An antimicrobial origin of targeting peptides to endosymbiotic organelles

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

Affiliations:

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

† These authors contributed equally to this work
* Corresponding author
Email: ingrid.lafontaine@ibpc.fr

Keywords
Eukaryogenesis, antimicrobial peptides, mitochondria, chloroplasts, targeting peptides

Author Contributions
Conceptualization: FAW, IL; Data Curation: CG, ODC, IL; Funding Acquisition: YC, FAW, IL; Investigation and Validation: CG and IL (computational), ODC (experimental); Software: CG, ODC; Supervision: YC, FAW, IL; All authors contributed to writing and editing the manuscript.

Abstract
Mitochondria and chloroplasts emerged from primary endosymbiosis. Most endosymbiont proteins have subsequently been expressed in the nucleo-cytosol of the host and organelle-targeted via the acquisition of N-terminal presequences, whose evolutionary origin remains enigmatic. Using a quantitative assessment of their physico-chemical properties, we show that organelle targeting peptides, which are distinct from signaling peptides targeting other subcellular compartments, group with a subset of antimicrobial peptides. We demonstrate that extant antimicrobial peptides target a fluorescent reporter to either the mitochondria or the chloroplast in the green alga *Chlamydomonas reinhardtii* and, conversely, that extant targeting peptides display antimicrobial activity. Thus, we provide conclusive computational and functional evidence for an evolutionary link between organelle-targeting and antimicrobial peptides. Our results support the view that resistance of bacterial progenitors of organelles to the attack of host antimicrobial peptides, has been instrumental in eukaryogenesis and in the subsequent emergence of photosynthetic eukaryotes.

Introduction
Mitochondria and chloroplasts are eukaryotic organelles that evolved from bacterial ancestors through endosymbiosis [see Archibald (2015), Martin et al. (2015) for recent reviews].

These endosymbiotic events were accompanied by a massive loss of genetic material from the bacterial ancestors, part of which were transferred to the host genome through what is known as Endosymbiotic Gene Transfer (EGT) (Timmis et al. 2004). Thus, to be successful, primary endosymbiosis required the establishment of efficient protein import machineries in the outer membranes of the proto-organelle to be able to re-import protein products of genes that had been transferred to the nuclear genome. As a result, most mitochondrial and chloroplast genomes encode less than 100 proteins and the majority of proteins localized therein (ca. 1000 in mitochondria and >2000 in the chloroplast) are now translated in the cytosol and subsequently imported back to the organelle. Most of the nuclear-encoded proteins found in organelles harbor a targeting peptide (TP) which is an N-terminal presequence functioning as an address tag, *i.e.* determining the subcellular localization of cargo proteins within endosymbiotic organelles (von Heijne et al. 1989). TPs are recognized by the main mitochondrial and chloroplast translocation pathways (Wiedemann and Pfanner 2017; Nakai 2018) and destroyed upon import into organelles (Neupert and Herrmann 2007; Chotewutmontri et al. 2017). The emergence of such TP-based import, despite being a key innovation enabling endosymbiosis and eukaryotism, remains poorly understood (Fukasawa et al. 2017; Reumann et al. 2005; Nakai 2018).

Here we provide evidence that support an antimicrobial origin of TPs as described in a proposed scenario for the emergence of the endosymbiotic protein import system (Wollman 2016), which is summarized in Figure 1. Archaea, bacteria and eukaryotes alike use antimicrobial peptides (AMPs) as part of their innate immune system to kill microbes, typically via membrane permeabilization (Maróti et al. 2011; Besse et al. 2015). Numerous studies have established that AMPs consistently play a role in most microbial interactions (Mergaert et al. 2017; Mergaert 2018), which argues for their involvement in the initial relationship between a host and a proto-endosymbiont. Extant heterotrophic protists employ AMPs to kill engulfed prey, which suggests that early eukaryotes likely used AMPs in a similar way against their cyanobacterial prey that ultimately became the chloroplast (Ball et al. 2016). Similarly, the archaeal host will have delivered AMPs against the α-proteobacterial ancestor of mitochondria, whether it was a prey or an intracellular pathogen akin to Rickettsiales (Zachar and Szathmáry 2017; Andersson et al. 1998). AMPs also would also have been instrumental when considering mutualism at the origin of endosymbiosis, since present-day hosts use non-lethal concentrations of AMPs to control the growth of symbionts and enable nutrient exchange to facilitate metabolic integration (Maróti et al. 2011; Andrä et al. 2003).

Studies of the various extant strategies for microbial defense against AMPs have revealed instances where AMPs are imported into bacterial cells via dedicated transporters and then degraded by cytoplasmic peptidases, hence referred to as an import and destroy mechanism (e.g. Shelton et al. 2011; Rinker et al. 2012), which is strikingly similar to TP-based import.

EGT would have thus started with incorporation of DNA fragments from lysed bacteria into the host genome, as observed in extant phagotrophic protists (Ford Doolittle 1998; Clarke et al. 2013). Endosymbiotic integration would then have been promoted by the acquisition by proto-endosymbiotic bacteria of an import and destroy mechanism to resist the AMP attack from the host. Indeed, the serendipitous insertion of a bacterial gene downstream of an AMP coding sequence in the host genome, whether right upon EGT or after chromosomal rearrangements, would then have allowed for import of the gene product back into the proto-organelle via the very same transporter system that had allowed the AMP-resistant endosymbiont to detoxify attacking peptides. Consequently, extant TPs continue to show structural similarities to a type of AMPs called Helical Amphiphilic-Ribosomal associated AMPs (HA-RAMPs) (Wollman 2016): chloroplast TPs (cTPs), mitochondrial TPs (mTPs) and HA-RAMPs all carry net positive charges, and display a domain

that folds into an amphiphilic α-helix upon contacting the target membrane (Bruce 2000; Bechinger et al. 1993; Wiedemann and Pfanner 2017).

Here, we tested this hypothesis that TPs have originated from host-delivered AMPs. If their common origin holds true, enduring similarities in the physico-chemical properties of extant TPs and HA-RAMPs should remain despite the large evolutionary distance, and at least a subset of TPs and HA-RAMPs may still display some dual antimicrobial and organelle targeting activities. We demonstrate that these two predictions are fulfilled, thus placing bacterial resistance to the host at the core of the emergence of eukaryotism.

We first re-assess in some depth the physico-chemical properties of the various families of HA-RAMPs using a consistent set of descriptors, thus providing a more robust classification of these peptides compared to the current AMP families. We then provide computational evidence for the extensive overlap of the physico-chemical properties of TPs with those of a cluster of HA-RAMPs. Finally, we demonstrate experimentally that extant antimicrobial peptides are able to target a fluorescent reporter to either the mitochondria or the chloroplast in a eukaryotic model system, the green alga *Chlamydomonas reinhardtii*. Conversely, we show that targeting peptides still display antimicrobial activity.

That targeting peptides are evolutionarily related to a subgroup of antimicrobial peptides sheds new light on the key event that led to eukaryogenesis and to the emergence of photosynthetic eukaryotes: in the two cases, primary endosymbiosis would result from a bacterial resistance to the antimicrobial defense of a host cell.

Results

TPs and HA-RAMPs share a common set of physico-chemical properties

Peptide families and their descriptors

We performed a comparative analysis of different functional groups of peptides. We selected TPs from *Chlamydomonas reinhardtii*, *Arabidopsis thaliana*, *Saccharomyces cerevisiae* and *Homo sapiens sapiens* for which both the subcellular location of the cargo protein and the cleavage site have been experimentally determined. HA-RAMP families and, as negative control, the cyclotide family of globular AMPs, were taken from the CAMP$_{R3}$ database (Waghu et al. 2016). We also considered a set of peptides, hereafter referred to as secretory signal peptides (SPs), which function as address tags just as TPs do, but targeting a different set of subcellular destinations. SPs have a well-established evolutionary link and all use Sec-type translocation systems (von Heijne 1990; Frain et al. 2016; Schatz and Dobberstein 1996). For this study, we consider the following peptides as belonging to the SP group: bacterial SPs (bSPs) that target proteins for secretion into their periplasm, their eukaryotic relatives (eSPs) that target cytosolic proteins to the endoplasmic reticulum, and (tSPs), more commonly referred to as thylakoid targeting peptides, that enable proteins to be targeted from the chloroplast stroma to the thylakoids (von Heijne et al. 1989; Schatz and Dobberstein 1996). Lastly, a set of random peptides was generated (Table S5).

All peptides under study were less than 100 amino acid long. On average they are comprised of 45 residues for TPs, 32 residues for the HA-RAMPs and 31 residues for SPs (Supplementary Figure 1). Given that TPs and HA-RAMPs are so short and share a low level of sequence similarity, classical phylogenetic inferences were not applicable (von Heijne et al. 1989). Thus, to evaluate

the likelihood of an evolutionary relationship between TPs and HA-RAMPs, we resorted to their physico-chemical properties rather than to their primary sequences. To do so, we used 'Z-scales', amino-acid descriptors defined by Hellberg et al. (1987). Each of the 20 amino acids is therein described as a set of three values, which correspond to the first three linear combinations (principal components) of 29 physico-chemical properties measured experimentally. These three 'z factors' reflect mostly hydrophobicity (z1), bulkiness of the side chain (z2) and electronic properties (z3). A comparative study of 13 types of amino acid descriptors showed that these three Z-scales are sufficient to explain the structure-activity variability of peptides (van Westen et al. 2013b, 2013a). To take interdependencies between residues into account – i.e the properties of the whole peptide- each peptide was defined by 36 terms corresponding to auto-cross covariances (ACC) between Z-scale values (Wold et al. 1993) within a 4-neighbor window, which mimics a single α-helix turn of 3.6 residues (see "Peptide description" in the Materials and Methods section).

HA-RAMPs can be divided into two distinct classes

We retrieved 31 AMP families from the CAMP$_{R3}$ database (Waghu et al. 2016) for which a documented amphipathic domain has been reported (see Material and Methods). Many of these families were defined on a rather descriptive basis and the criteria used to group peptides differ from one family to another. In order to have a more consistent picture of the diversity of HA-RAMPs throughout these families, we first performed a k-means clustering of the 686 HA-RAMPs that we retrieved from the CAMP$_{R3}$ database, together with 353 SPs and 433 TPs based on the Euclidean distances between their 36 ACC vectors (Figure 2). Among clustering with k varying from 2 to 10, we selected the clustering with k=2, as this gave the highest average of silhouette coefficients (Rousseeuw 1987) for all peptides (reflecting the consistency of the clustering). The HA-RAMPs distributed between the two clusters in a 2/3 vs. 1/3 proportion. The 68% of HA-RAMPs that grouped in cluster 1 will be hereafter referred to as class I HA-RAMPs whereas those in cluster 2 will be referred to as class II HA-RAMPs.

To get a more detailed picture of their similarity relationships, we performed a neighbor-joining clustering of all HA-RAMPs based on the Euclidean distances between their 36 ACC vectors (Figure 3). Class I HA-RAMPs (in blue on Figure 3) are found in different parts of the NJ clustering tree, sometimes together with a subset of class II HA-RAMPs (in red on Figure 3), but the two classes are not intermingled between each other: they tend to form homogeneous sub-trees. When ascribing the peptides in each cluster to the distinct antimicrobial families in the literature, their distribution along the tree was patchy (outer color circle on Figure 3). Indeed, when we assessed the family distribution between the two clusters (Supplementary Figure 2), only 60% of these families fitted well either into either cluster 1 or cluster 2, which supports their classification as families of distinct physico-chemical properties. However, 13 families displayed peptides belonging to either one or the other cluster, calling for further investigation regarding the rationale for their identification as members of a same family (see discussion section).

In order to get a better handle on the differences between class I and class II HA-RAMPs, we compared the following features (Figure 4): net charge, length, hydrophobicity and the number of residues that can theoretically adopt an amphipathic helical structure. The two classes indeed had rather distinctive traits, with significant differences for all of them: class I HA-RAMPs proved longer than class II HA-RAMPs, with higher net charges and displayed a lower hydrophobicity as well as longer amphipathic helices.

TPs and class I HA-RAMPs are characterized by a shared set of physico-chemical properties

Following this classification of HA-RAMPs into two classes with distinct traits, we further investigated the properties of TPs and SPs relative to HA-RAMPs. Figure 2 shows that all TPs, but for a few isolated mTPs, clustered with class I HA-RAMPs. The majority of SPs grouped together in the other cluster, which is consistent with their common evolutionary origin. The basis for this distinct clustering is illustrated in Fig 4: the features for TPs and class I HA-RAMPs follow the same trends (although having statistically different distributions for all of them), in opposite directions of the trends observed for SPs bearing a well-documented hydrophobic stretch, and the more hydrophobic class II HA-RAMPs, which is consistent with their distribution in the bimodal clustering of Figure 2. TPs and class I HA-RAMPs have the highest net positive charges, significantly higher than the charges for SPs and class II HA-RAMPs (Figure 4A). TPs and class I HA-RAMPs also are the longest, while class II HA-RAMPs are significantly shorter than all the other peptides (Figure 4B). Whereas TPs and class I HA-RAMPs are significantly less hydrophobic than random peptides, class II HA-RAMPs are as hydrophobic as expected by chance and SPs are significantly more hydrophobic than expected by chance (Figure 4C), meaning that the aggregation of class II HA-RAMPs with SPs is fortuitous and only explained by the difference in hydrophobicity relative to class I HA-RAMPs and TPs. Class I HA-RAMPs and TPs form the longest amphipathic helices (Figure 4D), significantly longer than do random peptides. It is of note that class I HA-RAMPs and TPs respectively have similar sizes as cyclotides and random peptides -generated following the length distribution of TPs-, demonstrating that their longer amphipathic helices are not explained by longer sizes. By contrast, amphipathic helices in class II HA-RAMPs and SPs are not longer than those of random peptides. Therefore, while TPs and class I HA-RAMPs always follow the same trends, class II HA-RAMPs and SPs cluster together because of their relatively lower net charges, shorter lengths and higher hydrophobicity, compared to class I HA-RAMPs and TPs, masking their significant differences in terms of hydrophobicity and length. As controls, globular cyclotides have significantly lower net charges than HA-RAMPs, TPs and SPs, not different from what would be expected by chance for random peptides (Figure 4A) but they display a high level of hydrophobicity (Figure 4C) and the shortest amphipathic helices, which is consistent with their globular nature. As expected, the net charges of random peptides are equally distributed around zero (Figure 4A). Random peptides also have a higher hydrophobicity than TPs and class I HA-RAMPs, but shorter amphipathic helices (Figure 4C,D).

We next built a Logistic Regression (LR) classifier trained on ACC terms of 72 human mTPs and 96 bSPs to assign each peptide a probability of either being mTP-like or bSP-like. All TPs and class I HA-RAMPs were overwhelmingly recognized at ~ 95% as human mTP-like, while tSPs and eSPs were classified at about ~ 80% as bSP-like (Figure 5). It is of note that class I HA-RAMPs classified more strongly with human mTPs than did tSPs and eSPs with bSPs, despite the latter sharing a recognized evolutionary link. Class II HA-RAMPs showed poor discrimination between the two training groups (Figure 5). In agreement with the suboptimal definition of several HA-RAMP families that was revealed by the NJ clustering above, several families contained peptides with contrasting properties in the LR classification (Supplementary Figure 3). The ACC terms most contributing to the LR model are cross-covariances between z1 (hydrophobicity) and z2 (bulkiness) for the positive coefficients; and between z2 and z3 (bulkiness and electronic properties) for the negative coefficients (Supplementary Figure 4A), as expected for a classification the aim of which is to distinguish hydrophobic from amphipathic peptides.

In order to better characterize the physico-chemical properties that are most discriminatory between these peptides, we performed a Principal Component Analysis (PCA) of the peptides described by their ACC vectors. Figure 6 presents a PCA without class II HA-RAMPs (see

Supplementary Figure 5 for a PCA with class II HA-RAMPs). The separation between all peptides is provided by a combination of the contributions of various ACC terms to the two principal components (PC1 and PC2). For a qualitative analysis of the physico-chemical significance of these two components, one has to keep in mind that an α-helix being 3.6 amino acids per turn, the n, n+3, n+4 residues (reflected by lag 3 and lag 4 terms) can be considered as lying on the same face of the α-helix, whereas the n+1 and n+2 residues (reflected by a lag 1 and lag 2 terms) rather lie on the opposite face relative to residue n (Supplementary Figure 4). As shown by the contributions of the various ACC terms in Supplementary Figure 4B, the coupling between electronic and steric properties (z2.3 and z3.2) of the residues from the two faces of the amphipathic helix (lag1 and lag3) are the main contributors to PC1 (largest distance from the origin along the x-axis) whereas the hydrophobic and steric properties and their coupling (z1.2 and z2.1) of the residues along the same face of the helix (lag1 and lag2) most contribute to PC2 (largest distance from the origin along the x-axis). Note that z3.lag4 and z1.lag2 are the points with the highest distance from the radius of the circle. They define a line that goes through the radius of the circle, meaning that z3.lag4 and z1.lag2 are the most anti-correlated terms. As explained above, when considering the amphipathic helical domain of a peptide, these terms respectively reflect the electronic constraints that residues have to match on the same face of the α-helix and the amphipathic constraints between residues from opposite faces.

