More than just a ticket canceller: The mitochondrial processing peptidase matures complex precursor proteins at internal cleavage sites

Jana Friedl[1], Michael R. Knopp[2], Carina Groh[1], Eyal Paz[3], Sven B. Gould[2], Felix Boos[1#], Johannes M. Herrmann[1]

[1]Cell Biology, Technische Universität Kaiserslautern, 67663 Kaiserslautern, Germany

[2]Molecular Evolution, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany

[3]Departments of Biochemistry and Molecular Biology, Tel Aviv University, Tel Aviv 6997801, Israel

[#], corresponding author, fboos@rhrk.uni-kl.de

Abstract

Most mitochondrial proteins are synthesized in the cytosol as precursors that carry N-terminal presequences. After import into mitochondria, these targeting signals are cleaved off by the mitochondrial processing peptidase MPP, giving rise to shorter mature proteins. Using the mitochondrial tandem protein Arg5,6 as a model substrate, we demonstrate that MPP has an additional role in preprotein maturation, beyond the removal of presequences. Arg5,6 is synthesized as a polyprotein precursor that is imported into the mitochondrial matrix and subsequently separated into two distinct enzymes that function in arginine biogenesis. This internal processing is performed by MPP, which cleaves the Arg5,6 precursor both at its N-terminus and at an internal site between the Arg5 and Arg6 parts. The peculiar organization and biogenesis of Arg5,6 is conserved across fungi and might preserve the mode of co-translational subunit association of the arginine biosynthesis complex of the polycistronic arginine operon in prokaryotic mitochondrial ancestors. Putative MPP cleavage sites are also present at the junctions in other mitochondrial fusion proteins from fungi, plants and animals. Our data suggest that, in addition to its role as “ticket canceller” for the removal of presequences, MPP exhibits a second, widely conserved activity as internal processing peptidase for complex mitochondrial precursor proteins.

Introduction

All cellular processes are carried out by proteins, linear chains of amino acids that fold into three-dimensional structures. While the amino acid sequence of a protein is primarily determined by its DNA sequence, many proteins are additionally modified by proteolytic cleavage after their synthesis. Processing of polypeptides at their N-terminus is pervasive in both prokaryotic and eukaryotic proteomes. For instance, the amino-terminal methionine is removed from many polypeptides when they emerge from the ribosome and the new N-terminus is a crucial determinant of protein stability (Bradshaw et al., 1998, Varshavsky, 2011, Frottin et al., 2006). The majority of intracellular protein targeting signals are located at the N-terminus and cleaved off upon arrival at the correct cellular destination. For example, around two thirds of the nuclear-encoded mitochondrial proteins are synthesized as precursors that carry an N-terminal mitochondrial targeting sequence (MTS) or presequence which directs them to mitochondrial surface receptors (Bykov et al., 2020, Becker et al., 2019, von Heijne, 1986). These preproteins are imported into mitochondria via the translocases of the outer (TOM complex) and inner mitochondrial membrane (TIM23 complex) (Pfanner et al., 2019, Chacinska et al., 2009). Their MTS is cleaved by the mitochondrial processing peptidase (MPP). In some cases, the thereby generated new N-terminus of the polypeptide is further shortened by cleavage of single amino acids or short peptides by the proteases Icp55 or Oct1 before the matured protein folds into its native structure (Poveda-Huertes et al., 2017, Vögtle et al., 2011, Vögtle et al., 2009, Calvo et al., 2017, Naamati et al., 2009). Correct processing of precursor proteins in the matrix is crucial to maintain mitochondrial function and proteostasis. Dysfunctional preprotein maturation in mitochondria was observed in models of Alzheimer's disease (Mossmann et al., 2014). It results in proteome instability, aggregation of incorrectly processed precursors and proteotoxic stress (Poveda-Huertes et al., 2020).

Internal cleavage of polypeptides is less frequent than processing from the N-terminus, but equally relevant for cellular physiology and organismal health. Examples of medically relevant polypeptides that undergo internal proteolytic processing during their biosynthesis include insulin or amyloid

precursor protein (APP) (Steiner and Oyer, 1967, Müller et al., 2017). In most cases, peptides are removed to yield the mature form of a single protein. However, some genes encode fusion proteins that are synthesized as a single precursor and are then separated into distinct, functional proteins by proteolytic cleavage. A prominent example is ubiquitin, which is encoded as a fusion with subunits of the ribosome or as head-to-tail repeats of several ubiquitin monomers which are rapidly separated by deubiquitinating proteases (Finley et al., 1989, Ozkaynak et al., 1984, Gemayel et al., 2017). Polyproteins also frequently occur in viral genomes, including HIV and SARS-CoV-2, where the cleavage products often form protein complexes (Yost and Marcotrigiano, 2013, Zhang et al., 2020, Krichel et al., 2020). In eukaryotic genomes, such an organization is rare, despite its obvious advantage of stoichiometric co-expression of functionally related proteins.

Here we report about a notable exception: The *ARG5,6* gene of *Saccharomyces cerevisiae* encodes both the acetylglutamate kinase (Arg6) and the acetylglutamyl-phosphate reductase (Arg5), two enzymes that catalyze the second and third step of arginine biosynthesis in the mitochondrial matrix (Minet et al., 1979). These two enzymes are synthesized as a single precursor protein that is post-translationally cleaved into separate polypeptides. Arg6 and Arg5 then form a complex together with the acetylglutamate synthase Arg2 (Pauwels et al., 2003, Abadjieva et al., 2001). Here, we investigate the biogenesis of Arg5,6 in more detail and identify MPP as the protease that is responsible for processing of the precursor into two functional enzymes. We demonstrate that Arg6 and Arg5 can be imported into mitochondria separately, where they still form a functional enzyme. However, its organization as a composite precursor that is matured by MPP is highly conserved across fungi. These findings broaden our view on MPP as processing peptidase of more general role in mitochondrial preprotein maturation, reaching beyond its canonical function of removing mitochondrial targeting signals.

Results

Arg5,6 is processed by MPP in the mitochondrial matrix

The acetylglutamate kinase (Arg6) and acetylglutamyl-phosphate reductase (Arg5) are located in the mitochondrial matrix, where they catalyze the second and third step of the biosynthesis of arginine from glutamate (Fig. 1A) (Minet et al., 1979, Morgenstern et al., 2017). The composite precursor protein that is synthesized from the *ARG5,6* gene contains a mitochondrial targeting sequence (MTS) (Vögtle et al., 2009), which was suggested to direct the single precursor into mitochondria where it is subsequently cleaved into two separate polypeptides (Boonchird et al., 1991b). Indeed, when we expressed Arg5,6 with a C-terminal hemagglutinin (HA) tag, immunoblotting of cell lysates against the HA epitope revealed only a single band at around 40 kDa, much less than the expected mass of 90 kDa of the composite precursor protein (Fig. 1B). Obviously, the precursor is rapidly and efficiently cleaved *in vivo*, which yields a C-terminal Arg5 fragment.

