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

Oliver D Caspari[1¶*†], Clotilde Garrido[1], Yves Choquet[1], Francis-André Wollman[1], Ingrid Lafontaine[1*]

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

[¶]Lead contact

*corresponding authors:

ODC: **odc20@cantab.net**

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

current addresses:

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

Summary

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

Keywords:

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

50 **Introduction**

51 Mitochondria and Chloroplasts arose through endosymbiosis and retain their own
52 genomes, yet the vast majority of organellar proteins is encoded in the nucleus,
53 translated in the cytoplasm and imported into the organelles[1,2]. N-terminal targeting
54 peptides (TPs) that are cleaved off upon import provide the information on targeting,
55 although their primary structures are very diverse[3]. Yet chloroplast transit peptides
56 (cTP) and mitochondrial presequences (mTP) have very similar physico-chemical
57 properties making reliable differentiation often challenging. A multitude of studies
58 has found sequence elements contributing to specificity determination, and prediction
59 programs have been improving[4,5], yet a mechanistic understanding of how targeting
60 information is encoded has remained elusive.

61 Here, we use antimicrobial peptides (AMP) as an original chassis, to test the
62 contribution of key TP features towards targeting efficiency and specificity. Part of
63 the innate immune system, AMPs are produced by virtually all types of cells in a bid
64 to kill or control microbial adversaries[6,7]. AMPs have recently been proposed to be at
65 the evolutionary origin of TPs[8,9]. The proposed evolutionary scenario posits that early
66 during endosymbiotic organellogenesis of first the mitochondrion and later the
67 chloroplast, the host cell used AMPs to attack the bacterial proto-endosymbiont. A
68 bacterial resistance mechanism, whereby the AMP is imported into the bacterial
69 cytoplasm, would have generated a pathway for host proteins to reach the bacterial
70 cytosol – a plausible first step in the evolution of a protein import machinery.
71 Cationic, Helical-Amphipathic Ribosomally-produced AMPs (HA-RAMPs) and TPs
72 share key physico-chemical properties and they were shown, in some instances, to
73 retain cross-functionality[10]: several TPs showed antimicrobial activity and selected
74 HA-RAMPs, fused to a cleavage-site containing TP element, were shown to promote
75 the import of a Venus fluorescent protein into either the mitochondria or the
76 chloroplast of the model green alga *Chlamydomonas reinhardtii*.

77 The main part of mTPs and the central element of cTPs most closely resemble HA-
78 RAMPs on account of a shared cationic, amphipathic helical structure that often
79 encompasses the entire length of the HA-RAMPs[9] and mTPs[11,12]. Although plant
80 cTPs have been described as unstructured[13], algal cTPs more closely resemble mTPs
81 in being helical[14]. Helices were observed by NMR in membrane-mimetic
82 environments in both algal and plant cTPs[3,15–18], and signatures of amphipathic
83 helices can be detected in a majority of cTPs[10].

84 In addition to the helices, both mTPs and cTPs contain recognition sites at the C-
85 terminus where processing peptidases in the mitochondrial matrix (MPP) and the
86 chloroplast stroma (SPP) cleave off the cargo protein[19]. These recognition sites
87 encompass some 10 residues upstream of the cleavage site and are structurally distinct
88 from the rest of the TPs, even showing a weak sequence conservation[4,12,20]. While
89 targeting information is usually contained within mTP sequences upstream of the
90 cleavage site, targeting by cTPs shorter than ca. 60 amino acids often requires
91 downstream unstructured sequence stretches in the N-terminal domain of the mature
92 protein[21,22]. Besides the amphipathic helical module and the C-terminal cleavage
93 module shared between mTPs and cTPs, it has been argued that distinct features at
94 their N-termini conferred organelle specificity to each set of targeting
95 peptides[2,12,20,23–28].
96 In this study, we systematically introduced modifications into a diversity of HA-
97 RAMPs in a bid to generate targeting to the mitochondria or chloroplast in
98 Chlamydomonas. This dataset provides new insights into how different TP elements
99 contribute to differential targeting and to the efficiency of protein import. Being
100 similar in physico-chemical properties to TPs[10], HA-RAMPs provide a privileged
101 vantage point from which to study sequence elements that govern targeting. In our
102 choice of HA-RAMPs, we aimed to reflect the diversity of available sequences by
103 choosing representatives of different HA-RAMP families based on similarity with
104 TPs. We show that some of our 13 HA-RAMPs natively contain TP-like sequence
105 elements, with some HA-RAMPs being prone to chloroplast targeting while others
106 show a preference for the mitochondria. Furthermore, we provide evidence for a
107 critical functional difference in cationic residues with Lysine (K) being used in HA-
108 RAMPs and Arginine (R) in TPs.

109 **Results**

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

111 Figure 1 shows the major sequence features of the 13 HA-RAMPs used in the present
112 study (Fig. 1a) together with those of a typical cTP and a typical mTP (Fig. 1b). On
113 the right-hand side, these peptides are represented according to their proportion of a-
114 helical amphipathic structure. The HA-RAMPs Brevinin-2ISb (B2I), Magainin 2
115 (MII), Ranatuerin-2G (R2G), Dermaseptin S4 (DS4), Dermadistinctin-M (DDM),

116 Brevinin-1E (B1E), Cecropin-P3 (CP3), Sarcotoxin-1D (S1D), Esculentin-1SEA
117 (E1S), Leucocin A (LCA), SI Moricin (SIM), Bacillocin 1580 (B15), and Enterocin
118 HF (EHF) were chosen such that different AMP families would be represented with
119 varying proximity to TPs (Table S1) and thus a range of physico-chemical properties
120 be explored (Fig. S1). As negative controls, two peptides that lack predicted
121 amphipathic helices were chosen from among randomly generated sequences (Fig.
122 1c).
123 To explore targeting, the 13 HA-RAMPs were systematically modified (Fig. 2),
124 notably by adding TP elements from the N- and/or C-terminal domains of
125 Chlamydomonas Rubisco activase (RBCA) cTP, or a C-terminal domain of similar
126 length from mitochondrial γ-carbonic anhydrase 2 (CAG2) mTP (Fig. 1b, Fig. 2). In a
127 bid to keep peptidase recognition sites intact, C-terminal elements were designed to
128 start 10 residues upstream of the cleavage site, even though this slightly truncates TP
129 helices. Peptides carrying the 15 amino-acid long cTP N-terminus (cTP-N) will be
130 denoted ^{c}P. Similarly, peptides with C-terminal elements from RBCA-cTP (cTP-C) or
131 CAG2-mTP (mTP-C) will be denoted P^{c} or P^{m} respectively. Micrographs for three
132 biological replicates per construct are displayed on Fig. S2-S19. The reader will be
133 asked to return to Fig. 2 throughout, with Fig. 3 and Fig. 4 providing selected
134 micrographs as examples highlighting particular points of interest.

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

136 In the absence of any modifications, B2I and MII were capable of organelle-targeting
137 visible in fluorescence microscopy (Fig. 2 row A, Fig. S2-S14). When equipped with
138 B2I, the fluorescent reporter Venus shows mitochondrial localization (Fig. 3a, Fig.
139 S2): the Venus signal is observed as a characteristic network pattern that co-localizes
140 with MitoTracker fluorescence. By contrast, in the case of MII only a small fraction
141 of fluorescence signal in the Venus channel shows co-localisation with the
142 mitotracker dye, signifying that targeting is only partial (Fig. 3a, Fig. S3). Note that in
143 these epifluorescence images, some autofluorescence emanating from the chloroplast
144 is present in the Venus channel (Fig. 3b). Where possible, the brightness of these
145 images was adjusted to remove this autofluorescence. Chosen brightness settings are
146 indicated next to each image. In cases where low Venus accumulation necessitated
147 brightness values low enough for autofluorescence to be visible in the Venus channel,
148 a black dotted line is visible in the brightness display.

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

150 We looked for differences in primary sequence that might impede the ability of the
151 remaining 11 HA-RAMPs to target. As illustrated on Fig. S20, extant HA-RAMPs
152 and organellar TPs display a few differences in their amino-acid content. HA-
153 RAMPs are richer in helix-breaking Glycines (G) but poorer in Prolines (P), whereas
154 TPs are enriched in Serines (S) compared to HA-RAMPs. Most of the other amino
155 acid frequencies are broadly similar between the two sets of peptides and within range
156 of the average value across all entries in the UNIPROT database (red lines). As
157 expected, both HA-RAMPs and TPs are poor in acidic residues (D, E) but enriched in
158 basic residues (K or R). However, their complement in basic residues is markedly
159 different (Fig. 3c, Fig. S20): HA-RAMPs are rich in Lysines (K) whereas TPs are
160 rich in Arginines (R).

161 In order to see whether these contrasting differences in K/R ratio contributed to the
162 functional divergence between HA-RAMPs and TPs, we substituted all instances of R
163 by K in TPs and of K by R in HA-RAMPs. In the rest of the text, the basic amino acid
164 mostly present in a given peptide P will be indicated by a subscript as P_R or P_K. HA-
165 RAMPs with K→R transition showed reduced antimicrobial activity, as illustrated on
166 Fig. 3d by the increased minimum inhibitory concentrations for MII, DS4 or B1E.

167 We used Chlamydomonas Rubisco activase (RBCA) cTP ($^{c}CH_{R}{}^{c}$) and mitochondrial
168 γ-carbonic anhydrase 2 (CAG2) mTP ($MH_{R}{}^{m}$) as positive controls for chloroplast and
169 mitochondrial targeting respectively (Fig. 3e, Fig. S17-S18). Note that TP helical
170 fragments, stripped of their N- and C-terminal domains, are denoted as MH for mTP
171 and CH for cTP (Fig. 1b). Chloroplast localization of RBCA-cTP ($^{c}CH_{R}{}^{c}$) is evident
172 from a bright Venus signal colocalizing with chlorophyll autofluorecence,
173 Furthermore, Venus accumulates in the pyrenoid, a proteinaceous chloroplast
174 subcompartment that contains a large majority of Rubisco. While small proteins like
175 Venus can enter the pyrenoid, most thylakoid membranes are excluded, allowing the
176 structure to be localized as a dip in chlorophyll autofluorescence. Venus channel
177 signal emanating from the pyrenoid is thus a useful visual guide to true chloroplast
178 localization of the fluorescent reporter[29]. Note that due to cTP-C containing a
179 Rubisco-binding motif[30], Venus accumulation in the Rubisco microcompartment, the
180 pyrenoid, is particularly pronounced in constructs carrying this element. When
181 natively R-rich RBCA-cTP and CAG2-mTP sequences were subjected to systematic
182 R→K substitutions ($^{c}CH_{K}{}^{c}$, $MH_{K}{}^{m}$), their ability to target were reduced or abolished

183 (Fig. 3e, Fig. S17-S18). These experiments demonstrate the respective functional
184 contributions of K and R residues to antimicrobial and organelle targeting activity.
185 We then systematically re-examined the organelle targeting ability of the set of 13
186 HA-RAMP that had undergone K→R substitutions (Fig. 2 row B, Fig. S2-S14). MII_R
187 now shows full mitochondrial targeting, and $R2G_R$ as well as $CP3_R$ show partial
188 targeting, as evidenced by significant overlap of Venus and mitotracker fluorescence
189 (Fig. 3f, Fig. S3, S4, S8).

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

191 Addition of mTP-C, the cleavage-site containing C-terminal element of CAG2-mTP
192 (Fig. 4a, Fig. 2 rows CD) enabled mitochondrial targeting for $E1S_R$ and $R2G_K$ and
193 improved the targeting already observed for $R2G_R$ and $CP3_R$, in addition to continued
194 targeting by B2I and MII. The impact of adding cTP-C, the C-terminal element
195 derived from RBCA-cTP, is even more important: cTP-C significantly enabled
196 chloroplast targeting, which could be seen in nine HA-RAMP constructs involving
197 $CP3_R$, S1D, LCA, B15 and EHF (Fig. 4b, Fig. 2 rows E,F, Fig. S8, S9, S11, S13,
198 S14). The addition of cTP-C to the helical part of the CAG2-mTP (MH_R^c) even
199 partially retargets the construct to the chloroplast (Fig. 4c, Fig. S18), and deletion of
200 cTP-C abolishes chloroplast targeting by RBCA-cTP (^cCH_R) (Fig. 4c, Fig. S17). By
201 comparison, deletion of mTP-C reduces but does not totally abolish mitochondrial
202 targeting of CAG2-mTP (MH_R) (Fig. 4c, Fig. S18), and addition of mTP-C fails to
203 rescue ^cCH_R ($^cCH_R^m$) (Fig. 4c, Fig. S17). Note that the addition of cTP-C was also
204 compatible with mitochondrial localization by B2I and MII (Fig. 4b, Fig. 2 rows E,F,
205 Fig. S2, S3). The addition of mTP-C also enabled chloroplast localization by EHF_R
206 (Fig. 2 row D), visible despite low Venus channel brightness as Venus signal
207 emanating from within the pyrenoid (Fig. 4a, Fig. S14).

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

209 The sole addition of cTP-N generated at least partial chloroplast targeting in 6 HA-
210 RAMP constructs, notably by SIM, but also $CP3_R$, $E1S_R$, $B15_R$ and EHF_R (Fig. 4d,
211 Fig. 2 rows G,H, Fig. S8, S10, S12, S13, S14). However, cTP-N also enabled
212 mitochondrial targeting in three HA-RAMPs that had not previously shown
213 mitochondrial localization, B1E, DDM_R and S1D (Fig. 4e, Fig. 2 rows G,H, Fig. S5,
214 S7, S9).

215 The importance of cTP-N as a chloroplast determinant becomes more evident when
216 combined with a C-terminal element: addition of cTP-N to constructs equipped with
217 mTP-C resulted in 10 HA-RAMP constructs targeting the chloroplast ($CP3_K$, E1S,
218 LCA_K, SIM, B15, EHF) as well as 6 constructs showing dual targeting to the
219 chloroplast and the mitochondria ($B2I_R$, MII_R, B1E, S1D) (Fig. 2 rows I,J, Fig. S2-S3,
220 S5, S8-S14). Combining cTP-N with cTP-C (Fig. 2 rows KL, Fig. S2-S5, S7-S14)
221 generates at least partial chloroplast targeting in 22 out of 26 HA-RAMP constructs
222 (including 1 instance of dual targeting by $R2G_R$; Fig. S4); only DS4 fails to show any
223 chloroplast targeting (Fig. S6). Consider ^{c}B2I (Fig. 2 column a rows G-L, Fig. S2) as
224 an example of the effect of combining cTP-N with C-termini, shown in Fig. 4f:
225 Equipped with mTP-C, $^{c}B2I_K^{m}$ targets the mitochondria whereas addition of cTP-C
226 ($^{c}B2I_K^{c}$) generates partial targeting to the chloroplast. Upon K→R, the presence of
227 mTP-C ($^{c}B2I_R^{m}$) now supports dual targeting and full chloroplast targeting is
228 observed in the presence of cTP-C ($^{c}B2I_R^{c}$).
229 Finally, Fig. 4g shows that while addition of cTP-N to the CAG2-mTP abolished
230 mitochondrial targeting ($^{c}MH_R^{m}$), the cTP-N/cTP-C combination also retargets MH to
231 the chloroplast, and enables partial chloroplast targeting of a control in which the
232 central element is composed of a poly-Alanine tract ($^{c}AA^{c}$) devoid of amphipathic
233 properties and of a no peptide control (^{c}-^{c}) (Fig. 2 columns q-s, rows KL). The two
234 latter experiments demonstrate that there are enough determinants for recognition of
235 the chloroplast translocon dispersed between the Nter and Cter of a cTP, to target
236 Venus to the chloroplast, albeit with a lower efficiency than when an amphipathic
237 helix is present in between.
238 To further explore the role of the N-terminal cTP domain, we performed some in
239 silico comparisons of the studied peptides. In Principal Component Analyses on ACC
240 Z-scales[10] comparing TP N- and C-termini with our 13 HA-RAMPs (Fig. S21), cTPs
241 and mTPs could be differentiated much better by N- compared to C-termini,
242 supporting N-termini as important specificity determinants. Furthermore, even though
243 algal cTP N-termini are more charged, less hydrophobic and more disordered than
244 higher plant cTPs (Fig. S22) in line with algal cTPs being more mTP-like[14],
245 Chlamydomonas cTP N-termini are recognized as chloroplast-targeting in a model
246 trained on Arabidopsis TP and vice versa (Fig. S22), confirming that essential
247 physico-chemical features of chloroplast determination in cTP N-termini are
248 maintained in both lineages.

249 ***HA-RAMPs dominate targeting specificity***

250 Considering Fig. 2 by columns reveals that organelle specificity is, to a large extent,
251 determined by HA-RAMPs. Only 3 HA-RAMPs (B2I, R2G, MII; referred to hereafter
252 as the mt-set) account for >70% of all constructs where mitochondrial targeting is
253 seen. Similarly, 5 HA-RAMPs (E1S, LCA, SIM, B15, EHF; referred to as the cp-set)
254 account for the majority (~57%) of all chloroplast targeting, and for two thirds when
255 excluding the cTP-N/cTP-C combination that generates chloroplast targeting across
256 most HA-RAMPs. In some instances, a set of peptide modifications may switch
257 targeting from the mitochondria to the chloroplast, yet the major effect of
258 modifications, i.e. exchanging K→R (Fig. 3, Fig. S23) or adding TP elements (Fig. 4,
259 6), is to enhance the targeting ability to an organelle determined by the HA-RAMP
260 primary sequence properties (Fig. 2).

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

262 To understand what differentiates mt-set from cp-set HA-RAMPs, we examined the
263 sequence characteristics of our 13 HA-RAMPs with those of well-characterized
264 Chlamydomonas targeting peptides (Fig. 5). In Chlamydomonas, mTPs are ca. 35
265 residues in length (Fig. 5a). By contrast, most cTPs are significantly longer
266 ($p=0.0017$). The difference is even greater than shown here given that many cTPs
267 require a contribution from post-cleavage site residues for successful targeting[10,21,22].
268 Consistently, cp-set HA-RAMPs (in green, Fig. 5a) are longer than mt-set HA-
269 RAMPs (in orange; $p=0.0184$) and require further elongation by addition of TP
270 elements before targeting can be observed (Fig. 2).

271 Most mTPs fold into an amphipathic helix for most of the sequence (~80% on
272 average, Fig. 5b) starting directly from their N-terminus (Fig. 5c). The fraction
273 dedicated to amphipathic helix formation in cTPs is significantly lower (Fig. 5b,
274 $p<0.0001$), with a longer stretch of non-helical sequence at the N-terminus (Fig. 5c,
275 $p<0.0001$). A PCA analysis based on these properties, plus the fraction of R and K,
276 recapitulates the combined properties within each peptide (Fig. S24). The PCA shows
277 that cp-set HA-RAMPs are localized with cTPs in the upper left area of the graph,
278 consistent with a lower helix fraction (Fig. 1, Fig. 5b) due to longer sequence
279 stretches upstream of the main helix than in the mt-set HA-RAMPs (Fig. 1, Fig. 5c,
280 $p=0.0199$). Consistent with the importance of these helical features, LCA, the most
281 helical among cp-set HA-RAMPs also shows the highest fraction of only partial

282 chloroplast targeting (3 out of 5 targeting constructs, Fig. 2 column j rows E,F,K).
283 Furthermore, SIM and E1S, the two cp-set HA-RAMPs with the shortest pre-helix
284 segments (Fig7 row C), both require addition of cTP-N for chloroplast targeting (Fig.
285 2 columns i,k).

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

287 cTPs appear more likely to interact with proteins than mTPs (Fig. S25), just like cp-
288 set HA-RAMPs. 'FGLK' motifs that are reportedly TOC interaction sites in plant
289 cTPs[2] appear shortened to 'GLK' motifs in Chlamydomonas but they are associated
290 with an increased protein interaction potential in cTPs compared to mTPs ($p<0.0001$),
291 including within cTP-N and cTP-C elements. However, Hsp70-interaction sites[25],
292 which are commonly found at cTP N-termini[2], also occur at a high frequency in mTP
293 sequences and HA-RAMPs (Fig. S25, S26).

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

295 To characterize the maturation of HA-RAMP–targeted proteins upon organellar
296 import, we performed immunoblotting experiments (Fig. 6). Indicative examples were
297 selected for clarity; a more comprehensive overview is provided in Fig. S27.
298 In the absence of a dedicated cleavage site (Fig. 6a), we found some cleavage did
299 occur but preproteins are also maintained: two bands are observed in constructs that
300 were localized in an organelle by fluorescence imaging (lanes d,f-j). The upper band
301 corresponds to unprocessed Venus, which is a fusion of the reporter with the HA-
302 RAMP construct and migrates at varying positions depending on the length of the
303 presequence. The lower band corresponds to the processed form migrating closer to
304 the "free" Venus position (lane c) depending on the exact site of cleavage. The cTP
305 and mTP controls (lanes a,b) contain 23 residues between cleavage site and Venus
306 (Venus+23), and maintain very little preprotein migrating above. In HA-RAMP
307 constructs by contrast, the majority of Venus remained in the unprocessed top band
308 for constructs targeting the mitochondria (lanes f-j). Nonetheless, the presence of a
309 faint processed form in the lower band just above free Venus shows that some
310 cleavage occurs. The proportion of processed versus unprocessed Venus preprotein is
311 much higher in partially chloroplast-targeted $^{c}SIM_{K}$ (lane d), hinting at more efficient
312 degradation of the unprocessed form in the chloroplast (see also Fig. S27).
313 Addition of cleavage sites improves cleavage across organelles: In the presence of
314 either cTP-C (Fig. 6b, Blot 1) or mTP-C (Fig. 6b, Blot 2), which contain the RBCA-

315 cTP and CAG2-mTP cleavage sites respectively, targeting constructs appear to be
316 cleaved at a site corresponding to TP controls (Venus+23). Constructs equipped with
317 either of the two cleavage sites deliver a processed Venus of similar size,
318 independently of whether the mitochondria or the chloroplast are targeted (Fig. 6b).
319 These results indicate that mTP and cTP cleavage sites are recognized in both
320 organelles. We note that in the case of HA-RAMP constructs equipped with mTP-C,
321 cleavage appears to occur not at the cleavage site previously identified for CAG2
322 (Fig. 6b Blot 2 lane a, top band, corresponding to Venus+23), but instead at or near a
323 second site closer to "free" Venus (compare Fig. 6b Blot 2 lane a, bottom band to
324 lanes b-g), implying processing within ten residues upstream of Venus. Unlike mTP-
325 C (Fig. 6b Blot 2 lane h), cTP-C appears to be processed in the absence of targeting
326 (Fig. 6b Blot 1 lanes a,b) at a site different to the one used in organelles (Fig. 6b Blot
327 1 lanes c-h).
328 A switch from K to R (Fig. 6c), which improved targeting, also increased the amount
329 of processed form relative to that of the unprocessed form in all cases, whether due to
330 a gain of targeting (lanes f-i) or an increase in efficiency (lanes b-e), suggesting that
331 cleavage can serve as a proxy for targeting. Indeed, there is a remarkable correlation
332 between which constructs show cleavage, as observed in immunoblots, and which
333 constructs show targeting, as judged from microscopy (Fig. S27). We note that faint
334 processed bands can be seen for several unmodified HA-RAMPs (Fig. S27),
335 suggesting further peptides in addition to $B2I_K$ and MII_K (Fig. 2) may be able to
336 support some amount of targeting that is, however, too partial to be seen in
337 fluorescence microscopy.

338 **Discussion**

339 ***How to make TPs from HA-RAMPs***

340 To gain insight into peptide features that govern targeting, we used a related but
341 separate class of peptides, HA-RAMPs, as chassis and stacked modifications to
342 generate targeting into mitochondria or chloroplasts, or in some cases into both (Fig.
343 7a).

344 ***Arginines target better than Lysines***

345 We showed that a K-to-R-switch increases targeting efficacy to both organelles across
346 many constructs. R is more common than K in cTPs across green algae, vascular

347 plants, red algae, glaucophytes and many secondary plastids[31], but the functional
348 significance for targeting efficacy had not previously been recognized. Our
349 observation is in line with a previous report that a R-to-K switch at the N-ter of an
350 Arabidopsis mTP abolished targeting (Lee at al., 2019). The underlying mechanism
351 warrants further research, since differences in bulkiness and pKa[32,33], or in trans-
352 acting factors that regulate targeting through ubiquination of K[34], may play a role.
353 Indeed a preference for R may help to protect preproteins from degradation while they
354 transit in the cytosol (Fig. 7a). It is of note that arginine features in consensus
355 sequences of the cleavage sites[4,35], which thus may have been generated by K→R
356 modification, potentially accounting for increased chloroplast-localization among
357 cTP-N bearing constructs (Fig. 2 rows GH).

358 ***Unstructured sequences contribute to chloroplast determination***

359 TP N-termini contribute to differential targeting[24]. In vascular plants, the presence of
360 a multi-Arginine motif within the N-terminus of the peptide prevents chloroplast
361 import[23]. Our data on a wider series of constructs suggest that chloroplast targeting
362 specificity is provided by an unstructured sequence upstream of the amphipathic helix
363 in cTP N-termini rather than by the mere absence of Arginines[23,36] (Fig. S22) or a
364 difference in Hsp70 binding sites[2,27,28] (Fig. S25). Indeed, several HA-RAMPs that
365 intrinsically contained unstructured N-termini proved able to support chloroplast
366 targeting, even in the absence of a cTP-N addition.

