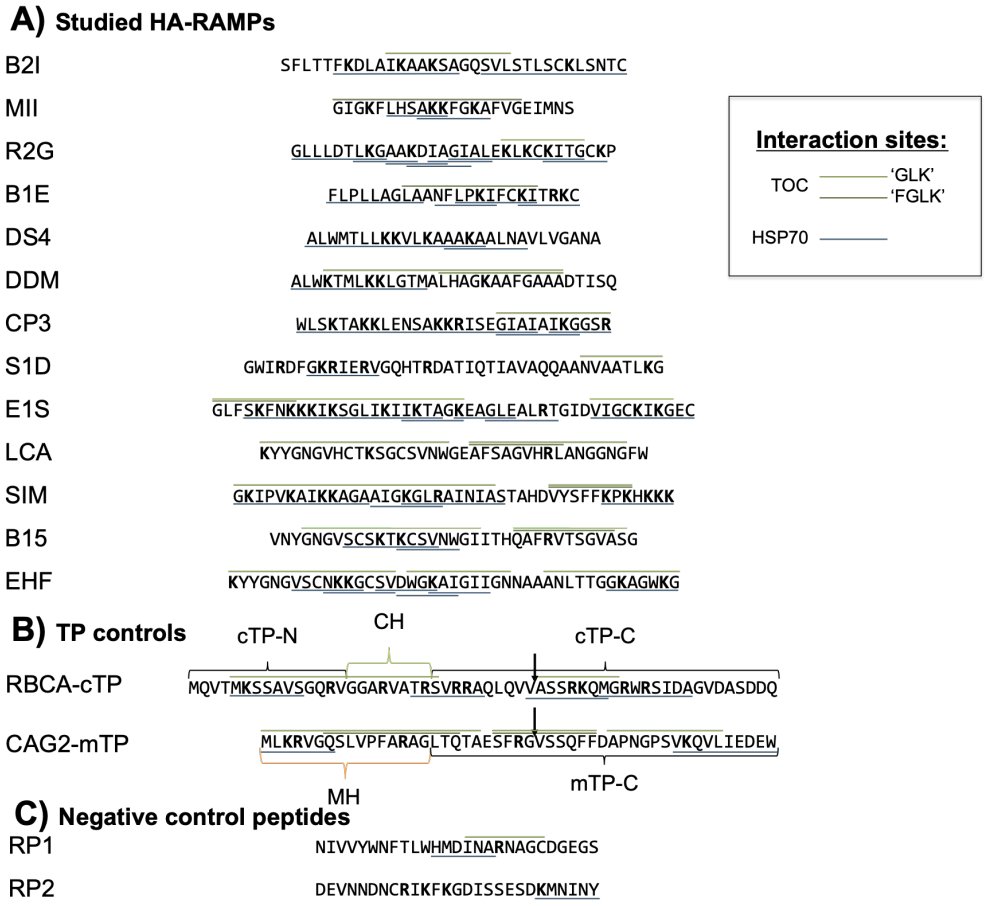
1124 protein interaction potential, and <2 points to membrane interaction. **B)** cTPs

1125 show increased protein interactivity (p<0.0001), and cp-set HA-RAMPs show a

1126 similar tendency (p=0.0638), as estimated with Anchor2[67], developed to predict