In order to analyze the mechanistic basis of this unusual biogenesis, we sought to reconstitute the biogenesis and processing of Arg5,6 in an *in vitro* system. To this end, we synthesized radiolabeled Arg5,6 precursor in reticulocyte lysate and incubated it with isolated yeast mitochondria. We observed that the precursor of around 90 kDa was efficiently processed to a slightly smaller intermediate form – indicating the removal of the N-terminal MTS – and further into several smaller fragments, most prominently two polypeptides of around 40 and 50 kDa, which correspond to the C-terminal Arg5 and the N-terminal Arg6, respectively (Fig. 1C). The intermediate as well as the mature polypeptides, but not the Arg5,6 precursor were protected from digestion with externally added proteinase K. This shows that these polypeptides were translocated across the mitochondrial outer membrane. Both import and processing were dependent on the mitochondrial inner membrane potential ($\Delta\psi$) as expected for an MTS-containing matrix protein (Sato et al., 2019, Schendzielorz et al., 2017, Garg and Gould, 2016). We conclude that Arg5,6 is imported into the mitochondrial matrix via the presequence pathway and cleaved into separate polypeptides inside mitochondria.

How is the Arg5,6 precursor processed to give rise to the Arg6 and Arg5 enzymes? A number of proteases in the mitochondrial matrix are described (Quiros et al., 2015, Veling et al., 2017). However, most of them are either implicated in degradation and turnover of proteins (such as the Lon protease Pim1) or known to remove short peptides or single amino acids from the N-terminus of mitochondrial precursor proteins (such as Oct1 or Icp55), but not for internal cleavage of proteins into two mature parts (Vögtle et al., 2009, Poveda-Huertes et al., 2017, Vögtle et al., 2011, Wagner et al., 1994, Suzuki et al., 1994). For MPP, the peptidase responsible for the cleavage of the N-terminal MTS, a notable exception was recently reported: The composite precursor protein Atp25 contains internal MPP cleavage sites at which the protein is split into two functionally unrelated polypeptides (Woellhaf et al., 2016). In Atp25, the internal MPP cleavage sites coincide with a sequence stretch that structurally mimics the properties of an N-terminal MTS. We asked whether MPP might also be involved in the processing of Arg5,6 and subjected its amino acid sequence to an *in silico* prediction of such internal MTS-like sequences (iMTS-Ls) using an adapted version of the TargetP algorithm (Boos et al., 2018, Emanuelsson et al., 2007). In fact, in addition to the N-terminal MTS, two internal regions with high TargetP score were detected around amino acid positions 344 and 503 (Fig. 1D). The latter one would fit to the molecular masses of the Arg6 and Arg5 proteins observed in our *in vitro* system. To directly test whether MPP can cleave the Arg5,6 precursor, we purified MPP from *E. coli* expressing His-tagged Mas1 and Mas2 (the two subunits of MPP). Incubation of radiolabeled Arg5,6 precursor protein with MPP resulted in the formation of smaller fragments whose size perfectly matched those that were generated after import into isolated mitochondria (Fig. 1E). Proper processing was blocked when EDTA was added to the reaction, which inhibits the metalloprotease MPP by chelating divalent cations (Suppl. Fig. 1A) (Luciano et al., 1998).

We conclude that Arg5,6 is imported into the mitochondrial matrix and processed twice by MPP. A first cleavage removes the N-terminal MTS and a second cleavage at an iMTS-L separates the Arg6 and Arg5 enzymes (Fig. 1F).

Arg5 and Arg6 can be imported into mitochondria separately and complement the *arg5,6* deletion mutant

The unusual biogenesis of Arg5,6 prompted us to ask whether Arg6 and Arg5 can also be imported separately. Therefore, we created truncated versions of the *ARG5,6* gene which contain only the N-terminal Arg6 with its MTS ($Arg6^{1-502}$) or only the C-terminal Arg5, starting at the first ($Arg5^{344-863}$) or the second iMTS-L ($Arg5^{503-863}$). For the latter two, we also generated variants which additionally carry the well-characterized presequence of ATP synthase subunit 9 from *Neurospora crassa* (Su9-$Arg5^{344-863}$ and Su9-$Arg5^{503-863}$) (Fig. 2A).

Radiolabeled proteins were synthesized *in vitro* and incubated with isolated mitochondria to test their import competence. As expected, $Arg6^{1-502}$ was efficiently imported and its MTS was cleaved (Fig. 2B). The shorter Arg5 variant ($Arg5^{503-863}$) did not reach a protease-protected compartment and, thus, was not imported into mitochondria (Fig. 2C). However, N-terminal fusion of the Su9 presequence completely restored import of $Arg5^{503-863}$ (Fig. 2D). Hence, import of Arg5 and Arg6 into mitochondria is in principle possible also for separated polypeptides, at least in the *in vitro* assay used here.

We next tested whether Arg6 and Arg5 can be imported separately *in vivo* and function in arginine biosynthesis. The deletion of the *ARG5,6* gene renders yeast cells auxotrophic for arginine. We expressed either the full length Arg5,6 precursor or combinations of separate Arg6 and Arg5 variants in a Δ*arg5,6* deletion mutant. If Arg6 and Arg5 make their way into mitochondria and acquire a functional conformation, arginine prototrophy should be restored. When we streaked out these cells on plates with minimal growth medium lacking arginine, we observed growth for the wildtype and the Δ*arg5,6* mutant complemented with full length Arg5,6, but not for Δ*arg5,6* carrying only an empty plasmid, as expected (Fig. 2E).

The mutant expressing both $Arg6^{1-502}$ and the shorter $Arg5^{503-863}$ variant was not able to grow without arginine (Fig. 2E). However, when the presequence of Su9 was fused to the short $Arg5^{503-863}$, cells

157 regained arginine prototrophy (Fig. 2F). All strains grew on plates containing arginine, showing that
158 the Arg5$^{503-863}$ protein has no toxic gain-of-function effect when residing in the cytosol (Suppl. Fig.
159 1B,C). Growing the strains in liquid medium lacking arginine confirmed the results obtained on plates
160 and additionally demonstrated that the growth rate of the strain expressing Arg6$^{1-502}$ and Su9-
161 Arg5$^{503-863}$ is comparable to that of the wildtype (Suppl. Fig. 1D), at least under the overexpression
162 conditions used here. This indicates that separate expression of Arg6 and Arg5 is, in principle,
163 possible without adverse effects on cellular fitness.

164 A truncated version of Arg6 (Arg6$^{1-343}$) did not complement the deletion mutant with any of the Arg5
165 variants, even though this variant contains the entire region of homology to the bacterial
166 acetylglutamate kinase argB from *E. coli* (Fig. 2F). Apparently, the amino acids 344 to 502 are
167 functionally relevant for enzyme activity and not merely a spacer or linker between Arg5 and Arg6.

168 Taken together, both Arg6 and Arg5 can be imported separately into mitochondria *in vitro* and *in vivo*
169 and are functional in arginine biosynthesis, as long as mitochondrial localization is conferred by
170 appropriate N-terminal targeting signals.