367 We found chloroplast targeting was further improved by an unstructured C-terminus,
368 in addition to the unstructured N-terminal element. Indeed mTP-C are more
369 structured, containing a long predicted amphipathic helix. The need to contain
370 unstructured sequences[13] as well as the increased length of cTPs, even extending
371 beyond the cleavage site in many cases[10,21,22], appear mechanistically related to the
372 import system (Fig. 7a): whereas mitochondrial import makes use of the proton
373 gradient to power uptake of positively charged presequences[37,38], energized
374 chloroplast import requires the cTP to stretch across both TOC and TIC and contact
375 the translocation motor[2,39,40]. Structured sequence elements can thus impede import,
376 including the helix at the Venus N-terminus[41]. HA-RAMPs with unstructured C-
377 termini, such as SIM, were thus most likely to generate chloroplast targeting in the
378 absence of an added TP-C. Unstructured elements also likely provide protein-protein
379 interaction motifs. For instance, TOC-interaction has been attributed to semi-
380 conserved 'FGLK' motifs[27,42–44], although the requirement for F appears to be relaxed

381 in Chlamydomonas (Fig. 4g)[45]. The presence of GLK sites with high predicted
382 interactivity in cTP-N and cTP-C should contribute to the high chloroplast targeting
383 potential of constructs equipped with both elements.

384 ***Amphipathic helices contribute to differential targeting***

385 While N- and C-terminal elements aid specificity determination, particularly for
386 chloroplast import, their influence fails to explain why some HA-RAMPs target
387 preferentially the mitochondria or exclusively the chloroplast. This observation
388 suggests that targeting specificity also is determined by some sequence properties of
389 the amphipathic helical elements, which in the case of HA-RAMPs serve to mediate
390 insertion into specific target membranes. It is of note that individual mTPs and cTPs
391 also interact with membrane bilayers[12,18]. For instance, the Rubisco small subunit cTP
392 interacts with chloroplast-mimetic membranes only in presence of the chloroplast-
393 specific galactolipids[46]. Taken together these observations suggest that amphipathic
394 helices may interact specifically with the membranes of the targeted organelle[6]. A
395 direct interaction with the membrane bilayer, before that with proteins of the
396 translocons, would provide a basic mechanism for a first step in differential organelle
397 targeting (Fig. 7a). It could explain how TPs can be functionally specific whilst
398 diverse in sequence.

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

400 A small subset of randomly chosen sequences is able to inefficiently deliver proteins
401 into extant mitochondria, relying on amphipathic helices[47,48]. Thus, the existing
402 protein import machinery recognizes amphipathic helical peptides. However, an
403 origin of TPs from random sequences does not explain how protein import into
404 organelles would have spontaneously occurred in absence of the extant translocons.
405 Here, our use of antimicrobial peptides that harbor an amphipathic helix to further
406 understand organelle targeting specificity, was inspired by the hypothesis that HA-
407 RAMPs may have given rise to TPs during endosymbiotic organellogenesis because
408 of an "import and destroy" mechanism from HA-RAMP-resistant bacteria[8,9]. In this
409 context, combining the organelle-targeting behavior of all those constructs that
410 convert antimicrobial into targeting peptides can be translated into a temporal
411 evolutionary scenario (Fig. 7b) that adds to previously proposed models[37,48,50]. First,
412 we found that several unaltered HA-RAMPs can deliver cargo, suggesting the
413 evolution of mTPs from HA-RAMPs would have been straightforward. However,

414 K→R was found to reduce toxicity and increase targeting, a dual effect that would
415 have produced a strong selection pressure favoring this exchange as an early step.
416 Since mitochondria appear to tolerate the presence of preproteins containing unaltered
417 HA-RAMPs, the addition of cleavage sites that allow presequences to be degraded
418 separately from the cargo protein[49] would have come as a second step.
419 That chloroplast targeting requires additional discriminating elements is consistent
420 with cTPs evolving in a cell that already had mTPs. There are two possible scenarios
421 for the origin of cTPs. Firstly, cTPs might have evolved directly from HA-RAMPs
422 that already contained unstructured sequence elements (Fig. 7b), as seen for our cp-set
423 HA-RAMPs. Secondly, as has been suggested earlier[50], cTPs might have co-opted
424 existing mTPs. According to the present scenario, these mTPs already contained R
425 and a cleavage site, yet would still have been recognized by cyanobacterial HA-
426 RAMP importers. In this case, the key innovation that generated cTPs may simply
427 have been the addition of unstructured, possibly TOC-interacting elements to mTPs,
428 which we showed here to be sufficient to retarget the mt-set HA-RAMPs equipped
429 with TP C-termini at least partially to the chloroplast. The cTP N-terminus likely
430 originally started with an F, given that in glaucophytes and rhodophytes (and many
431 derived secondary plastids), a conserved N-terminal F plays a role in chloroplast
432 import[20,31,51]. This observation led to the idea that cTPs may have originated from a
433 re-use of a C-terminal F-based motif involved in secretion via OMP85 beta-barrel
434 proteins[52,53] in (cyano)bacteria, from which TOC75 evolved. To this end, periplasmic
435 POTRA domains responsible for substrate-recognition were proposed to have flipped
436 orientation and now point into the host cytoplasm[54,55]. However, several studies of
437 plant TOC75 have since consistently found POTRA-domains to be in the inter-
438 membrane space, not the cytoplasm[56–58].
439 Note that the proposed scenario makes no prediction about whether host/proto-
440 organelle interactions were mutualistic or antagonistic. Either way, antimicrobial
441 peptides are likely to have been part of the suite of tools used in these host/proto-
442 symbiont interactions. In particular, the role of antimicrobial peptides in mutualistic
443 symbioses[59] includes one of the best documented cases for defensive antimicrobial
444 peptide import into bacteria[60].

445 ***Conclusion***

446 Investigating the steps required to generate TPs from HA-RAMPs has allowed us to
447 uncover a number of novel mechanistic insights. Notably, we discovered a role for
448 Arginines for targeting efficacy, and delineated the contributions of N- and C-
449 terminal elements in targeting specificity. Our work also suggests that whether due to
450 common descent or convergence, the similarities between TPs and HA-RAMPs point
451 to TPs interacting with membrane lipids as an early targeting step. As any
452 evolutionary scenario, including ours describing a plausible pathway from HA-
453 RAMPs to TPs, it must be considered as a working hypothesis that will need to be
454 assessed further by a series of bioinformatics and laboratory-controlled evolutionary
455 experiments. Further understanding of peptide-lipid interactions and the phylogeny of
456 import machinery components should shed new light on the evolution and functioning
457 of organelle TPs.

458 **Materials and Methods**

459 ***Construct generation***

460 Venus expression constructs were designed in SnapGene (v4.3.11) and generated by
461 integrating PCR-amplified (Q5 hot start high-fidelity, #M0515, New England
462 Biolabs) DNA fragments into plasmid pODC53[61] upstream of Venus using Gibson
463 assembly (NEBuilder HiFi DNA assembly, #E5520S, New England Biolabs).
464 Chlamydomonas TP sequences were amplified from genomic DNA extracted from
465 strain T222+ (CC-5101). Templates for codon optimized HA-RAMP, RP and R→K
466 modified TP sequences were obtained by gene synthesis (Eurofins Genomics).
467 Correct assembly was verified by sequencing (Eurofins Genomics). Linear
468 transformation cassettes were generated through restriction digest with *Eco*RV (New
469 England Biolabs).

470 ***Transformation and fluorescence screen conditions***

471 Constructs were transformed into wild-type strain T222+ (CC-5101) using the
472 protocol described previously[62], except using 4µl of DNA at 2µg/µl. Transformants
473 (≥24 per construct) selected for Paromomycin resistance were grown in 200µl TAP in
474 96-well plates under 50µmol photons m^{-2} s^{-2} for 3-5 days and then screened for Venus

475 expression in a fluorescence plate reader (CLARIOstar, BMG Labtech) as described
476 previously[61].

477 ***Microscopy***

478 Cells were grown in 96-well plates and epifluorescence microscopy was carried out as
479 described before[10]. Strains with suspected mitochondrial targeting were subjected to
480 0.1 µM MitoTracker Red CMXRos (ThermoFisher) in growth medium for 30 min in
481 the dark and washed with TAP prior to taking epifluorescence images with the same
482 settings as before[10], using a white light LED and a beam splitter at 535nm (423052-
483 0104-000, Zeiss) in combination with a custom-made filter set (559/34 BrightLine
484 HC, Beamsplitter T 585 LP, 607/36 BrightLine HC; AHF Analysentechnik AG) for
485 detecting the MitoTracker signal. Images were adjusted in Fiji[63] (ImgageJ version
486 2.0.0) as described before[10], and final figures assembled in PowerPoint (Microsoft
487 PowerPoint for Mac 2011, Version 14.6.3).

488 ***Western Blots***

489 Cells were grown up in liquid culture (10ml TAP, 30µmol photons m^{-2} s^{-1}, 160RPM)
490 until late mid-log phase. 2ml Aliquots were resuspended in 30µl storage buffer (1×
491 Roche cOmplete™ Mini proteinase inhibitor cocktail, 10 mM NaF, 0.2 M DTT, 0.2
492 M $NaCO_3$) and stored at -20°C. 20µl boiling buffer (1× Roche cOmplete™ Mini
493 proteinase inhibitor cocktail, 10mM NaF, 50 g/L SDS, 200 g/L sucrose) was added,
494 then aliquots were boiled (50 seconds). Cell debris was removed (tabletop centrifuge,
495 maximum speed, 15 minutes, 4°C). Chlorophyll content was estimated
496 spectrophotometrically: 1 µg chlorophyll μl^{-1} = 0.11 × (OD_{680} – OD_{770}) for 5 µl
497 diluted in 1 ml water. Samples (10 µg chlorophyll, equal volumes) were run
498 (overnight, room temperature, 18A) on large gels (35×27cm; resolving gel: 12%
499 Acrylamide, 0.32% Bisacrylamide, 0.25 µl/ml Tetramethylethylenediamine, 250
500 mg/L Ammonium Persulfate, 375 mM Tris/HCl pH 8.8; stacking gel: 5%
501 Acrylamide, 0.133% Bisacrylamide, 0.666 µl/ml Tetramethylethylenediamine, 666
502 mg/L Ammonium Persulfate, 125 mM Tris/HCl pH 6.8). Proteins (<25kDa, >75kDa)
503 were transferred (0.1 µm nitrocellulose membranes, 1h, 0.8 A cm^{-2}) as follows:
504 cathode – 5 filter papers (FP, 3 mm, Whatman) soaked in transfer buffer 1 (40 mM
505 Aminocaproic acid, 20% Isopropanol, 25 mM Tris/HCl pH 9.4) – gel – membrane – 2
506 FPs soaked in transfer buffer 2 (20% Isopropanol, 25 mM Tris/HCl pH 10.4) – 3 FPs
507 soaked in transfer buffer 3 (20% Isopropanol, 300 mM Tris/HCl pH 10.4) – anode.

508 Membranes were fixed (Ponceau Red), incubated (1h, room temperature) in block (30
509 g/L skimmed milk powder, 0.1% tween-20, 1x PBS: 140 mM NaCl, 30 mM KCl, 100
510 mM Na_2HPO_4, 15 mM KH_2PO_4), immunolabelled (overnight, 4°C) using α-FLAG 1°
511 (Sigma Aldrich F1804, diluted 1:10 000 in block), washed (0.1% tween-20, 1x PBS),
512 treated (1h, room temperature) with horseradish peroxidase-conjugated α-mouse 2°
513 (diluted 1:10 000 in block), washed, revealed (ECL; ChemiDoc, Biorad). Blots were
514 processed using ImageLab (Version 6.0.0 build 26, BioRad), final figures were
515 assembled in PowerPoint (Microsoft).

516 ***Antimicrobial activity assay***

517 Minimum inhibitory concentration assays were performed as described before[10].

518 ***Sequence Data Set***

519 TPs with experimentally-confirmed cleavage sites were recovered from proteomic
520 studies: *Chlamydomonas reinhardtii* cTP (Terashima 2011), mTP (Tardif 2012);
521 *Arabidopsis thaliana* cTP (Ge 2008, Rowland 2015), mTP (Huang 2009). For each
522 peptide, we recovered the full-length protein sequence from NCBI and UNIPROT.
523 Cytoplasmic control sequences were generated by taking N-terminal sequence
524 stretches of random length, matching the distribution of peptide lengths observed in
525 our Chlamydomonas TP dataset, from a random subset of Chlamydomonas proteins
526 with a validated cytoplasmic location in UNIPROT. For PCAs and calculating amino
527 acid frequencies, the same HA-RAMP, signal peptide and non-Chlamydomonas TP
528 sequences were used as before[10].

529 ***Amphipathic helix prediction***

530 Amphipatic α-helices are predicted as described previously[10], following the principle
531 of the Heliquest algorithm[64]. Briefly, the approach aims to establish the longest
532 sequence stretch that contains identifiable hydrophilic and hydrophobic faces. The
533 algorithm is iterated such that multiple non-overlapping helices can be found within a
534 given peptide (Fig. 1). Consequently, helix fractions are calculated as number of
535 residues within all predicted helices divided by the total number of residues in the
536 peptide; for evaluating the number of upstream residues, only the longest helix was
537 considered (Fig. 6, Fig. S24).

538 ***ACC terms***

539 In order to evaluate the physicochemical properties of our peptides we used the
540 approach we previously described[10]. Briefly, each amino acid is described in terms of
541 3 'Z-scale' values[65] that can be interpreted as representing a residue's hydrophobic,
542 steric and electronic properties. Auto-cross-covariances (ACC) between nearby
543 residues are calculated up to a distance of four amino acids, generating a quantitative
544 representation of a given peptide in terms of 36 ACC terms. Euclidian distances
545 between HA-RAMP ACC term vectors and the barycenter of Chlamydomonas TPs
546 were used as a measure of similarity (Table S1).

547 ***Visualization***

548 We performed principal component analysis (PCA) to visualize the relationships
549 between peptides as described by their 36 ACC terms (Fig. S1, S21) or by their 5
550 salient TP properties (Fig. S24) using python package sklearn v0.22.1[66].

551 ***Analysis of TP N-termini***

552 TP N-termini, defined as the N-terminal 15 amino acids, were analysed (Fig. S22) as
553 follows: Charge profiles were generated as described in the literature[27].
554 Hydrophobicity of TP N-termini was estimated using the Heliquest standalone
555 application[64]. To evaluate disorder, we used IUpred2A, a software that calculates the
556 probability for each residue of being in a disordered region[67]. We used the default
557 'Long' setting, which has been reported to be more accurate than the alternative
558 'Short' setting[68]. The disorder of a given sequence was taken as the mean of the
559 probability values for each residue (average residue disorder probability).

560 ***Statistical prediction***

561 To evaluate the predictive power of ACC terms[10] obtained for 15 residue-long TP N-
562 termini with regards to localization (Fig. S22) we used a binomial logistic regression
563 classifier. We performed one hundred 5-fold cross validation runs. In each set of one
564 hundred runs, we randomly selected sequences such that the same number of mTPs
565 and cTPs were used. For *C. reinhardtii* we used 33 mTP and 33 cTP and for *A.*
566 *thaliana* we use 29 mTP and 29 cTP sequences. We used an elastic net penalty in a
567 saga solver with an l1-ratio (a-parameter) of 0, which is equivalent to using only ridge
568 penalty, where all features take part in the model, and a C parameter ($1/\lambda$) of 0.1. The
569 a and $1/\lambda$ parameters were optimised with 10-fold cross validation. Firstly, when $1/\lambda$ =

570 1, the best accuracy (0.82) was obtained with a between 0 and 0.09. Secondly, with a
571 = 0, the best accuracy (0.82) was obtained with 1/λ = 0.1. A Logistic Regression
572 model with elastic net penalty ratio of 0.15 (scikit-learn Python package v0.22.1)
573 trained on class I HA-RAMPs[10] and Chlamydomonas TPs was used to evaluate how
574 similar potential HA-RAMP candidates are to TPs (Table S1). Custom scripts were
575 written in python (v3.7.6).

576 ***Interaction site prediction***

577 Values for the Boman index, a quantitative proxy for whether a peptide is more likely
578 to interact with proteins (high values) or lipids (low values), were calculated as
579 described in the literature[69]. Anchor2 interactivity values, a second proxy for protein
580 interaction potential developed for disordered sequences[70], were calculated using the
581 IUPred2A standalone application and webserver[67,71]. Interaction sites for Hsp70 were
582 predicted as described by Ivey and coworkers[25] based on experimental affinity values
583 for individual amino acids. Putatively TOC-interacting 'FGLK' sites were established
584 by searching for presence of F and [P or G] and [K or R] and [A or L or V], and
585 absence of [D and E], within each 8-residue window of a sequence, corresponding to
586 rule 22 by Chotewutmontri and co-workers[27] that was recommended by the authors in
587 personal communication. 'FGLK-1' sites were established the same way, but
588 requiring the presence of only three of the four positive determinants. Custom scripts
589 were implemented in R[72] (v3.6.1) using RStudio[73] (v1.2.1335).

590 ***Statistical analysis***

591 Chlamydomonas TP distributions (Fig. 5, Fig. S25) were compared using two-sided t-
592 tests (n=34 for mTPs and n=85 for cTPs), and associated mt-set and cp-set HA-
593 RAMPs were compared using one-sided t-tests based on the trends set by TPs (n=3
594 for mt-set and n=5 for cp-set HA-RAMPs) in R[72] using RStudio[73]. Multiple Kruskal
595 statistical tests are performed (same Chlamydomonas mTPs and cTPs as above, plus
596 n=382 HA-RAMPs) to evaluate the distribution of amino acids (Fig. 3c, Fig. S20) in
597 the different groups, followed by Dunn post-hoc tests (scipy v1.4.1).

598 **Acknowledgements**

599 Thank you to Tiffina Benhamou and Gabriel Chemin, who worked on aspects of this
600 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.

The following financial support is gratefully acknowledged: the Centre National de la Recherche Scientifique and Sorbonne University for annual funding to UMR7141; the Agence National de la Recherche for (a) "ChloroMitoRAMP" ANR grant (ANR-19-CE13-0009) and (b) "LabEx Dynamo" (ANR-LABX-011), both of which provided postdoctoral support to ODC, and (c) "MATHTEST" grant (ANR-18-CE13-0027) which provided doctoral support to CG; finally, the Fondation Edmond Rothschild which provided complementary financial support to ODC and CG. 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.

Author contributions

FAW, IL and YC conceptualized the project and acquired funds. ODC devised, conducted and visualized the wet-lab experimental part of the investigation, and was chiefly responsible for assigning subcellular localization interpretations based on microscopic evidence. CG carried out the dry-lab/bioinformatic part of the investigation and associated visualization under the supervision of IL, writing custom software to do so (ODC contributed to the analysis of FGLK and Hsp70-binding sites). ODC wrote the original draft of the manuscript. FAW, IL, ODC and YC reviewed and edited the manuscript. All authors approved the final manuscript. The authors declare no competing interests.

Data availability

All data are available in the main text or the supplementary materials. Custom code generated in the course of this project will be made available without restrictions upon request to the authors. All constructs will be made available through the Chlamydomonas resource centre.

References

1. Wiedemann, N., and Pfanner, N. (2017). Mitochondrial Machineries for Protein Import and Assembly. Annu. Rev. Biochem. *86*, 685–714.

631 2. Chotewutmontri, P., Holbrook, K., and Bruce, B.D. (2017). Plastid Protein
632 Targeting: Preprotein Recognition and Translocation. Int. Rev. Cell Mol. Biol.
633 *330*, 227–294.
634 3. Bruce, B.D. (2001). The paradox of plastid transit peptides: Conservation of
635 function despite divergence in primary structure. Biochim. Biophys. Acta -
636 Mol. Cell Res. *1541*, 2–21.
637 4. Tardif, M., Atteia, A., Specht, M., Cogne, G., Rolland, N., Brugière, S.,
638 Hippler, M., Ferro, M., Bruley, C., Peltier, G., et al. (2012). Predalgo: A new
639 subcellular localization prediction tool dedicated to green algae. Mol. Biol.
640 Evol. *29*, 3625–3639.
641 5. Armenteros, J.J.A., Salvatore, M., Emanuelsson, O., Winther, O., Von Heijne,
642 G., Elofsson, A., and Nielsen, H. (2019). Detecting sequence signals in
643 targeting peptides using deep learning. Life Sci. Alliance *2*, 1–14.
644 6. Lazzaro, B.P., Zasloff, M., and Rolff, J. (2020). Antimicrobial peptides:
645 Application informed by evolution. Science *368*.
646 7. Joo, H.S., Fu, C.I., and Otto, M. (2016). Bacterial strategies of resistance to
647 antimicrobial peptides. Philos. Trans. R. Soc. B Biol. Sci. *371*.
648 8. Wollman, F.-A. (2016). An antimicrobial origin of transit peptides accounts for
649 early endosymbiotic events. Traffic *17*, 1322–1328.
650 9. Caspari, O.D., and Lafontaine, I. (2021). The role of antimicrobial peptides in
651 the evolution of endosymbiotic protein import. PLoS Pathog. *17*, e1009466.
652 10. Garrido, C., Caspari, O.D., Choquet, Y., Wollman, F.A., and Lafontaine, I.
653 (2020). Evidence Supporting an Antimicrobial Origin of Targeting Peptides to
654 Endosymbiotic Organelles. Cells *9*, 1795.
655 11. von Heijne, G. (1986). Mitochondrial targeting sequences may form
656 amphiphilic helices. EMBO J. *5*, 1335–1342.
657 12. von Heijne, G., Steppuhn, J., and Herrmann, R.G. (1989). Domain structure of
658 mitochondrial and chloroplastic targeting peptides. Eur. J. Biochem. *180*, 535–
659 545.
660 13. von Heijne, G., and Nishikawa, K. (1991). Chloroplast transit peptides. The
661 perfect random coil? FEBS Lett. *278*, 1–3.
662 14. Franzén, L.-G.G., Rochaix, J.-D.D., and von Heijne, G. (1990). Chloroplast
663 transit peptides from the green alga Chlamydomonas reinhardtii share features
664 with both mitochondrial and higher plant chloroplast presequences. FEBS Lett.

665 *260*, 169–172.
666 15. Lancelin, J.M., Gans, P., Bouchayer, E., Bally, I., Arlaud, G.J., and Jacquot,
667 J.P. (1996). NMR structures of a mitochondrial transit peptide from the green
668 alga Chlamydomonas reinhardtii. FEBS Lett. *391*, 203–208.
669 16. Krimm, I., Gans, P., Hernandez, J.F., Arlaud, G.J., and Lancelin, J.M. (1999).
670 A coil-helix instead of a helix-coil motif can be induced in a chloroplast transit
671 peptide from Chlamydomonas reinhardtii. Eur. J. Biochem. *265*, 171–180.
672 17. Wienk, H.L.J., Wechselberger, R.W., Czisch, M., and De Kruijff, B. (2000).
673 Structure, Dynamics, and Insertion of a Chloroplast Targeting Peptide in
674 Mixed Micelles. Biochemistry *39*, 8219–8227.
675 18. Bruce, B.D. (1998). The role of lipids in plastid protein transport. Plant Mol.
676 Biol. *38*, 223–46.
677 19. Teixeira, P.F., and Glaser, E. (2013). Processing peptidases in mitochondria
678 and chloroplasts. Biochim. Biophys. Acta - Mol. Cell Res. *1833*, 360–370.
679 20. Köhler, D., Dobritzsch, D., Hoehenwarter, W., Helm, S., Steiner, J.M., and
680 Baginsky, S. (2015). Identification of protein N-termini in Cyanophora
681 paradoxa cyanelles: transit peptide composition and sequence determinants for
682 precursor maturation. Front. Plant Sci. *6*, 1–11.
683 21. Bionda, T., Tillmann, B., Simm, S., Beilstein, K., Ruprecht, M., and Schleiff,
684 E. (2010). Chloroplast import signals: The length requirement for translocation
685 in vitro and in vivo. J. Mol. Biol. *402*, 510–523.
686 22. Caspari, O.D. (2021). Chloroplast transit peptides often require downstream
687 unstructured sequence in Chlamydomonas reinhardtii. bioRxiv,
688 doi:10.1101/2021.11.26.470094.
689 23. Lee, D.W., Lee, S., Lee, J., Woo, S., Razzak, M.A., Vitale, A., and Hwang, I.
690 (2019). Molecular Mechanism of the Specificity of Protein Import into
691 Chloroplasts and Mitochondria in Plant Cells. Mol. Plant.
692 24. Bhushan, S., Kuhn, C., Berglund, A., Roth, C., and Glaser, E. (2006). The role
693 of the N-terminal domain of chloroplast targeting peptides in organellar protein
694 import and miss-sorting. FEBS *580*, 3966–3972.
695 25. Ivey, R. a, Subramanian, C., and Bruce, B.D. (2000). Identification of a Hsp70
696 recognition domain within the rubisco small subunit transit peptide. Plant
697 Physiol. *122*, 1289–1299.
698 26. Ivey, R. a, and Bruce, B.D. (2000). In vivo and in vitro interaction of DnaK

699 and a chloroplast transit peptide. Cell Stress Chaperones *5*, 62–71.
700 27. Chotewutmontri, P., Reddick, L.E., McWilliams, D.R., Campbell, I.M., and
701 Bruce, B.D. (2012). Differential Transit Peptide Recognition during Preprotein
702 Binding and Translocation into Flowering Plant Plastids. Plant Cell *24*, 3040–
703 3059.
704 28. Chotewutmontri, P., and Bruce, B.D. (2015). Non-native, N-terminal Hsp70
705 molecular motor recognition elements in transit peptides support plastid protein
706 translocation. J. Biol. Chem. *290*, 7602–7621.
707 29. Caspari, O.D. (2022). Chloroplast transit peptides often require downstream
708 unstructured sequence in Chlamydomonas reinhardtii. Front. Plant Sci. *13*,
709 825797.
710 30. Meyer, M., Itakura, A., Patena, W., Wang, L., He, S., Emrich-Mills, T., Lau,
711 C.S., Yates, G., Mackinder, L.C.M., and Jonikas, M.C. (2020). Assembly of
712 the algal CO2-fixing organelle, the pyrenoid, is guided by a Rubisco-binding
713 motif. Sci. Adv. *6*, eabd2408.
714 31. Patron, N.J., and Waller, R.F. (2007). Transit peptide diversity and divergence :
715 a global analysis of plastid targeting signals. BioEssays *29*, 1048–1058.
716 32. Li, L., Vorobyov, I., and Allen, T.W. (2013). The different interactions of
717 lysine and arginine side chains with lipid membranes. J. Phys. Chem. B *117*,
718 11906–11920.
719 33. Li, J., Koh, J.-J., Liu, S., Lakshminarayanan, R., Verma, C.S., and Beuerman,
720 R.W. (2017). Membrane Active Antimicrobial Peptides: Translating
721 Mechanistic Insights to Design. Front. Neurosci. *11*, 1–18.
722 34. Mattiroli, F., and Sixma, T.K. (2014). Lysine-targeting specificity in ubiquitin
723 and ubiquitin-like modification pathways. Nat. Struct. Mol. Biol. *21*, 308–316.
724 35. Calvo, S.E., Julien, O., Clauser, K.R., Shen, H., Kamer, K.J., Wells, J.A., and
725 Mootha, V.K. (2017). Comparative analysis of mitochondrial N-termini from
726 mouse, human, and yeast. Mol. Cell. Proteomics *16*, 512–523.
727 36. Lee, D.W., Lee, S., Min, C.-K., Park, C., Kim, J.-M., Hwang, C.-S., Park, S.K.,
728 Cho, N.-H., and Hwang, I. (2020). Cross-Species Functional Conservation and
729 Possible Origin of the N-Terminal Specificity Domain of Mitochondrial
730 Presequences. Front. Plant Sci. *11*, 1–7.
731 37. Garg, S.G., and Gould, S.B. (2016). The Role of Charge in Protein Targeting
732 Evolution. Trends Cell Biol. *26*, 894–905.