The evolutionarily-linked and hydrophobic tSPs, bSPs and eSPs co-localize on the top left of the graph, away from TPs and class I HA-RAMPs (Figure 6). Class I HA-RAMPs occupy the bottom of the graph with an amphipathic gradient from left to right, overlapping with TPs on the left side. TPs form a single overlapping spread, but mTPs show a tendency for higher values along PC1 than cTPs, suggesting mTPs tend to be more amphipathic. These observations are in line with k-means clustering (Figure 2), where TPs group with class I HA-RAMPs, apart from SPs. Despite different weighting of the ACC terms, this PCA furthermore mirrors the LR classification in that SPs and TPs are discriminated based on their contrasting hydrophobic and amphipathic characters. As control groups, we display on figure 6 the distribution of globular AMPs that occupy a separate part of the physico-chemical space to the left of the graph, reflecting their widely different structure, despite a shared antimicrobial function. Also, random peptide sequences generated *in silico* displayed a distribution, distinct from that of the other groups, that partially overlapped with HA-RAMPs, TPs and SPs, as expected.

When considering all HA-RAMPs together with SPs and TPs in the plane defined by PC1 and PC2 (Supplementary Figure 5A), TPs are enclosed within the convex area of class I HA-RAMPs. The partial overlap of SPs with class II HA-RAMPs stems from the hydrophobic character of SPs together with the higher hydrophobicity of class II relatively to class I HA-RAMP, both of which are reflected by the z1.lag1 term which has the largest contribution in PC2 (Supplementary Figure 4D). By contrast, in the plane defined by PC1 and PC3, class II HA-RAMPs grouped closer with class I HA-RAMPs and away from SPs (Supplementary Figure 5B) because the ACC term z1.lag1 has as a much lower contribution to PC3 than the now prominent ACC term z1.lag4 reflecting the amphipathic character of the helix (Supplementary Figure 4E). It is of note that TPs still overlapped with class I HA-RAMPs in the PC1/PC3 plane, which reflects their tighter physico-chemical relatedness than that of SPs with class II HA-RAMPs, as already observed when comparing the general features of the peptides (Figure 4).

HA-RAMPs and TPs show dual targeting and antimicrobial activities

A TP cleavage-site fragment is required for import of the Venus reporter

To assess TP-activity of AMPs we used a bicistronic expression system based on ribosome re-initiation as described by Onishi and Pringle (2016) with coding sequences for candidate peptides inserted upstream of a Venus fluorescent reporter (Figure 7 and Figure 9). In this bicistronic system, the STOP-codon of the fluorescent reporter and the initiation codon of the selectable marker are separated by only six nucleotides (TAGCAT), which is sufficient to ensure robust expression of both genes in *Chlamydomonas reinhardtii*. Compared to classical expression systems where the selectable marker is driven by a separate promoter, bicistronic expression results in a much higher fraction of recovered transformants showing expression of the gene of interest (Onishi and Pringle 2016).

Mitochondria and chloroplasts were imaged respectively using a MitoTracker dye and chlorophyll autofluorescence. In the absence of Venus (expression of the selectable marker only), some crosstalk is visible in the Venus channel (Figure 7A), which appears to originate from thylakoid localized pigments and in particular from the eyespot. In the absence of a presequence (Figure 7B) Venus remains cytosolic.

Surprisingly, the fluorescent reporter was equally cytosolic when the Rubisco activase cTP (RBCA-cTP) up to the cleavage site was included upstream of Venus (Figure 7C). For import into the chloroplast, a stretch of 23 downstream residues was required (RBCA-cTP+), reconstituting a native cleavage site (Figure 7D). This finding is in line with previous efforts to target fluorescent reporters to the chloroplast in Chlamydomonas (Razzak et al. 2017) and led us to include the cleavage site encompassing residues -10 to +23 in subsequent constructs. This cleavage site fragment (RBCA-cs) by itself displayed no capacity for directing the Venus reporter into either organelle (Figure 7E). We note that in addition to colocalizing with chlorophyll (Supplementary Figure 6), the Venus reporter driven by RBCA-cTP+ was abundant around the pyrenoid, the native location of RBCA (Fig.7D). The pyrenoid is a proteinaceous structure of importance to the algal carbon-concentrating mechanism that contains a lower density of thylakoid membranes than the rest of the chloroplast. As a result, the pyrenoid is visible as a characteristic dark zone in chlorophyll auto-fluorescence at the apex of the chloroplast (Mackinder et al. 2017), making Venus accumulating at this site easy to spot. Mitochondrial localization of Venus driven by a native Chlamydomonas mTP, including post-cleavage site residues, is characterized by a tell-tale pattern (Mackinder et al. 2017) and colocalisation with the MitoTracker signal (Figure 7F, Supplementary Figure 6).

HA-RAMPs target Venus to endosymbiotic organelles

To assess the organelle targeting ability of HA-RAMPs, we selected five peptide candidates drawn from class I that clustered with TPs based on their Euclidean distances (Figure 2): Bacillocin 1580 and Enterocin HF from the Bacteriocin IIA family, the Cecropin Sarcotoxin-1D, Brevinin-2ISb from the Brevinin-2 family, and the well-studied Magainin II. These candidates are recognized as TPs in the logistic regression analysis (Figure 5 and Supplementary Figure 3B) and they localize next to TPs in our PCA analysis (Figure 8A). When fused alongside RBCA-cs upstream of Venus and expressed in Chlamydomonas (Figure 9), both Bacillocin 1580 (Figure 9A) and Enterocin HF (Figure 9B) give rise to a fluorescence signal that is co-localized with chlorophyll auto-fluorescence (Supplementary Figure 6), in line with their proximity to cTPs in the PCA (Figure 8A). There is also a marked accumulation around the pyrenoid, particularly for Bacillocin 1580. Although closer to mTPs in our PCA analysis (Figure 8A), Sarcotoxin-1D was equally found to target Venus to the chloroplast (Figure 9C), in line with the fact that some cTPs are found in the

vicinity of our Sarcotoxin-1D construct (Figure 8A). Brevinin-2Isb on the other hand, proximal both to mTPs and cTPs in the PCA, resulted in Venus fluorescence showing the typical pattern of mitochondrial localization, co-localizing with the MitoTracker dye (Figure 9D, Supplementary Figure 6). Magainin II also targeted Venus to the mitochondria (Figure 9E), as may be expected from the construct most distal to cTPs in our PCA (Figure 8A). Class I HA-RAMPs are thus capable of targeting a cargo protein to either type of endosymbiotic organelle.

For an independent assessment of Venus localisation, we isolated intact chloroplasts and mitochondria from whole cells of *C. reinhardtii* harboring one chloroplast and one mitochondrial targeting construct (Figure 10). It is of note that chloroplast-enriched fractions still show some mitochondrial contamination due to the presence of a subpopulation of mitochondria firmly bound to chloroplasts in this microalga (Willmund and Schroda 2005). In agreement with fluorescence imaging observations, Bacillocin-driven Venus-FLAG was absent from isolated mitochondria but present in whole-cell and chloroplast-fractions, just like chloroplast markers OEE2 and RBCS (Figure 10A). Magainin-driven Venus-FLAG behaved like mitochondrial markers COXIIb and F1β, being present in all three fractions (Figure 10B).

For both constructs, most of the Venus-FLAG reporter was protected from degradation by proteinase K in isolated organelles unless treated with detergents, again mirroring the behavior of organelle-specific controls (Figure 10C,D). This observation confirms that Bacillocin 1580 and Magainin II act as *bona fide* TPs, with the majority of the cargo protein localized inside the respective target organelle. A larger-sized fraction of Venus-FLAG did show sensitivity to proteinase K in the absence of detergents, just like the tubulin and BIP, which represent minor contaminants in mitochondrial and chloroplast fractions respectively. These proteins were thus not protected within organelles and digested readily irrespective of the presence of a detergent. Thus, a subpopulation of AMP-reporter pre-proteins remains associated with the outer membrane of either organelle, likely as a result of incomplete or aborted translocation.

TPs show antimicrobial activity

To determine the AMP-activity of TPs, we performed the symmetrical experiment. Several chemically synthesized TPs, chosen for their variable proximity to class I HA-RAMPs (Figure 8B), were used to challenge *Bacillus subtilis* in a standard assay (Wiegand et al. 2008) using Magainin II as a positive control (Figure 11). *B. subtilis* was chosen as target organism rather than *E. coli*, the other readily available standard laboratory bacterial specimen, because the former proved more sensitive to HA-RAMP activity (Supplementary Figure 8). A high sensitivity was deemed a useful feature for an antimicrobial assay in this proof-of-principle experiment, since TPs were suspected of having a lower activity than *bona fide* AMPs by virtue of having been selected for targeting and not for impeding microbial growth. All four tested Chlamydomonas TPs showed antimicrobial activity, as did the yeast pF1β-mTP for which antimicrobial activity had previously been reported (Hugosson et al. 1994). By contrast, neither the Chlamydomonas TL16-tSP, which targets proteins to the lumen of thylakoid membranes, nor the small hormone peptide cholecystokinin-22 (cck-22), here used as negative controls, inhibited growth demonstrating that antimicrobial activity is not simply an inherent feature shared by all peptides.

Discussion

We here provide *in silico* and *in vivo* evidence that TPs and AMPs share significant physico-chemical properties and functionalities, in support of their common evolutionary origin (Wollman 2016).

The diversity of HA-RAMPs

In our *in silico* analysis, we first showed that HA-RAMPs can be grouped into distinct subtypes that do not always line up with the classification into the antimicrobial families described in the literature (Figure 2 and 3). Rather than reflecting structural relationships between peptides, AMP families have often been named after the source of novel peptides or a defining characteristic. For example, gaegurins were named after the Korean word "Gaegury", which means frog, and regroup AMPs isolated from the Korean frog *Glandirana emeljanovi* (formerly *Rana rugosa*) even though the peptides in question are closer in sequence to other HA-RAMP families than to each other (Won et al. 2009). Unsurprisingly, gaegurins are spread across both HA-RAMP classes (Supplementary Figure 1). However, not all idiosyncratically named families lack structural systematics. As more and more AMPs were being characterized, investigators started to class novel peptides that were deemed related in primary sequence in families named after the first described peptide of that type. Thus cecropins, named after the giant silk moth *Hyalopphora cecropia* from which the first members were extracted, are now a well-established family including peptides that carry names such as Sarcotoxin or Bactericidin. Consequently, cecropins form a relatively uniform group to the top-right of the NJ tree (Figure 3). Similarly, brevinin-1 and brevinin-2 are two separate families, each expanded by the addition of homologous peptides (Conlon 2008), that happen to carry the same name because they were originally isolated together from the Japanese frog *Pelohylax porosus* (formerly *Rana brevidopda porsa*) (Morikawa et al. 1992). The brevinin-1 family in particular forms a fairly homogeneous branch towards the bottom of the tree (Figure 3), interspersed only by gaegurin and ranatuerin peptides that had already been recognized as brevinin-1 homologues (Conlon 2008; Won et al. 2009). brevinin-2, by contrast, is split into three relatively distinct subgroups in our analysis, implying that this family might reasonably be split up. It is interesting to note that Rana-box containing brevinin, esculentin and ranatuerin families span both class I and class II HA-RAMPs (Supplementary Figure 1). The Rana-box is a C-terminal disulphide bridge motif that is common among ranid AMPs that appears to indicate a common origin, implying that related HA-RAMPs can rapidly diversify in terms of physico-chemical properties.

A systematic reclassification of HA-RAMP families is beyond the scope of this paper. We hence decided not to implement any robustness analysis with regard to NJ tree branches, the relative positioning of which should therefore be interpreted with caution. Nonetheless, our analysis indicates that a systematic classification according to physico-chemical properties would be possible upon further investigation, which should prove of interest to the AMP community.

Argument for a common origin of TPs with a class of HA-RAMPs

It has been proposed that a fraction of random sequences (between 20 and 30% of those that were tested) could function as mitochondrial and secretory targeting peptides (Lemire et al. 1989; Dunn and Paavilainen 2019; Kaiser et al. 1987). However, it has been shown that the random sequences that allowed functional targeting were strongly biased for the positioning of hydrophobic residues and certain amino-acids, suggesting the requirement of very specific properties for targeting function (Lemire et al. 1989). This is in line with our observation of an overlap between some random peptides and SPs, class I HA-RAMPs and TPs in our PCA analysis (Figure 6).

However, whether by k-means clustering, LR classification or PCA analysis, TPs consistently grouped together with class I HA-RAMPs, and away from the three types of SPs acting on the bacterial periplasmic membrane, the ER or thylakoid membranes. This should be distinguished from the partial overlap of a subset of class II HA-RAMPs with SPs, as observed in k-means clustering and when using the first two components of the PCA analysis. These two families differ by the amphipathic versus hydrophobic character of their helical domain. As expected, their overlap, partly accounted for by their relatively high hydrophobic character, was much less robust, and disappeared when using the third principal component of the PCA (Supplementary Figure 5). By contrast, TPs still overlap with class I HA-RAMPs in the third principal component of the PCA, further emphasizing their extensive similarities. Their shared physico-chemical properties suggest that a large subset of class I HA-RAMPs could be evolutionarily linked with TPs. This evolutionary relationship is exemplified here by the targeting abilities of the HA-RAMPs we probed experimentally, such as bacillocin 1580 targeting the chloroplast and Magainin II targeting the mitochondria (Figure 9, Figure 10).

One could argue that convergent evolution of class I HA-RAMPs and TPs, driven by strong selective constraints, could have led them to the same optimum of physico-chemical properties. However, our results show that cyclotides as well as class II HA-RAMPs are clearly distinct from class I HA-RAMPs despite a shared antimicrobial function. Analogously, SPs function as cleavable N-terminal targeting sequences like TPs but do not group with class I HA-RAMPs (Figure 5, Figure 6) nor do they show antimicrobial activity, as the tested TPs do (Figure 11). Thus, the respective antimicrobial and intracellular targeting functions of class I HA-RAMPs and TPs do not *per se* constitute a selective pressure for convergent evolution.

Although evolutionary scenarios necessarily give rise to conflicting views, it should be kept in mind that neither the scenario of convergent evolution nor that of a spontaneous generation as an explanation for the existence of extant TPs account for the emergence of the ancestral mitochondrial and chloroplast translocation systems. In contrast, considering an antimicrobial origin of TPs implies a more parsimonious scenario, in which the import and destroy resistance mechanism (Guefrachi et al. 2015) from the bacterial ancestor of the endosymbiont is at the root of the translocation systems (Wollman 2016). Further support for the view that both mitochondrial and chloroplast primary endosymbioses result from bacteria that became resistant to the attack of AMPs via an 'import and destroy' mechanism comes from recent studies with the amoeba *Paulinella chromatophora*. *P. chromatophora* acquired a novel primary endosymbiotic organelle called "chromatophore" approximately 100 million years ago (Delaye et al. 2016). Proteomic analysis of these chromatophores identified a large set of imported AMP-like peptides as well as chromatophore-imported proteins harboring common N-terminal sequences containing AMP-like motifs (Singer et al. 2017). These findings thus provide an independent example of a third primary endosymbiosis that is accompanied by the evolution of an AMP-derived protein import process. The detailed evolutionary histories of extant organelle translocons and bacterial transmembrane channels involved in AMP-resistance mechanisms should provide a means to further assess the antimicrobial origin of organelle-targeting peptides.