171

172 **Internal processing of Arg5,6 by MPP requires an N-terminal presequence**

173 Surprisingly, cells expressing Arg6$^{1-502}$ and the longer Arg5$^{344-863}$ could grow without arginine,
174 indicating that this longer Arg5 variant can be imported even without fusion of an additional
175 presequence (Fig. 2E). Indeed, Arg5$^{344-863}$ was imported into isolated mitochondria, albeit with low
176 efficiency (Fig. 3A). Addition of proteinase K to this import reaction resulted in the appearance of
177 lower-running bands that were absent without protease, indicating that a portion of the precursor
178 translocated only partially across the outer membrane. This is in agreement with earlier observations
179 that iMTS-Ls can confer targeting to mitochondria when presented at the N-terminus, but are not
180 necessarily able to drive complete translocation (Backes et al., 2018, Baker and Schatz, 1987). As

expected, fusion of the Su9 presequence to Arg5$^{344-863}$ resulted in more efficient translocation (Fig. 3B).

Interestingly, even the fully imported Arg5$^{344-863}$ was not processed inside mitochondria, while for Su9-Arg5$^{344-863}$, prominent lower-running bands were observed. Their sizes fit those expected for the intermediate form after removal of the Su9 presequence (around 60 kDa), the completely matured Arg5 after cleavage at the second MPP site (around 40 kDa) and the cleaved "linker" between the Su9 presequence and Arg5 (around 25 kDa). This suggests that an internal MPP cleavage at an iMTS-L requires a *bona fide* N-terminal MTS.

To test whether this holds true also *in vivo*, we expressed HA-tagged Arg5,6 as well as Arg5 variants in yeast cells and analyzed cell lysates by immunoblotting against the HA epitope. In agreement with the *in vitro* import experiments, both Su9-Arg5$^{344-863}$ and Su9-Arg5$^{503-863}$ were processed and yielded a band at the same molecular weight as full length Arg5,6. In contrast, Arg5$^{344-863}$ and Arg5$^{503-863}$ were exclusively present in their unprocessed forms (Fig. 3C). Hence, even though Arg5$^{344-863}$ is imported into mitochondria, its iMTS-L is not recognized by MPP. Obviously, internal MPP cleavage sites are dependent not only on the presence of a particular motif and its immediate context within the amino acid sequence, but also on more distant features of the polypeptide, such as the N-terminal presequence (Fig. 3D).

The *ARG5,6* genes are fused in fungi, separated in algae, and encoded polycistronically in gamma-proteobacteria

Tandem organization of functionally related genes is rare in eukaryotes. In contrast, genomic co-localization of functionally related genes is pervasive in prokaryotes. The *E. coli* proteins argB and argC are homologous to Arg6 and Arg5, respectively, and are organized in an operon, i.e. they are transcribed polycistronically (Piette et al., 1982). We used the sequences of argB and argC of *E. coli K-12* as query sequences to search a prokaryotic database of 5,655 organisms (Suppl. Table 1A).

206 3,666 species were identified that encode exactly one copy each of argB and argC (Suppl. Table 1B),
207 mainly distributed among Proteobacteria, Firmicutes, Actinobacteria and Cyanobacteria. Among the
208 3,666 prokaryotes, 882 strains (848 of which were gamma-proteobacteria) encode the two open
209 reading frames in immediate proximity, and hence presumably express them on a polycistronic
210 mRNA, (Suppl. Table 1C).

211 Even though we cannot directly trace back the evolutionary history of Arg5,6 of *Saccharomyces*
212 *cerevisiae*, these findings support the idea that this eukaryotic two-gene cluster evolved from syntenic
213 genes in the eukaryotic ancestor. Nevertheless, Arg5,6 represents an exceptional case. Only very few
214 proteins with a similar fusion structure are known in baker's yeast, and for none of them do the
215 separate proteins function in the same pathway (Woellhaf et al., 2016, Finley et al., 1989, Ozkaynak
216 et al., 1987). However, the Arg5,6 homolog *arg-6* in *Neurospora crassa* also encodes both
217 acetylglutamate kinase and acetylglutamyl-phosphate reductase which are separated by proteolytic
218 cleavage inside mitochondria (Parra-Gessert et al., 1998, Gessert et al., 1994). We therefore asked
219 whether the composite structure of Arg5,6 is generally conserved across eukaryote species. To this
220 end, we searched a dataset of 150 eukaryotes for homologs of the yeast composite Arg5,6 precursor
221 and of its post-translationally processed proteins Arg5 and Arg6 (Suppl. Table 1D). While 82
222 genomes returned no hits, we identified specific patterns within the diamond BLASTp hits of 36
223 eukaryotes (Fig. 4A, Suppl. Fig. 2, Suppl. Table 1E). In 28 species, all three query sequences hit the
224 same subject, indicating the presence of an *ARG5,6* gene fusion homolog. With the sole exception of
225 *Phytophthora sojae*, the fusion homologs were exclusively found among fungi. In eight species, Arg5
226 and Arg6 each hit a single independent sequence, while Arg5,6 hit both of those sequences, which
227 indicates that separate genes encode for Arg5 and Arg6. Again with a single exception (*Fusarium*
228 *graminearum* PH-1), all of these organisms were algae, comprising both green and red algae. In 32
229 species we identified mixed blast patterns (Suppl. Table 1F) that were not further investigated. We
230 also predicted the intracellular localization of all proteins via TargetP. Strikingly, mitochondrial

localization was assigned to almost all fusion proteins of fungi, whereas most separate proteins in algae were predicted to be imported into chloroplasts (Fig. 4B).

In summary, in fungi the acetylglutamate kinase and acetylglutamyl-phosphate reductase are generally encoded as a fusion protein, which is imported into mitochondria and processed twice by MPP to remove its presequence and gives rise to two functional enzymes. In contrast, algae encode two separate proteins which are individually imported into chloroplasts. Gamma-proteobacteria express the genes from one polycistronic RNA (Fig. 4C).

Internal precursor processing by MPP is conserved among eukaryotes

Since mitochondrial localization correlates well with the tandem structure of Arg5,6 homologs, we wondered whether the intramitochondrial processing into two separate enzymes by MPP might also be conserved in these species. To address this, we calculated the iMTS-L profiles of all fusion homologs of Arg5,6. In fact, we observed a common pattern across nearly all species: Besides the N-terminal MTS, our algorithm detected one conserved iMTS-L around amino acid position 500 (Fig. 5A). Hence, all these fusion proteins harbor a potential MPP cleavage site precisely at the junction between the Arg5 and Arg6 parts.

This remarkable conservation of protein structure and processing inspired us to ask whether MPP could be responsible for internal cleavage of other mitochondrial proteins. Apart from Arg5,6 and its homologs, some other mitochondrial proteins with composite structure have been described in different species. These include proteins from different fungi (Atp25 from *S. cerevisiae* and *E. nidulans* and Etp1, Rsm22-Cox11 and the uncharacterized SPAC22A12.08c from *S. pombe*) and plants (RPS14 from *O. sativa subsp. japonica*) (Oshima et al., 2005, Woellhaf et al., 2016, Khalimonchuk et al., 2006). We subjected these mitochondrial fusion proteins to our iMTS-L profiling analysis and indeed found prominent iMTS-Ls at each of their junction sites (Fig. 5B).