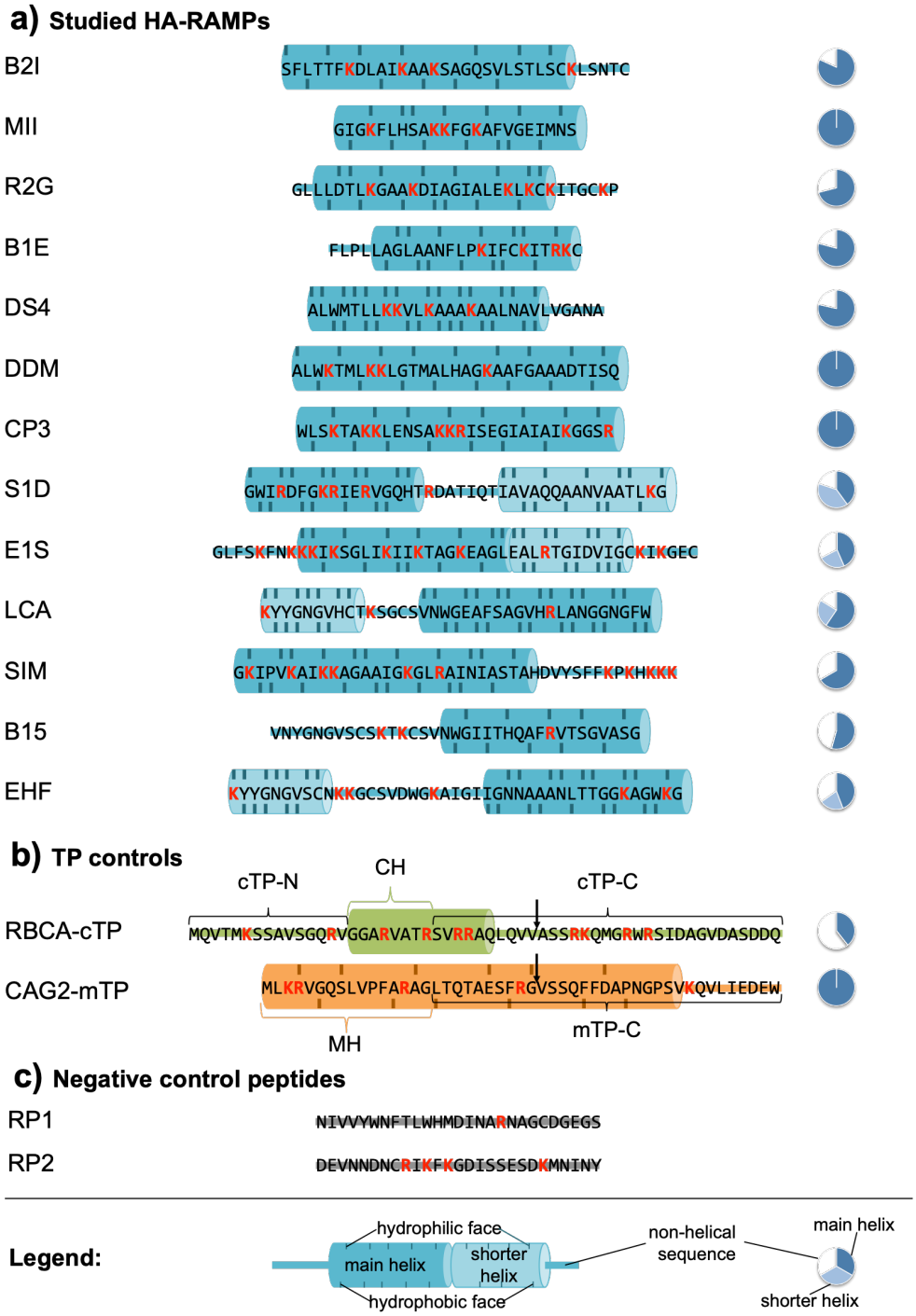
733 38. Martin, J., Mahlke, K., and Pfanner, N. (1991). Role of an energized inner
734 membrane in mitochondrial protein import: ΔΨ drives the movement of
735 presequences. J. Biol. Chem. *266*, 18051–18057.
736 39. Nakai, M. (2018). New Perspectives on Chloroplast Protein Import. Plant Cell
737 Physiol. *59*, 1111–1119.
738 40. Richardson, L.G.L., Small, E.L., Inoue, H., and Schnell, D.J. (2018). Molecular
739 topology of the transit peptide during chloroplast protein import. Plant Cell *30*,
740 1789–1806.
741 41. Rekas, A., Alattia, J.R., Nagai, T., Miyawaki, A., and Ikura, M. (2002). Crystal
742 structure of venus, a yellow fluorescent protein with improved maturation and
743 reduced environmental sensitivity. J. Biol. Chem. *277*, 50573–50578.
744 42. Holbrook, K., Subramanian, C., Chotewutmontri, P., Reddick, L.E., Wright, S.,
745 Zhang, H., Moncrief, L., and Bruce, B.D. (2016). Functional Analysis of Semi-
746 conserved Transit Peptide Motifs and Mechanistic Implications in Precursor
747 Targeting and Recognition. Mol. Plant *9*, 1286–1301.
748 43. Pilon, M., Wienk, H., Sips, W., De Swaaf, M., Talboom, I., Van't Hof, R., De
749 Korte- Kool, G., Demel, R., Weisbeek, P., and De Kruijff, B. (1995).
750 Functional domains of the ferredoxin transit sequence involved in chloroplast
751 import. J. Biol. Chem. *270*, 3882–3893.
752 44. Lee, D.W., Lee, S., Oh, Y.J., and Hwang, I. (2009). Multiple sequence motifs
753 in the rubisco small subunit transit peptide independently contribute to Toc159-
754 dependent import of proteins into chloroplasts. Plant Physiol. *151*, 129–141.
755 45. Razzak, A., Lee, D.W., Yoo, Y., and Hwang, I. (2017). Evolution of rubisco
756 complex small subunit transit peptides from algae to plants. Sci. Rep. *7*, 9279.
757 46. Pinnaduwage, P., and Bruce, B.D. (1996). In vitro interaction between a
758 chloroplast transit peptide and chloroplast outer envelope lipids is sequence-
759 specific and lipid class-dependent. J. Biol. Chem. *271*, 32907–32915.
760 47. Baker, A., and Schatz, G. (1987). Sequences from a prokaryotic genome or the
761 mouse dihydrofolate reductase gene can restore the import of a truncated
762 precursor protein into yeast mitochondria. Proc. Natl. Acad. Sci. *84*, 3117–
763 3121.
764 48. Lemire, B.D., Fankhauser, C., Baker, A., and Schatz, G. (1989). The
765 mitochondrial targeting function of randomly generated peptide sequences
766 correlates with predicted helical amphiphilicity. J. Biol. Chem. *264*, 20206–

767 20215.
768 49. Kmiec, B., Teixeira, P.F., and Glaser, E. (2014). Shredding the signal: targeting
769 peptide degradation in mitochondria and chloroplasts. Trends Plant Sci. *19*,
770 771–8.
771 50. Lee, D.W., and Hwang, I. (2021). Understanding the evolution of
772 endosymbiotic organelles based on the targeting sequences of organellar
773 proteins. New Phytol. *230*, 924–930.
774 51. Wunder, T., Martin, R., Löffelhardt, W., Schleiff, E., and Steiner, J.M. (2007).
775 The invariant phenylalanine of precursor proteins discloses the importance of
776 Omp85 for protein translocation into cyanelles. BMC Evol. Biol. *7*, 236.
777 52. Robert, V., Volokhina, E.B., Senf, F., Bos, M.P., Van Gelder, P., and
778 Tommassen, J. (2006). Assembly factor Omp85 recognizes its outer membrane
779 protein substrates by a species-specific C-terminal motif. PLoS Biol. *4*, 1984–
780 1995.
781 53. Knopp, M., Garg, S.G., Handrich, M., and Gould, S.B. (2020). Major Changes
782 in Plastid Protein Import and the Origin of the Chloroplastida. iScience *23*,
783 100896.
784 54. Sommer, M.S., Daum, B., Gross, L.E., Weis, B.L.M., Mirus, O., Abram, L.,
785 Maier, U.G., Kühlbrandt, W., and Schleiff, E. (2011). Chloroplast Omp85
786 proteins change orientation during evolution. Proc. Natl. Acad. Sci. U. S. A.
787 *108*, 13841–13846.
788 55. Bullmann, L., Haarmann, R., Mirus, O., Bredemeier, R., Hempel, F., Maier,
789 U.G., and Schleiff, E. (2010). Filling the gap, evolutionarily conserved Omp85
790 in plastids of chromalveolates. J. Biol. Chem. *285*, 6848–6856.
791 56. Paila, Y.D., Richardson, L.G.L., Inoue, H., Parks, E.S., McMahon, J., Inoue,
792 K., and Schnell, D.J. (2016). Multi-functional roles for the polypeptide
793 transport associated domains of Toc75 in chloroplast protein import. Elife *5*, 1–
794 29.
795 57. Chen, Y.L., Chen, L.J., and Li, H.M. (2016). Polypeptide transport-associated
796 domains of the Toc75 channel protein are located in the intermembrane space
797 of chloroplasts. Plant Physiol. *172*, 235–243.
798 58. Gross, L.E., Spies, N., Simm, S., and Schleiff, E. (2020). Toc75-V/OEP80 is
799 processed during translocation into chloroplasts, and the membrane-embedded
800 form exposes its POTRA domain to the intermembrane space. FEBS Open Bio

801 *10*, 444–454.
802 59. Mergaert, P. (2018). Role of antimicrobial peptides in controlling symbiotic
803 bacterial populations. Nat. Prod. Rep. *35*, 336–356.
804 60. Guefrachi, I., Pierre, O., Timchenko, T., Alunni, B., Barrière, Q., Czernic, P.,
805 Villaécija-Aguilar, J.-A., Verly, C., Bourge, M., Fardoux, J., et al. (2015).
806 Bradyrhizobium BclA Is a Peptide Transporter Required for Bacterial
807 Differentiation in Symbiosis with Aeschynomene Legumes. Mol. Plant-
808 Microbe Interact. *28*, 1155–1166.
809 61. Caspari, O.D. (2020). Introduction of a leaky stop codon as molecular tool in
810 Chlamydomonas reinhardtii. PLoS One *15*, e0237405.
811 62. Onishi, M., and Pringle, J.R. (2016). Robust Transgene Expression from
812 Bicistronic mRNA in the Green Alga *Chlamydomonas reinhardtii*. G3
813 (Bethesda). *6*, 4115–4125.
814 63. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M.,
815 Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012).
816 Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*,
817 676–682.
818 64. Gautier, R., Douguet, D., Antonny, B., and Drin, G. (2008). HELIQUEST: a
819 web server to screen sequences with specific alpha-helical properties.
820 Bioinformatics *24*, 2101–2102.
821 65. Hellberg, S., Sjostrom, M., Skagerberg, B., and Wold, S. (1987). Peptide
822 Quantitative Structure-Activity Relationships, a Multivariate Approach. J.
823 Med. Chem. *30*, 1126–1135.
824 66. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O.,
825 Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-
826 learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.
827 67. Erdős, G., and Dosztányi, Z. (2020). Analyzing Protein Disorder with
828 IUPred2A. Curr. Protoc. Bioinforma. *70*, 1–15.
829 68. Nielsen, J.T., and Mulder, F.A.A. (2019). Quality and bias of protein disorder
830 predictors. Sci. Rep. *9*, 1–11.
831 69. Boman, H.G. (2003). Antibacterial peptides: basic facts and emerging
832 concepts. J. Intern. Med. *254*, 197–215.
833 70. Mészáros, B., Simon, I., and Dosztányi, Z. (2009). Prediction of Protein
834 Binding Regions in Disordered Proteins. PLoS One *5*, e1000376.

71. Mészáros, B., Erdös, G., and Dosztányi, Z. (2018). IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res. *46*, W329–W337.
72. R Core Team (2019). R: A Language and Environment for Statistical Computing.
73. RStudio Team (2020). RStudio: Integrated Development for R.

843 **Figures and legends**



844

845 **Fig. 1. Peptide sequences under study.** Amino acid sequences are shown using the
846 one-letter code. Positively charged residues are highlighted in red. The fraction of the
847 sequence predicted to fold into amphipathic helices (see Methods) is provided by a
848 pie-chart to the right of the sequence; for TPs in (**b**) this was calculated up to the
849 cleavage site indicated by a downward arrow. Predicted amphipathic helices are
850 highlighted using a cylinder cartoon, with residues contributing to the
851 hydrophilic/hydrophobic face indicated on the top/bottom. No helix could be

852 predicted within RBCA-cTP, and thus the indicated helix is taken from a published
853 NMR-structure obtained in membrane-mimetic conditions[16]. Note that the two helices
854 of E1S are at an angle to each other and therefore cannot form a single continuous
855 amphipathic helix. Abbreviations: (**a**) B2I: Brevinin-2ISb, R2G: Ranatuerin-2G, MII:
856 Magainin 2, B1E: Brevinin-1E, DS4: Dermaseptin S4, DDM: Dermadistinctin-M,
857 CP3: Cecropin-P3, S1D: Sarcotoxin-1D, E1S: Esculentin-1SEA, LCA: Leucocin-A,
858 SIM: SI Moricin, B15: Bacillocin 1580, EHF: Enterocin HF; (**b**) TP – targeting
859 peptide, cTP: chloroplast TP, mTP: mitochondrial TP, RBCA: Rubisco activase,
860 CAG2: γ-carbonic anhydrase, cTP-N: cTP N-terminal element, CH: cTP helix, MH:
861 mTP helix, cTP-C: cTP C-terminal element, mTP-C: mTP C-terminal element; (**C**)
862 RP: random peptide.
863

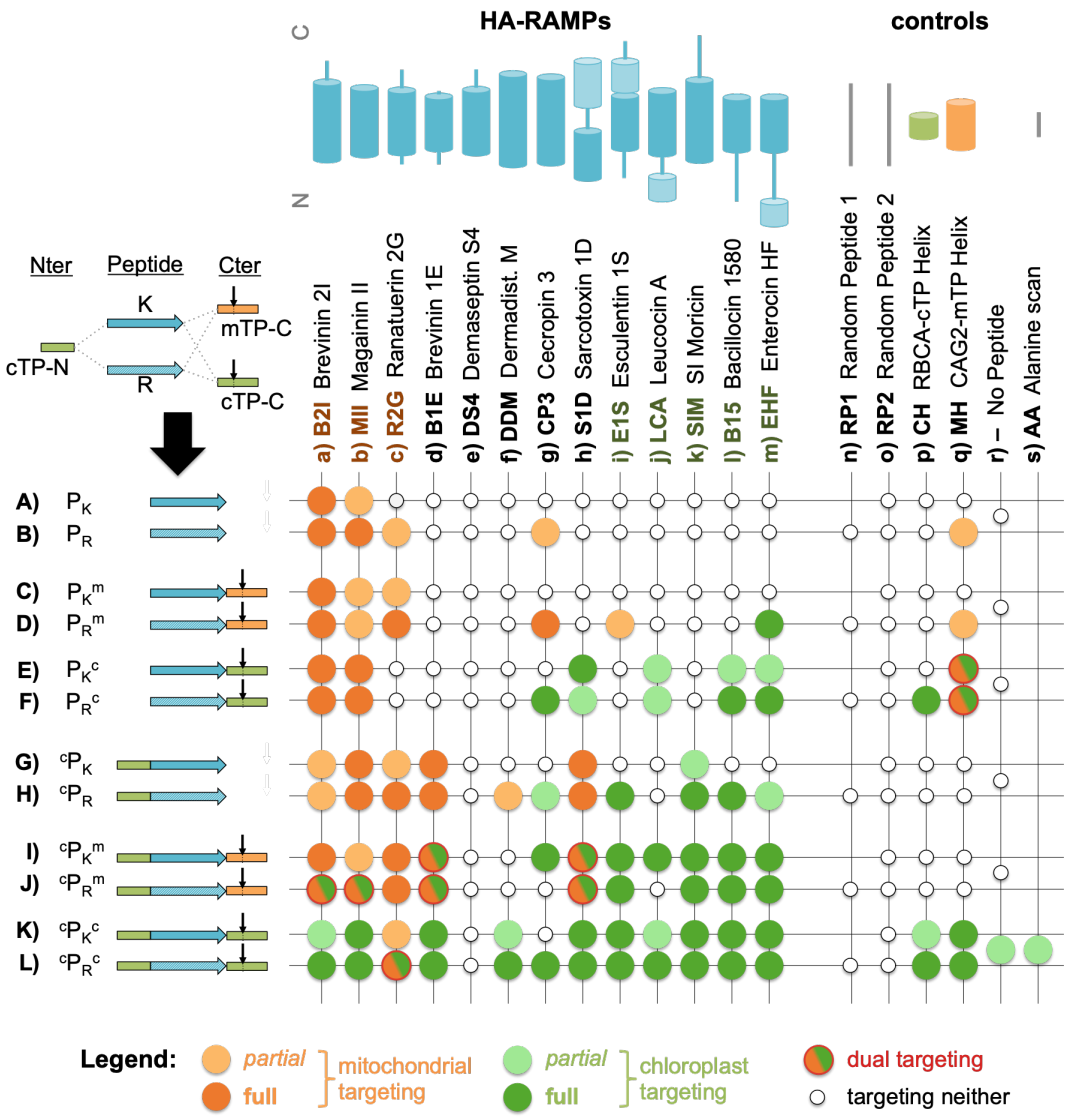


Fig. 2. TP modifications enable HA-RAMP targeting. Chimeric constructs were generated by combining TP elements with HA-RAMPs (see Fig. 1 for sequences; in the ${}^cAA^c$ construct in column s, all residues of the helical element "CH" within RBCA-cTP are replaced by Alanines). The overview graph shows in each column (**a-s**) one of the peptides with a cartoon indicating the position of the predicted amphipathic helices within the sequence; and in each row (**A-L**) a combination of peptide, K/R modification, addition of cTP-N (the 15 N-terminal residues upstream of the helix in RBCA-cTP) and/or a C-terminal TP element (cTP-C or mTP-C, which include -10 residues upstream and +23 residues downstream of the cleavage site for RBCA-cTP and CAG2-mTP respectively) indicated by a cartoon and the following shorthand: P = peptide, $_K$ = contains mostly Lysines, $_R$ = contains mostly Arginines, ^{m} = mTP element, ^{c} = cTP element (${}^cP$ = cTP-N, P^c = cTP-C). In each case, an overview of observed targeting is provided by a colour code (see figure legend). Images for all

878 constructs are shown in Fig. S2-S19. Note that the present results for $B2I_K^c$, MII_K^c,
879 $S1D_K^c$, $B15_K^c$ and EHF_K^c (row E, columns a,b,h,l,m) confirmed our previous report
880 on these strains[10].
881

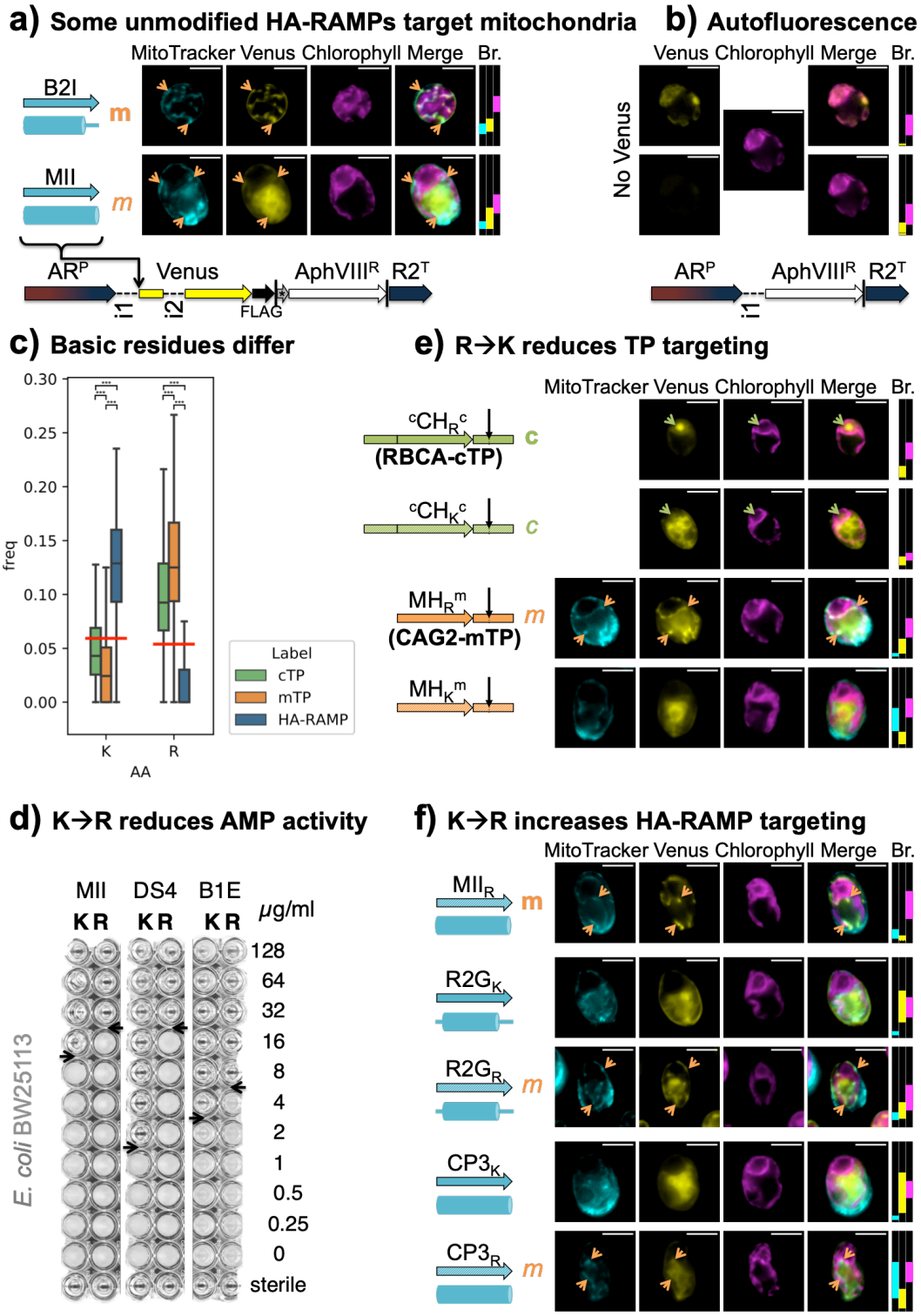


Fig. 3. K is for killing, R is for targeting. (**a**) Indicated peptides were inserted upstream of a Venus fluorescent protein reporter in a bicistronic expression system for Chlamydomonas[61]. AR^P: hybrid *HSP70a-RBCS2* promoter, i1: *RBCS2* intron 1, i2: *RBCS2* intron 2, FLAG: FLAG-tag, |: stop codon, *: bicistronic bridge sequence tagcat, $AphVIII^R$: Paromomycin resistance gene, $R2^T$: *RBCS2* terminator. Epifluorescence microscopy images of selected examples are shown. False-coloured