Methods and Materials

Sequence Data Set

The groups of peptides used in this study and the corresponding data sources are given in Table S1. Detailed information for each of the targeting peptides, signaling peptides, antimicrobial peptides and random peptides are given in Tables S2 to S5. Antimicrobial peptides were extracted from the CAMPR3 database (Waghu et al. 2016). We selected the families of antimicrobial peptides corresponding to the following criterion : i) activity experimentally validated, ii) documented amphipathic alpha helix structure, with at least one family member with a resolved 3D structure available in the Protein Data Bank (Berman et al. 2000), iii) documented activity of bacterial membrane destabilization. For 3 families comprising peptides with very different structures, additional filtering was applied. We recovered only the bacteriocin of type IIa that are characterized by an amphipathic helix (Rea et al.) in BACTIBASE (Hammami et al. 2010). We selected only those defensins with a resolved 3D structure with at least 5 consecutive residues in a alpha-helix and only those cathelicidin with a resolved structure or defined as an amphipathic helix peptide in Mookherjee et al. (2013) via UNIPROT. As a negative control, we retrieved the cyclotides of globular structure. TPs with experimentally-confirmed cleavage sites were recovered from proteomic studies (see Table S1). 200 eSPs were randomly extracted among 4707 confirmed eSPs from the Signal Peptide Website (http://www.signalpeptide.de). eSPs were selected so as to follow the length distributions of the peptides from HA-RAMP class I families. 200 random peptides were generated following the amino-acid frequencies observed in the Uniprot database and the length distribution of TPs. Data sets were curated to exclude sequences shorter than 12 amino acids or longer than 100 amino acids, following the length distributions of the peptides from HA-RAMP class I families.

Peptide description and Auto-Cross Covariance (ACC) Terms

As amino acid descriptors, we used the Z-scales established by Hellberg et al. (1987). ACC terms between Z-scales were computed as described previously in Wold et al. (1993). ACC terms combine auto-covariance (same z-scale i = j) and cross-covariance (different z-scale i ≠ j) of neighboring amino acids in a window of 4 residues (lags (l) ranging from 1 to 4). There are thus nine nearest neighbor ACC terms for the 3 z-scales factors, yielding 36 ACC terms per peptide of length N. Each ACC term is defined for a given Z-scale couple (i,j) as follows:

$$ACC_{i,j,l} = \sum_{n=1}^{N} -l \frac{z_{i,n} \times z_{j,n+l}}{N - l}$$

The Z-scales values were retrieved from the AAindex database (https://www.genome.jp/aaindex/) via the R package protr (v1.5-0). ACC terms were calculated with the acc function from the same R package and pre-processed by mean-centering and scaling to unit variance.

To assign a hydrophobicity value to a given peptide, we select the highest value obtained for a sliding window of 12 resiudes along the peptide. We used the hydrophobicity indices of amino acids estimated by octanol/water partitionin (Fauchère and Pliska 1983) to determine the mean hydrophobicity of the 12-residues window. The net charge of a peptide is the sum of the positively charged residues (arginin and lysine) and of the negatively charged residues (glutamate and aspartate) at pH 7.4.

The number of residues along the peptide that can theoretically adopt an amphipathic helical structure is calculated as follows: The peptide is drawn along an α helical wheel and the longest

region (of at least 12 residues) of the peptide that can adopt an amphipathic helix is searched following the same criterion as in Heliquest (Gautier et al. 2008).

K-means clustering

Peptides were clustered based on the Euclidean distance defined above by k-means (scikit-learn Python package version 0.21.2). Centroid initialization was performed with the 'k-means++' method with the best inertia among 100 runs for k ranging from 2 to 10. The selected k value (2) is the one for which the average silhouettes coefficient -a measure of the clustering quality- (Rousseeuw 1987) over all peptides is the best.

Distance trees

Euclidean distances between pairs of 36-dimensional vectors defining each peptide with their ACC terms were used to compute a distance tree between the studied HA-RAMPs, with the neighbor-joining implementation of the scikit-bio Python library, version 0.5.5. Tree annotation and display was performed with iTOL v5.

Statistical classification of peptide properties

The binomial logistic regression classifier was trained on ACC terms from 72 human mTPs and 96 bSPs. We used an elastic net penalty in a saga solver with an l1-ratio (α-parameter) of 0, which is equivalent to using only ridge penalty, where all features take part in the model (Supplementary Figure 4A), and a C parameter ($1/\lambda$ of 0.1. The classifier provides for each studied peptide a probability to be classified as human mTP or bSP. The α and $1/\lambda$ parameters were optimised with 10-fold cross validation. Firstly, when $1/\lambda = 1$, the best accuracy (0.92) was obtained with α between 0 and 0.13. Secondly, with $\alpha = 0$, the best accuracy (0.93) was obtained with $1/\lambda = 0.1$. Analyses were performed with the scikit-learn Python package version 0.21.2. The two principal components of a Principal Component Analysis (PCA) of the peptides defined by their 36 ACC terms were used for visualization of the peptides properties. The weight of each of the variables in the PCA are summarized by correlation circles in Supplementary Figure 4B-E.

Strains and culture conditions

Chlamydomonas reinhardtii cells derived from wild-type strain T222+ were grown in mixotrophic conditions in Tris-acetate-phosphate (TAP) medium (Harris 1989) under ~30 µmol photons m-2 s-1 at 25°C, either in 200 µL in 96-well plates for 3-4 days or in agitated flasks of 200 mL. Constructs were transformed into strain T222+ by electroporation and transformants, selected for paromomycin resistance, were screened for high Venus expression in a fluorescence plate reader (CLARIOstar, BMG labtech) as described in Onishi and Pringle (2016).

Generation of constructs

Constructs were made by inserting sequences coding for candidate peptides directly upstream of the Venus start codon in plasmid pMO611, kindly provided by the Pringle lab. pMO611 is a

derivative of the published bicistronic expression plasmid pMO449 (Onishi and Pringle 2016) in which translation of the eight first RBCS codons ahead of the Venus coding sequence was prevented by mutating the start codon to CTG. To increase expression of the AMP constructs, we modified plasmid pMO611 further by inserting RBCS2 intron 2 within Venus, 73 bases downstream of the initiation codon. Introns do not influence targeting. Native Chlamydomonas TP sequences were amplified from strain T222+ genomic DNA, while codon optimized AMP gene sequences were synthesized by Eurofins Genomics. Sequences used are detailed in Table S6. Peptide constructs were assembled and integrated upstream of Venus using the NEBuilder HiFi Assembly kit (New England Biolabs). Correct assembly was verified by sequencing of inserts and flanking regions. Linear transformation cassettes were excised from plasmids with EcoRV (New England Biolabs) prior to transformation.

Microscopy

For confocal imaging (Figure 7, Figure 9, Supplementary Figure 7), cells were subjected to 0.1μM MitoTracker Red CMXRos (ThermoFisher) for 30 minutes in the dark and washed with TAP prior to imaging on an upright SP5 confocal microscope (Leica). Venus (excitation 514 nm/532-555 nm emission) and MitoTracker (561 nm/573-637 nm) were imaged sequentially to avoid crosstalk, each alongside chlorophyll autofluorescence (670-750nm emission) to eliminate cells that had moved between images (data from 514 nm excitation shown in figures). Epifluorescence images (Supplementary Figure 7) were taken on an Axio Observer.Z1 inverted microscope (Zeiss) equipped with an ORCA-flash4.0 digital camera (Hamamatsu) and a Colibri.2 LED system (Zeiss) for excitation at 505 nm for Venus (filter 46HE YFP shift free, 520-550nm emission) and 470 nm for chlorophyll autofluorescence (filter set 50, 665-715nm emission) with cells in poly-L-lysine (Sigma Aldrich) coated 8-well μ-slides (Ibidi). Image brightness was adjusted for presentation in figures and cyan, yellow and magenta linear lookup tables, assigned to MitoTracker, Venus and chlorophyll channels respectively, in Fiji (http://fiji.sc/Fiji, version 2.0.0-rc-69/1.52p). To quantify colocalisation (Supplementary Figure 6), cells were cropped out of larger fields-of-view using a standard region-of-interest quadratic box with 13µm side length in Fiji. Fiji was also used to measure image background intensities from outside the cell for the Venus channel, or from within the cell but outside the organelle in the case of the MitoTracker and chlorophyll channels. Backgrounds were then substracted in R (version 3.6.1). Pearson correlation coefficients were used as a measure of colocalization (Dunn et al. 2011).

Biochemistry

Chloroplasts and mitochondria were isolated essentially as described in Willmund and Schroda (2005), with the following modifications. Protease assay data (Figure 10C,D) and purity data (Figure 10A,B) came from two different extractions (a) and (b) respectively, which were done as follows. 2 L cultures inoculated either into TAP 48h before experiment and grown at 50 µEm-2s-1 (a), or in Minimal Medium 72h before experiment and grown aerated under at 200 µEm-2s-1 (b), were kept in darkness the night preceding the experiment. Cells were washed twice in 20 mM HEPES-KOH pH 7.2 (a), or washed once in 10 mM HEPES-KOH pH 7.2 and then subjected to an acid shock to remove flagella as described in Craige et al. (2013) (1min at pH 4.5 by adding 1M acetic acid, then neutralising back to pH 7.2 with 1M KOH) (b). Working in a cold room, cells were then resuspended in 50 mL ice cold breaking buffer (300 mM Sorbitol, 3 mM MgCl2, 5 mM EDTA, 0.5% Polyvinylpyrrolidone PVP40, in 67 mM HEPES-KOH pH 7.2) and disrupted by

nebulization with Argon using either one pass at 40 psi (a), or four passes at 80 psi (b), after which an aliquot was set aside. A brief centrifugation up to 4000 g (8 °C, manual stop when reaching 6000 RPM, Beckman JA-12) was used to separate chloroplast and mitochondrial fractions in pellet and supernatant respectively. Chloroplasts were recovered at either the 45/70% (a) or the 20/40% (b) interphase of a discontinuous Percoll gradient (in chloroplast wash buffer: 300 mM Sorbitol, 3 mM MgCl2 in 67 mM HEPES-KOH pH 7.2) after 2 h of ultracentrifugation (4 °C, 5000 RPM, Beckman SW41, slow acceleration and deceleration) and washed in 5 volumes of chloroplast wash buffer (8 °C, manual stop when reaching 3000 RPM, Beckman JA-20, acceleration: 5, deceleration: 0). Mitochondria were purified using a series of differential centrifugations: 20 min at 1500 g to pellet residual chloroplasts (8 °C, 3500 RPM, Beckman JA-12), 20 min at 20 000 g to pellet mitochondria (8 °C, 11 300 RPM, Beckman JA-13.1), 40 min at 20 000 g (8 °C, 11 300 RPM, Beckman JA-13.1) to run a 20% Percoll purification (in 250 mM Sorbitol, 1 mM EDTA, 0.5% PVP40 in 10 mM MOPS-KOH pH 7.2) of which the bottom ~3 mL were kept, and 20 min at 20 000 g (8 °C, 11 300 RPM, Beckman JA-13.1) to pellet mitochondria after percoll removal by dilution with 40 mL mitochondrial wash buffer (250 mM Sorbitol, 1mM EDTA in 10 mM phosphate buffer pH 7.2). Isolated organelle pellets were resuspended in 100 µL SEM (250 mM Sucrose, 1 mM EDTA, in 10 mM MOPS-KOH, pH 7.2). For the proteinase assay (Willmund and Schroda 2005), one 22 µL aliquot of isolated organelles was treated with 1.1 µL 25 % Triton X-100 for 5 min on ice, another with an equal volume of water. Each sample was split into two aliquots of 10.5 µL, one of which was treated with 0.5 µL proteinase K 20x stock in SEM (final concentration: 150 µg/mL) for 15 min on ice, and the other with an equal volume of SEM. All aliquots were then treated with 0.25 µL 100 mM PMSF and 2.8 µL 5x storage buffer (5x Roche cOmplete Mini proteinase inhibitor cocktail, 50 mM NaF, 1 M DTT, 1 M Na2CO3) and stored at -20 °C. Of each isolated organelle and whole cell sample, an aliquot was precipitated overnight at 4°C by adding 900 µL 80% Acetone, pelleted for 10min (4°C, max speed, Eppendorf 5415 D benchtop centrifuge), dried for ~10min under vacuum, resuspended in protein resuspension buffer (2%SDS, 10mM NaF, 1x Roche cOmplete Mini proteinase inhibitor cocktail in 100mM Na2CO3) and used for protein quantification by BCA assay (ThermoFisher Scientific). Samples were then run on either 4-15% (a) or 8-16% (b) precast gels (Biorad) and used for immunoblot detection. To be able to probe a very limited amount of sample with multiple antibodies, membranes were cut horizontally into separately treated strips. Primary antibodies raised against the following proteins were used in block buffer (5% BSA, 0.1% Tween-20 in PBS) at the indicated dilutions: α-Tubulin (Sigma Aldrich T5168, 1:50,000), FLAG (Agrisera AS15 2871, 1:10,000), COXIIb (Agrisera AS06 151, 1:10,000), BIP (Agrisera AS09 481, 1:2000), CF1β (homemade, rabbit, 1:50,000) (Chua and Blomberg 1979), OEE2 (homemade, rabbit, 1:2000) (de Vitry et al. 1989) RBCS (kindly supplied by Spencer Whitney, rabbit, 1:20,000) (Whitney and John Andrews 2001), NAB1 (Agrisera AS08 333, 1:10,000). ECL signals were recorded on a ChemiDoc Touch (Biorad). Blots were cropped and final figures assembled in powerpoint.

Antimicrobial activity assays

Standard minimum inhibitory concentration broth microdilution assays in the presence of BSA/acetic acid were performed in triplicate as described in (Wiegand et al. 2008) using peptides chemically synthesized to ≥95% purity (Proteogenix). Dilution series are based on net peptide content, calculated by multiplying the dry weight by %N and by purity. %N is a measure of the peptide (rather than salt) fraction in the lyophilized product, while purity, provided by the

manufacturer, is the fraction of the peptide with the desired sequence among all supplied peptides. Peptide sequences are listed in Table S7.

Statistical analysis

Chi^2 tests were used to analyse the distribution of peptides according to their functional group among the different k-means clusters (Figure 2). One-way analysis of variance (ANOVA) and Tukey post-hoc were used to analyse the distributions of peptide features (Figure 4 and Supplementary Figure 2) and the Pearson correlation coefficients of fluorescence intensities (Supplementary Figure 6). AP-value threshold of 0.05 was used for all tests. All statistical calculations were performed with R version 3.6.1 and with functions from the Python scipy module.

Code avaibility

All Python and R in-house scripts are available upon request.

Acknowledgments

We thank Richard Kuras, Ulrike Endesfelder, Michael Schroda and Moritz Meyer for fruitful discussions and advice, Masayuki Onishi and the Pringle lab, for sharing their bi-cistronic expression plasmids, Pierre Crozet for advice on plasmid design, Zhou Xu and Simon Desjardin for help with microscopy. This work was supported by the annual funding from the Centre National de la Recherche Scientifique and Sorbonne University to UMR 7141. ODC was supported by The Rothschild Foundation, the LabEx Dynamo (ANR-LABX-011) and the ChloroMitoRAMP grant (ANR-19-CE13-0009). CG was supported by the MATHTEST grant (ANR-18-CE13-0027).

References

Andersson SGE, Zomorodipour A, Andersson JO, Sicheritz-Pontén T, Alsmark UCM, Podowski RM, Näslund AK, Eriksson A-S, Winkler HH, Kurland CG. 1998. The genome sequence of Rickettsia prowazekii and the origin of mitochondria. *Nature* **396**: 133–140.

Andrä J, Herbst R, Leippe M. 2003. Amoebapores, archaic effector peptides of protozoan origin, are discharged into phagosomes and kill bacteria by permeabilizing their membranes. *Developmental and Comparative Immunology* **27**: 291–304.

Archibald JM. 2015. Endosymbiosis and Eukaryotic Cell Evolution. *Current Biology* **25**: R911–R921.

Ball SG, Bhattacharya D, Weber APM. 2016. Pathogen to powerhouse. *Science* **351**: 659–60.
Bechinger B, Zasloff M, Opella SJ. 1993. Structure and orientation of the antibiotic peptide magainin in membranes by solid-state nuclear magnetic resonance spectroscopy. *Protein Science* **2**: 2077–2084.
Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE.

2000. The Protein Data Bank. *Nucleic Acids Res* **28**: 235–242.

Besse A, Peduzzi J, Rebuffat S, Carré-Mlouka A. 2015. Antimicrobial peptides and proteins in the face of extremes: Lessons from archaeocins. *Biochimie* **118**: 344–355.

Bruce BD. 2000. Chloroplast transit peptides: Structure, function and evolution. *Trends in Cell Biology* **10**: 440–447.