Discussion

The organization as fusion protein is an elegant solution to confer mitochondrial targeting of two enzymes that reside in the same compartment and even act in subsequent steps of a biochemical pathway. It is still remarkable that this organization of Arg5,6 was retained during evolution even in distantly related organisms, indicating that there exists a strong constraint that maintained this organization for more than a billion years of evolution. To our knowledge, this is the only example of a fusion of two functionally related proteins whose organization is so widely conserved across eukaryote species. Eukaryotic genomes typically strongly disfavor even "milder" variants of physical coupling of genes, such as an operon-like organization which is pervasively present in prokaryotes, but basically absent in most eukaryotes. Interestingly, also events of horizontal gene transfer from bacteria to eukaryotes were accompanied by progressive loss of the polycistronic organization of the genes (Kominek et al., 2019). What might be the reason that the peculiar tandem structure of Arg5,6 survived several million years of evolution?

Since Arg6 and Arg5 form a complex, it is conceivable that the tandem structure of the precursor might facilitate their assembly. Many cytosolic protein complexes assemble cotranslationally, and this early-onset interaction between the partner subunits is crucial for function and, more generally, for maintenance of proteostasis (Shiber et al., 2018, Schwarz and Beck, 2019). Cotranslational assembly of nuclear encoded mitochondrial proteins is hindered by the additional translocation step across two mitochondrial membranes. Therefore, coupling the assembly with the import of proteins into the mitochondrial matrix might represent the closest approximation to cotranslational assembly that is physically possible.

Besides its canonical role as "ticket canceller" that clips targeting signals, the mitochondrial processing peptidase MPP obviously also possesses a "tailor" activity for internal processing of several precursor proteins (Fig. 5C). This property is conserved across the fungi, plant and potentially also animal kingdoms. Internal MPP cleavage requires a proximal recognition motif which appears to be an iMTS-L, but remarkably also a strong N-terminal presequence. This suggests that in order to

282 access internal cleavage sites, MPP has to be loaded onto a precursor as soon as it emerges from the
283 TIM23 channel. MPP might then "scan" the still unfolded polypeptide for cleavage sites. A protein
284 with a strong presequence will efficiently recruit MPP, which enables subsequent internal cleavage
285 at an iMTS-L before this is buried by protein folding. In analyses of the N-proteome from yeast,
286 mouse and human mitochondria, a surprising variability in the N-termini of many proteins were
287 observed, sometimes more than 100 amino acids downstream of the annotated start (Calvo et al.,
288 2017, Vögtle et al., 2009, Vaca Jacome et al., 2015). Some of these isoforms might be generated by
289 "leaky" internal MPP cleavage. For instance, the mouse protein Sdha has an alternative N-terminus
290 at position 172 which coincides with an iMTS-L (Fig. 5B). It will be exciting to elucidate the
291 biogenesis and physiological role of such protein isoforms in future research. Its unusual biogenesis
292 might render Arg5,6 a valuable model substrate to explore the mechanisms that confer specificity of
293 internal MPP cleavage.

294

295

296

Materials and Methods

Yeast strains and plasmids

All yeast strains used in this study were based on the WT strain BY4742 (Winston et al., 1995). Unless indicated differently, strains were grown on synthetic medium (0.17% yeast nitrogen base and 0.5% $(NH_4)_2SO_4$) containing 2% glucose.

The Arg5,6-coding region or a fragment of it was amplified by PCR and cloned into pGEM4 (Promega, Madison, WI) using the EcoRI and BamHI restriction sites. For construction of the Arg5,6-HA, Arg5$^{344-863}$-HA and Arg5$^{503-863}$-HA versions, the corresponding sequence without stop codon was cloned into the expression plasmid pYX142, which harbors a constitutive *TPI* promoter upstream and the sequence of a hemagglutinin (HA) tag downstream of the multiple cloning site. For expression of Arg6$^{1-343}$-HA and Arg6$^{1-502}$-HA, pYX122 vectors were used which differ from pYX142 in the selectable marker. The sequence of the Su9 presequence was amplified from a Su9-DHFR plasmid and cloned into the EcoRI site of the pYX142 plasmid carrying the *ARG5* inserts to yield the Su9-Arg5$^{344-863}$-HA and Su9-Arg5$^{503-863}$-HA variants.

Isolation of yeast mitochondria

Isolation of mitochondria was performed essentially as described (Saladi et al., 2020). Yeast cells were grown in YPGal medium (1% yeast extract, 2% peptone, 2% galactose) to an OD_{600} of 0.7-1.3, harvested, washed with water and resuspended in MP1 buffer (10 mM DTT, 100 mM Tris). Cells were incubated for 10 min at 30°C, pelleted (5 min at 4,000 x g at RT) and washed with 1.2 M sorbitol. The cell was digested in MP2 buffer (1.2M sorbitol, 20mM KH_2PO_4 pH 7.4) supplemented with 3mg zymolyase/g wet weight (Seikagaku Corporation) for 1 h at 30°C. The spheroplasts were harvested, resuspended in ice-cold homogenization buffer (0.6 M sorbitol, 1 mM EDTA pH 8, 1 mM phenylmethanesulfonyl fluoride (PMSF), 10 mM Tris-HCl pH 7.4, 0.2% fatty acid-free BSA) and lysed by douncing 10 times in a cooled potter homogenizer.

The homogenate was centrifuged for 5 min at 3,500 x g at 4°C to separate cell debris and nuclei from organelles. The mitochondrial fraction was isolated by centrifugation of the supernatant from the

previous step for 12 min at 12,000 x g at 4°C. The crude mitochondrial pellet was gently resuspended in SH buffer (0.6M sorbitol, 20mM HEPES/KOH pH 7.4), centrifuged for 5 min at 4,000 x g at 4°C, recovered from the supernatant by centrifugation for 12 min at 12,000 x g at 4°C and finally resuspended in SH buffer. The protein concentration of the mitochondrial suspension was determined by a Bradford assay and mitochondria were diluted to a final concentration of 10 mg/ml protein with ice-cold SH buffer, aliquoted, frozen in liquid nitrogen and stored at -80°C.

Import of radiolabeled proteins into isolated mitochondria

Import reactions were essentially performed as described previously (Peleh et al., 2015) in the following import buffer: 500 mM sorbitol, 50 mM Hepes, pH 7.4, 80 mM KCl, 10 mM magnesium acetate, and 2 mM KH_2PO_4. Mitochondria were energized by addition of 2 mM ATP and 2 mM NADH before radiolabeled precursor proteins were added. To dissipate the membrane potential, a mixture of 1 µg/ml valinomycin, 8.8 µg/ml antimycin, and 17 µg/ml oligomycin was added to the mitochondria. Precursor proteins were incubated with mitochondria for different times at 25°C before nonimported protein was degraded by addition of 100 µg/ml proteinase K.

Growth Assays

Growth curves were performed automated in a 96 well plate in technical triplicates using the ELx808 Absorbance Microplate Reader (BioTek). Precultures of 100 µl were inoculated at an OD_{600} of 0.1 in microtiter plates and sealed with an air-permeable membrane (Breathe-Easy; Sigma-Aldrich, St. Louis, MO). The growth curves started at OD_{600} 0.1 and incubated at 30°C for 72 h under constant shaking. The OD_{600} was measured every 10 min.