889 yellow fluorescence from the Venus channel reports on the subcellular localization of
890 the fluorescent reporter. MitoTracker fluorescence, false-coloured in cyan, indicates
891 the position of mitochondria (although parts of the cell exterior are sometimes also
892 stained), with salient features highlighted with orange arrows to indicate co-
893 localisation with the Venus channel. Chlorophyll autofluorescence, shown in
894 magenta, indicates the location of the chloroplast. Scale bars are 5µm. Refer to Fig. 1
895 for sequences and Fig. S2 and S3 for biological replicates. Where a construct was
896 interpreted as generating reporter localization in mitochondria or chloroplast, this is
897 indicated by an orange 'm' or a green 'c' respectively, in bold for full targeting or in
898 italics for partial targeting. Brightness (Br.) was adjusted for clarity: fluorescence
899 intensity values were restricted to the range shown for each channel by matching
900 coloured rectangles. Intensity scales to 0 at the bottom of the panel, and to 65535 at
901 the top. (**b**) A 'no Venus' control strain, expressing an empty vector, is shown with
902 two different Venus channel brightness settings to visualize chloroplast
903 autofluorescence in the Venus channel. Autofluorescence intensity is typically below
904 2000, indicated by a black dotted line in Brightness rectangles. Therefore, if Venus
905 channel brightness is adjusted below 2000, autofluorescence originating from the
906 chloroplast may be misinterpreted as Venus located in the chloroplast. **(c)** Lysine (K)
907 and Arginine (R) frequencies for Chlamydomonas TPs and HA-RAMPs are shown as
908 boxplots (center line: median; box limits: upper and lower quartiles; whiskers:
909 min/max values within 1.5x interquartile range). To give a baseline for comparison,
910 their average across the UNIPROT database is given as red horizontal line.
911 Statistically significant differences are indicated with stars: (Multiple Kruskal-Wallis
912 tests followed by Dunn post-hoc tests, ***$p<0.0001$). See Fig. S20 for all amino
913 acids. (**d**) Grayscale photos of antimicrobial activity growth assays in the presence of
914 dilutions of three selected antimicrobial peptides against *E. coli* strain BW25113
915 show reduced activity when natively K-rich sequences (K, left hand columns) are
916 altered by replacing all Ks by Rs (R, right hand columns). (**e,f**) Epifluorescence
917 images of selected chimeric (**e**) TP control or (**f**) HA-RAMP constructs are shown as
918 in (a). MH: mTP helical fragment, CH: cTP helical fragment.
919

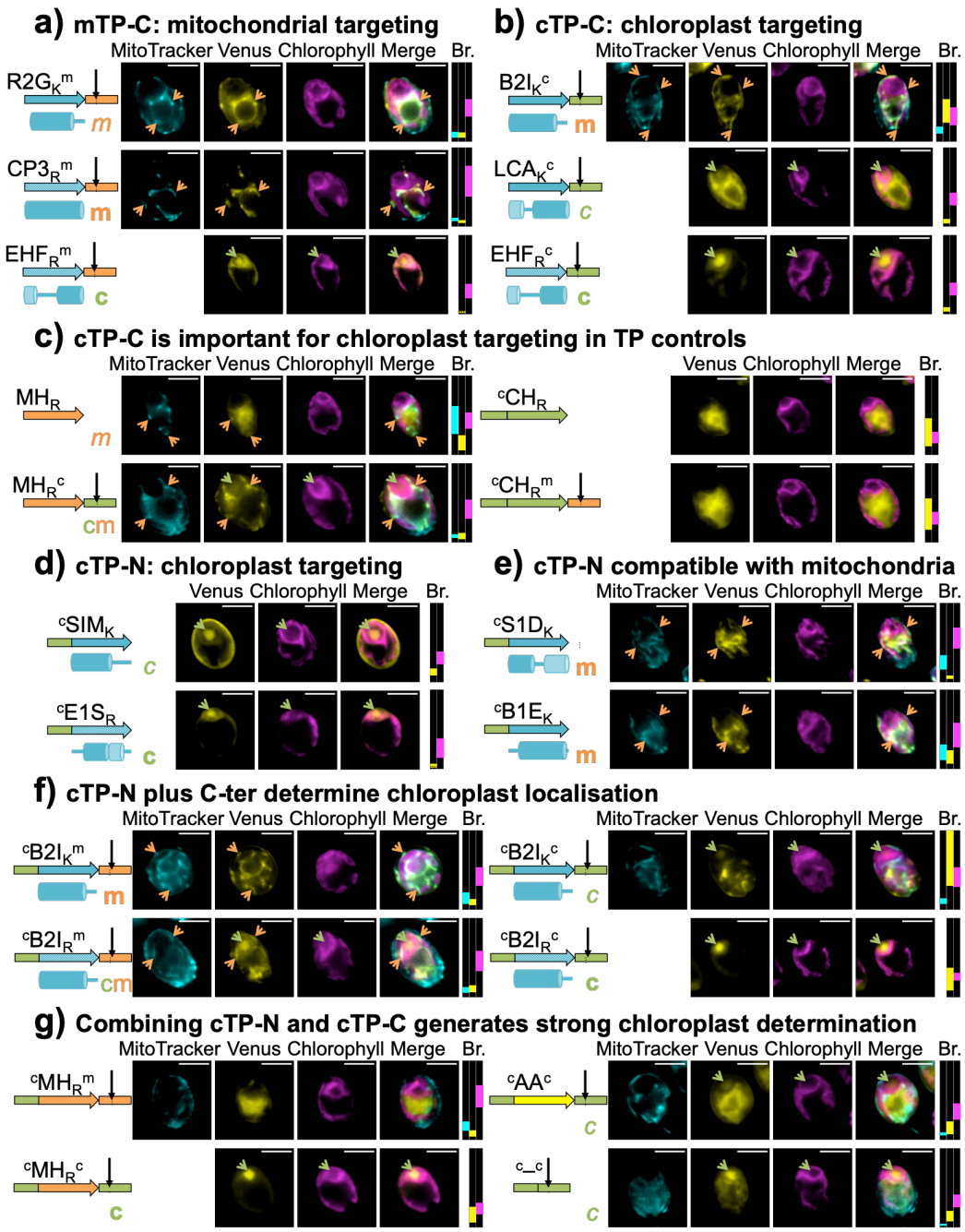


Fig. 4. TP N- and C-termini enable chloroplast targeting by HA-RAMPs. Exemplary epifluorescence images are shown as in Fig. 3; same convention as in Fig. 3 for Venus localisation. The result for $B2I_K^c$ in (**b**) confirms our previous report on this construct[10].

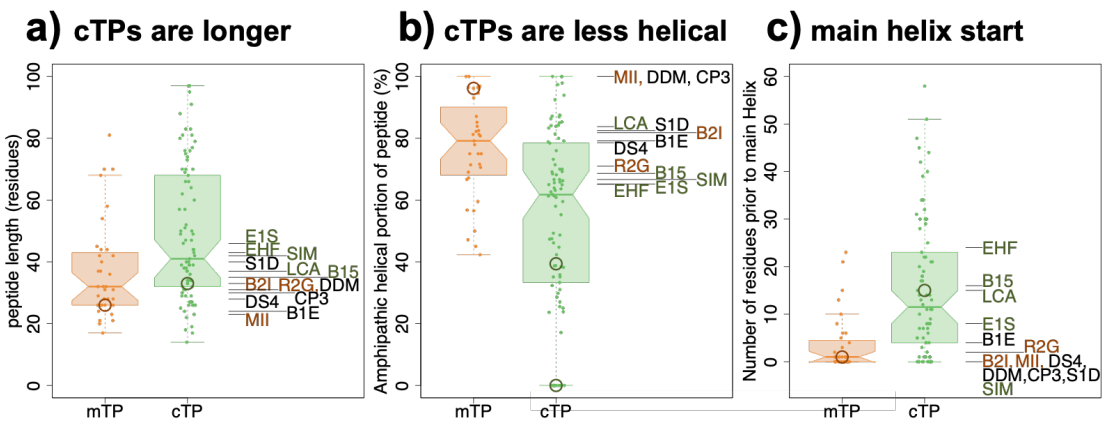


Fig. 5. Chloroplast and mitochondrial targeting HA-RAMPs match cTPs and mTPs respectively. For salient properties, Chlamydomonas mTPs and cTPs are compared to our 13 HA-RAMPs. TP distributions are shown as boxplots (center line: median; box limits: upper and lower quartiles; whiskers: min/max values 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[16], since no helix could be predicted by our approach. HA-RAMPs are colour-coded by preferred targeting: orange = mt-set; green = cp-set (cf. Fig. 2). (**a**) cTPs are significantly longer than mTPs (p=0.0017), and cp-set are longer than mt-set HA-RAMPs (p=0.0184). (**b**) Helices make up a significantly smaller fraction of cTPs than mTPs (p<0.0001), and HA-RAMPs show a similar trend (0.1122). (**c**) cTPs contain significantly longer sequence stretches upstream of the longest predicted helix than mTPs (p<0.0001), as do cp-set compared to mt-set HA-RAMPs (p=0.0205). Reported p-values were obtained through two-way t-tests for TPs and one-way t-tests for HA-RAMPs.

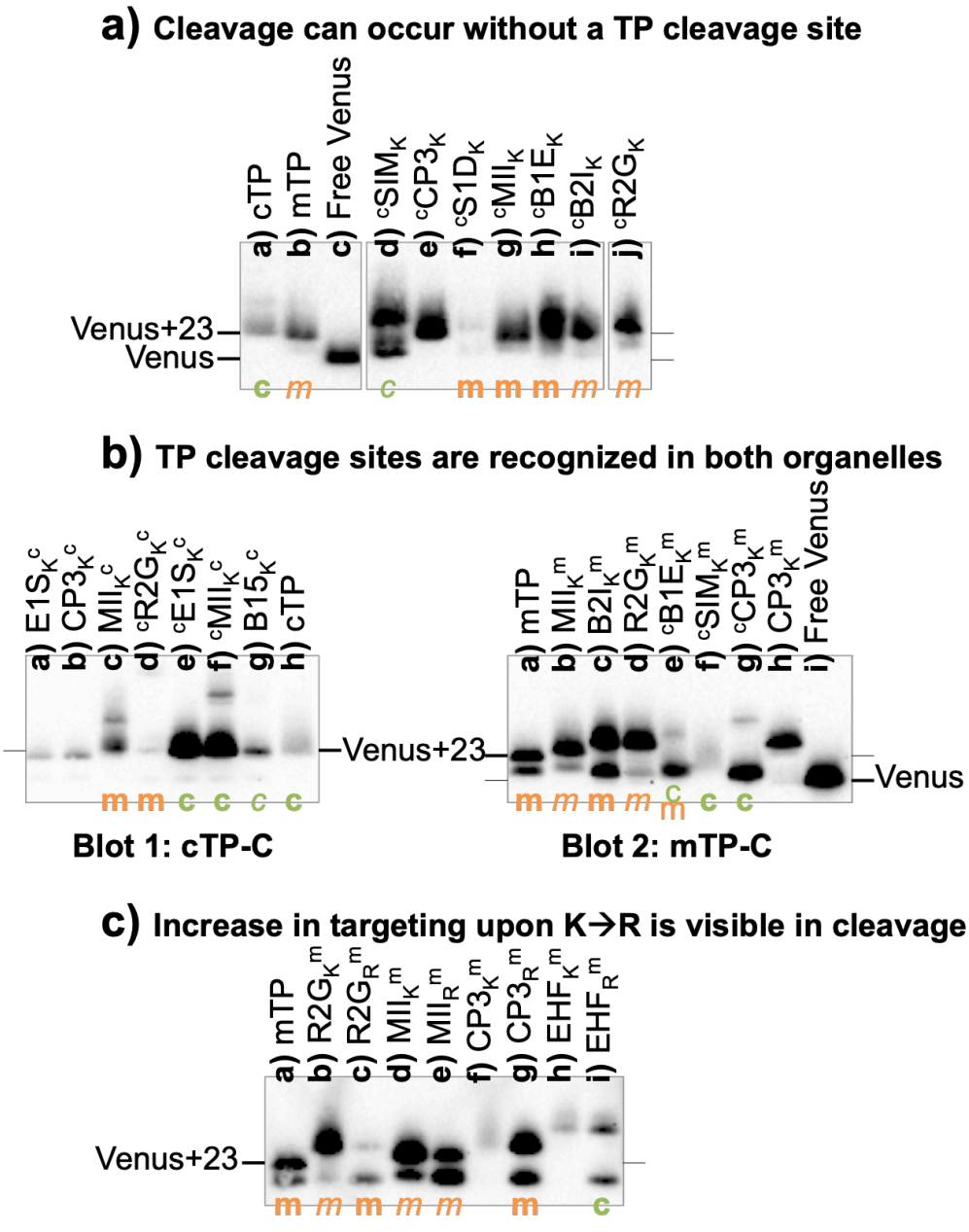


Fig. 6. Import is associated with maturation of the preprotein. Western Blots used an α-FLAG antibody on selected constructs as indicated above blots. Where a construct was interpreted as generating reporter localization by fluorescence microscopy (cf. Fig. 2, Fig. S2-S19) in mitochondria or chloroplast, this is indicated by an orange 'm' or a green 'c' respectively, in bold for full targeting or in italics for partial targeting. In (**a**) some lanes were spliced for clarity; the uncropped blot is provided in Fig. S27.

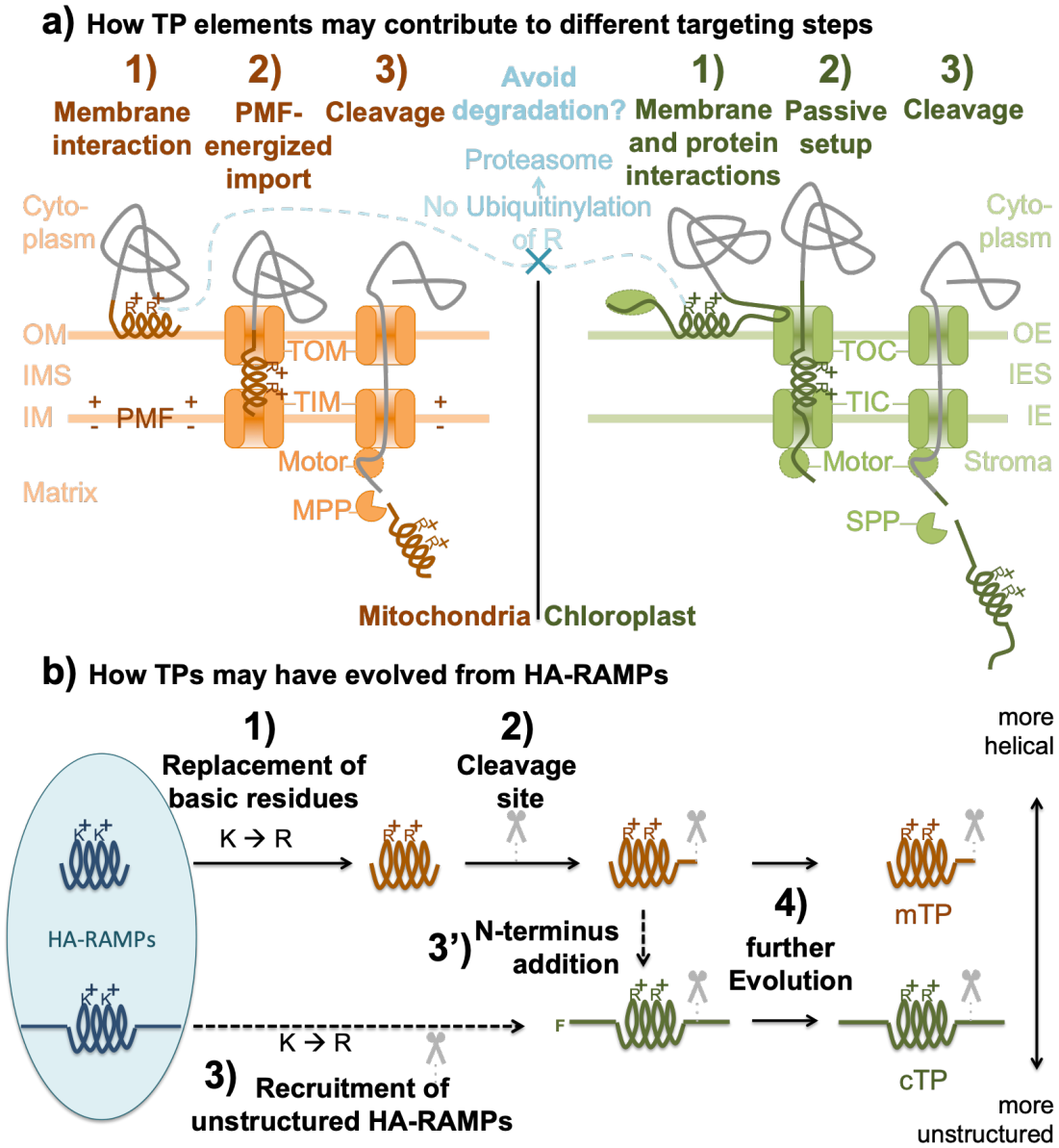


Fig. 7. Proposed functioning and evolution of TPs. Abbreviations: OM – outer membrane, IMS – inter-membrane space, IM – inner membrane, OE – outer envelope, IES – inter-envelope space, IE – inner envelope, PMF – proton motive force, TOM – translocon of the outer mitochondrial membrane, TIM – translocon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

965 interactions may play a role in enabling differential targeting due to organelle-specific
966 lipid preferences of TP helices. Protein interactions by cTP N- and C-terminal
967 elements, e.g. with cytosolic factors or TOC-components (subunits are not shown for
968 simplicity), may also play a role in specific targeting. (2) Import of positively-charged
969 mTPs across IM is energized by the PMF. By contrast, a passive setup is required for
970 cTPs to allow N-termini to reach into the stroma and contact the motor complex,
971 likely contributing to increased length and relatively unstructured N- and C-termini of
972 cTPs. (3) Sequence elements contributing to targeting may be present downstream of
973 cleavage sites in cTPs. (**b**) If TPs evolved from AMPs, 1) HA-RAMPs would have
974 first changed K to R and 2) acquired a cleavage site to become mTPs. To generate
975 cTPs, either 3) more unstructured HA-RAMPs were recruited directly to become
976 cTPs by undergoing a K to R shift and acquiring a cleavage site separately, or 3')
977 mTPs acquired an N-terminal non-helical domain. Early cTP N-termini likely
978 contained a starting F^{51}. 4) Further evolution would have reinforced the differences
979 between cTPs and mTPs to limit mis-targeting.
980

981 **Supporting Information**

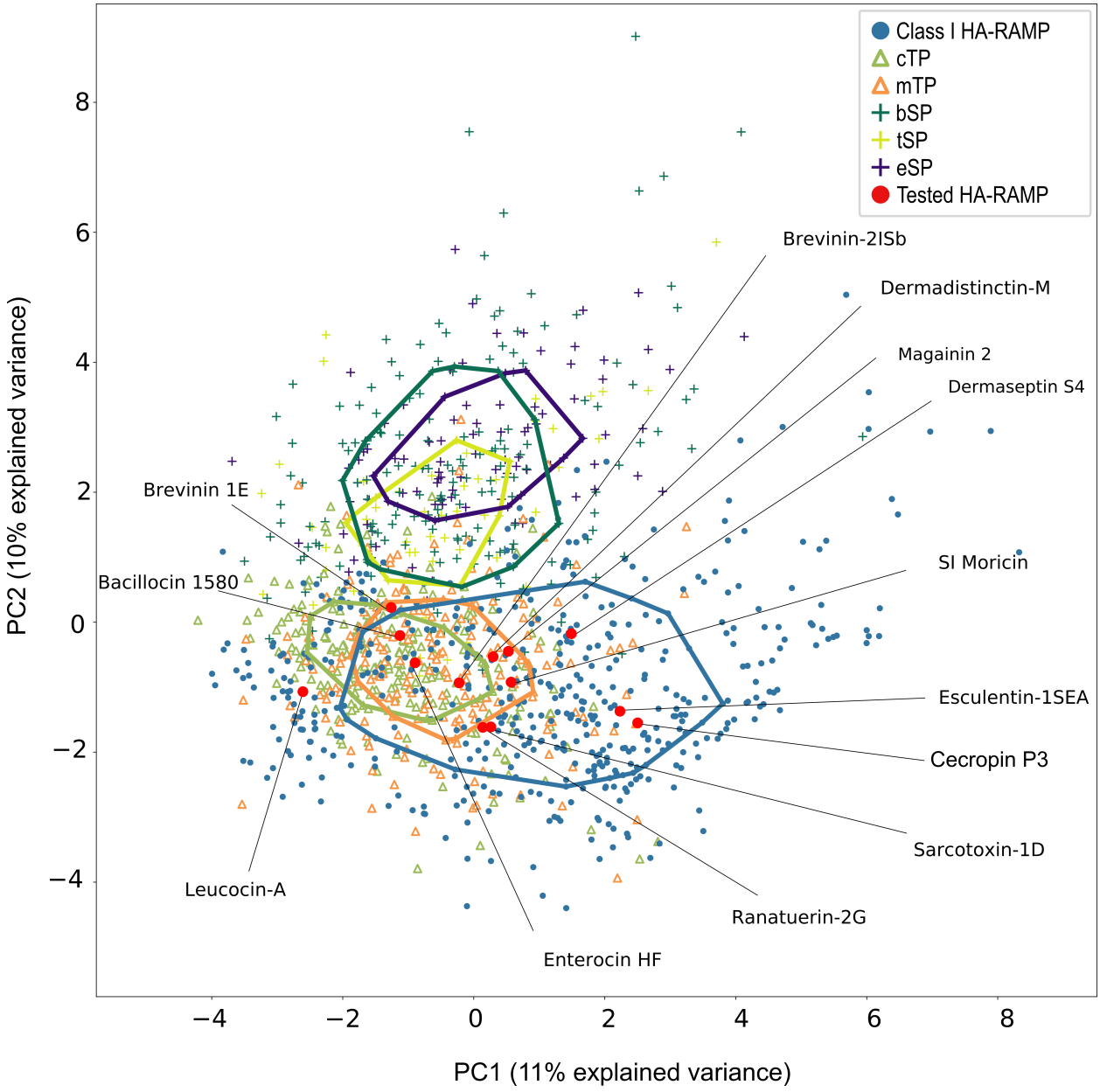
982 **Table S1** Similarity of HA-RAMPs to TPs quantified (in Excel file)

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

984

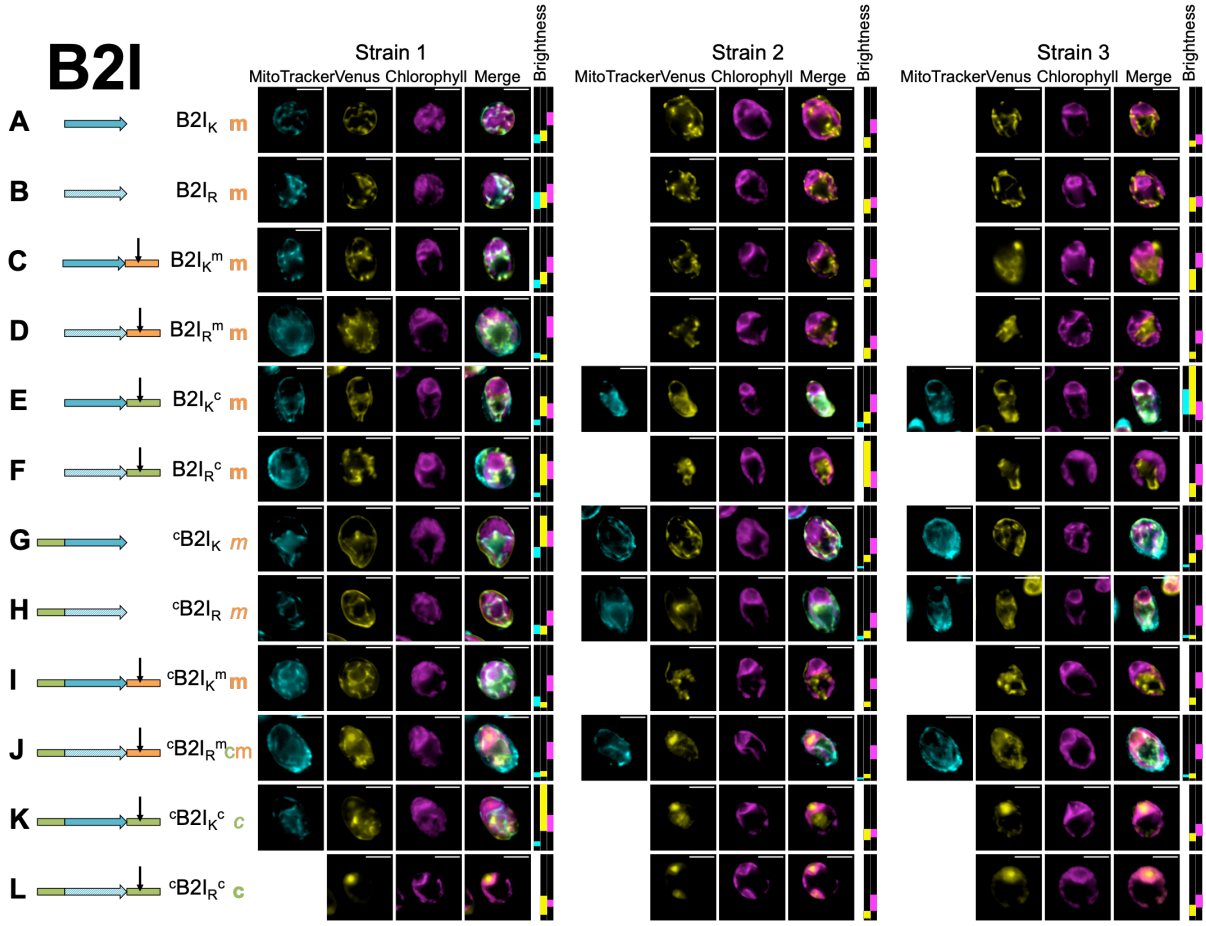
985

986 **Fig. S1. HA-RAMP candidates cover a diversity of physico-chemical properties.**
987 A Principal component analysis (PCA) based on auto-cross-correllated (ACC) Z-scale
988 values reflects divergent physico-chemical properties of Signal Peptides (SP: bacterial
989 secretory peptides – bSP, thylakoid signal peptides – tSP, eukaryotic signal peptides –
990 eSP) relative to Targeting Peptides (TP: chloroplast transit peptides – cTP,
991 mitochondrial targeting peptides – mTP) and helical-amphipathic ribosomally-
992 produced antimicrobial peptides (HA-RAMP, shown here are class I HA-RAMPs
993 *sensu* Garrido *et al.*[10]). Axes are principal components (PC) 1 and 2. Each dot
994 represents one peptide; the 13 HA-RAMP candidates studied in this article are
995 highlighted in red. Convex areas include the 50% of peptides at the centre of each
996 distribution.