Chotewutmontri P, Holbrook K, Bruce BD. 2017. Chapter Six - Plastid Protein Targeting: Preprotein Recognition and Translocation. In *International Review of Cell and Molecular Biology* (ed. L. Galluzzi), Vol. 330 of, pp. 227–294, Academic Press

Chua NH, Blomberg F. 1979. Immunochemical studies of thylakoid membrane polypeptides from spinach and Chlamydomonas reinhardtii. A modified procedure for crossed immunoelectrophoresis of dodecyl sulfate.protein complexes. *J Biol Chem* **254**: 215–223.

Clarke M, Lohan AJ, Liu B, Lagkouvardos I, Roy S, Zafar N, Bertelli C, Schilde C, Kianianmomeni A, Bürglin TR, et al. 2013. Genome of Acanthamoeba castellanii highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. *Genome Biology* **14**: R11.

Conlon JM. 2008. Reflections on a systematic nomenclature for antimicrobial peptides from the skins of frogs of the family Ranidae. *Peptides* **29**: 1815–1819.
Craige B, Brown JM, Witman GB. 2013. Isolation of chlamydomonas flagella. *Current Protocols in Cell Biology* 1–9.

de Vitry C, Olive J, Drapier D, Recouvreur M, Wollman FA. 1989. Posttranslational events leading to the assembly of photosystem II protein complex: a study using photosynthesis mutants from Chlamydomonas reinhardtii. *Journal of Cell Biology* **109**: 991–1006.

Delaye L, Valadez-Cano C, Pérez-Zamorano B. 2016. How Really Ancient Is Paulinella Chromatophora? *PLoS Curr* **8**. 10.1371/currents.tol.e68a099364bb1a1e129a17b4e06b0c6b

Dunn CD, Paavilainen VO. 2019. Wherever I may roam: organellar protein targeting and evolvability. *Current Opinion in Genetics & Development* **58–59**: 9–16.

Dunn KW, Kamocka MM, McDonald JH. 2011. A practical guide to evaluating colocalization in biological microscopy. *American Journal of Physiology-Cell Physiology* **300**: C723–C742.

Fauchère J-L, Pliska V. 1983. Hydrophobic parameters πof amino-acid side chains from the partitioning of N-acetyl-amino-acid amides. *Eur J Med Chem* **18**: 369–375.

Ford Doolittle W. 1998. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends in Genetics* **14**: 307–311.

Frain KM, Gangl D, Jones A, Zedler JAZ, Robinson C. 2016. Protein translocation and thylakoid biogenesis in cyanobacteria. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1857**: 266–

273.

Fukasawa Y, Oda T, Tomii K, Imai K. 2017. Origin and Evolutionary Alteration of the Mitochondrial Import System in Eukaryotic Lineages. *Mol Biol Evol* **34**: 1574–1586.

Gautier R, Douguet D, Antonny B, Drin G. 2008. HELIQUEST: a web server to screen sequences with specific alpha-helical properties. *Bioinformatics* **24**: 2101–2102.

Guefrachi I, Pierre O, Timchenko T, Alunni B, Barrière Q, Czernic P, Villaécija-Aguilar J-A, Verly C, Bourge M, Fardoux J, et al. 2015. Bradyrhizobium BclA Is a Peptide Transporter Required for Bacterial Differentiation in Symbiosis with Aeschynomene Legumes. *MPMI* **28**: 1155–1166.

Hammami R, Zouhir A, Le Lay C, Ben Hamida J, Fliss I. 2010. BACTIBASE second release: a database and tool platform for bacteriocin characterization. *BMC Microbiology* **10**: 22.
Harris E. 1989. *The Chlamydomonas Sourcebook - 1st Edition*. Academic Press, San Diego.

Hellberg S, Sjoestroem M, Skagerberg B, Wold S. 1987. Peptide quantitative structure-activity relationships, a multivariate approach. *J Med Chem* **30**: 1126–1135.

Hugosson M, Andreu D, Boman HG, Glaser E. 1994. Antibacterial peptides and mitochonrial presequences affect mitochonrial coupling, respiration and protein import. *European Journal of Biochemistry* **223**: 1027–1033.

Kaiser CA, Preuss D, Grisafi P, Botstein D. 1987. Many random sequences functionally replace the secretion signal sequence of yeast invertase. *Science* **235**: 312–317.

Lemire BD, Fankhauser C, Baker A, Schatz G. 1989. The mitochondrial targeting function of randomly generated peptide sequences correlates with predicted helical amphiphilicity. *J Biol Chem* **264**: 20206–20215.

Mackinder LCM, Chen C, Leib RD, Patena W, Blum SR, Rodman M, Ramundo S, Adams CM, Jonikas MC. 2017. A Spatial Interactome Reveals the Protein Organization of the Algal CO2-Concentrating Mechanism. *Cell* **171**: 133–147.

Maróti G, Kereszt A, Kondorosi É, Mergaert P, Maro G. 2011. Natural roles of antimicrobial peptides in microbes , plants and animals. *Research in Microbiology* **162**: 363–374.

Martin WF, Garg S, Zimorski V. 2015. Endosymbiotic theories for eukaryote origin. *Philos Trans R Soc Lond B Biol Sci* **370**.

Mookherjee N, Brown KL, Hancock REW. 2013. Chapter 15 - Cathelicidins. In *Handbook of Biologically Active Peptides (Second Edition)* (ed. A.J. Kastin), pp. 77–84, Academic Press, Boston

Morikawa N, Hagiwara K, Nakajima T. 1992. Brevinin-1 and -2, unique antimicrobial peptides from the skin of the frog, Rana brevipoda porsa. *Biochemical and Biophysical Research*

Communications **189**: 184–190.

Nakai M. 2018. New Perspectives on Chloroplast Protein Import. *Plant Cell Physiol* **59**: 1111–1119.

Neupert W, Herrmann JM. 2007. Translocation of proteins into mitochondria. *Annual Review of Biochemistry* **76**: 723–49.

Onishi M, Pringle JR. 2016. Robust Transgene Expression from Bicistronic mRNA in the Green Alga Chlamydomonas reinhardtii. *G3 (Bethesda, Md)* **6**: 4115–4125.

Razzak A, Lee DW, Yoo Y, Hwang I. 2017. Evolution of rubisco complex small subunit transit peptides from algae to plants. *Scientific Reports* **7**: 9279.

Rea M, Ross R, Cotter D, Hill C. Part II Classification of Prokaryotic Antimicrobial Peptides, Chapter 3 Classification of Bacteriocins from Gram-Positive Bacteria. In *Prokaryotic antimicrobial peptides: from genes to applications*, Springer.

Reumann S, Inoue K, Keegstra K. 2005. Evolution of the general protein import pathway of plastids (review). *Molecular Membrane Biology* **22**: 73–86.

Rinker SD, Gu X, Fortney KR, Zwickl BW, Katz BP, Janowicz DM, Spinola SM, Bauer ME. 2012. Permeases of the Sap Transporter Are Required for Cathelicidin Resistance and Virulence of Haemophilus ducreyi in Humans. *J Infect Dis* **206**: 1407–1414.

Rousseeuw PJ. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* **20**: 53–65.

Schatz G, Dobberstein B. 1996. Common Principles of Protein Translocation Across Membranes. *Science* **271**: 1519–1526.

Shelton CL, Raffel FK, Beatty WL, Johnson SM, Mason KM. 2011. Sap transporter mediated import and subsequent degradation of antimicrobial peptides in Haemophilus. *PLoS Pathogens* **7**: e1002360.

Singer A, Poschmann G, Mühlich C, Valadez-Cano C, Hänsch S, Hüren V, Rensing SA, Stühler K, Nowack ECM. 2017. Massive Protein Import into the Early-Evolutionary-Stage Photosynthetic Organelle of the Amoeba Paulinella chromatophora. *Current Biology* **27**: 2763–2773.

Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet* **5**: 123–135.

van Westen GJ, Swier RF, Cortes-Ciriano I, Wegner JK, Overington JP, IJzerman AP, van Vlijmen HW, Bender A. 2013a. Benchmarking of protein descriptor sets in proteochemometric modeling (part 2): modeling performance of 13 amino acid descriptor sets. *Journal of Cheminformatics* **5**: 42.

van Westen GJ, Swier RF, Wegner JK, IJzerman AP, van Vlijmen HW, Bender A. 2013b. Benchmarking of protein descriptor sets in proteochemometric modeling (part 1): comparative study of 13 amino acid descriptor sets. *J Cheminform* **5**: 41.

von Heijne G. 1990. The signal peptide. *The Journal of Membrane Biology* **115**: 195–201.

von Heijne G, Steppuhn J, Herrmann RG. 1989. Domain structure of mitochondrial and chloroplast targeting peptides. *Eur J Biochem* **180**: 535–545.

Waghu FH, Barai RS, Gurung P, Idicula-Thomas S. 2016. CAMPR3: A database on sequences, structures and signatures of antimicrobial peptides. *Nucleic Acids Research* **44**: D1094–D1097.

Whitney SM, John Andrews T. 2001. The gene for the ribulose-1,5-bisphosphate carboxylase/ oxygenase (Rubisco) small subunit relocated to the plastid genome of tobacco directs the synthesis of small subunits that assemble into Rubisco. *Plant Cell* **13**: 193–205.

Wiedemann N, Pfanner N. 2017. Mitochondrial Machineries for Protein Import and Assembly. *Annual Review of Biochemistry* **86**: 685–714.

Wiegand I, Hilpert K, Hancock REW. 2008. Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nature Protocols* **3**: 163–175.

Willmund F, Schroda M. 2005. HEAT SHOCK PROTEIN 90C is a bona fide Hsp90 that interacts with plastidic HSP70B in Chlamydomonas reinhardtii. *Plant physiology* **138**: 2310–22.

Wold S, Jonsson J, Sjörström M, Sandberg M, Rännar S. 1993. DNA and peptide sequences and chemical processes multivariately modelled by principal component analysis and partial least-squares projections to latent structures. *Analytica Chimica Acta* **277**: 239–253.

Wollman F-A. 2016. An antimicrobial origin of transit peptides accounts for early endosymbiotic events. *Traffic* **17**: 1322–1328.

Won HS, Kang SJ, Lee BJ. 2009. Action mechanism and structural requirements of the antimicrobial peptides, gaegurins. *Biochimica et Biophysica Acta - Biomembranes* **1788**: 1620–1629.

Zachar I, Szathmáry E. 2017. Breath-giving cooperation: critical review of origin of mitochondria hypotheses. *Biology Direct* **12**: 19.

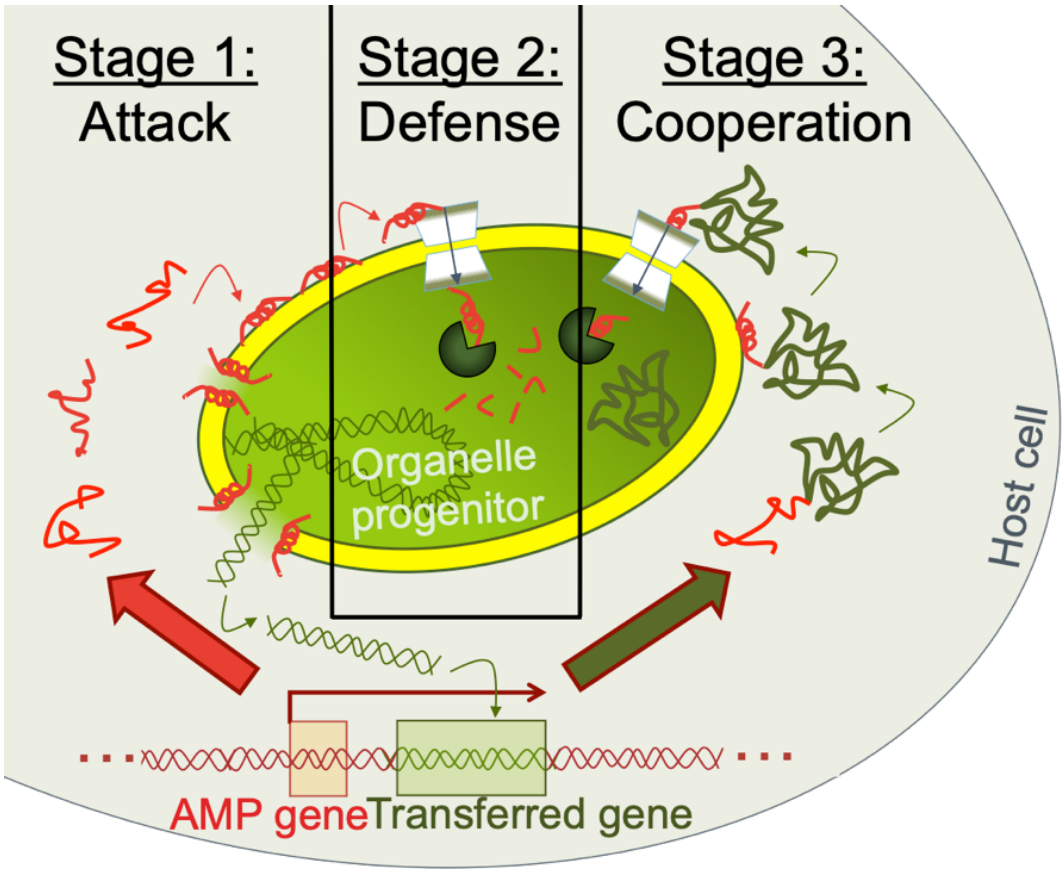


Figure 1. Emergence of TPs from AMPs implies a three-stage scenario for the evolution of endosymbiotic protein targeting systems. In stage 1, the host attacks the proto-endosymbiont using ribosomally synthesized AMPs. Cell lysis releases genetic material, which may occasionally be integrated into the host genome. In stage 2, the proto-endosymbiont acquires an import and destroy mechanism to resist host attacks, comprising a dedicated AMP transporter and a cytosolic peptidase. In stage 3, this detoxification mechanism is co-opted to import any proteins that result from serendipitous fusions of genes downstream of AMP coding sequences.

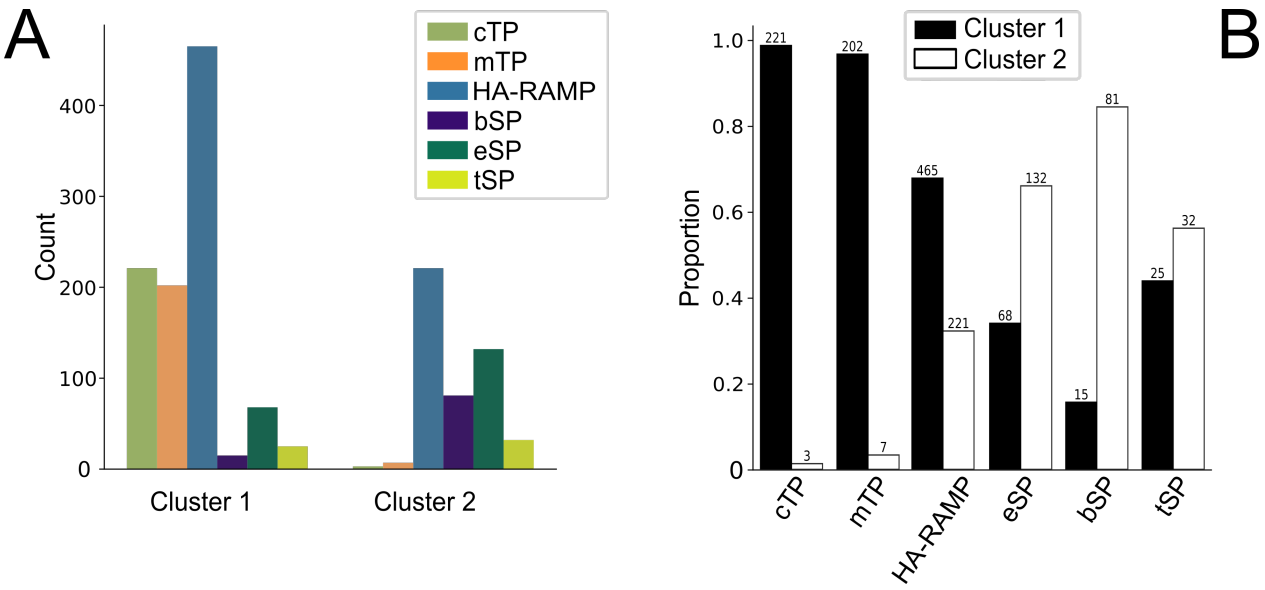


Figure 2. TPs cluster with certain types of HA-RAMPs. (A) Contribution of the different groups of peptides, as described by 36 ACC terms, to the 2 clusters obtained by k-means. (Chi^2 Pearson test, $p < 4.94 \cdot 10^{-324}$) (B) Proportion of the peptides among the 2 clusters for each group. See figure panels for color code.