Antibodies

The antibody against Sod1 was raised in rabbits using purified recombinant protein. The secondary antibody was ordered from Biorad (Goat Anti-Rabbit IgG (H+L)-HRP Conjugate #172-1019). The horseradish peroxidase-coupled HA antibody was ordered from Roche (Anti-HA-Peroxidase, High Affinity (3F10), #12 013 819 001). Antibodies were diluted in 5% (w/v) nonfat dry milk-TBS (Roth T145.2) with the following dilutions: Anti-Sod1 1:1,000, Anti-HA 1:500, Anti-Rabbit 1:10,000.

MPP purification

The *E. coli* strain expressing histidine-tagged MPP subunits Mas1 and Mas2 from an expression plasmid was a gift of Vincent Géli (Luciano et al., 1997). The cells were grown at 37°C to an OD_{600} of 1 and induced with 0.5 mM IPTG over night at 30°C. The bacteria were harvested and resuspended in buffer A (250 mM NaCl, 10 mM imidazole, 2 mM β-mercaptoethanol, 0.2% NP-40, lysozyme), incubated at room temperature for 15 min and snap-frozen in liquid nitrogen. After thawing, DNAseI was added and the cells were sonified 20 times for 1 s at 60% duty level with a Branson sonifier 250. The lysate was cleared by centrifugation, loaded on buffer A-equilibrated Ni-NTA Sepharose resin (Aminitra; Expedeon, San Diego, CA) and washed with buffer A without detergent. A second wash was performed with buffer A adjusted to 1 M NaCl. The enzyme was eluted in buffer C (250 mM NaCl, 300 mM imidazole, 5% glycerol) and stored at -80°C.

MPP in vitro cleavage assay

The *in vitro* cleavage reactions were performed in 150 mM NaCl, 10% glycerol, 100 µM $MnCl_2$, and 50 mM Tris, pH 7.5, in the presence of 250 µg of purified MPP and 5% reticulocyte lysate containing the substrate. If not otherwise stated, the incubation time was 1.5 h at 30°C. To inhibit MPP, the same reaction was performed in the presence of 2.5 mM EDTA and without addition of $MnCl_2$.

Prediction of iMTS-Ls

iMTS-Ls were predicted essentially as described (Backes et al., 2018, Boos et al., 2018). Briefly, multiple truncated sequences were generated by sequentially removing amino acids, one by one, N-terminally from the protein of interest. These sequences were submitted to TargetP with appropriate choice of "plant" or "non-plant" organism and without cutoffs. The mTP scores obtained were plotted against the corresponding amino acid position. A Savitzky–Golay filtering step with a window size of 21 (the expected value of the length distribution of known MTSs) was used to smooth the iMTS-L profile. The iMTS-L profiles can also be calculated using the iMTS-L predictor service online tool iMLP (http://imlp.bio.uni-kl.de/).

Arg5,6 homologs, intracellular localization and genomic distance

Amino acid sequences of yeast Arg5,6 and both post-translationally processed proteins Arg5 and Arg6 were retrieved from (Boonchird et al., 1991a) and used as query sequences for a Diamond BLASTp (Buchfink et al., 2015) against a dataset of 150 eukaryotic genomes (Suppl. Table 1D). Subject organisms that exhibited hits (at least 25% local identity and a maximum E-value of 1E-10) were grouped by the number and pattern of the identified homologs, corresponding to either encoding Arg5,6 as one gene or as two genes (Suppl. Table 1E), by a python script. The intracellular localization of all homologs was checked via TargetP 2.0 (Almagro Armenteros et al., 2019). Results were plotted on a eukaryotic reference tree generated in a previous analysis (Brueckner and Martin, 2020). For the identification of prokaryotic homologs, argB and argC amino acid sequences of *Escherichia coli K-12* were used as queries to search our dataset of 5,655 complete prokaryotic genomes (Refseq, Suppl. Table 1A) (O'Leary et al., 2016). Query sequences were identified as homologs to yeast Arg5 and Arg6 via diamond BLASTp. Only prokaryotic subject organisms exhibiting exactly one homolog to each of the two query sequences (at least 25% local identity and a maximum E-value of 1E-10) were further investigated and the nucleotide distance between the subject genes was calculated (Suppl. Table 1C). Sequence pairs with a maximum distance of 30 nucleotides were suspected to be encoded polycistronically. All genomic distances in nucleotides were derived from Refseq genome feature tables. Sequence pairs with overlapping start and end position were given a distance of one nucleotide.

Author contributions

F.B. and J.M.H. conceived and supervised the study. J.F. generated constructs and strains and performed *in vivo* experiments. J.F. and E.P. purified recombinant proteins and performed *in vitro* MPP digestions. J.F. and C.G. performed *in vitro* import assays. J.F. and F.B. performed *in silico* prediction of iMTS-Ls. M.R.K. and S.B.G. analyzed the organization of *ARG5,6* homologs in different species. J.F., J.M.H. and F.B. analyzed the data. F.B. wrote the manuscript to which all authors contributed.

Acknowledgements

We thank Sabine Knaus, Alexander Grevel and Thomas Becker for assistance with the experiments, and Abdussalam Azem and Katja Hansen for helpful discussions and critical reading of the manuscript. This project was funded by grants from the Deutsche Forschungsgemeinschaft (DIP MitoBalance and 2803/10-1 to J.M.H. and 267205415 – SFB 1208 to S.B.G.), the Volkswagen Stiftung (Life) to S.B.G., the Minerva Stiftung (to J.F.) and the Joachim Herz Stiftung (to F.B.).

Competing interests

The authors declare that they have no competing interests.

References

ABADJIEVA, A., PAUWELS, K., HILVEN, P. & CRABEEL, M. 2001. A new yeast metabolon involving at least the two first enzymes of arginine biosynthesis: acetylglutamate synthase activity requires complex formation with acetylglutamate kinase. *J Biol Chem,* 276**,** 42869-80.

ALMAGRO ARMENTEROS, J. J., SALVATORE, M., EMANUELSSON, O., WINTHER, O., VON HEIJNE, G., ELOFSSON, A. & NIELSEN, H. 2019. Detecting sequence signals in targeting peptides using deep learning. *Life Sci Alliance,* 2**,** e201900429.

BACKES, S., HESS, S., BOOS, F., WOELLHAF, M. W., GODEL, S., JUNG, M., MÜHLHAUS, T. & HERRMANN, J. M. 2018. Tom70 enhances mitochondrial preprotein import efficiency by binding to internal targeting sequences. *J Cell Biol,* 217**,** 1369-1382.

BAKER, A. & SCHATZ, G. 1987. Sequences from a prokaryotic genome or the mouse dihydrofolate reductase gene can restore the import of a truncated precursor protein into yeast mitochondria. *Proc Natl Acad Sci U S A,* 84**,** 3117-21.

BECKER, T., SONG, J. & PFANNER, N. 2019. Versatility of Preprotein Transfer from the Cytosol to Mitochondria. *Trends Cell Biol,* 29**,** 534-548.

BOONCHIRD, C., MESSENGUY, F. & DUBOIS, E. 1991a. Characterization of the yeast ARG5,6 gene: determination of the nucleotide sequence, analysis of the control region and of ARG5,6 transcript. *Mol Gen Genet,* 226**,** 154-66.