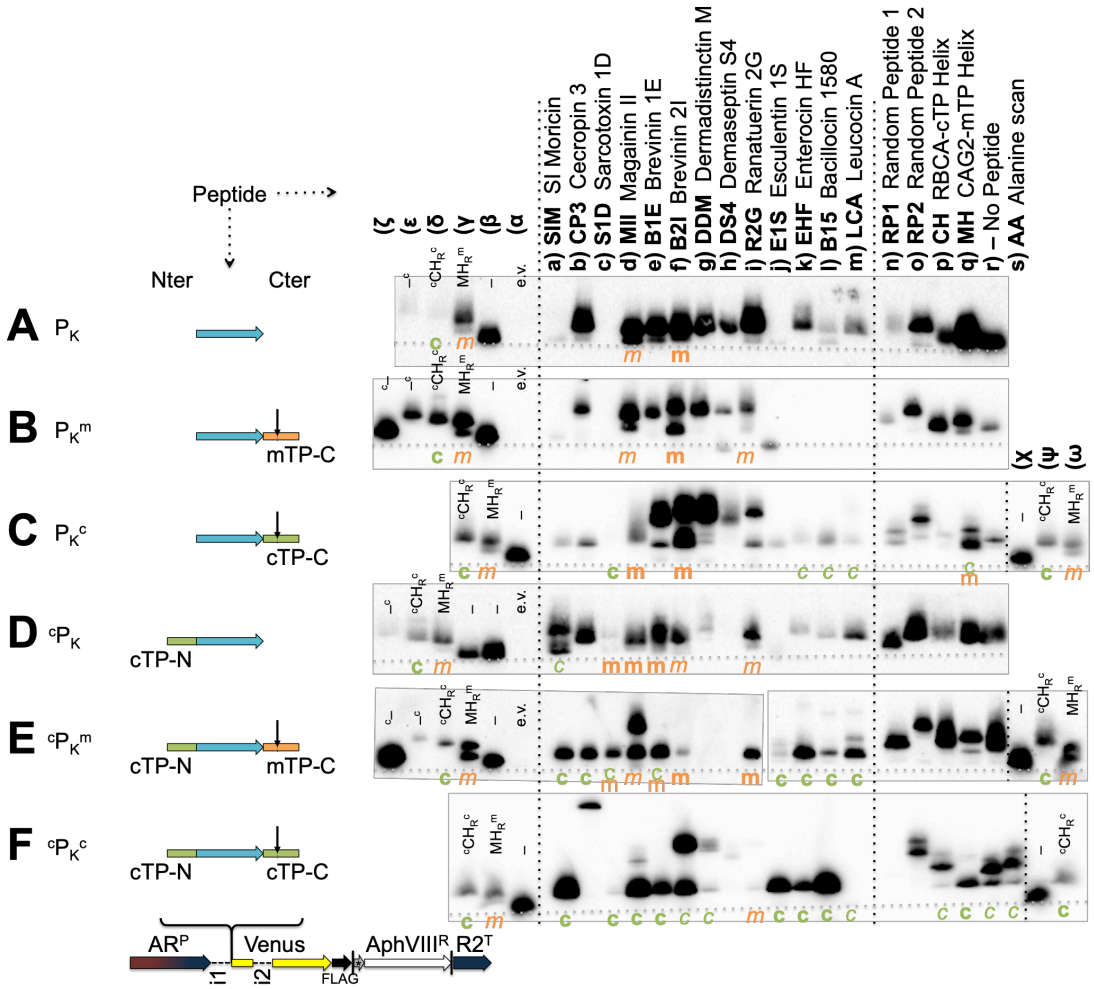
1127 the protein interaction potential of disordered sequences. **C)** Both cTPs and
1128 mTPs show a high fraction of peptides (~50%) that have an Hsp70 interaction
1129 site over a 6-residue window at any given position along the sequence in the top
1130 graph, with a peak towards the N-terminus. HA-RAMP constructs shown in the
1131 bottom graph also display a peak in predicted Hsp70-binding sites at the N-
1132 terminus, independently of construct localisation. The value obtained after
1133 randomizing the position of residues is given as dashed grey line to provide an
1134 estimate of how often sites would be expected to occur given amino acid
1135 frequencies (~50%). Randomization was done over the entire sequence
1136 including the cargo protein. **D)** For Arabidopsis cTPs in the top graph, the
1137 percentage of peptides that have a full 'FGLK' site[25] (black curve; presence of F
1138 and G/P and L/V/A and K/R and absence of D/E within an 8-residue window)
1139 far exceeds the value obtained after randomizing the position of residues
1140 (dashed black line for full 'FGLK' sites), a proxy for the frequency of the motif
1141 expected at random, for most positions along the sequence up to ~60 residues.
1142 Randomization was done over the entire sequence including the cargo protein.
1143 Reduced 'FGLK'-1 sites containing three out of four elements appear to occur
1144 mostly in the context of full FGLK sites, except for 'GLK' sites for which
1145 randomization is also shown (dashed red line). Arabidopsis mTPs (bottom
1146 graph) also contain 'FLGK' sites at a frequency higher than expected at random
1147 between residues 20-40, and 'GLK' sites upstream of residue 40, but motifs are
1148 less prevalent than in cTPs. **E)** Equivalent graphs (as D) for Chlamydomonas
1149 show that full 'FGLK' sites (black lines) are rare in cTPs (top graph) and actually
1150 more common in mTPs (bottom graph) upstream of residue 40. Reduced 'GLK'
1151 sites (red lines) by contrast far exceed the prevalence expected at random

(dashed red line) ca. up to residue 40. **F)** Protein interactivity, estimated through Boman values, for only those parts of peptides that correspond to FGLK-1 sites is increased for cTPs ($p<0.0001$), and cp-set HA-RAMPs show a similar tendency ($p= 0.0627$). Note that cTP-C and cTP-N, but not mTP-C, contain 'FGLK-1' sites with predicted protein interactivity. *DS4 contains no 'FGLK-1' sites. Reported p-values were obtained through two-way t-tests for TPs and one-way t-tests for HA-RAMPs.



Supplementary Fig. 25. Hsp70 and (F)GLK sites are present in HA-RAMPs. Predicted Hsp70 binding sites are underlined in blue. Putative TOC interaction sites are overlined in green: 'FGLK' sites in dark green and 'GLK' sites in light green. See Fig. 1 for detailed annotations. Notably, both TOC and Hsp70-

interaction sites are contained within native HA-RAMPs (A) and within our TP controls (B).



Supplementary Fig. 26. Western Blots for K-bearing constructs. Samples were immunolabelled using an α-FLAG primary antibody. Constructs representing different combinations of TP-element additions are shown in A-F, as indicated by a shorthand description and a cartoon (cf. Fig. 2). HA-RAMPs are consistently arranged in lanes a-m. Note that the order of HA-RAMPs differs from the one in Fig. 2. Control constructs in lanes o-r carry the R→K modification. Additional technical controls are present in lanes labeled with greek letters. Sections containing HA-RAMPs, control constructs and technical controls are separated by dotted black lines. Technical controls (cf. Fig. 1): – is no peptide,

1176 $^{c}CH_{R}^{c}$ is Rubisco activase cTP, MH_{R}^{m} is mitochondrial γ-carbonic anhydrase 2
1177 mTP, ^{c}– is Rubisco activase cTP N-terminal element (15 residues), –^{c} is Rubisco
1178 activase cTP C-terminal element (33 residues, -10 to +23 relative to cleavage
1179 site), e.v. is empty vector (no Venus expression). Constructs that generated
1180 chloroplast and/or mitochondrial localization are marked with a green 'c' and/or
1181 an orange 'm' respectively, in bold for full or in italics for partial targeting. Note
1182 that E) contains two separate blots for the left-hand section up to lane i and the
1183 right-hand section following lane j.