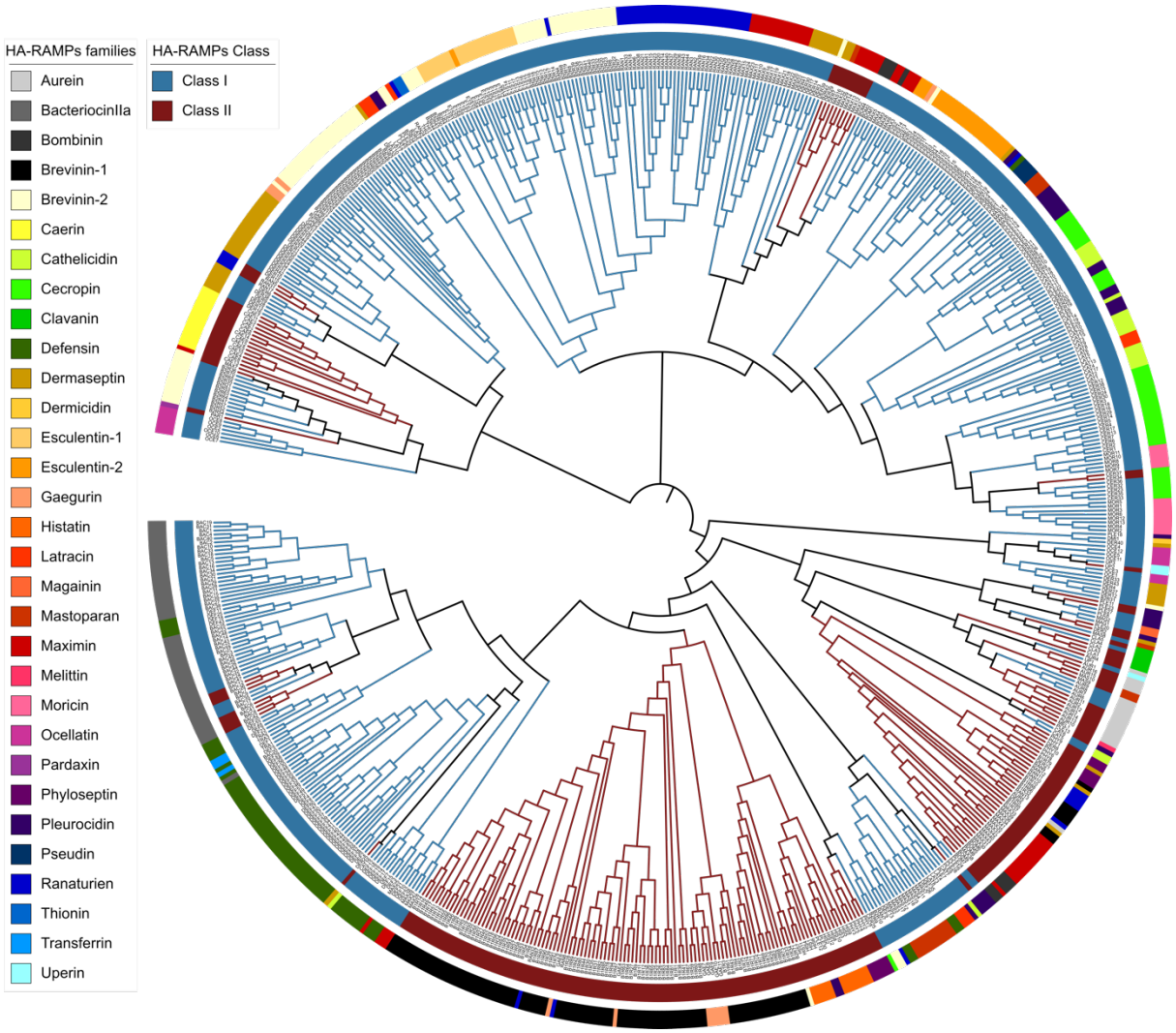


Figure 3. HA-RAMPs families and classes are spread all over the tree. Neighbor-joining tree based on Euclidean distances of HA-RAMPs described by 36 ACC terms. The branches are colored according to HA-RAMP classes I (blue) and II (red), as defined by the k-means clustering (Figure 2). When all children of a node have the same color, the color propagate inwards towards the root. The inner circle around the NJ tree indicates the HA-RAMP class of the peptide. The outer circle indicates the family of the peptide, as described in the literature. See figure box, for color code.

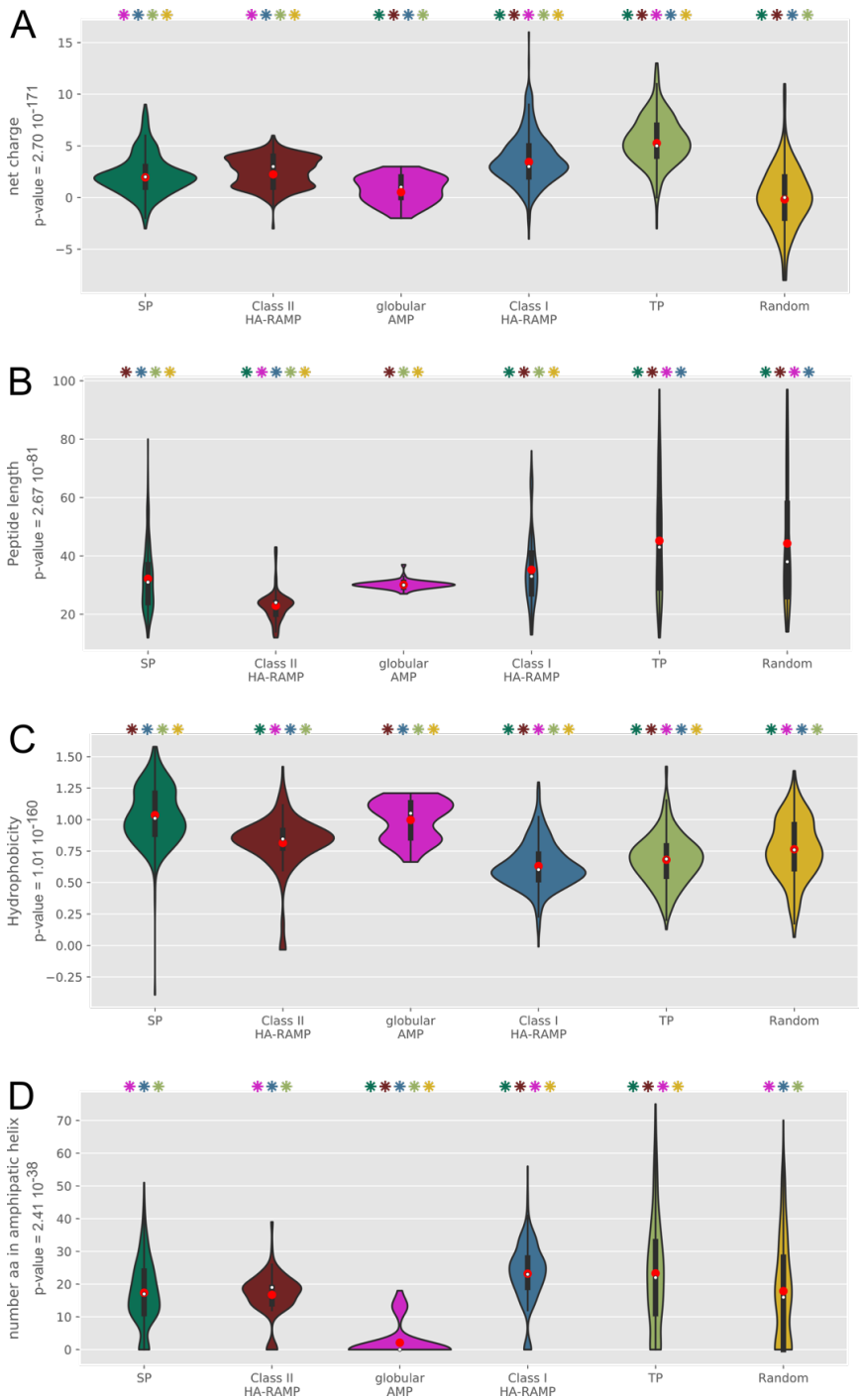


Figure 4. TPs and HA-RAMPs display similar general features. Distributions are represented by violon plots. The box plot is represented in black, white and red dots indicate median and mean of distributions, respectively. (A) net charge of the peptide, (B) peptide length in amino acids. (C) maximum hydrophobicity for a 12-residues window along the peptide. (D) Number of amino acids that can formally adopt a amphipatic alpha helical structure within the peptide. ANOVA P-values are indicated for each feature along the y-axis. Stars above distributions indicate significant differences (Tukey post-hoc tests, p-value < 0.05) between that distribution and the distribution with the same color as the one of the star.

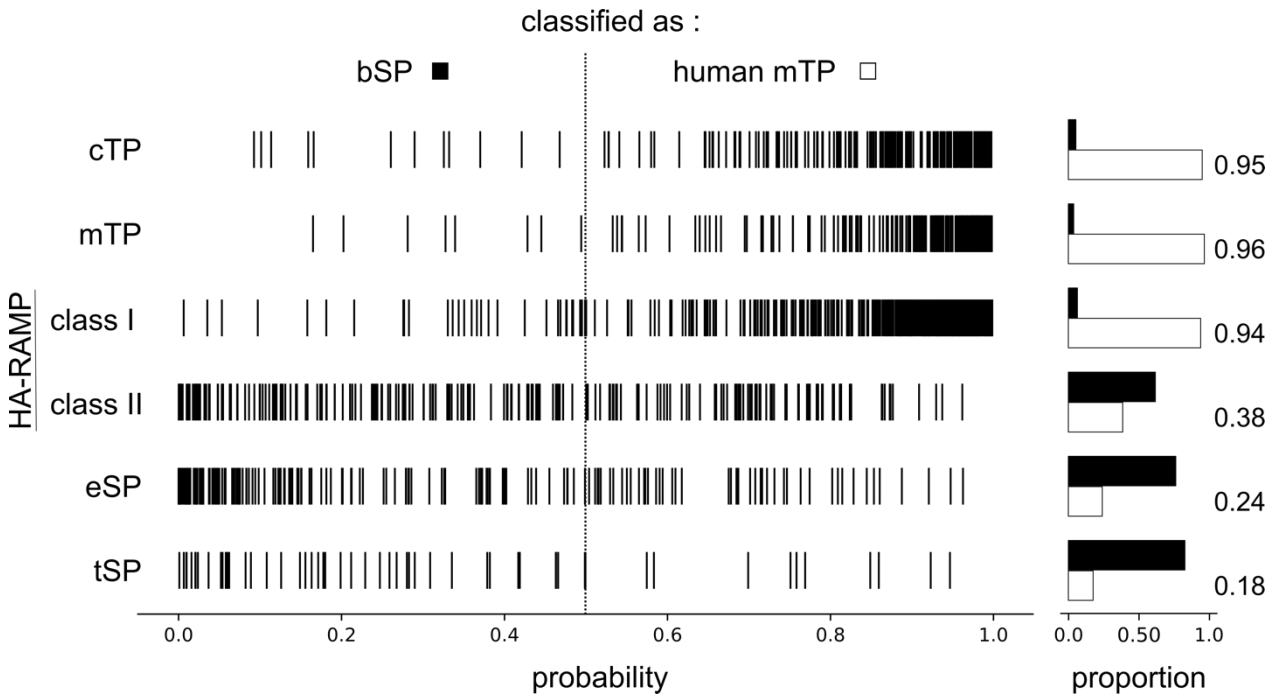


Figure 5. HA-RAMPs are recognized as TPs. Logistic regression classification trained on ACC terms from human mTPs and bSPs for peptide groups listed on the left. Left panel: probability distributions of being classified as human mTP-like ($P > 0.5$) or as bSP-like ($P < 0.5$). Right panel: proportions of peptides classified as human mTP-like (white rectangles) or as bSP-like (black rectangles) in each peptide group. See Supplementary Figure 4A for the coefficient attributed to each ACC term in the model.

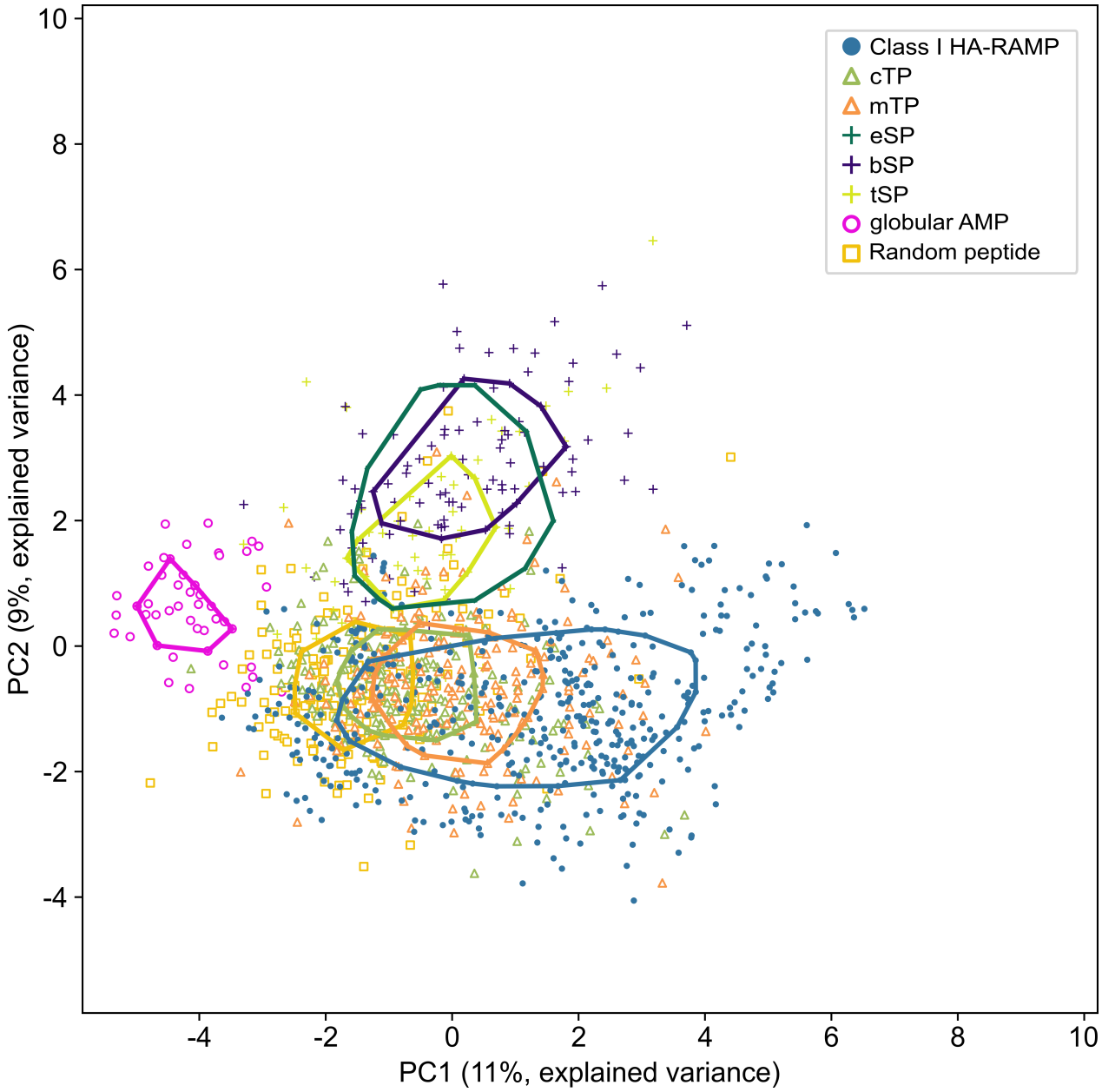


Figure 6. TPs and class I HA-RAMPs share similar physico-chemical properties. PCA is performed on peptides described by 36 ACC terms. Peptide positions are plotted along the first (PC1) and second (PC2) principal component on the x and y axes respectively, with explained variance in parenthesis. Solid lines represent the convex areas containing the 50% most central peptides of each group. HA-RAMPs are separated in class I (blue), class II (burgundy) as described in Figure 2; TPs (mTPs, orange; cTPs, green) and SPs (eSPs, dark green; bSPs, indigo; tSPs, light green). See Supplementary Figure 4B for the contribution of ACC terms PC1 and PC2 and Supplementary Figure 5 for PCA with class II HA-RAMPs.