BOONCHIRD, C., MESSENGUY, F. & DUBOIS, E. 1991b. Determination of amino acid sequences involved in the processing of the ARG5/ARG6 precursor in Saccharomyces cerevisiae. *Eur J Biochem,* 199**,** 325-35.

BOOS, F., MÜHLHAUS, T. & HERRMANN, J. M. 2018. Detection of Internal Matrix Targeting Signal-like Sequences (iMTS-Ls) in Mitochondrial Precursor Proteins Using the TargetP Prediction Tool. *Bio-protocol,* 8**,** e2474.

BRADSHAW, R. A., BRICKEY, W. W. & WALKER, K. W. 1998. N-terminal processing: the methionine aminopeptidase and N alpha-acetyl transferase families. *Trends Biochem Sci,* 23**,** 263-7.

BRUECKNER, J. & MARTIN, W. F. 2020. Bacterial Genes Outnumber Archaeal Genes in Eukaryotic Genomes. *Genome Biol Evol,* 12**,** 282-292.

BUCHFINK, B., XIE, C. & HUSON, D. H. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods,* 12**,** 59-60.

BYKOV, Y. S., RAPAPORT, D., HERRMANN, J. M. & SCHULDINER, M. 2020. Cytosolic Events in the Biogenesis of Mitochondrial Proteins. *Trends in Biochemical Sciences*.

CALVO, S. E., JULIEN, O., CLAUSER, K. R., SHEN, H., KAMER, K. J., WELLS, J. A. & MOOTHA, V. K. 2017. Comparative Analysis of Mitochondrial N-Termini from Mouse, Human, and Yeast. *Mol Cell Proteomics,* 16**,** 512-523.

CHACINSKA, A., KOEHLER, C. M., MILENKOVIC, D., LITHGOW, T. & PFANNER, N. 2009. Importing mitochondrial proteins: machineries and mechanisms. *Cell,* 138**,** 628-44.

EMANUELSSON, O., BRUNAK, S., VON HEIJNE, G. & NIELSEN, H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc,* 2**,** 953-71.

FINLEY, D., BARTEL, B. & VARSHAVSKY, A. 1989. The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis. *Nature,* 338**,** 394-401.

FROTTIN, F., MARTINEZ, A., PEYNOT, P., MITRA, S., HOLZ, R. C., GIGLIONE, C. & MEINNEL, T. 2006. The proteomics of N-terminal methionine cleavage. *Mol Cell Proteomics,* 5**,** 2336-49.

GARG, S. G. & GOULD, S. B. 2016. The Role of Charge in Protein Targeting Evolution. *Trends Cell Biol,* 26**,** 894-905.

GEMAYEL, R., YANG, Y., DZIALO, M. C., KOMINEK, J., VOWINCKEL, J., SAELS, V., VAN HUFFEL, L., VAN DER ZANDE, E., RALSER, M., STEENSELS, J., VOORDECKERS, K. & VERSTREPEN, K. J. 2017. Variable repeats in the eukaryotic polyubiquitin gene ubi4 modulate proteostasis and stress survival. *Nat Commun,* 8**,** 397.

GESSERT, S. F., KIM, J. H., NARGANG, F. E. & WEISS, R. L. 1994. A polyprotein precursor of two mitochondrial enzymes in Neurospora crassa. Gene structure and precursor processing. *J Biol Chem,* 269**,** 8189-203.

KHALIMONCHUK, O., OTT, M., FUNES, S., OSTERMANN, K., RODEL, G. & HERRMANN, J. M. 2006. Sequential processing of a mitochondrial tandem protein: insights into protein import in Schizosaccharomyces pombe. *Eukaryot Cell,* 5**,** 997-1006.

KOMINEK, J., DOERING, D. T., OPULENTE, D. A., SHEN, X. X., ZHOU, X., DEVIRGILIO, J., HULFACHOR, A. B., GROENEWALD, M., MCGEE, M. A., KARLEN, S. D., KURTZMAN, C. P., ROKAS, A. & HITTINGER, C. T. 2019. Eukaryotic Acquisition of a Bacterial Operon. *Cell,* 176**,** 1356-1366 e10.

KRICHEL, B., FALKE, S., HILGENFELD, R., REDECKE, L. & UETRECHT, C. 2020. Processing of the SARS-CoV pp1a/ab nsp7-10 region. *Biochem J,* 477**,** 1009-1019.

LUCIANO, P., GEOFFROY, S., BRANDT, A., HERNANDEZ, J. F. & GELI, V. 1997. Functional cooperation of the mitochondrial processing peptidase subunits. *J Mol Biol,* 272**,** 213-25.

LUCIANO, P., TOKATLIDIS, K., CHAMBRE, I., GERMANIQUE, J. C. & GELI, V. 1998. The mitochondrial processing peptidase behaves as a zinc-metallopeptidase. *J Mol Biol,* 280**,** 193-9.

MINET, M., JAUNIAUX, J. C., THURIAUX, P., GRENSON, M. & WIAME, J. M. 1979. Organization and expression of a two-gene cluster in the arginine biosynthesis of Saccharomyces cerevisiae. *Mol Gen Genet,* 168**,** 299-308.

MORGENSTERN, M., STILLER, S. B., LUBBERT, P., PEIKERT, C. D., DANNENMAIER, S., DREPPER, F., WEILL, U., HOSS, P., FEUERSTEIN, R., GEBERT, M., BOHNERT, M., VAN DER LAAN, M., SCHULDINER, M., SCHUTZE, C., OELJEKLAUS, S., PFANNER, N., WIEDEMANN, N. & WARSCHEID, B. 2017. Definition of a High-Confidence Mitochondrial Proteome at Quantitative Scale. *Cell Rep,* 19**,** 2836-2852.

MOSSMANN, D., VÖGTLE, F. N., TASKIN, A. A., TEIXEIRA, P. F., RING, J., BURKHART, J. M., BURGER, N., PINHO, C. M., TADIC, J., LORETH, D., GRAFF, C., METZGER, F., SICKMANN, A., KRETZ, O., WIEDEMANN, N., ZAHEDI, R. P., MADEO, F., GLASER, E. & MEISINGER, C. 2014. Amyloid-beta peptide induces mitochondrial dysfunction by inhibition of preprotein maturation. *Cell Metab,* 20**,** 662-9.

MÜLLER, U. C., DELLER, T. & KORTE, M. 2017. Not just amyloid: physiological functions of the amyloid precursor protein family. *Nat Rev Neurosci,* 18**,** 281-298.

NAAMATI, A., REGEV-RUDZKI, N., GALPERIN, S., LILL, R. & PINES, O. 2009. Dual targeting of Nfs1 and discovery of its novel processing enzyme, Icp55. *J Biol Chem,* 284**,** 30200-8.