997

998 **Fig. S2. Microscopy for biological replicates of Brevinin 2ISb.** Three independent
999 transformant lines (Strains 1-3) are shown for each combination of modifications (A-
1000 L), represented by a cartoon and a shorthand description (cf. Fig. 2). Where a
1001 construct was interpreted as generating reporter localization in mitochondria or
1002 chloroplast, this is indicated by an orange 'm' or a green 'c' respectively, in bold for
1003 full targeting or in italics for partial targeting. Scale bars are 5µm.



1004
1005

Fig. S3. Microscopy for biological replicates of Magainin 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.

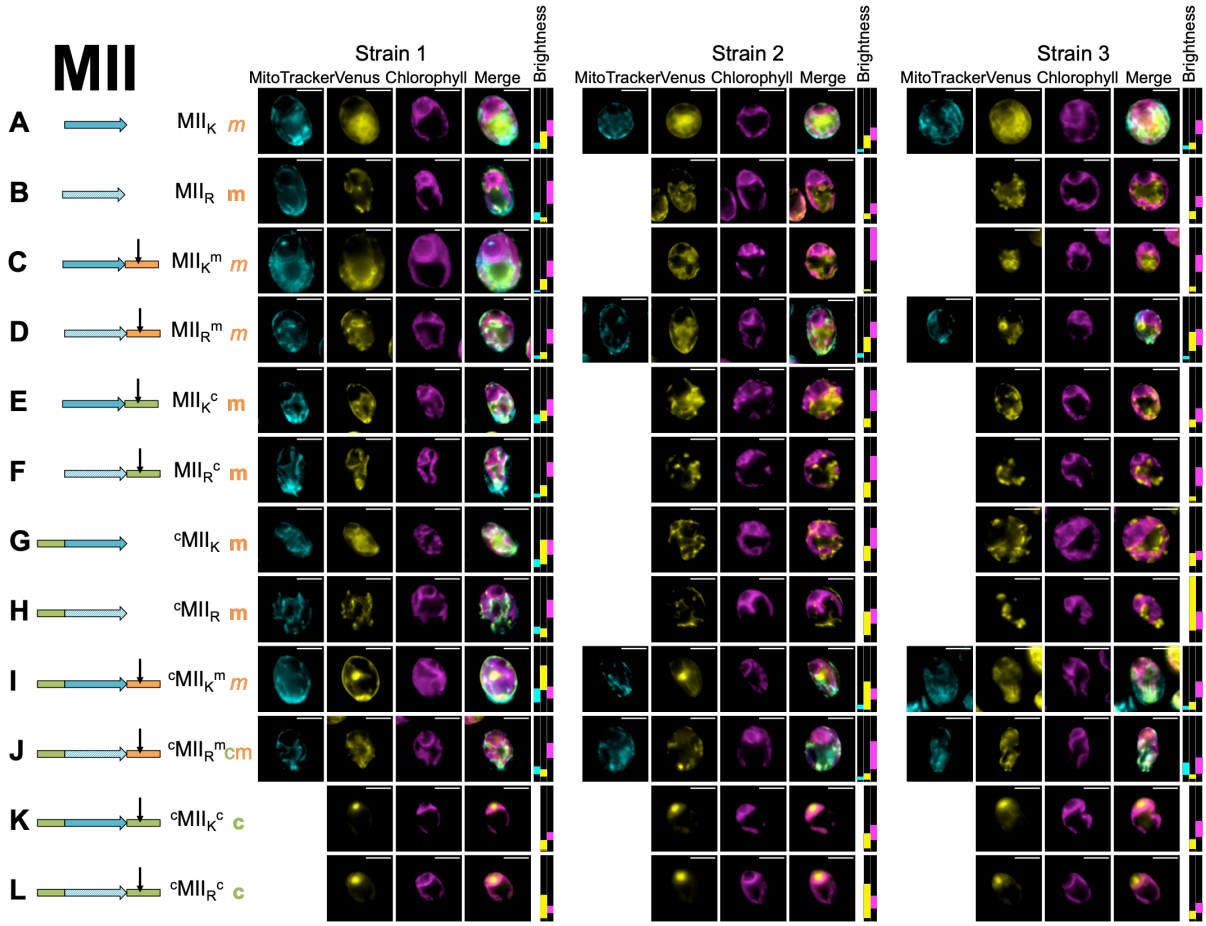


Fig. S4. Microscopy for biological replicates of Ranatuerin 2G. Three independent transformant lines (Strains 1-3) are shown for each combination of modifications (A-L), represented by a cartoon and a shorthand description (cf. Fig. 2). Where a construct was interpreted as generating reporter localization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

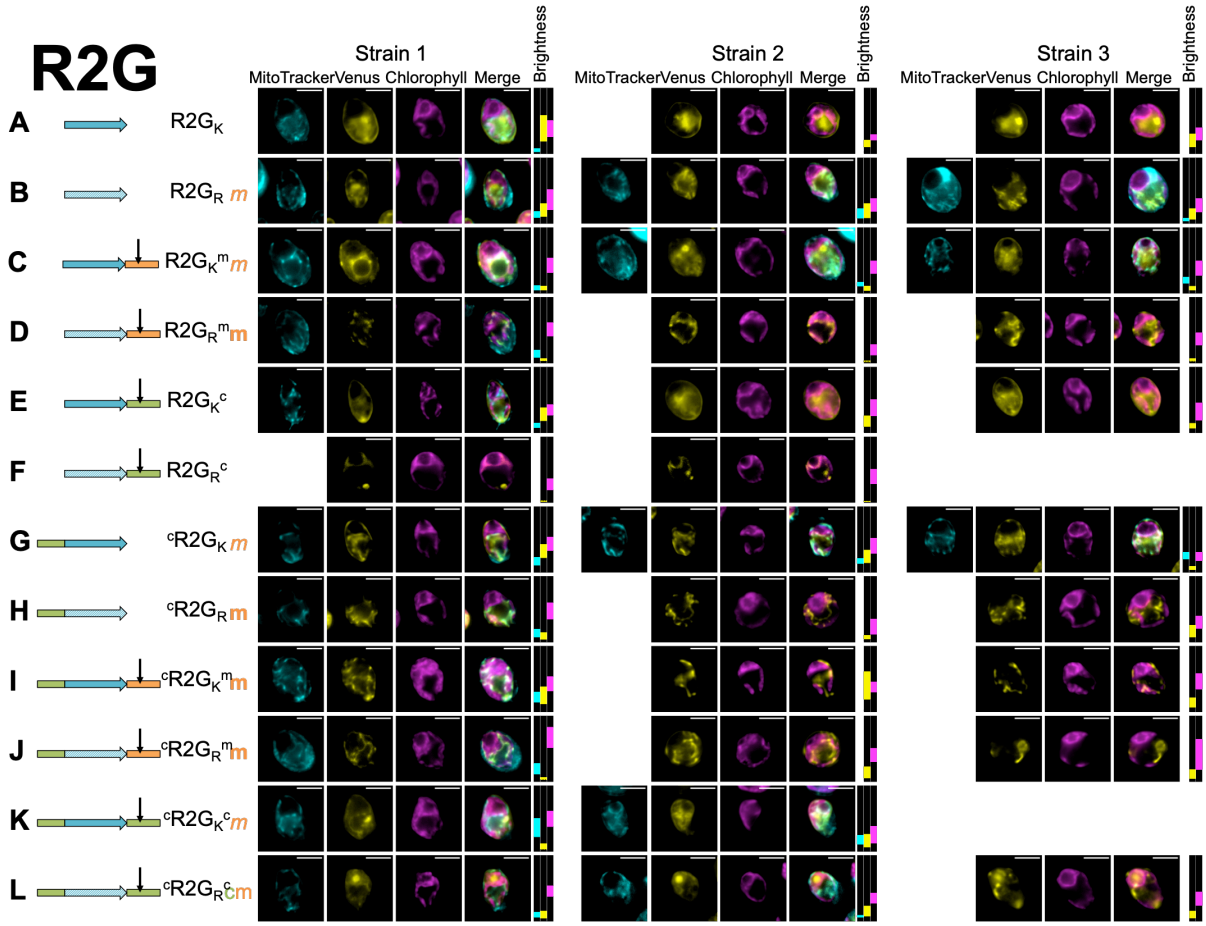
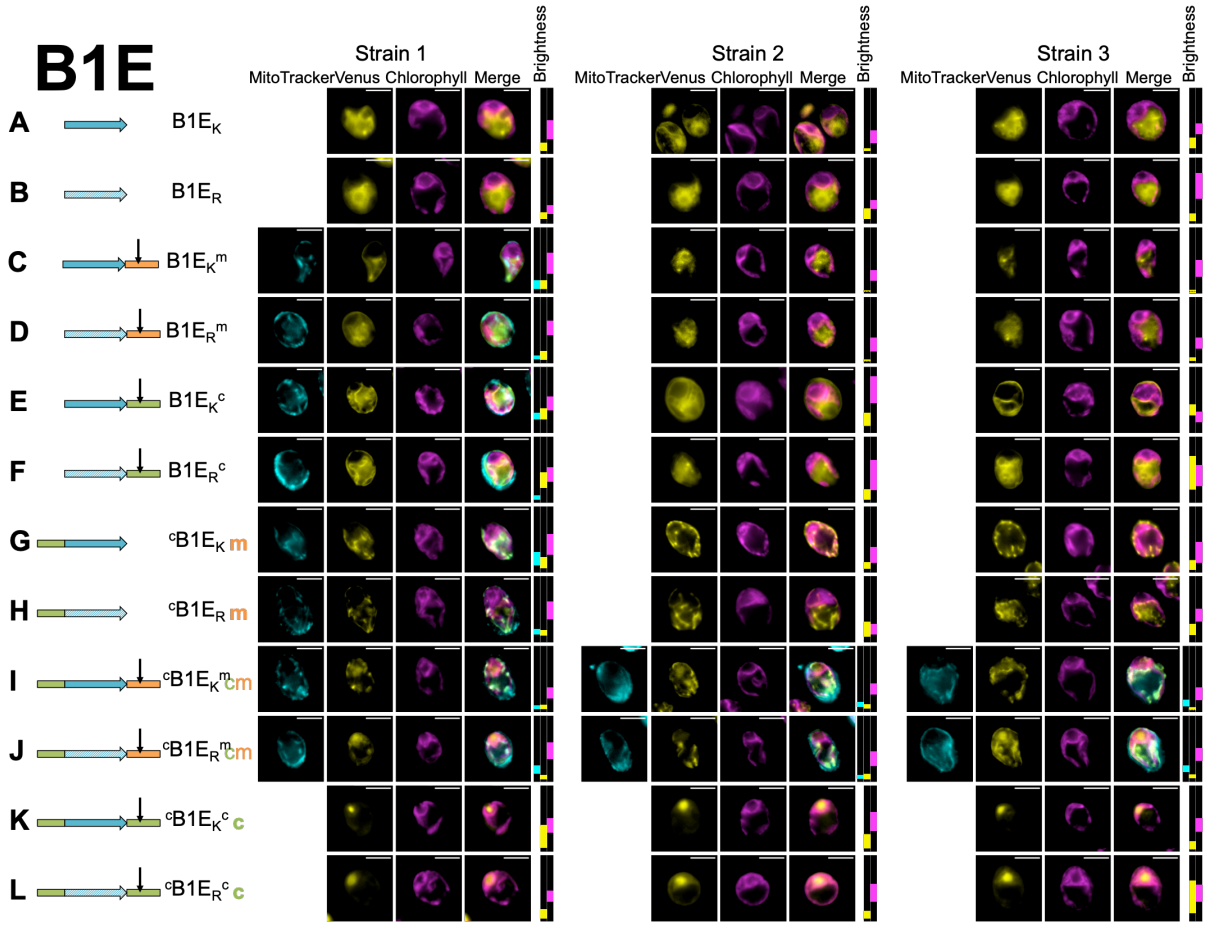
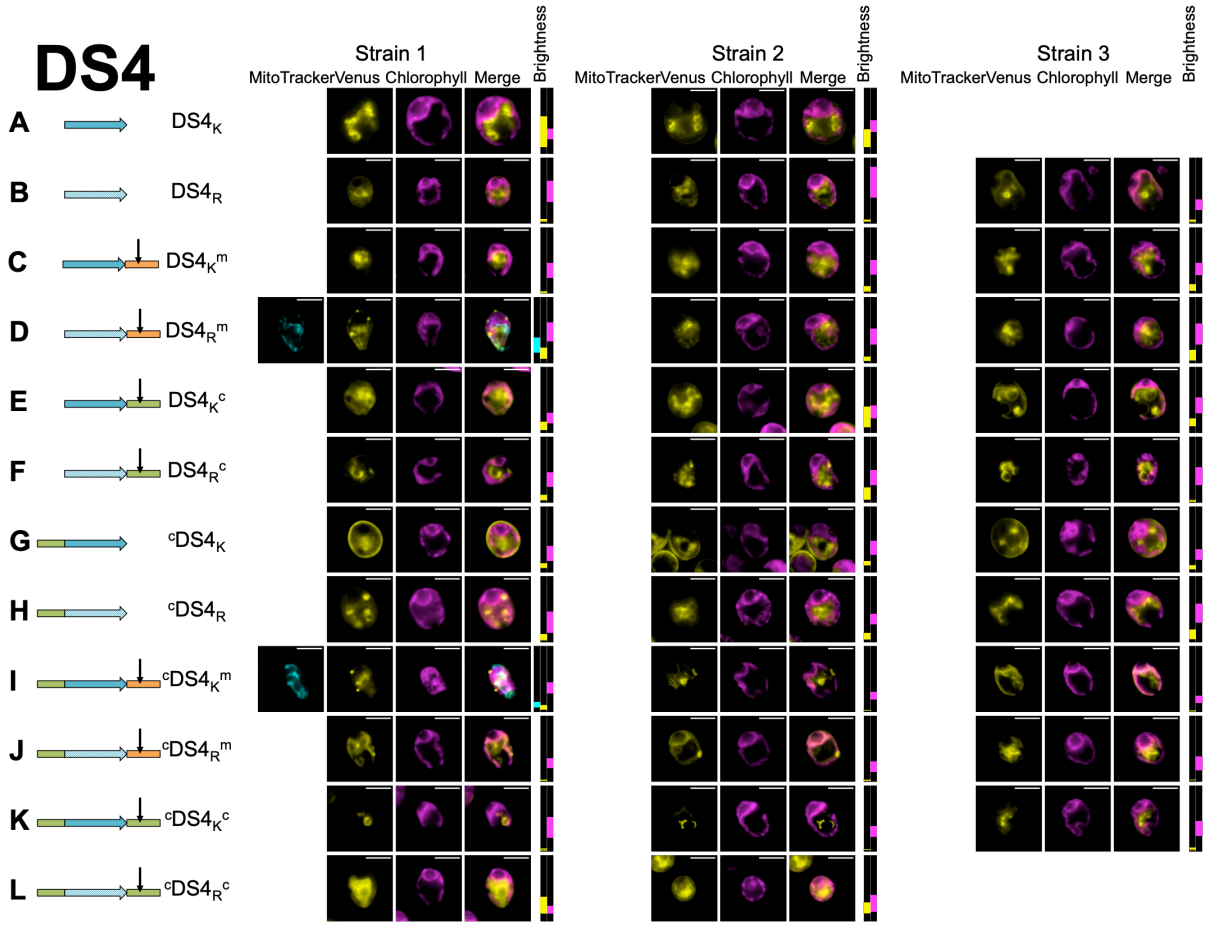


Fig. S5. Microscopy for biological replicates of Brevinin 1E. Three independent transformant lines (Strains 1-3) are shown for each combination of modifications (A-L), represented by a cartoon and a shorthand description (cf. Fig. 2). Where a construct was interpreted as generating reporter localization 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.



1033

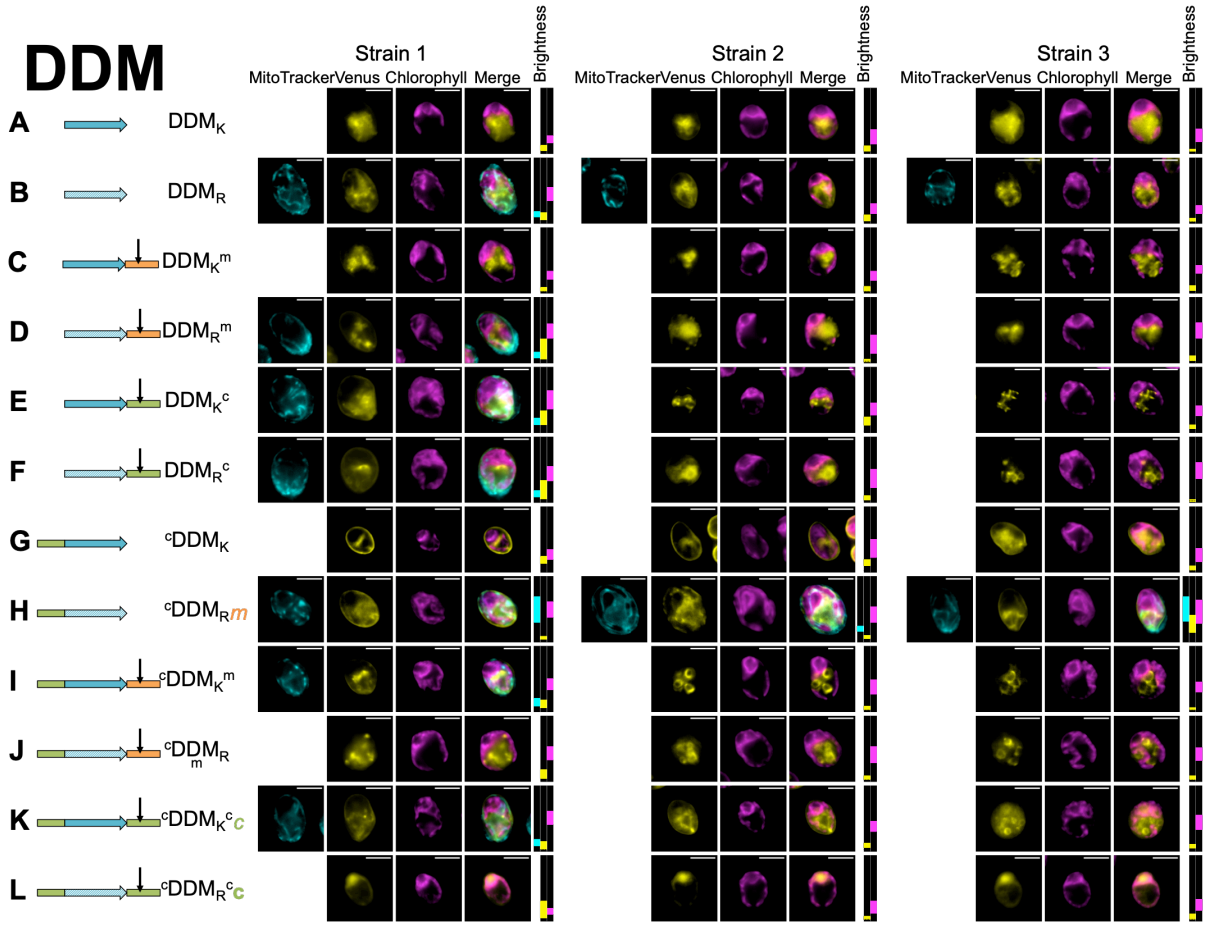
1034 **Fig. S6. Microscopy for biological replicates of Dermaseptin S4.** Three
1035 independent transformant lines (Strains 1-3) are shown for each combination of
1036 modifications (A-L), represented by a cartoon and a shorthand description (cf. Fig. 2).
1037 Where a construct was interpreted as generating reporter localization in mitochondria
1038 or chloroplast, this is indicated by an orange 'm' or a green 'c' respectively, in bold
1039 for full targeting or in italics for partial targeting. Scale bars are 5µm.



1040
1041

1042

1043 **Fig. S7. Microscopy for biological replicates of Dermadistinctin M.** Three
1044 independent transformant lines (Strains 1-3) are shown for each combination of
1045 modifications (A-L), represented by a cartoon and a shorthand description (cf. Fig. 2).
1046 Where a construct was interpreted as generating reporter localization in mitochondria
1047 or chloroplast, this is indicated by an orange 'm' or a green 'c' respectively, in bold
1048 for full targeting or in italics for partial targeting. Scale bars are 5µm.



1049

1050

Fig. S8. Microscopy for biological replicates of Cecropin P3. 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.