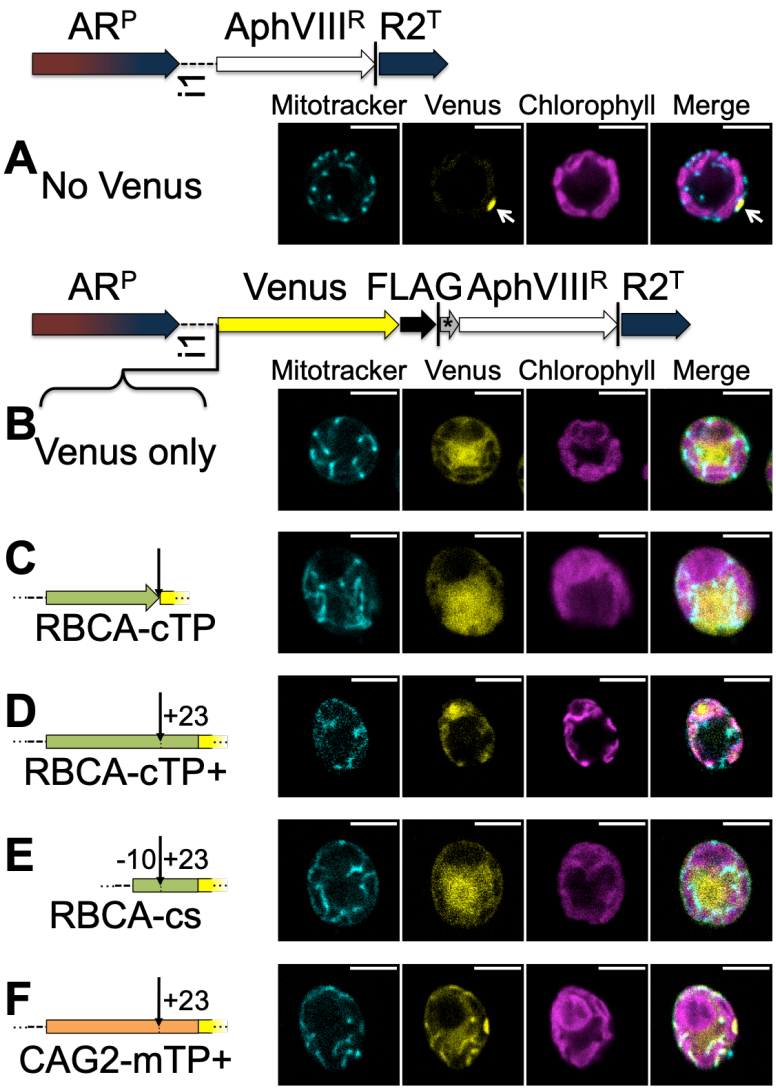


Figure 7. A RBCA cleavage site including downstream residues is necessary but not sufficient for targeting. False-color confocal images of representative Chlamydomonas cells show mitochondria as indicated by mitotracker fluorescence in cyan, the localisation of Venus in yellow and chlorophyll autofluorescence in magenta. Scale bars are 5μm. Expression constructs use the chimaeric HSP70-RBCS2 promotor (AR^P), RBCS2 intron 1 (i1) in the 5' UTR, the paromomycin resistance gene ($AphVIII^R$) as selectable marker, and the RBCS2 terminator ($R2^P$). Vertical lines represent STOP codons. (A) shows expression of the selectable marker only (the white arrow marks the eyespot). A Venus fluorescent reporter carrying a FLAG-tag was introduced upstream of $AphVIII^R$, and bicistronic expression ensured by connecting the two genes with a STOP-TAGCAT sequence (*). Candidate peptides to be assayed for targeting were then introduced upstream of Venus. Venus fluorescence is shown in the absence of a presequence (B), when fused to Rubisco activase (RBCA) cTP up to the cleavage site (C), including 23 residues downstream of the cleavage site (D), the RBCA cleavage site (RBCA-cs) fragment alone encompassing residues -10 to +23 (E), or fused to γ-carbonic anhydrase 2 (CAG2) mTP (F). In each case, the site of cleavage is indicated by a downward arrow. See Supplementary Figure 6 for a quantification of colocalisation, Supplementary Figure 7 for replicates, and Table S6 for a description of peptide sequences.

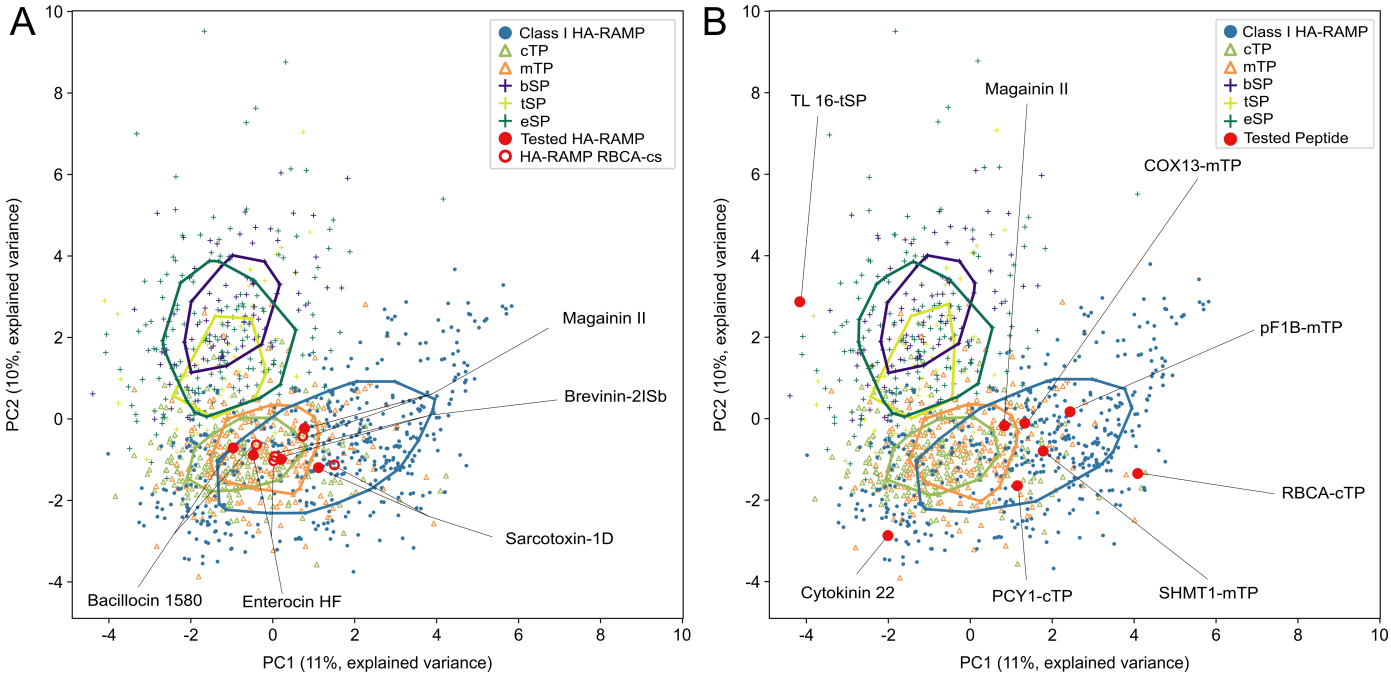


Figure 8. Position of class I HA-RAMPs and TPs used for experimental analysis. Same principle as in Figure 6, considering TP, class I HA-RAMPs and SP. (A) HA-RAMPs used for targeting assays are indicated (red: original, red circle: construct including the Rubisco activase cTP cleavage site, RBCA-cs). (B) TPs used for antimicrobial assays are indicated in red, as well as the hormone Cytokinin 22 and the thylakoid signal peptide TL 16-tSP. See Supplementary Figure 4C for the contribution of ACC terms to the principal components PC1 and PC2.

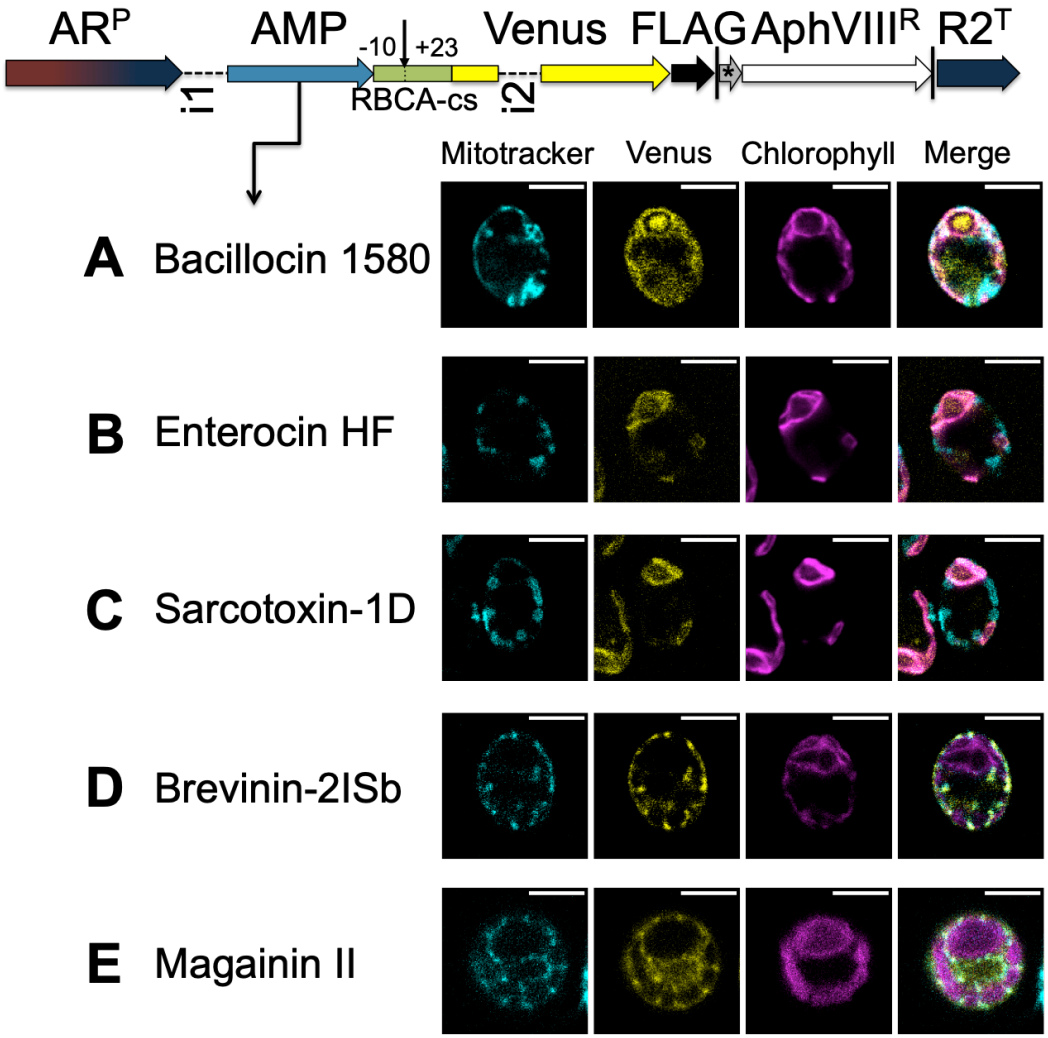


Figure 9. AMPs function as TPs. Constructs assay the targeting ability of candidate peptides fused to the Venus-FLAG reporter, driven by the chimaeric HSP70-RBCS promoter (AR^P) and RBCS2 terminator ($R2^T$), and expressed bicistronically via the STOP-TAGCAT (*) sequence with the paromomycin resistance marker ($AphVIII^R$). Vertical lines indicate STOP-codons. Expression levels in Chlamydomonas are increased by the use of introns: RBCS2 intron 1 (i1) in the 5' UTR and RBCS2 intron 2 (i2) within the Venus coding sequence. Candidate HA-RAMPs Bacillocin 1580 (A), Enterocin HF (B), Sarcotoxin-1D (C), Brevinin-2ISb (D) and Magainin II (E) were fused to the RBCA cleavage site fragemnet encompassing residues -10 to +23 (RBCA-cs) and inserted upstream of Venus. The site of cleavage is indicated by a downward arrow. False-color confocal images of representative cells show mitochondria as indicated by mitotracker fluorescence in cyan, the localisation of Venus in yellow and chlorophyll autofluorescence in magenta. Scale bars are 5μm. See Supplementary Figure 6 for a quantification of colocalisation, Supplementary Figure 7 for replicates, and Table S6 for a description of peptide sequences.

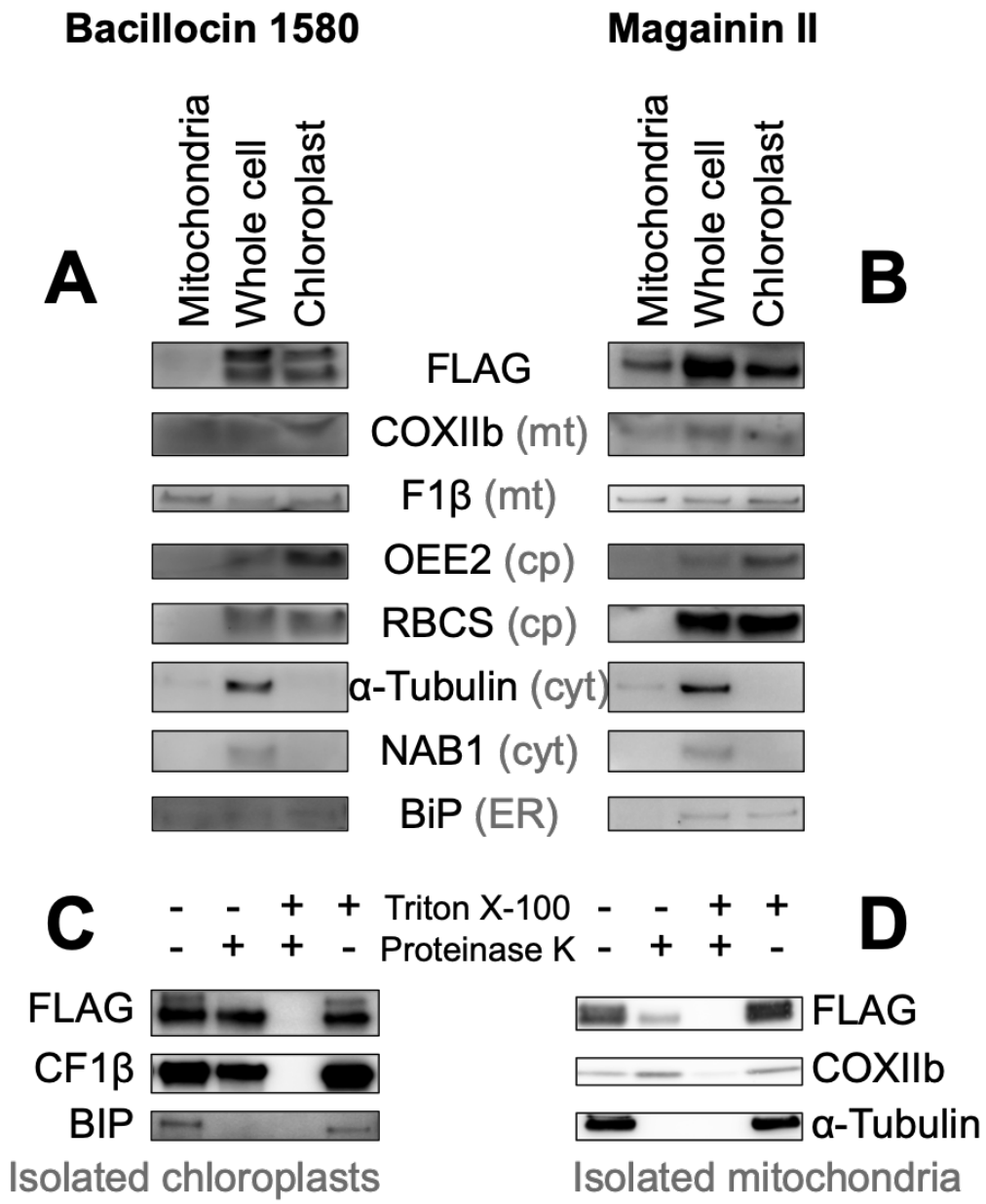


Figure 10. AMP targeting activity confirmed biochemically. Mitochondrial, whole cell and chloroplast fractions (1μg protein per well) isolated from Chlamydomonas strains where Venus localization is driven by Bacillocin 1580 (A) or Magainin II (B), each fused to the RBCA cleavage site, were immunolabelled with antibodies raised against markers for different cellular compartments: Cytochrome Oxidase subunit IIb (COXIIb) and ATPsynthase subunit F1β for mitochondria (mt), Photosystem II Oxygen evolving enhancer 2 (OEE2) and Rubisco small subunit (RBCS) for chloroplasts (cp), α-Tubulin and nucleic-acid binding protein 1 (NAB1) for the cytoplasm (cyt) and lumenal binding protein (BiP) for the endoplasmic reticulum (ER). Isolated chloroplasts from the Bacillocin 1580 strain (C) and isolated mitochondria from the Magainin II strain (D) were subjected to a proteinase assay, where aliquots were variously treated with 150μg/ml proteinase K and/or 1% Triton X-100, a membrane solubilizing detergent. Aliquots were subsequently immunolabelled with antibodies against chloroplast ATPsynthase subunit CF1β and other organelle-markers as described above.