O'LEARY, N. A., WRIGHT, M. W., BRISTER, J. R., CIUFO, S., HADDAD, D., MCVEIGH, R., RAJPUT, B., ROBBERTSE, B., SMITH-WHITE, B., AKO-ADJEI, D., ASTASHYN, A., BADRETDIN, A., BAO, Y., BLINKOVA, O., BROVER, V., CHETVERNIN, V., CHOI, J., COX, E., ERMOLAEVA, O., FARRELL, C. M., GOLDFARB, T., GUPTA, T., HAFT, D., HATCHER, E., HLAVINA, W., JOARDAR, V. S., KODALI, V. K., LI, W., MAGLOTT, D., MASTERSON, P., MCGARVEY, K. M., MURPHY, M. R., O'NEILL, K., PUJAR, S., RANGWALA, S. H., RAUSCH, D., RIDDICK, L. D., SCHOCH, C., SHKEDA, A., STORZ, S. S., SUN, H., THIBAUD-NISSEN, F., TOLSTOY, I., TULLY, R. E., VATSAN, A. R., WALLIN, C., WEBB, D., WU, W., LANDRUM, M. J., KIMCHI, A., TATUSOVA, T., DICUCCIO, M., KITTS, P., MURPHY, T. D. & PRUITT, K. D. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res,* 44**,** D733-45.

OSHIMA, T., YAMASAKI, E., OGISHIMA, T., KADOWAKI, K., ITO, A. & KITADA, S. 2005. Recognition and processing of a nuclear-encoded polyprotein precursor by mitochondrial processing peptidase. *Biochem J,* 385**,** 755-61.

OZKAYNAK, E., FINLEY, D., SOLOMON, M. J. & VARSHAVSKY, A. 1987. The yeast ubiquitin genes: a family of natural gene fusions. *EMBO J,* 6**,** 1429-39.

OZKAYNAK, E., FINLEY, D. & VARSHAVSKY, A. 1984. The yeast ubiquitin gene: head-to-tail repeats encoding a polyubiquitin precursor protein. *Nature,* 312**,** 663-6.

PARRA-GESSERT, L., KOO, K., FAJARDO, J. & WEISS, R. L. 1998. Processing and function of a polyprotein precursor of two mitochondrial proteins in neurospora crassa. *J Biol Chem,* 273**,** 7972-80.

PAUWELS, K., ABADJIEVA, A., HILVEN, P., STANKIEWICZ, A. & CRABEEL, M. 2003. The N-acetylglutamate synthase/N-acetylglutamate kinase metabolon of Saccharomyces cerevisiae allows co-ordinated feedback regulation of the first two steps in arginine biosynthesis. *Eur J Biochem,* 270**,** 1014-24.

PELEH, V., RAMESH, A. & HERRMANN, J. M. 2015. Import of proteins into isolated yeast mitochondria. *Methods Mol Biol,* 1270**,** 37-50.

PFANNER, N., WARSCHEID, B. & WIEDEMANN, N. 2019. Mitochondrial proteins: from biogenesis to functional networks. *Nat Rev Mol Cell Biol,* 20**,** 267-284.

PIETTE, J., CUNIN, R., BOYEN, A., CHARLIER, D., CRABEEL, M., VAN VLIET, F., GLANSDORFF, N., SQUIRES, C. & SQUIRES, C. L. 1982. The regulatory region of the divergent argECBH operon in Escherichia coli K-12. *Nucleic Acids Res,* 10**,** 8031-48.

POVEDA-HUERTES, D., MATIC, S., MARADA, A., HABERNIG, L., LICHEVA, M., MYKETIN, L., GILSBACH, R., TOSAL-CASTANO, S., PAPINSKI, D., MULICA, P., KRETZ, O., KUCUKKOSE, C., TASKIN, A. A., HEIN, L., KRAFT, C., BUTTNER, S., MEISINGER, C. & VÖGTLE, F. N. 2020. An Early mtUPR: Redistribution of the Nuclear Transcription Factor Rox1 to Mitochondria Protects against Intramitochondrial Proteotoxic Aggregates. *Mol Cell,* 77**,** 180-188 e9.

POVEDA-HUERTES, D., MULICA, P. & VÖGTLE, F. N. 2017. The versatility of the mitochondrial presequence processing machinery: cleavage, quality control and turnover. *Cell Tissue Res,* 367**,** 73-81.

QUIROS, P. M., LANGER, T. & LOPEZ-OTIN, C. 2015. New roles for mitochondrial proteases in health, ageing and disease. *Nat Rev Mol Cell Biol,* 16**,** 345-59.

SALADI, S., BOOS, F., POGLITSCH, M., MEYER, H., SOMMER, F., MUHLHAUS, T., SCHRODA, M., SCHULDINER, M., MADEO, F. & HERRMANN, J. M. 2020. The NADH Dehydrogenase Nde1 Executes Cell Death after Integrating Signals from Metabolism and Proteostasis on the Mitochondrial Surface. *Mol Cell,* 77**,** 189-202 e6.

SATO, T. K., KAWANO, S. & ENDO, T. 2019. Role of the membrane potential in mitochondrial protein unfolding and import. *Sci Rep,* 9**,** 7637.

SCHENDZIELORZ, A. B., SCHULZ, C., LYTOVCHENKO, O., CLANCY, A., GUIARD, B., IEVA, R., VAN DER LAAN, M. & REHLING, P. 2017. Two distinct membrane potential-dependent steps drive mitochondrial matrix protein translocation. *J Cell Biol,* 216**,** 83-92.

SCHWARZ, A. & BECK, M. 2019. The Benefits of Cotranslational Assembly: A Structural Perspective. *Trends Cell Biol,* 29**,** 791-803.

SHIBER, A., DÖRING, K., FRIEDRICH, U., KLANN, K., MERKER, D., ZEDAN, M., TIPPMANN, F., KRAMER, G. & BUKAU, B. 2018. Cotranslational assembly of protein complexes in eukaryotes revealed by ribosome profiling. *Nature,* 561**,** 268-272.

STEINER, D. F. & OYER, P. E. 1967. The biosynthesis of insulin and a probable precursor of insulin by a human islet cell adenoma. *Proc Natl Acad Sci U S A,* 57**,** 473-80.

SUZUKI, C. K., SUDA, K., WANG, N. & SCHATZ, G. 1994. Requirement for the yeast gene LON in intramitochondrial proteolysis and maintenance of respiration. *Science,* 264**,** 273-6.

VACA JACOME, A. S., RABILLOUD, T., SCHAEFFER-REISS, C., ROMPAIS, M., AYOUB, D., LANE, L., BAIROCH, A., VAN DORSSELAER, A. & CARAPITO, C. 2015. N-terminome analysis of the human mitochondrial proteome. *Proteomics,* 15**,** 2519-24.

VARSHAVSKY, A. 2011. The N-end rule pathway and regulation by proteolysis. *Protein Sci,* 20**,** 1298-345.

VELING, M. T., REIDENBACH, A. G., FREIBERGER, E. C., KWIECIEN, N. W., HUTCHINS, P. D., DRAHNAK, M. J., JOCHEM, A., ULBRICH, A., RUSH, M. J. P., RUSSELL, J. D., COON, J. J. & PAGLIARINI, D. J. 2017. Multi-omic Mitoprotease Profiling Defines a Role for Oct1p in Coenzyme Q Production. *Mol Cell,* 68**,** 970-977 e11.

VÖGTLE, F. N., PRINZ, C., KELLERMANN, J., LOTTSPEICH, F., PFANNER, N. & MEISINGER, C. 2011. Mitochondrial protein turnover: role of the precursor intermediate peptidase Oct1 in protein stabilization. *Mol Biol Cell,* 22**,** 2135-43.