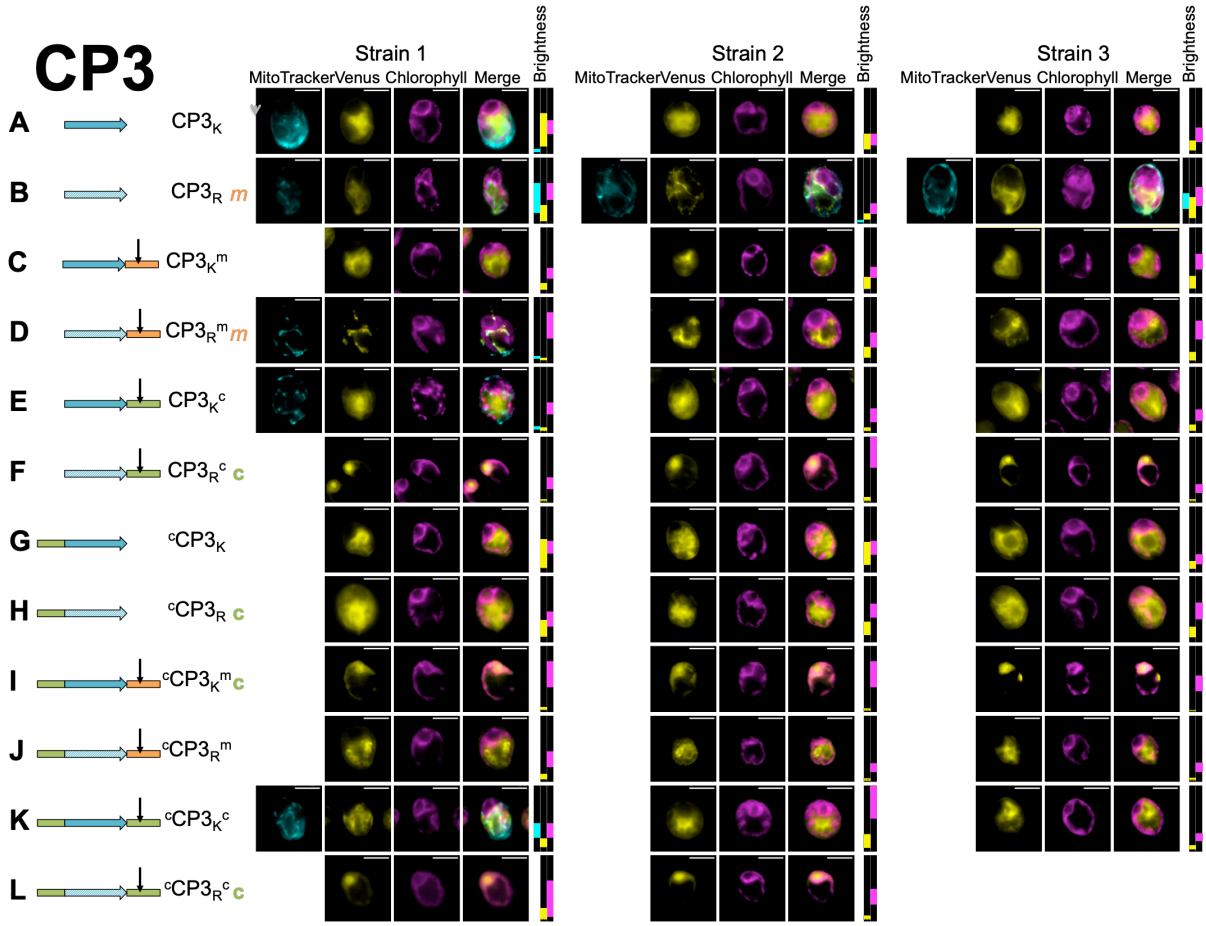
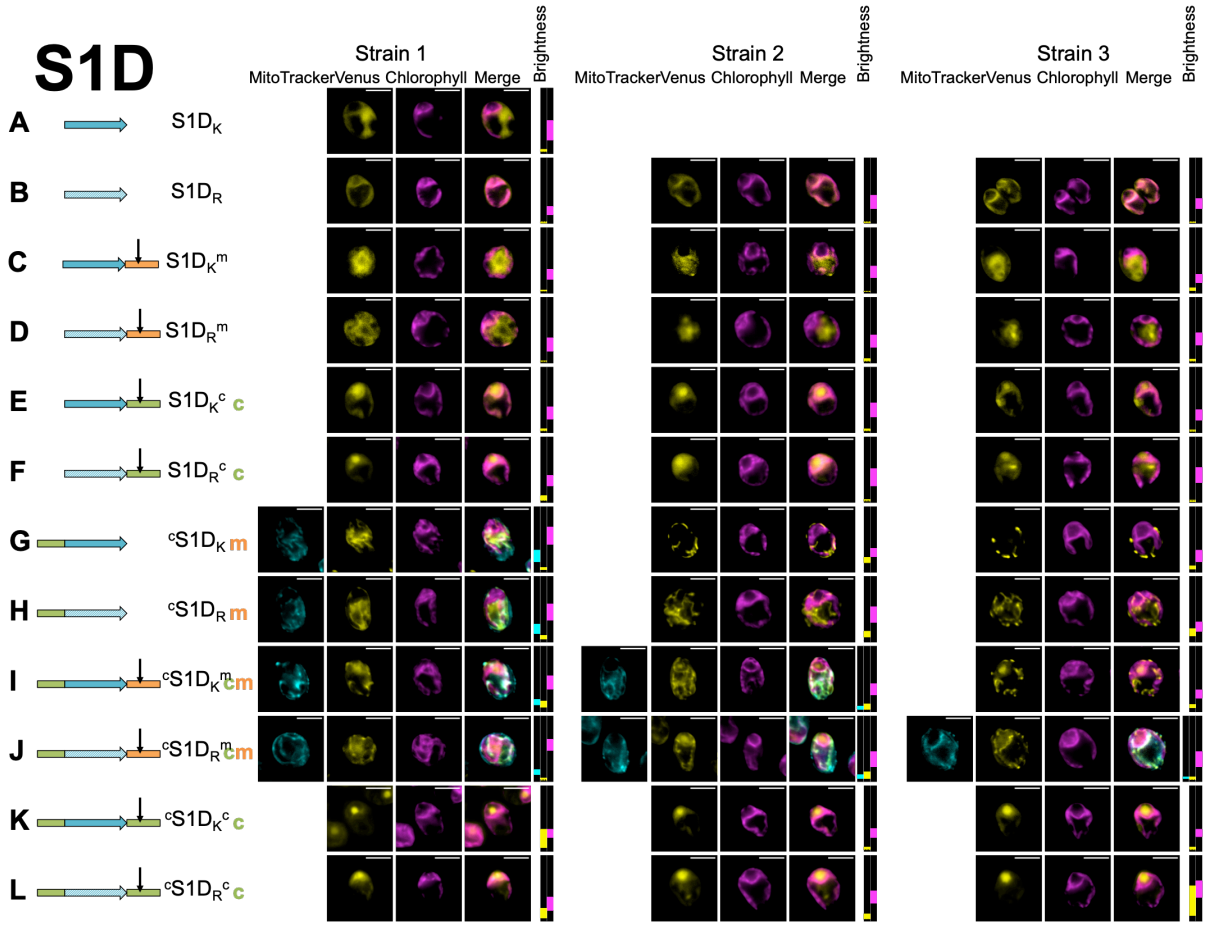
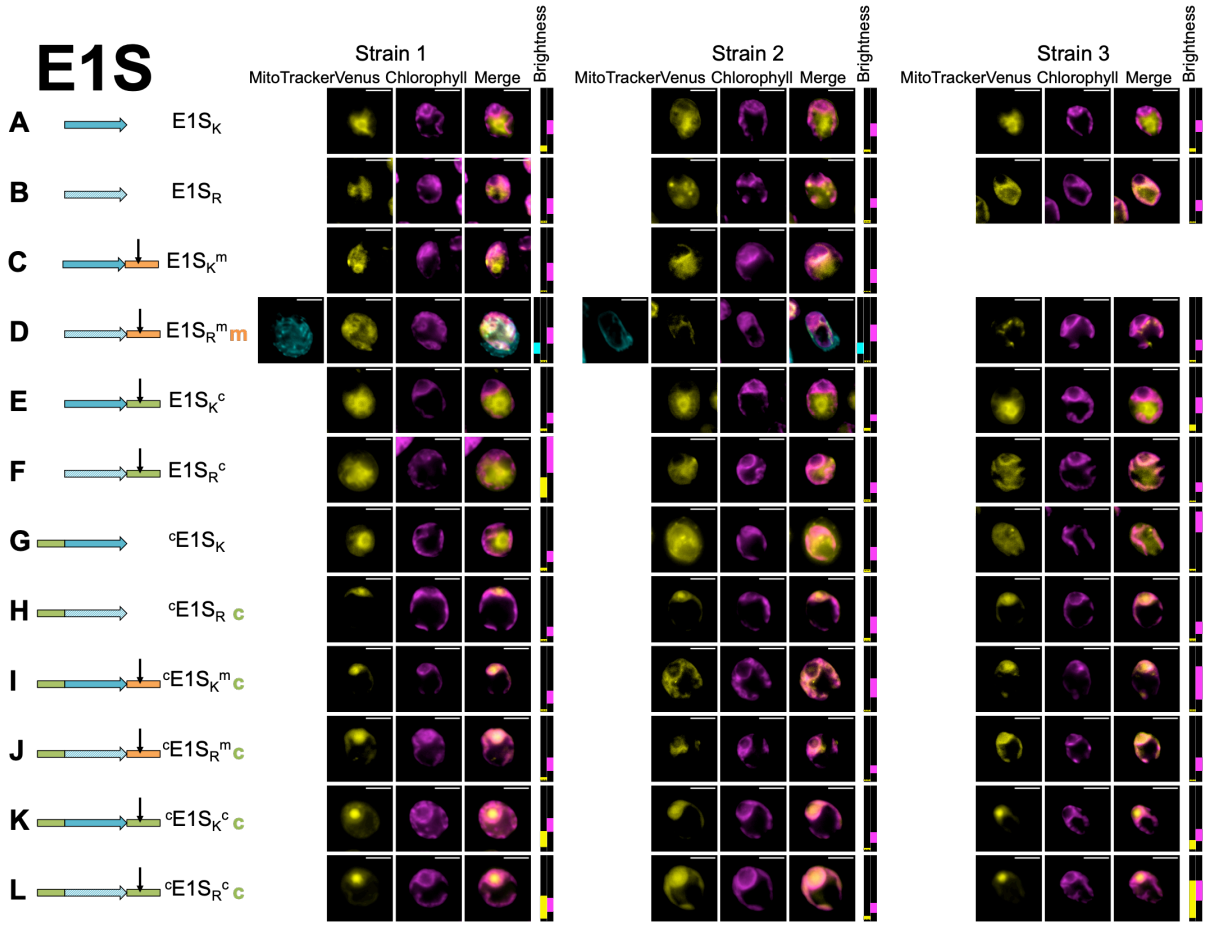


Fig. S9. Microscopy for biological replicates of Sarcotoxin 1D. Three independent transformant lines (Strains 1-3) are shown for each combination of modifications (A-L), represented by a cartoon and a shorthand description (cf. Fig. 2). Where a construct was interpreted as generating reporter localization 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.



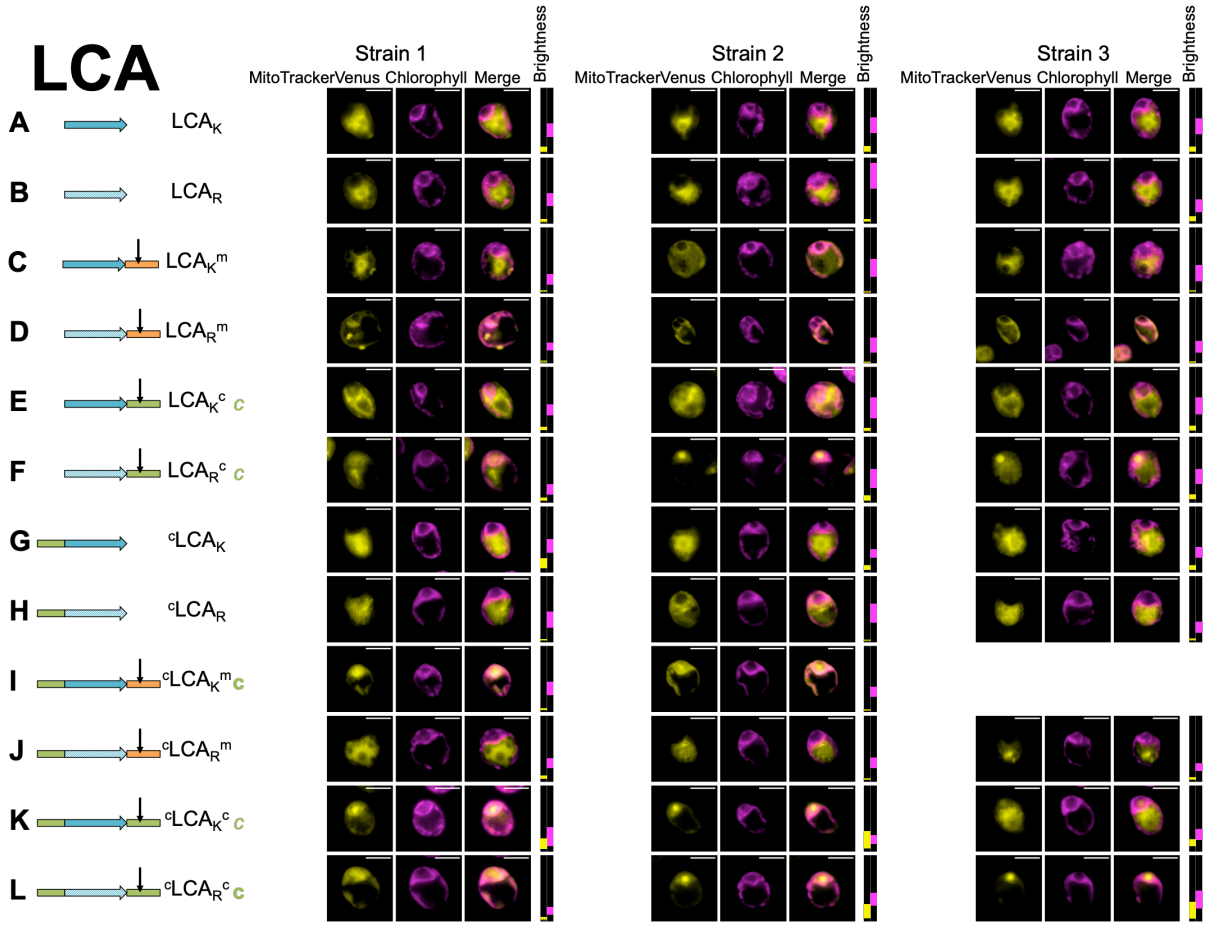
1069

1070 **Fig. S10. Microscopy for biological replicates of Esculentin 1S.** Three independent
1071 transformant lines (Strains 1-3) are shown for each combination of modifications (A-
1072 L), represented by a cartoon and a shorthand description (cf. Fig. 2). Where a
1073 construct was interpreted as generating reporter localization in mitochondria or
1074 chloroplast, this is indicated by an orange 'm' or a green 'c' respectively, in bold for
1075 full targeting or in italics for partial targeting. Scale bars are 5µm.



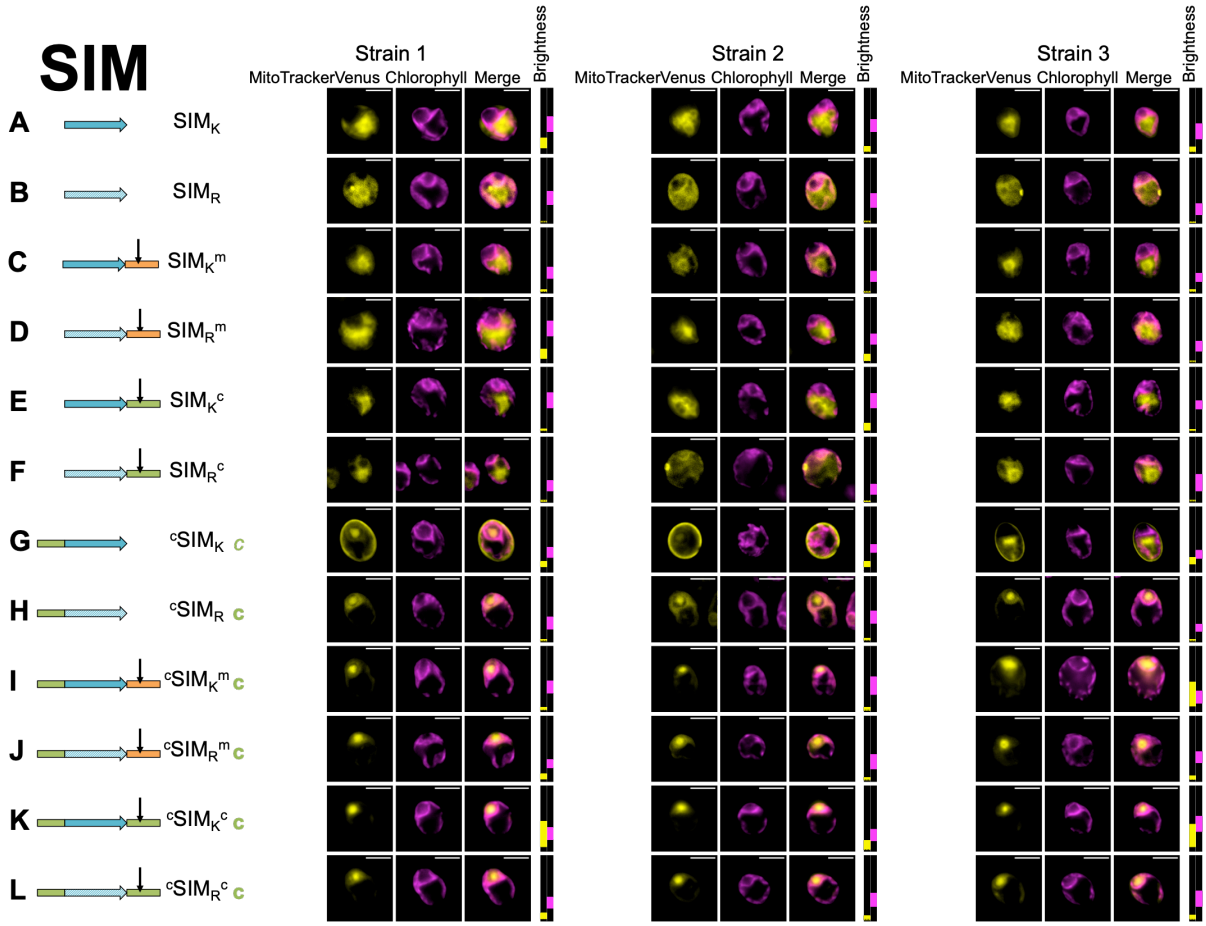
1076
1077

Fig. S11. Microscopy for biological replicates of Leucocin A. 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.



1087

1088 **Fig. S12. Microscopy for biological replicates of SI Moricin.** Three independent
1089 transformant lines (Strains 1-3) are shown for each combination of modifications (A-
1090 L), represented by a cartoon and a shorthand description (cf. Fig. 2). Where a
1091 construct was interpreted as generating reporter localization in mitochondria or
1092 chloroplast, this is indicated by an orange 'm' or a green 'c' respectively, in bold for
1093 full targeting or in italics for partial targeting. Scale bars are 5µm.



1094
1095

Fig. S13. 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.

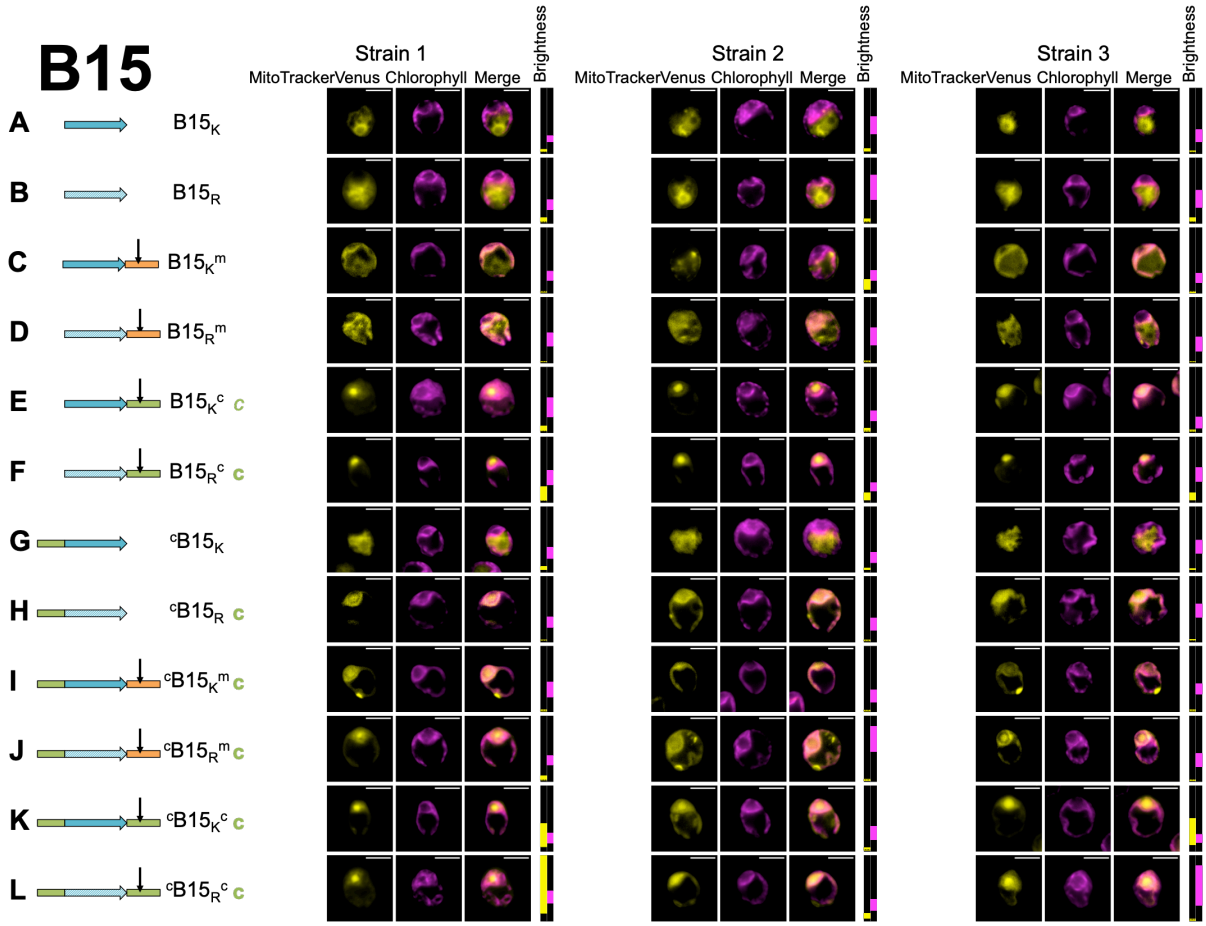
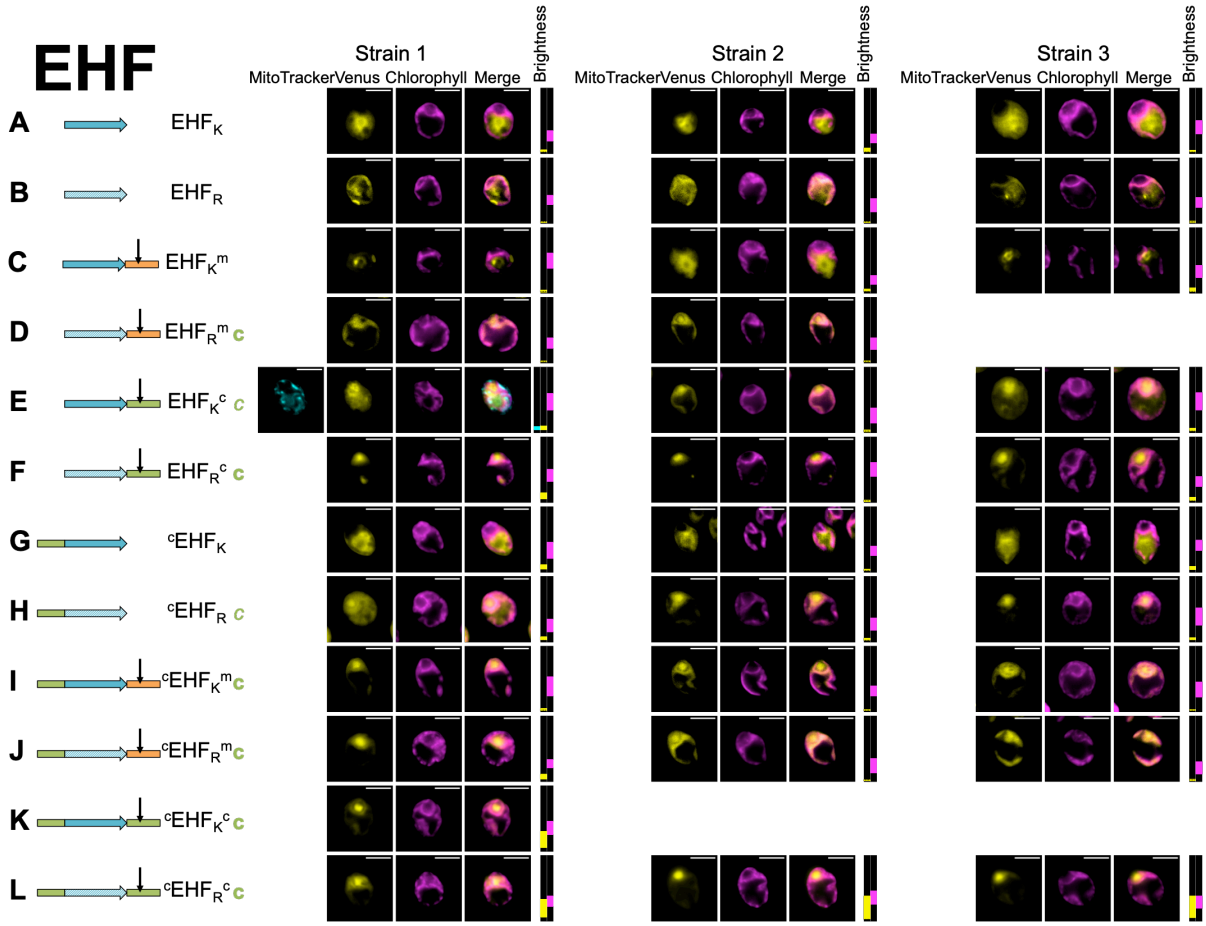
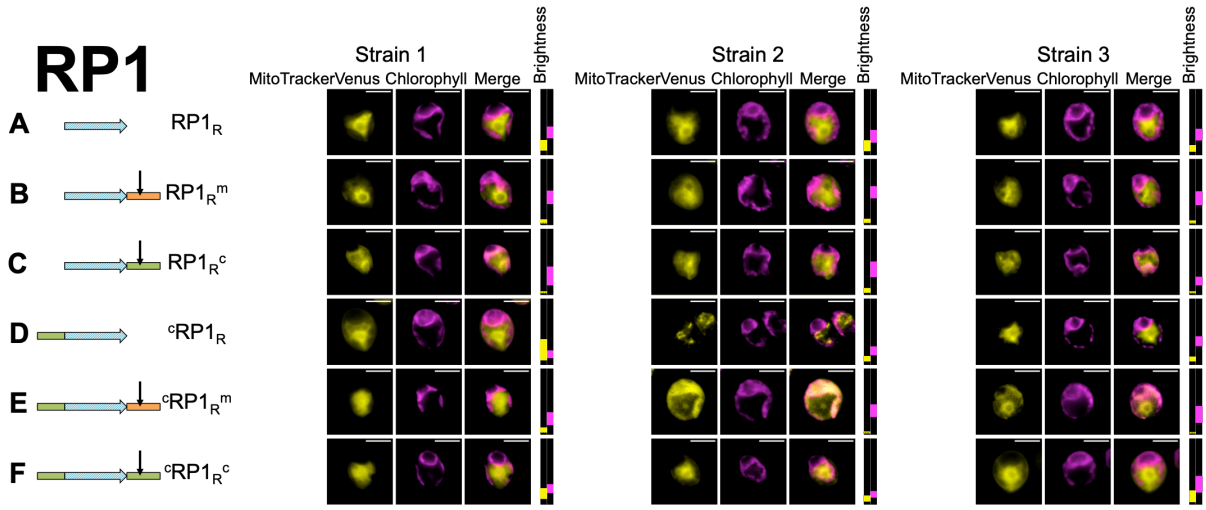


Fig. S14. Microscopy for biological replicates of Enterocin HF. Three independent transformant lines (Strains 1-3) are shown for each combination of modifications (A-L), represented by a cartoon and a shorthand description (cf. Fig. 2). Where a construct was interpreted as generating reporter localization 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.