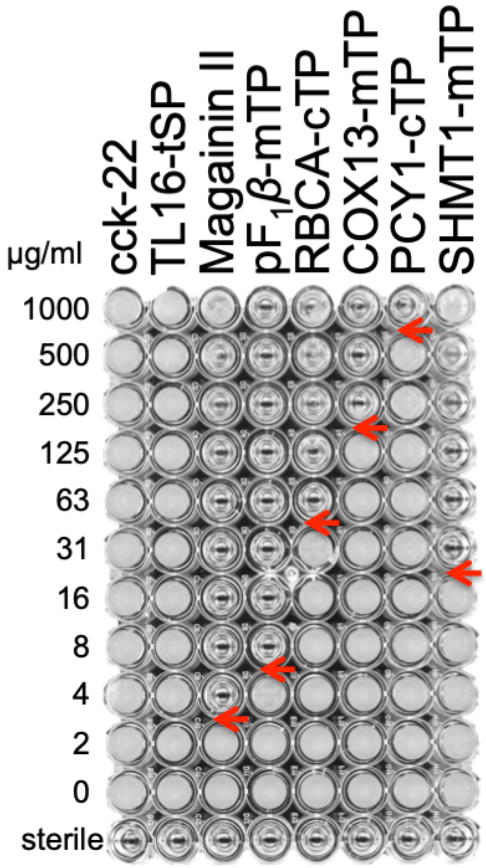
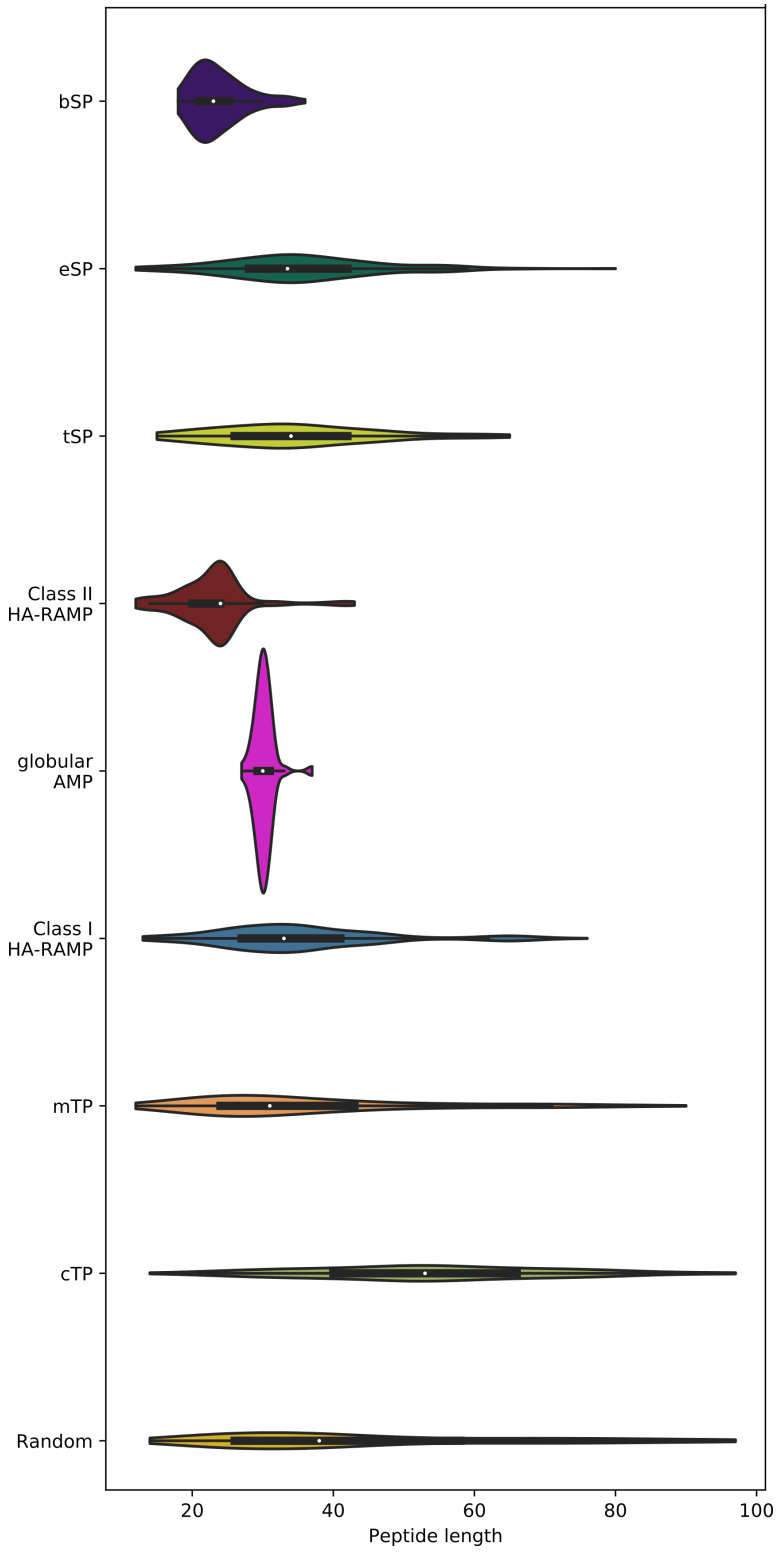
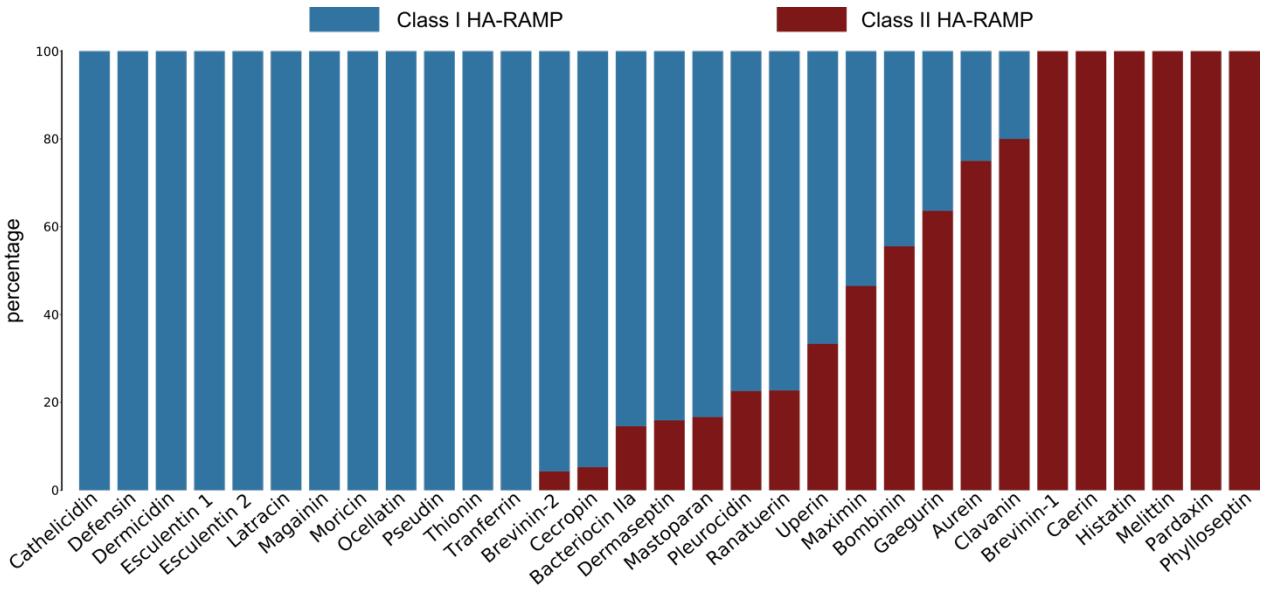


Figure 11. TPs exhibit antimicrobial activity. *B. subtilis* was challenged with serial dilutions of synthetic peptides (See Table S7 for sequences): the Rattus norvegicus peptide hormone cytokinin 22, the tSP of lumenal 16.5 kDa protein, the HA-RAMP Magainin II, yeast mTP of ATP synthase F1β subunit and Chlamydomonas cTPs from RBCA and plastocyanin (PCY1) or mTPs from cytochrome c oxidase 12 kDa subunit (COX13) and Serine HydroxylMethyl Transferase 1 (SHMT1). Transparent wells illustrate absence of growth. Red arrows point to the minimal peptide inhibiting concentration.

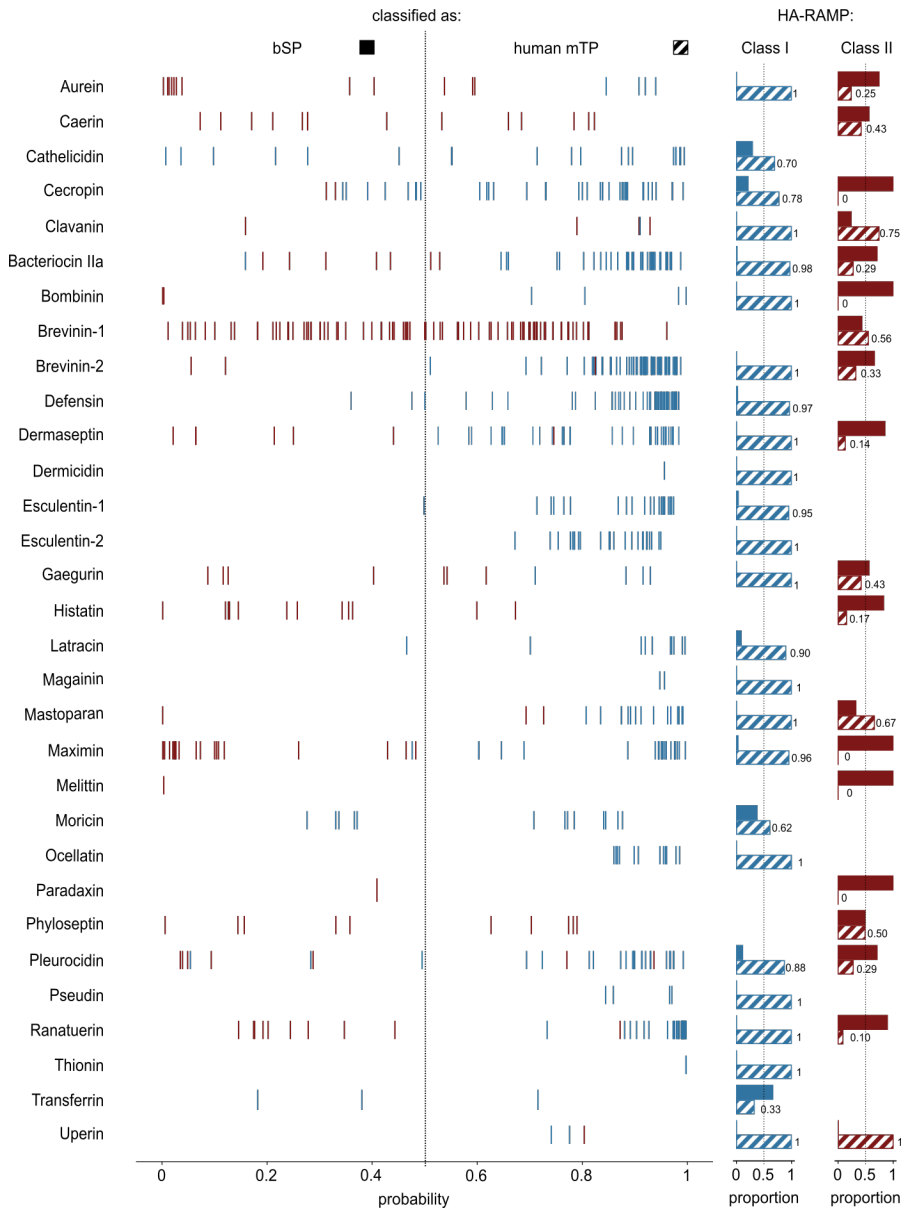
Supplementary Figures



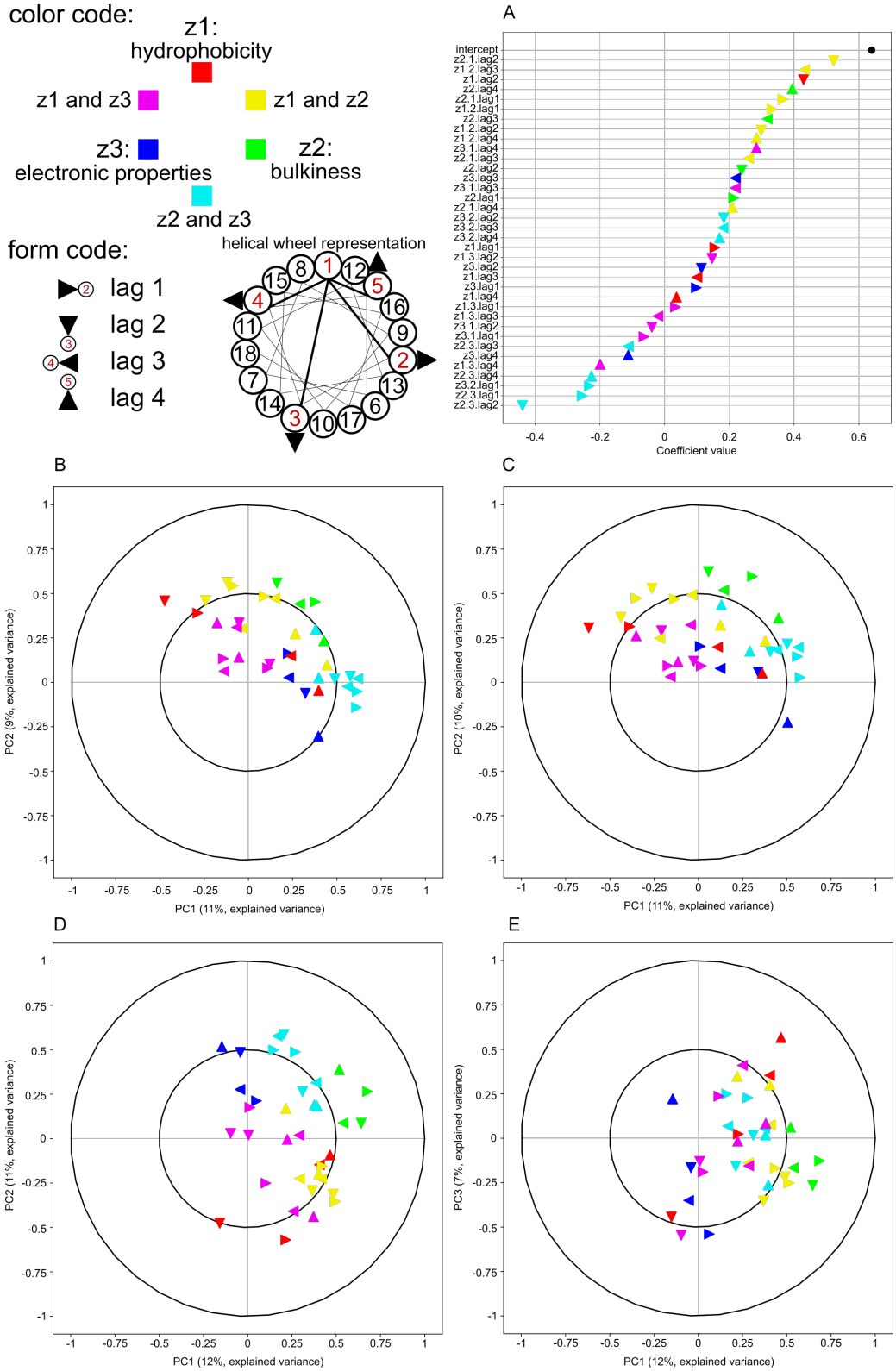
Supplementary Figure 1. Distribution of the peptides length (aa) in each group.



Supplementary Figure 2. **Percentage of peptides of HA-RAMP families described in the literature among classes I and II.** Classes are as defined by k-means clustering (Figure 2).

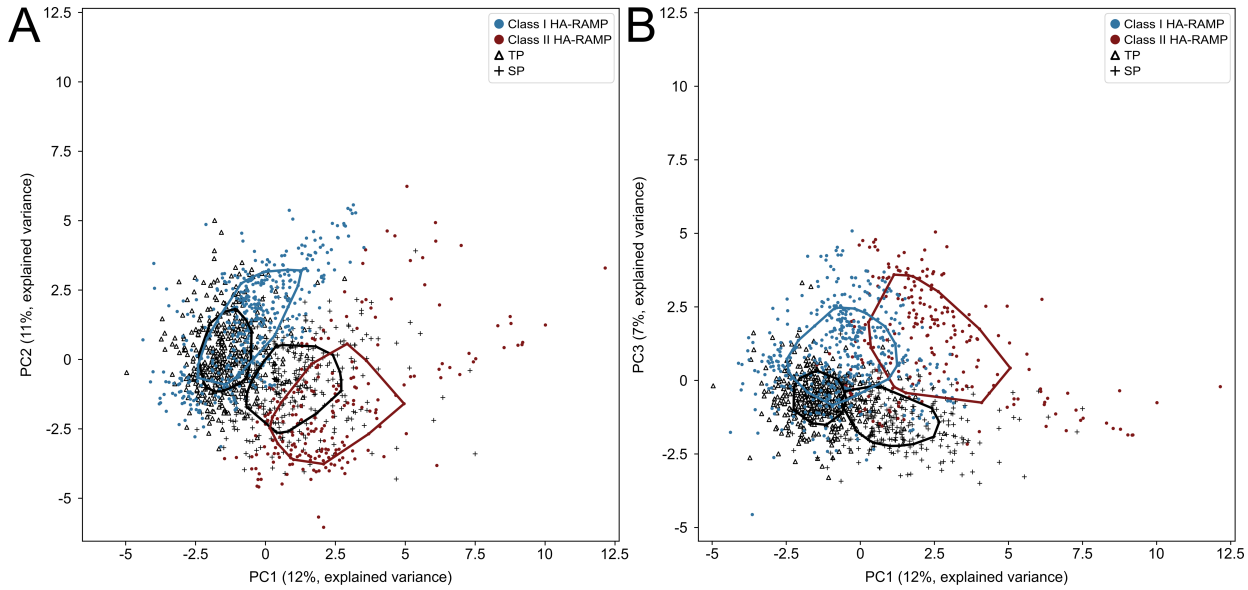


Supplementary Figure 3. Classification of HA-RAMPs grouped by families. Individual peptides are recognizable as belonging to a different group through the colour code: class I HA-RAMPs, blue; class II HA-RAMPs, red. Left panel: probability distributions of being classified as human mTP-like ($P > 0.5$) or as bSP-like ($P < 0.5$). Middle panel: proportions of class I peptides classified as human mTP-like (striped blue rectangles) or as bSP-like (blue rectangles). Right panel: proportions of class II peptides classified as human mTP-like (striped red rectangles) or as bSP-like (red rectangles).

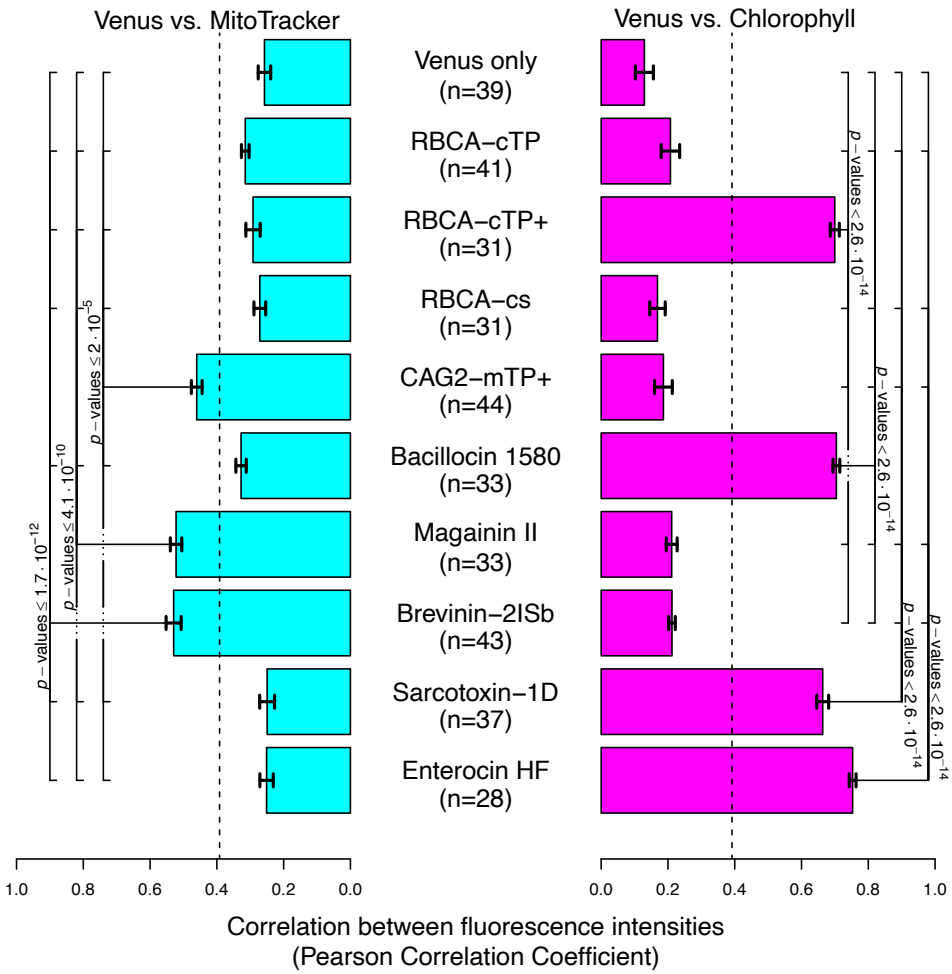


Supplementary Figure 4. Separation of peptides in the logistic regression and PCA analyses reflects hydrophobic and amphipathic features. Each triangle represents an ACC term, where the color corresponds to the z-scales involved as described in the legend, and the orientation

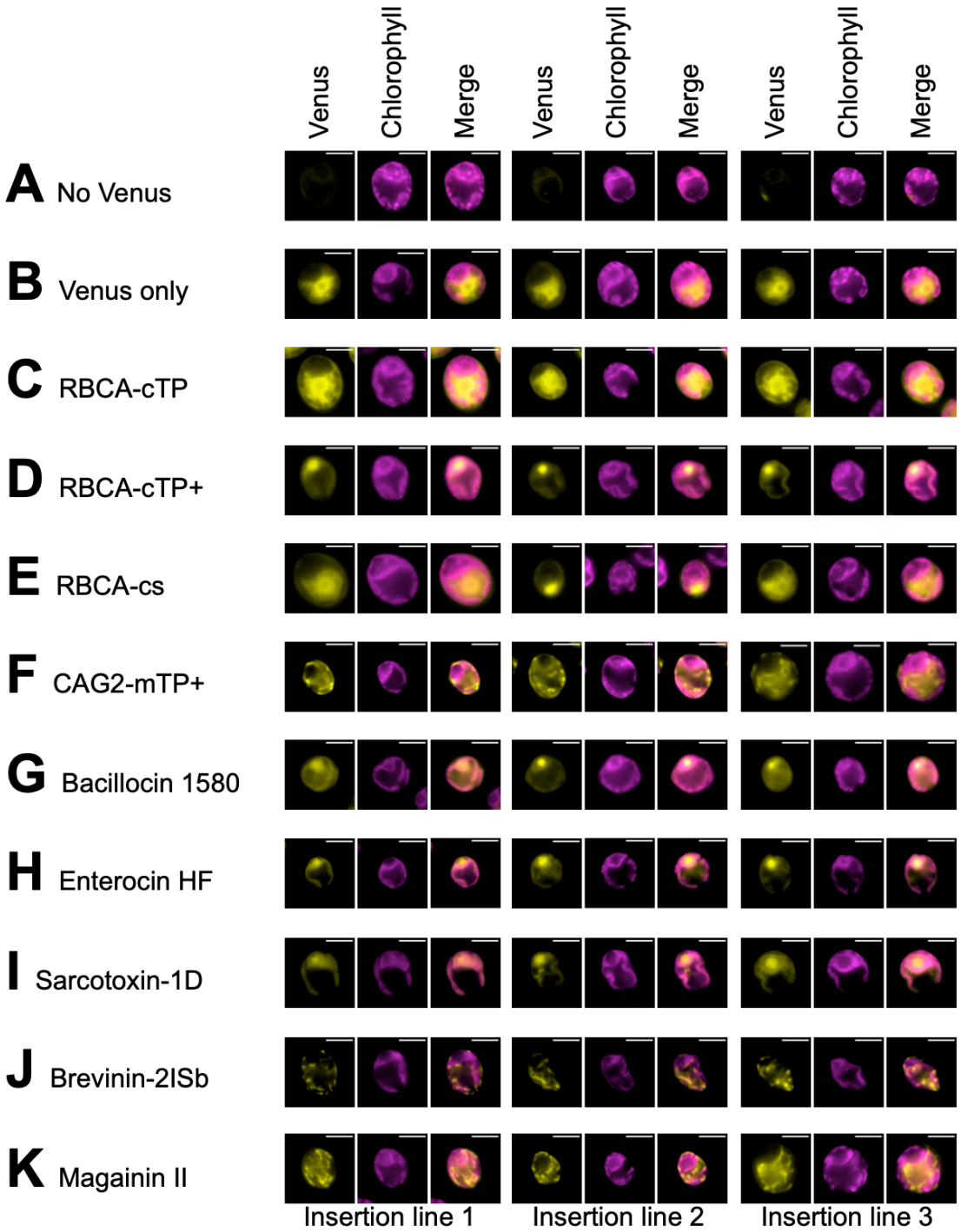
corresponds to the lag as indicated on the helical wheel representation: a lag of 1 thus describes the relation between residue 1 and residue 2; a lag of 2 between residue 1 and residue 3 etc. The maximum considered is lag 4, thus only residues 1-5 are concerned. The red numbers in the circles next to the triangles indicate the residues considered relative to the first one, which are also highlighted in red on the helical wheel representation. (A) Ranking and values of the ACC term coefficients in the model of the logistic regression classifier (Figure 5). Note that intercept is the constant of the linear model. (B) Correlation circles for the PCA presented in Figure 6. (C) Correlation circles for the PCA presented in Figure 8. (D) Correlation circles for the PCA presented in Supplementary Figure 5A. (E) Correlation circles for the PCA presented in Supplementary Figure 5B. For (B-E) x-axis and y-axis are principal components used for representation. Unit circle and half unit circles are represented.



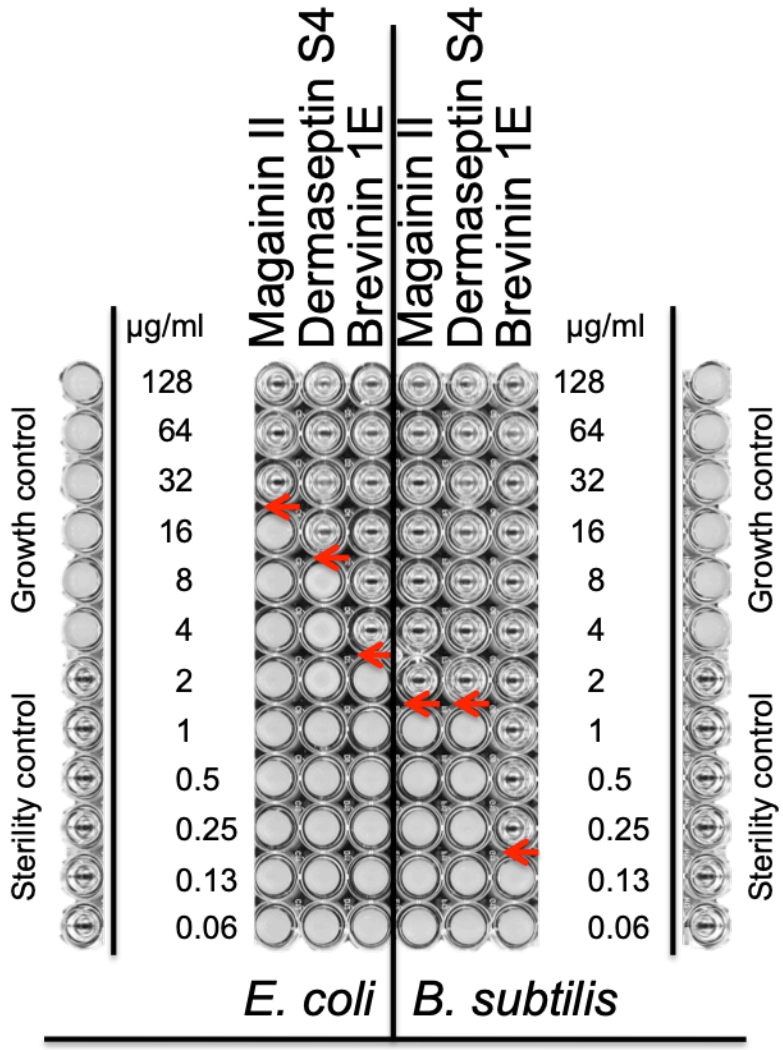
Supplementary Figure 5. The third principal component of the PCA analysis confirms proximity between class I HA-RAMPs and TPs, while improving the separation between class II HA-RAMPs and SPs. (A) PCA on normalized ACC terms for class I HA-RAMP (blue circle), class II HA-RAMP (dark red circle), targeting peptides (black triangle) and signal peptides (black cross). peptide positions are plotted along the first (X) and second (Y) principal components. (B) Same PCA as in (A) with peptides positions plotted along the first (X) and third (Y) principal components. Explained variance are indicated in parenthesis for each axis. Solid lines represent the convex areas of the 50% most central peptides in each group. ACC term contribution in the principal components are given in Supplementary Figure 4D,E.



Supplementary Figure 6. A quantitative assessment of colocalisation backs up targeting interpretations. Pearson correlation coefficients (PPC) were calculated for each strain between the Venus channel and either the MitoTracker or the chlorophyll channel of n cells, and plotted as mean ± standard error. Perfect colocalisation would give a value of 1, whereas 0 would indicate completely uncorrelated localisation. The dashed line indicates the upper limit of a 95% confidence interval for PCCs between MitoTracker and chlorophyll channels across all strains, which serves as a control for random colocalisation that arises simply due to proximity of compartments within a cell. Strains CAG2-mTP+, Brevinin-2Isb and Magainin II show signi1cantly higher signal correlation of Venus with MitoTracker than the other strains, supporting the localisation of Venus in the mitochondria. Similarly, strains RBCA-cTP+, Enterocin HF, Bacillocin 1580 and Sarcotoxin-1D show significantly higher correlation of Venus with chlorophyll fluorescence than the other strains, supporting the localisation of Venus in the chloroplast. See Figure 7 and Figure 9 for description of constructs and exemplary images. The "no Venus" strain was excluded from this analysis, as in the absence of Venus, the Venus channel intensities within the cell are on the order of background values substracted from other strains strains (except for the eyespot). Significance was tested via a one-way ANOVA followed by a Tukey post-hoc analysis. All p-values <0.05 from the Tukey analysis are indicated.



Supplementary Figure 7. Three biological replicates all show the same phenotype for each construct. Epifluorescence false-colour images show representative cells from three independent insertion lines for each construct. See Figure 7 for description of constructs (A) to (F), and Figure 9 for description of constructs (G) to (K).



Supplementary Figure 8. *B. subtilis* is a more sensitive probe for antimicrobial activity of peptides than *E. coli.* Gram negative *E. coli* and gram positive *B. subtilis* were challenged with serial dilutions of synthetic HA-RAMPs Magainin II, Dermaseptin S4 and Brevinin 1E (See Table S7 for sequences). Transparent wells illustrate absence of growth. Red arrows point to the minimal peptide inhibiting concentration. All wells shown here are part of the same 96-well plate; control rows are shown separately on either side for improved clarity.