1114

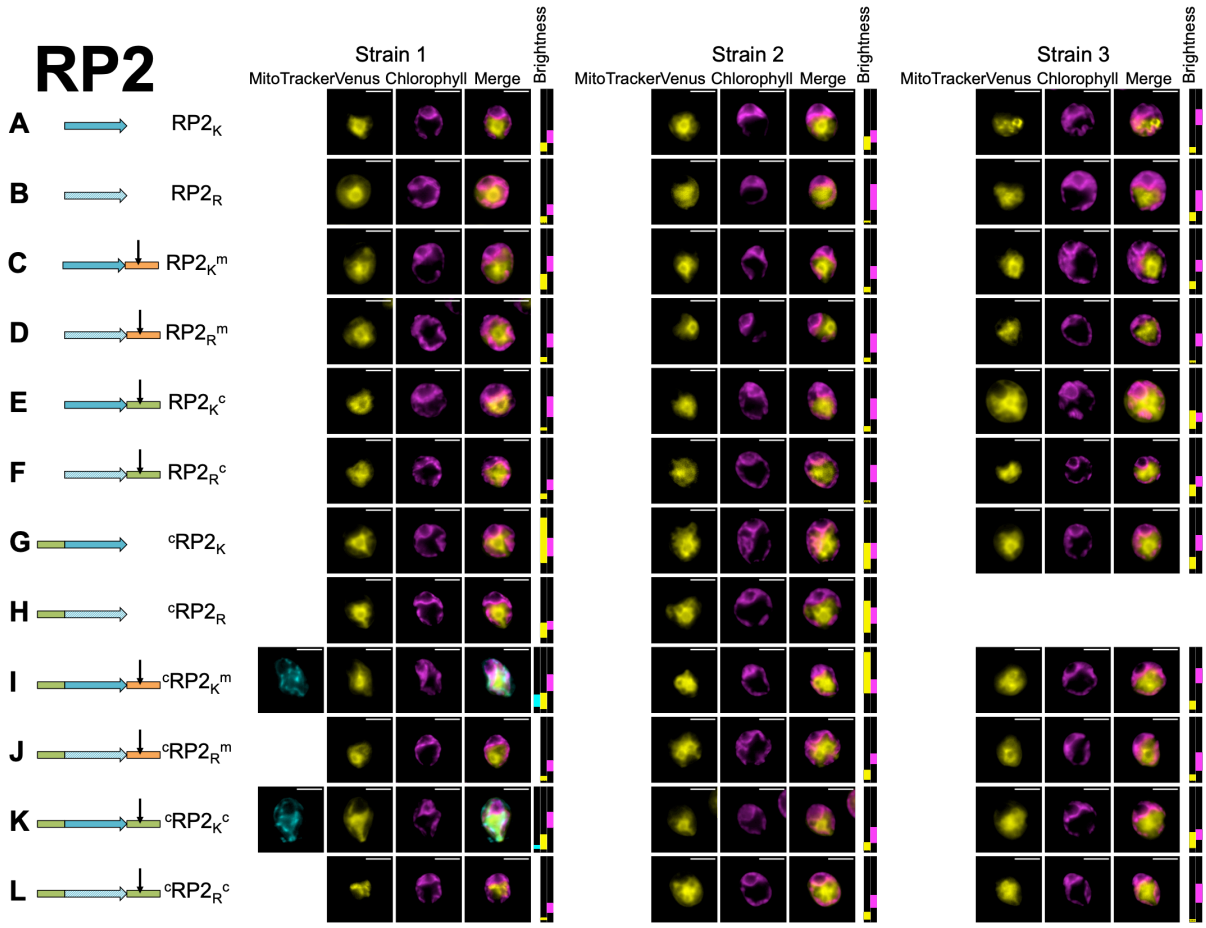
1115 **Fig. S15. Microscopy for biological replicates of negative control Random**
1116 **Peptide 1.** Three independent transformant lines (Strains 1-3) are shown for each
1117 combination of modifications (A-F), represented by a cartoon and a shorthand
1118 description (cf. Fig. 2). Where a construct was interpreted as generating reporter
1119 localization in mitochondria or chloroplast, this is indicated by an orange 'm' or a
1120 green 'c' respectively, in bold for full targeting or in italics for partial targeting. Scale
1121 bars are 5µm.



1122
1123

1124

1125 **Fig. S16. Microscopy for biological replicates of negative control Random**
1126 **Peptide 2.** Three independent transformant lines (Strains 1-3) are shown for each
1127 combination of modifications (A-L), represented by a cartoon and a shorthand
1128 description (cf. Fig. 2). Where a construct was interpreted as generating reporter
1129 localization in mitochondria or chloroplast, this is indicated by an orange 'm' or a
1130 green 'c' respectively, in bold for full targeting or in italics for partial targeting. Scale
1131 bars are 5µm.



1132
1133

Fig. S17. 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.

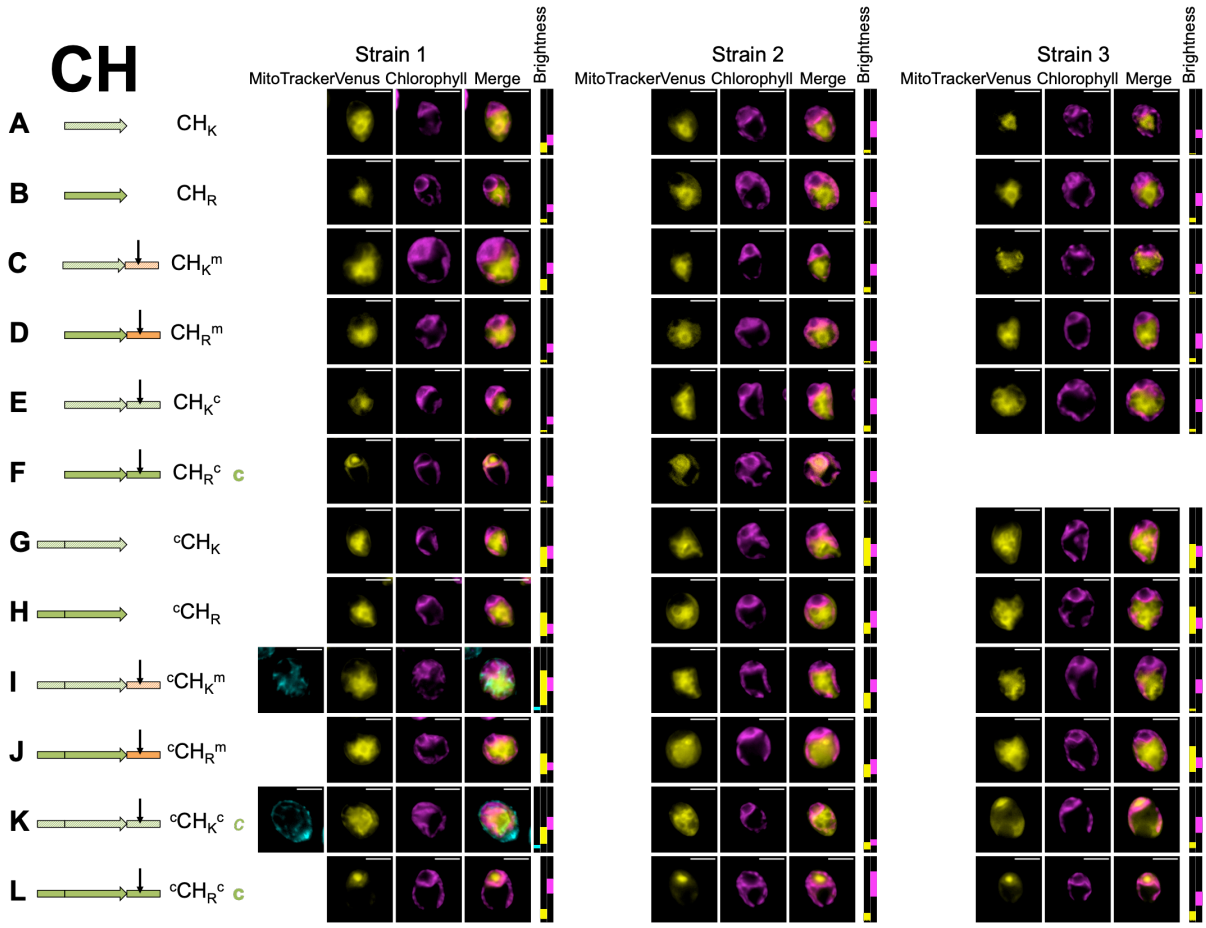


Fig. S18. Microscopy for biological replicates of TP control mitochondrial γ-carbonic anhydrase 2 mTP helical element (MH). Three independent transformant lines (Strains 1-3) are shown for each combination of modifications (A-L), represented by a cartoon and a shorthand description (cf. Fig. 2). Where 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.

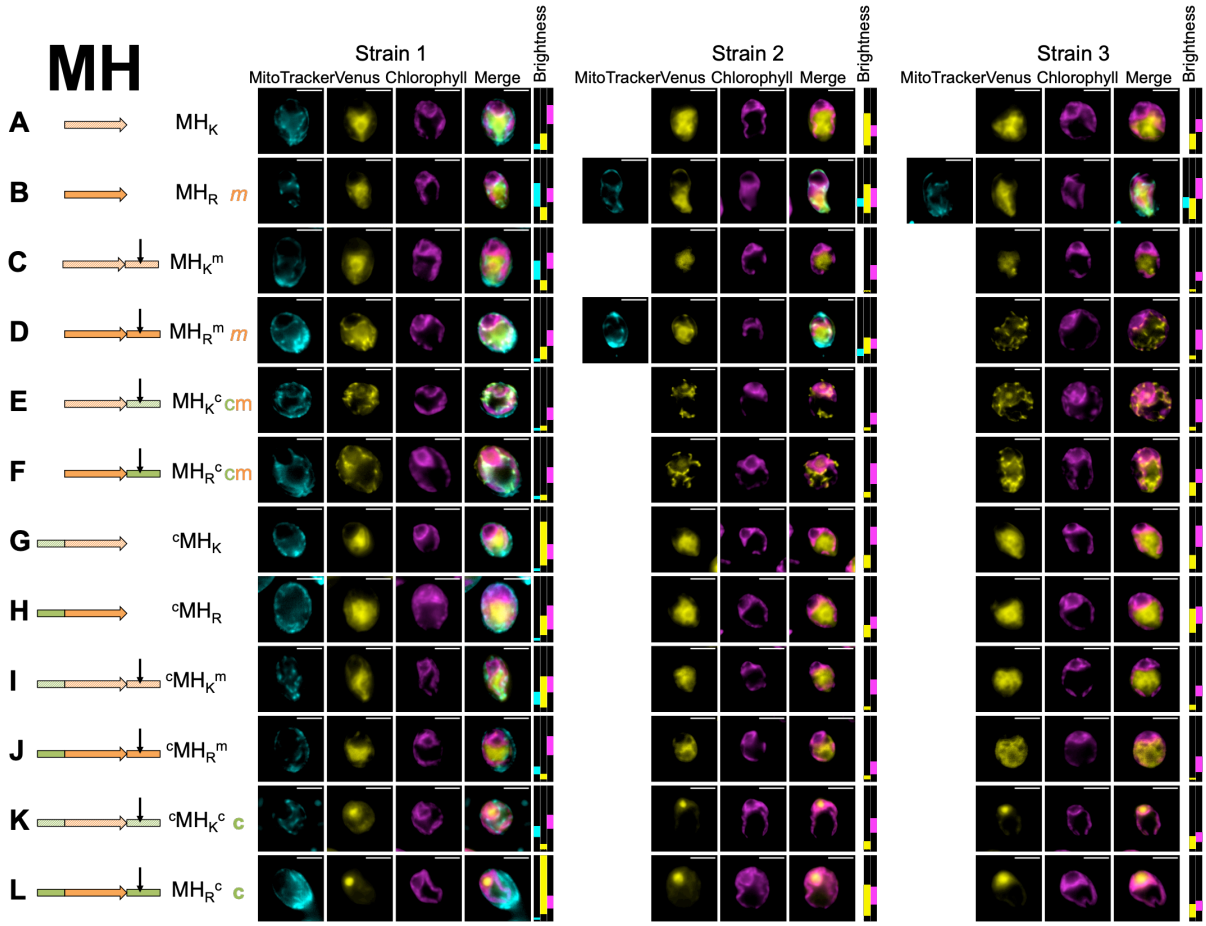


Fig. S19. Microscopy for biological replicates of no-peptide and Alanine-screen controls. Three independent transformant lines (Strains 1-3) are shown for each combination of elements (A-G), represented by a cartoon and a shorthand description (cf. Fig. 2). Where a construct was interpreted as generating 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.

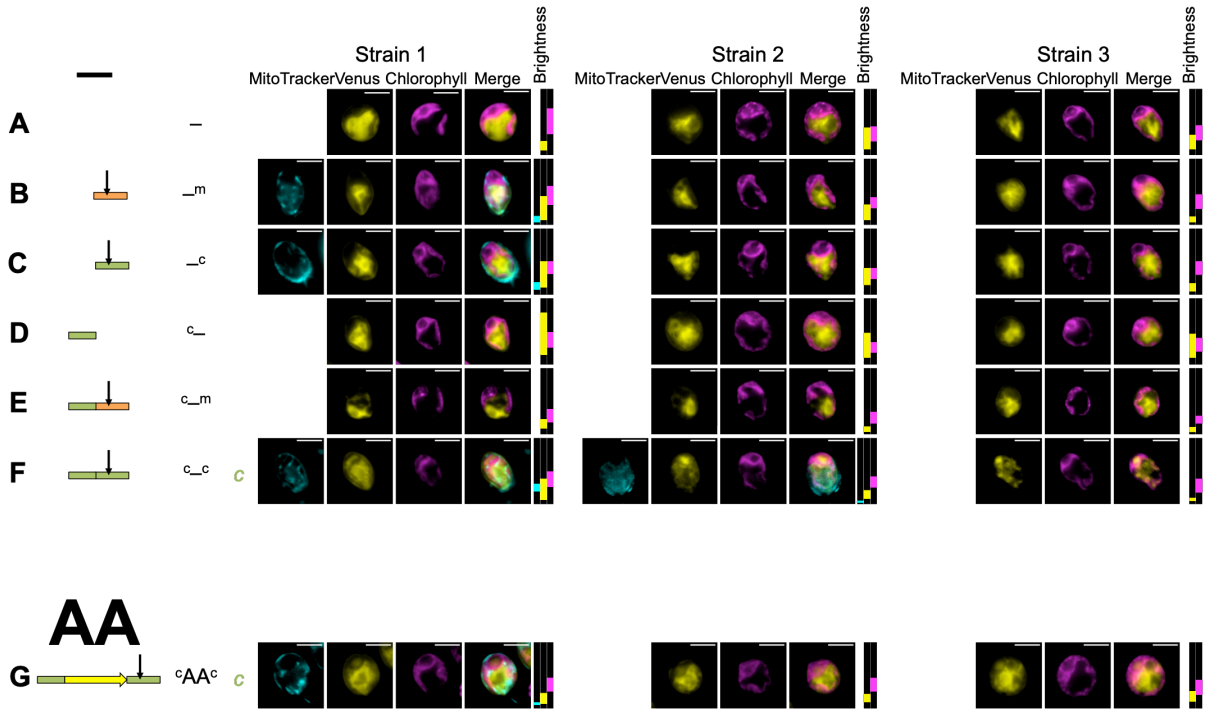


Fig. S20. Comparison of amino acid frequencies reveals K/R shift. (**a**) Amino acid frequencies are shown as boxplots (center line: median; box limits: upper and lower quartiles; whiskers: min/max values within 1.5x interquartile range) for Chlamydomonas cTPs in green, Chlamydomonas mTPs in orange and HA-RAMPs in blue. To give a baseline for comparison, the average across UNIPROT is given for each amino acid as red horizontal line. For each residue, different letters underneath distributions indicate a significant difference: groups that share the same letter are significantly different at $p<0.05$ (Multiple Kruskal Wallis tests followed by Dunn post-hoc tests) from groups attributed a different letter. (**b**) Amino acid frequencies are shown as heatmap for human, plant, algal and yeast TPs.

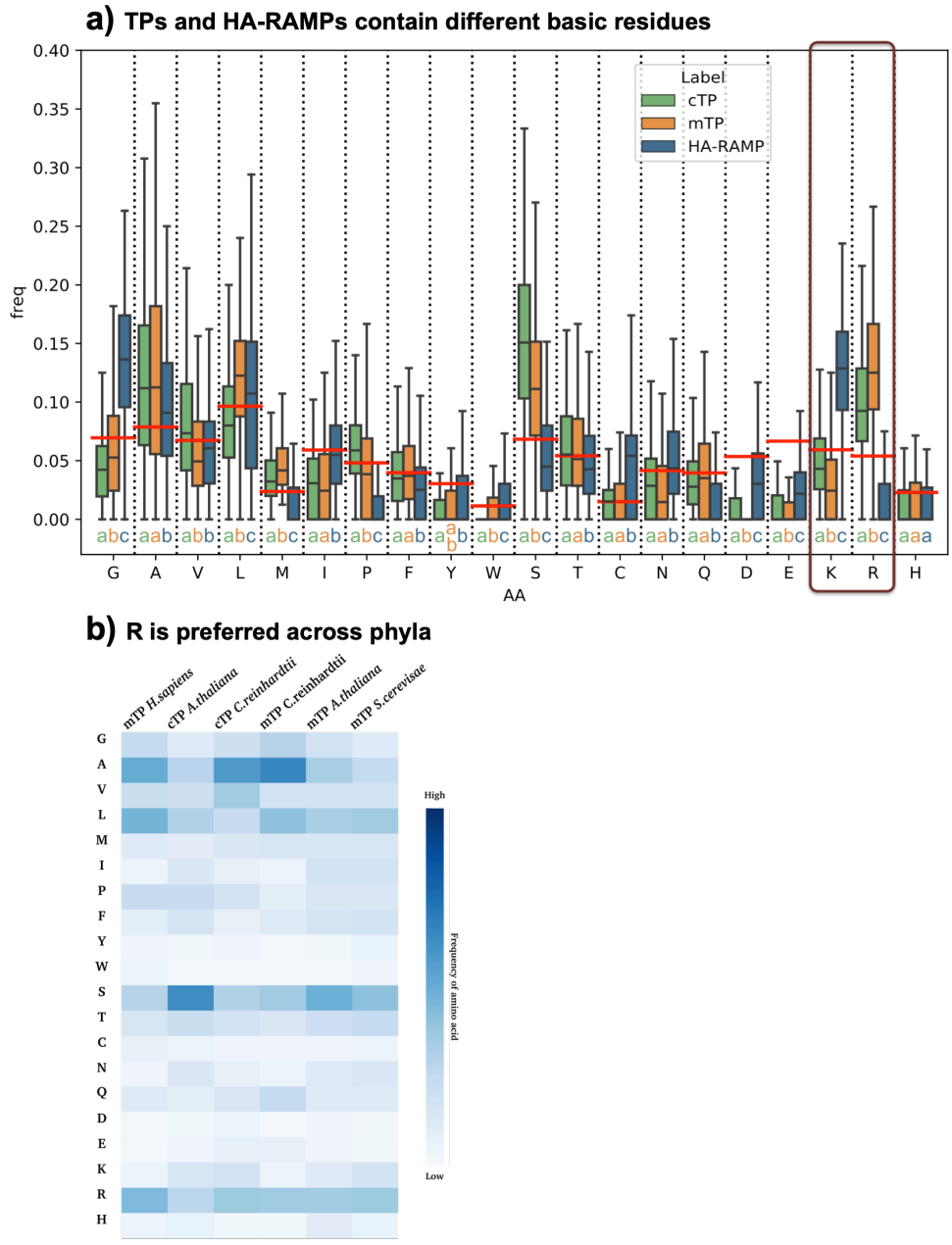
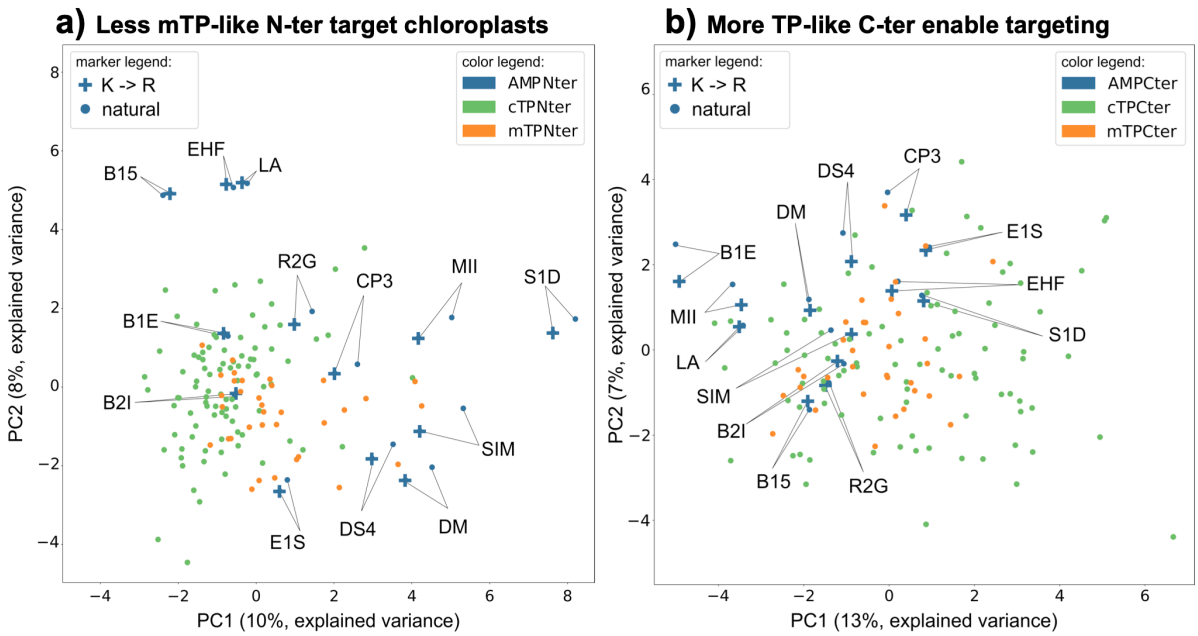
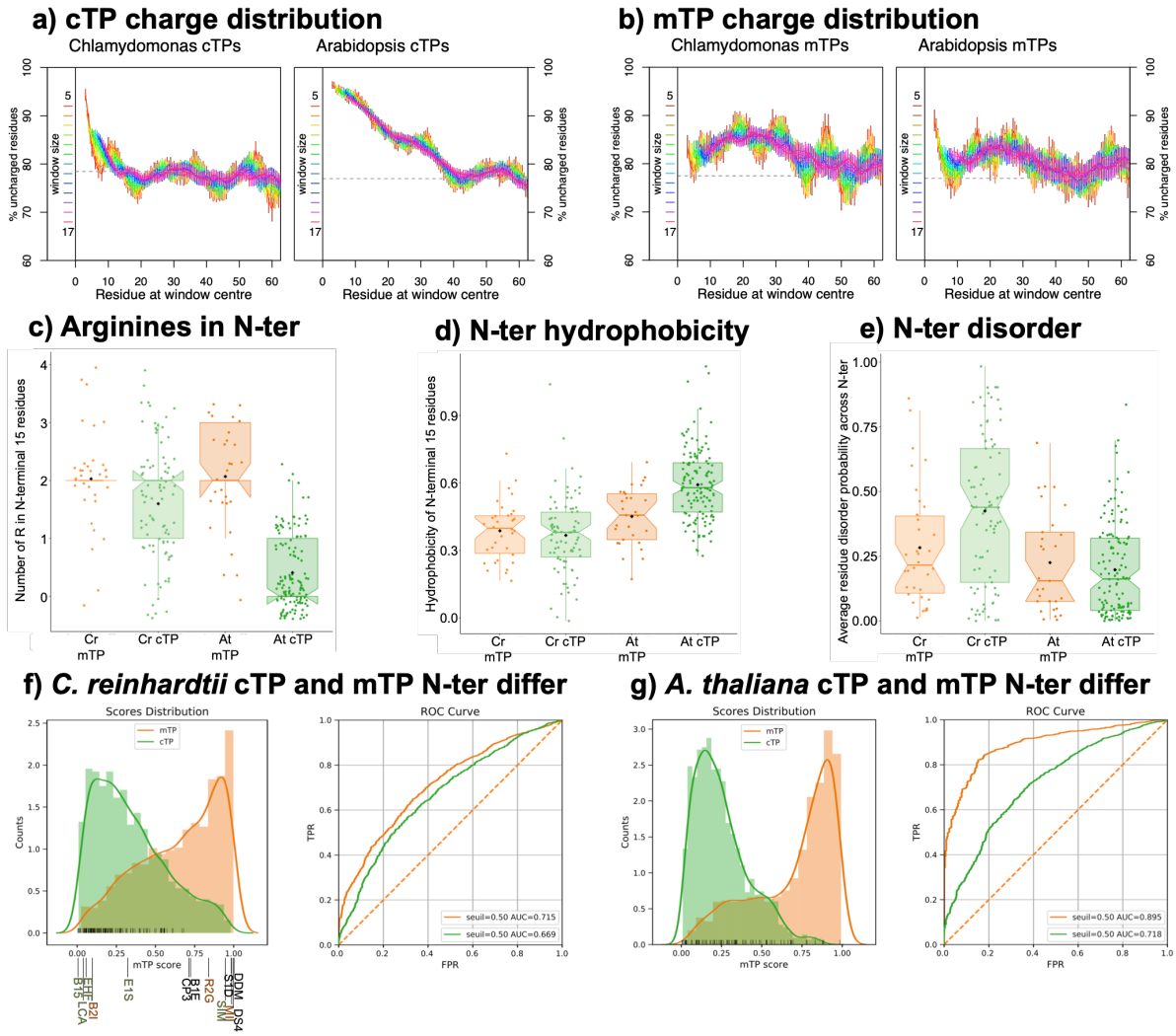


Fig. S21. Principal component analyses reveal that N- but not C-termini differ between cTPs and mTPs. Principal component (PC) analyses of auto-cross-correlated Z-scale values[10] for (**a**) N-termini (15 residues) and (**b**) C-termini (33 residues, encompassing -10 to +23 relative to the cleavage site for TPs) respectively of Chlamydomonas cTPs (in green) and mTPs (in orange) as well as the 13 HA-RAMPs under study (in blue; crosses denote HA-RAMPs after K→R).



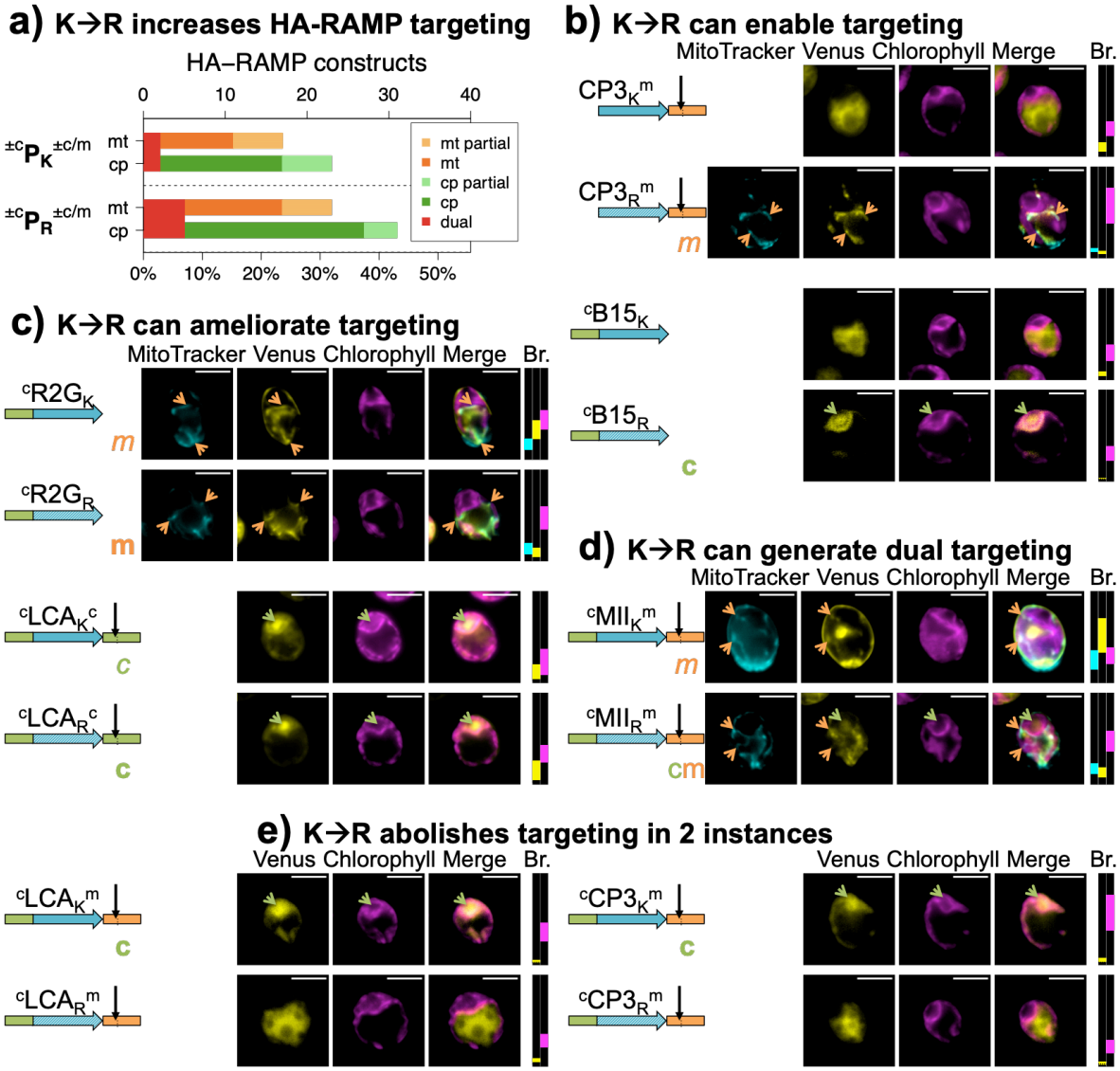
1184

1185 **Fig. S22. Unusual algal cTP N-termini share physicochemical differences against**
1186 **mTP N-termini with plant counterparts.** (**a**) Plots detailing the percentage of
1187 uncharged residues (i.e. excluding K/R and D/E, as in[27]) for window sizes ranging
1188 from 5 to 17 residues (refer to legend for colour code) show that Chlamydomonas
1189 cTPs (on the left) are less uncharged at the N-terminus than Arabidopsis cTPs (on the
1190 right)**.** Dashed grey lines indicate the average value for randomized sequences
1191 (including mature proteins), to provide an estimate of what would be expected by
1192 chance. (**b**) Equivalent graphs (as in A) for mTPs show a similar charge profile across
1193 Chlamydomonas and Arabidopsis. (**c-e**) Distributions of salient values as boxplots
1194 (center line: median; box limits: upper and lower quartiles; whiskers: min/max values
1195 within 1.5x interquartile range) for mTPs in orange and cTPs in green with
1196 *Chlamydomonas reinhardtii* (Cr) on the left and *Arabidopsis thaliana* (At) on the
1197 right show that: (**c**) Chlamydomonas cTP N-termini contain R at almost the same
1198 frequency as mTP N-termini, whereas Arabidopsis cTP N-termini contain fewer R;
1199 (**d**) Chlamydomonas TP N-termini are less hydrophobic than Arabidopsis TP N-
1200 termini, with plant cTP N-termini showing the highest hydrophobicity; (**e**)
1201 Chlamydomonas cTP N-termini are more disordered than mTP or Arabidopsis cTP N-
1202 termini. Points represent individual peptides (note that for integer values, point
1203 positions are randomized within ±0.5 in y as well as in x to increase point separation),
1204 population means are shown as black diamonds. (**f**) Binomial logistic regression
1205 classifier separating TPs using auto-cross-covariance of Z-scales[10] of the N-terminal
1206 15 residues for *C. reinhardtii*. The left-hand graph shows the distribution of cTPs in
1207 green and mTPs in orange over the model output 'mTP score'. Black bars at the
1208 bottom of the graph represent scores for cTPs of Arabidopsis, showing that plant cTPs
1209 are recognized as cTPs by the model trained on algal cTP N-termini (89% recognized
1210 as cTPs, with an mTP score <0.5). Values for our 13 HA-RAMPs are given below the
1211 graph. The right-hand graph shows receiver operating characteristic (ROC) curves,
1212 plotting the true positive rate (TPR) against the false positive rate (FPR), where the
1213 area under the curve (AUC) serves as estimate of the model quality with values above
1214 0.5 indicating that the model is better than random. (**g**) Equivalent model (as in F)
1215 trained on *A. thaliana* TPs, with black bars now representing Chlamydomonas cTPs
1216 showing that the majority of algal cTPs are recognized as cTPs by the plant model
1217 (75% recognized as cTPs, with an mTP score <0.5).


a) cTP charge distribution
Chlamydomonas cTPs
Arabidopsis cTPs
b) mTP charge distribution
Chlamydomonas mTPs
Arabidopsis mTPs
% uncharged residues
window size
Residue at window centre
c) Arginines in N-ter
Number of R in N-terminal 15 residues
d) N-ter hydrophobicity
Hydrophobicity of N-terminal 15 residues
e) N-ter disorder
Average residue disorder probability across N-ter
Cr mTP
Cr cTP
At mTP
At cTP
f) C. reinhardtii cTP and mTP N-ter differ
Scores Distribution
mTP
cTP
Counts
mTP score
ROC Curve
TPR
FPR
seuil=0.50 AUC=0.715
seuil=0.50 AUC=0.669
g) A. thaliana cTP and mTP N-ter differ
Scores Distribution
mTP
cTP
Counts
mTP score
ROC Curve
TPR
FPR
seuil=0.50 AUC=0.895
seuil=0.50 AUC=0.718

1221

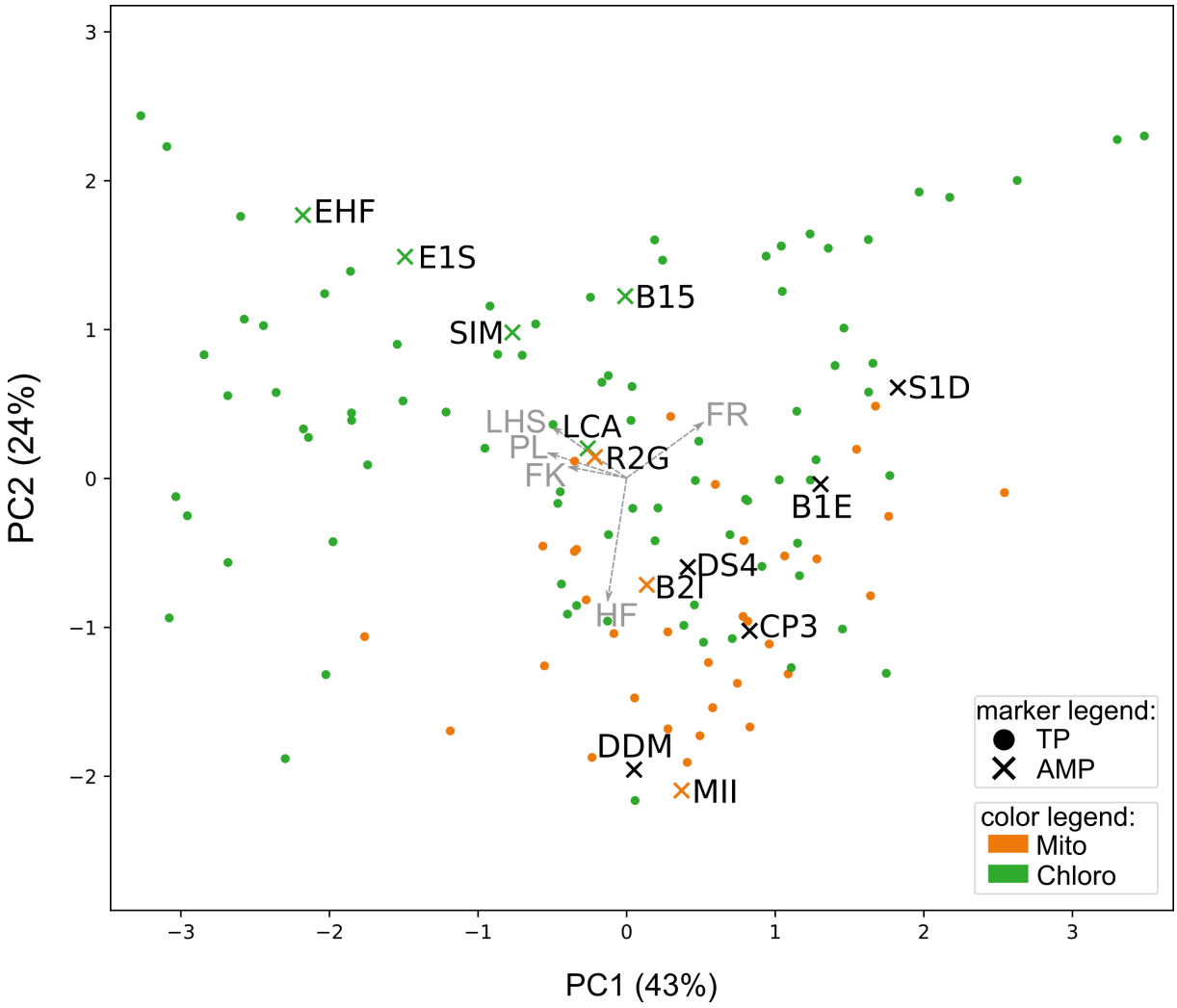
1222 **Fig. S23. K→R generally improves HA-RAMP targeting.** (**a**) An overview graph
1223 showing how many, out of 13 HA-RAMPs studied, show targeting to either
1224 mitochondria (mt) or chloroplast (cp), or both (dual) before and after K→R
1225 represented. (**b-e**) Epifluorescence microscopy images of selected examples are
1226 shown as in Fig. 3.



1227

1228

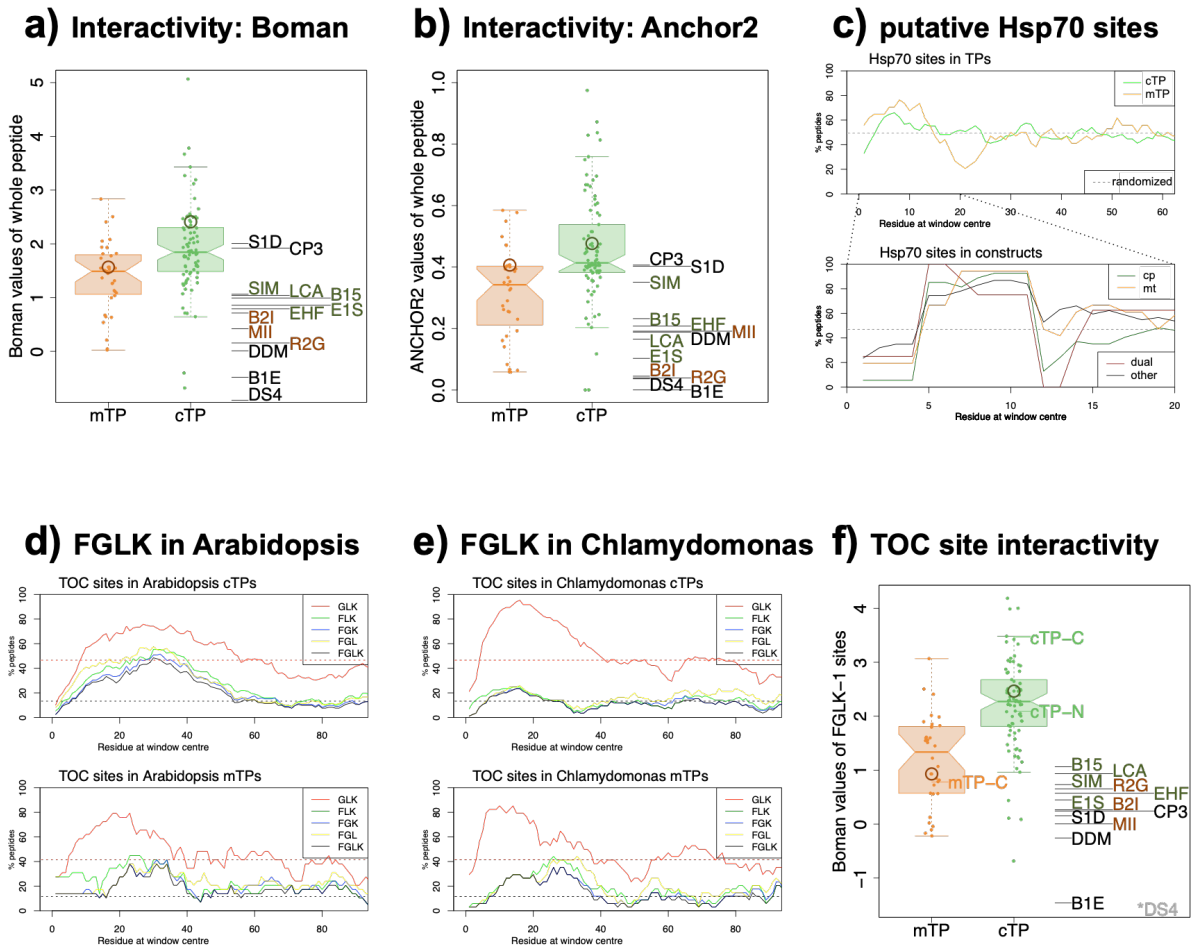
Fig. S24. HA-RAMP properties determine their targeting propensities. PCA analysis of cTPs (green dots), mTPs (orange dots), cp-set HA-RAMPs (green crosses), mt-set HA-RAMPs (orange crosses) and other HA-RAMPs (dark crosses) based on their length (Peptide Length: PL), number of residues before the longest predicted amphipathic helix (Long Helix Start: LHS), fraction of residues forming the predicted amphipathic helix (Helix Fraction: HF) and fraction of K (Fraction of Lysines: FK) and R residues (Fraction of Arginine: FR). 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 FK, PL, LHS (pointing left) and FR residues (pointing right). The second principal component explains 24% of the variance and distinguishes peptides principally according to HF. The cp-set HA-RAMPs group with the most distinguishable cTPs, being longer, with a higher LHS and a lower HF compared to mt-set HA-RAMPs and mTPs.



1244

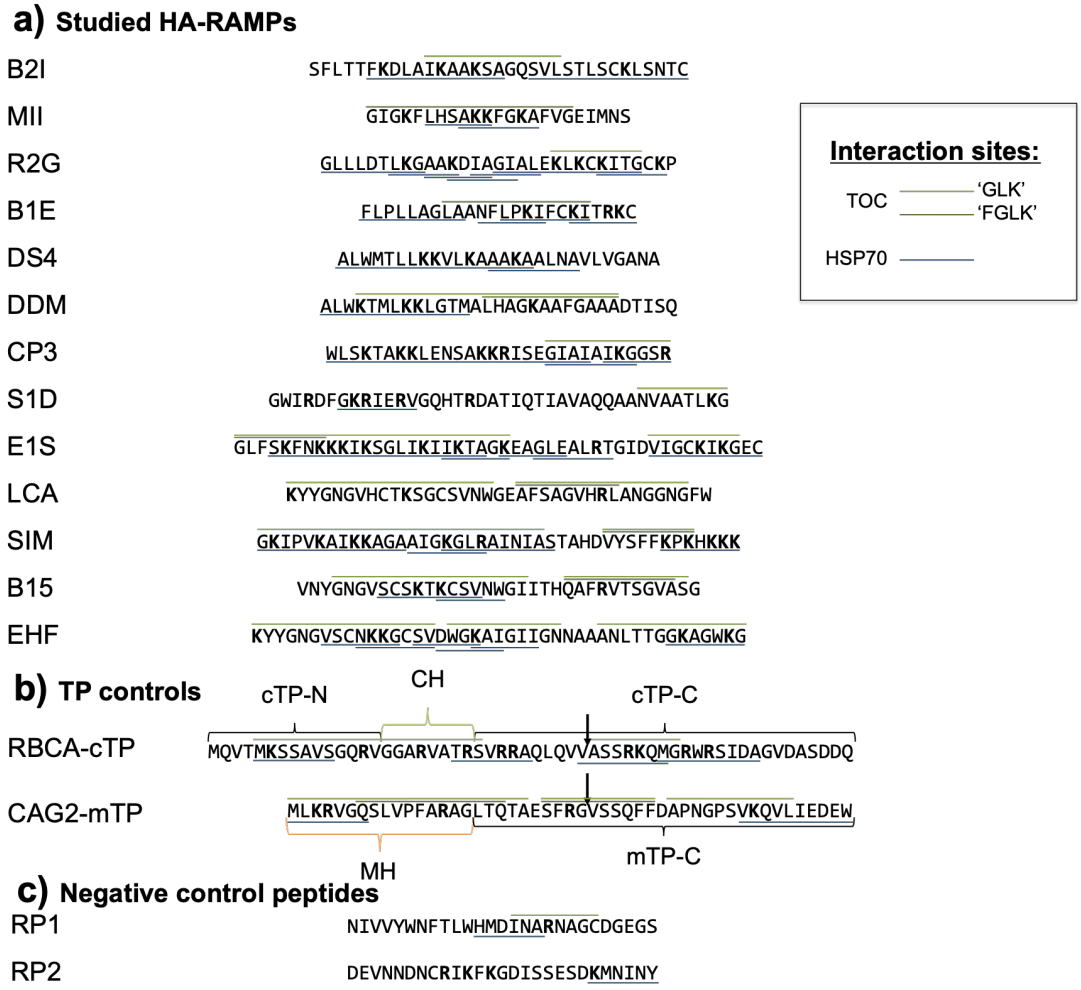
1245 **Fig. S25. Higher protein interactivity predicted for cTPs than mTPs.** (**a,b,f**) For
1246 salient properties, Chlamydomonas mTPs and cTPs are compared to HA-RAMPs (cp-
1247 set in green, mt-set in orange). Distributions are shown as boxplots (center line:
1248 median; box limits: upper and lower quartiles; whiskers: min/max values within 1.5x
1249 interquartile range), coloured points represent individual peptides (**a**) cTPs show
1250 increased protein interactivity (p=0.0003), as do cp-set HA-RAMPs (p= 0.0355), as
1251 estimated through Boman values[69], a proxy developed for AMPs where a value of ca.
1252 >2 indicates increased protein interaction potential, and <2 points to membrane
1253 interaction. (**b**) cTPs show increased protein interactivity (p<0.0001), and cp-set HA-
1254 RAMPs show a similar tendency (p=0.0638), as estimated with Anchor2[70], developed
1255 to predict the protein interaction potential of disordered sequences. (**c**) Both cTPs and
1256 mTPs show a high fraction of peptides (~50%) that have an Hsp70 interaction site
1257 over a 6-residue window at any given position along the sequence in the top graph,
1258 with a peak towards the N-terminus. HA-RAMP constructs shown in the bottom
1259 graph also display a peak in predicted Hsp70-binding sites at the N-terminus,
1260 independently of construct localisation. The value obtained after randomizing the
1261 position of residues is given as dashed grey line to provide an estimate of how often
1262 sites would be expected to occur given amino acid frequencies (~50%).
1263 Randomization was done over the entire sequence including the cargo protein. (**d**) For
1264 Arabidopsis cTPs in the top graph, the percentage of peptides that have a full 'FGLK'
1265 site[27] (black curve; presence of F and G/P and L/V/A and K/R and absence of D/E
1266 within an 8-residue window) far exceeds the value obtained after randomizing the
1267 position of residues (dashed black line for full 'FGLK' sites), a proxy for the
1268 frequency of the motif expected at random, for most positions along the sequence up
1269 to ~60 residues. Randomization was done over the entire sequence including the
1270 cargo protein. Reduced 'FGLK'-1 sites containing three out of four elements appear
1271 to occur mostly in the context of full FGLK sites, except for 'GLK' sites for which
1272 randomization is also shown (dashed red line). Arabidopsis mTPs (bottom graph) also
1273 contain 'FLGK' sites at a frequency higher than expected at random between residues
1274 20-40, and 'GLK' sites upstream of residue 40, but motifs are less prevalent than in
1275 cTPs. (**e**) Equivalent graphs (as D) for Chlamydomonas show that full 'FGLK' sites
1276 (black lines) are rare in cTPs (top graph) and actually more common in mTPs (bottom
1277 graph) upstream of residue 40. Reduced 'GLK' 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.



1288

1289 **Fig. S26. Hsp70 and (F)GLK sites are present in HA-RAMPs.** Predicted Hsp70
1290 binding sites are underlined in blue. Putative TOC interaction sites are overlined in
1291 green: 'FGLK' sites in dark green and 'GLK' sites in light green. See Fig. 1 for
1292 detailed annotations. Notably, both TOC and Hsp70-interaction sites are contained
1293 within native HA-RAMPs (**a**) and within our TP controls (**b**).



1294

1295

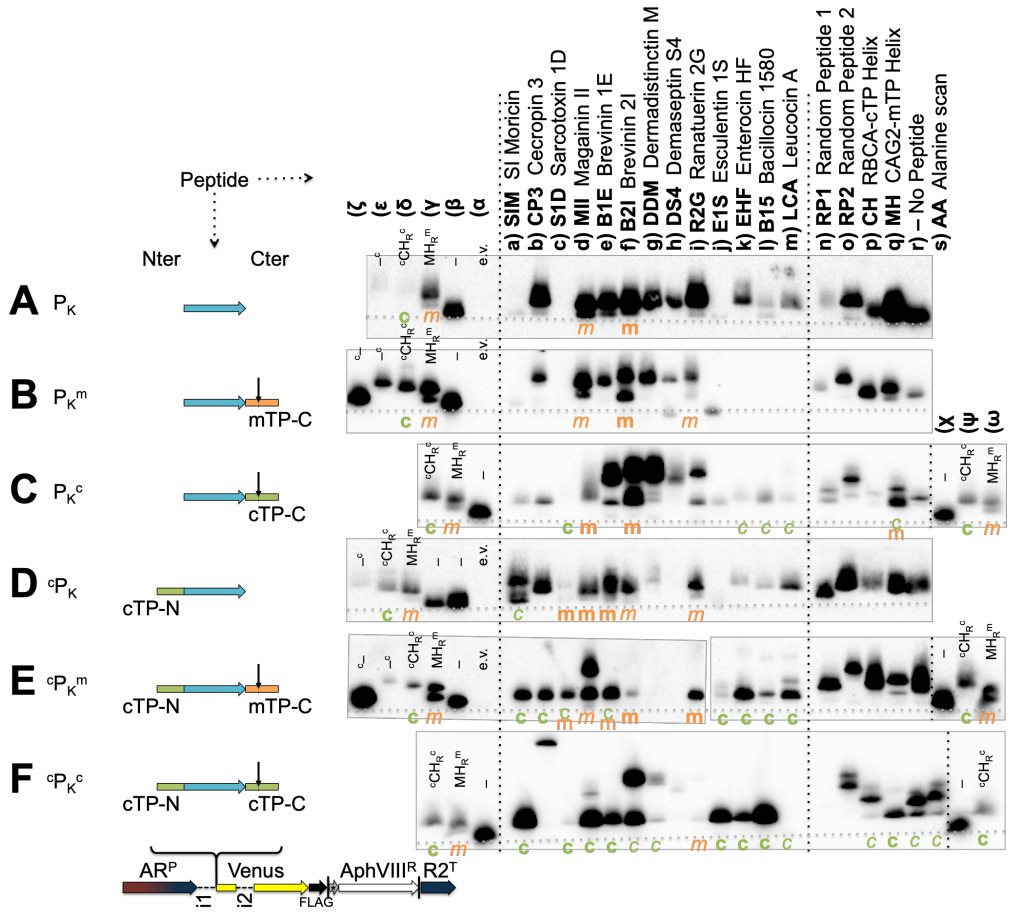



Fig. S27. Western Blots for K-bearing constructs. Samples were immunolabelled using an α-FLAG primary antibody. Constructs representing different combinations of TP-element additions are shown in rows A-F, as indicated by a shorthand description and a cartoon (cf. Fig. 2). HA-RAMPs are consistently arranged 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, $^{c}CH_{R}{}^{c}$ is Rubisco activase cTP, $MH_{R}{}^{m}$ is mitochondrial γ-carbonic anhydrase 2 mTP, ^{c}– is Rubisco activase cTP N-terminal element (15 residues), –^{c} is Rubisco activase cTP C-terminal element (33 residues, -10 to +23 relative to cleavage site), e.v. is empty vector (no Venus expression). Constructs that generated chloroplast and/or mitochondrial localization are marked with a green 'c' and/or an orange 'm' respectively, in bold for full or in italics for partial targeting. Note that row E contains two separate blots for the left-hand section up to lane i and the right-hand section following lane j.