VÖGTLE, F. N., WORTELKAMP, S., ZAHEDI, R. P., BECKER, D., LEIDHOLD, C., GEVAERT, K., KELLERMANN, J., VOOS, W., SICKMANN, A., PFANNER, N. & MEISINGER, C. 2009. Global analysis of the mitochondrial N-proteome identifies a processing peptidase critical for protein stability. *Cell,* 139**,** 428-39.

VON HEIJNE, G. 1986. Mitochondrial targeting sequences may form amphiphilic helices. *EMBO J,* 5**,** 1335-42.

WAGNER, I., ARLT, H., VAN DYCK, L., LANGER, T. & NEUPERT, W. 1994. Molecular chaperones cooperate with PIM1 protease in the degradation of misfolded proteins in mitochondria. *EMBO J,* 13**,** 5135-45.

WINSTON, F., DOLLARD, C. & RICUPERO-HOVASSE, S. L. 1995. Construction of a set of convenient Saccharomyces cerevisiae strains that are isogenic to S288C. *Yeast,* 11**,** 53-5.

WOELLHAF, M. W., SOMMER, F., SCHRODA, M. & HERRMANN, J. M. 2016. Proteomic profiling of the mitochondrial ribosome identifies Atp25 as a composite mitochondrial precursor protein. *Mol Biol Cell,* 27**,** 3031-3039.

YOST, S. A. & MARCOTRIGIANO, J. 2013. Viral precursor polyproteins: keys of regulation from replication to maturation. *Curr Opin Virol,* 3**,** 137-42.

572 ZHANG, L., LIN, D., SUN, X., CURTH, U., DROSTEN, C., SAUERHERING, L., BECKER, S., ROX, K. & HILGENFELD,
573 R. 2020. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-
574 ketoamide inhibitors. *Science,* 368**,** 409-412.

575

Figure Legends

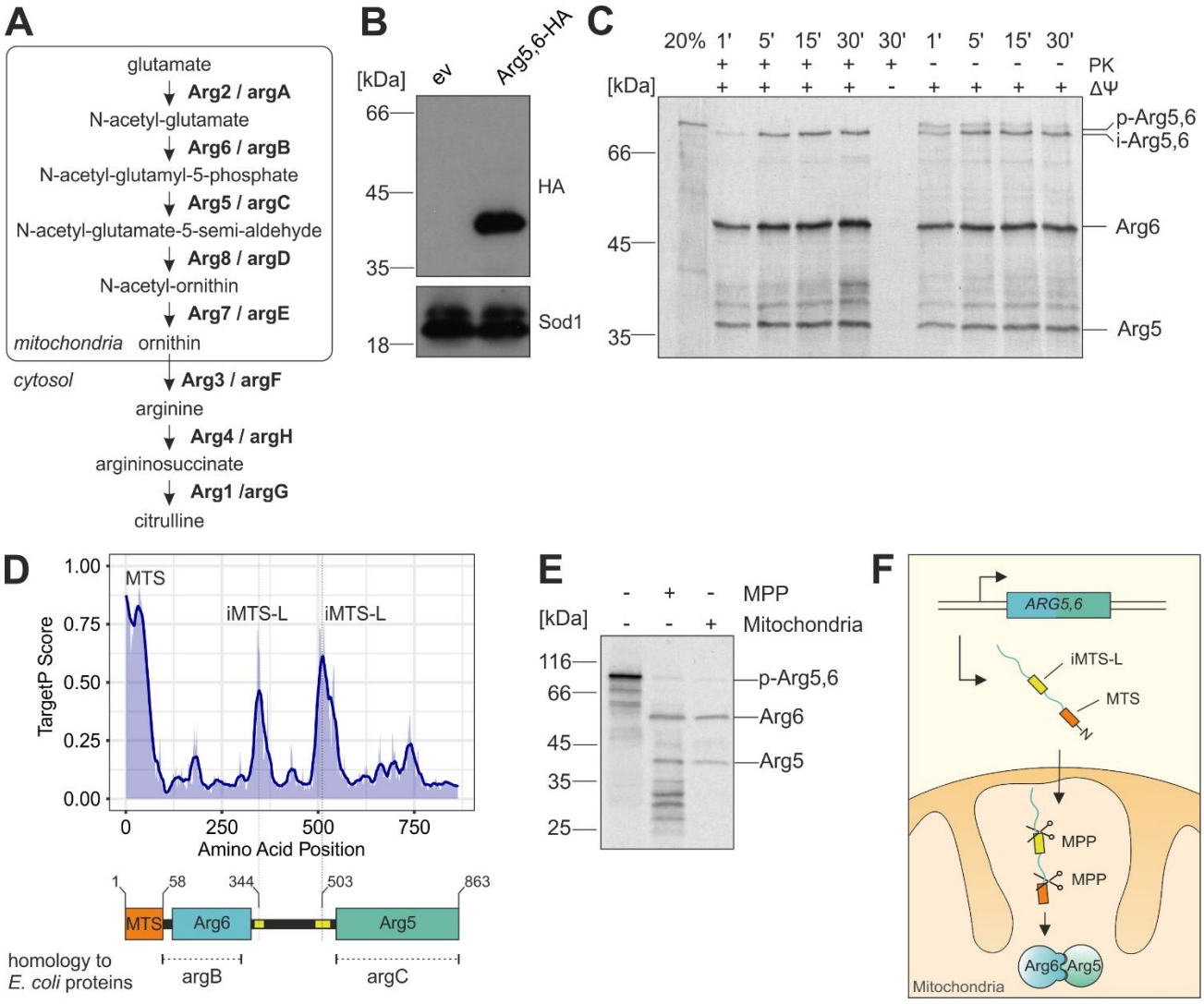


Figure 1. Arg5,6 is a composite mitochondrial precursor that is processed twice by MPP in the mitochondrial matrix. A, Schematic representation of arginine biosynthesis in *S. cerevisiae*. Shown in bold are the enzymes that catalyze the respective step as well as their *E. coli* orthologs. **B,** When Arg5,6 was C-terminally HA-tagged, immunoblotting revealed a single band at 40 kDa, indicating proteolytic cleavage of the 90 kDa precursor protein. **C,** Radiolabeled Arg5,6 precursor was incubated with isolated mitochondria for the indicated times and analyzed by SDS-PAGE and autoradiography. Non-imported material was digested with proteinase K (left half). 20% of the total lysate used per import lane was loaded for control. The membrane potential ($\Delta\psi$) was dissipated with VAO (valinomycin, antimycin, oligomycin). p, precursor, i, intermediate. **D,** Arg5,6 was subjected to TargetP profiling. High values indicate regions within the protein which structurally resemble

588 mitochondrial presequences. **E,** His-tagged MPP was expressed and purified from *E. coli*.
589 Radiolabeled Arg5,6 precursor was incubated with isolated mitochondria for 15 min or purified MPP
590 for 90 min. The processing of Arg5,6 was analyzed by SDS-PAGE, Western blotting and
591 autoradiography. **F,** Arg5,6 is imported into the mitochondrial matrix and cleaved twice by MPP:
592 once at the N-terminus to remove the presequence, and once internally at an iMTS-L to separate Arg5
593 and Arg6.

594

595

596

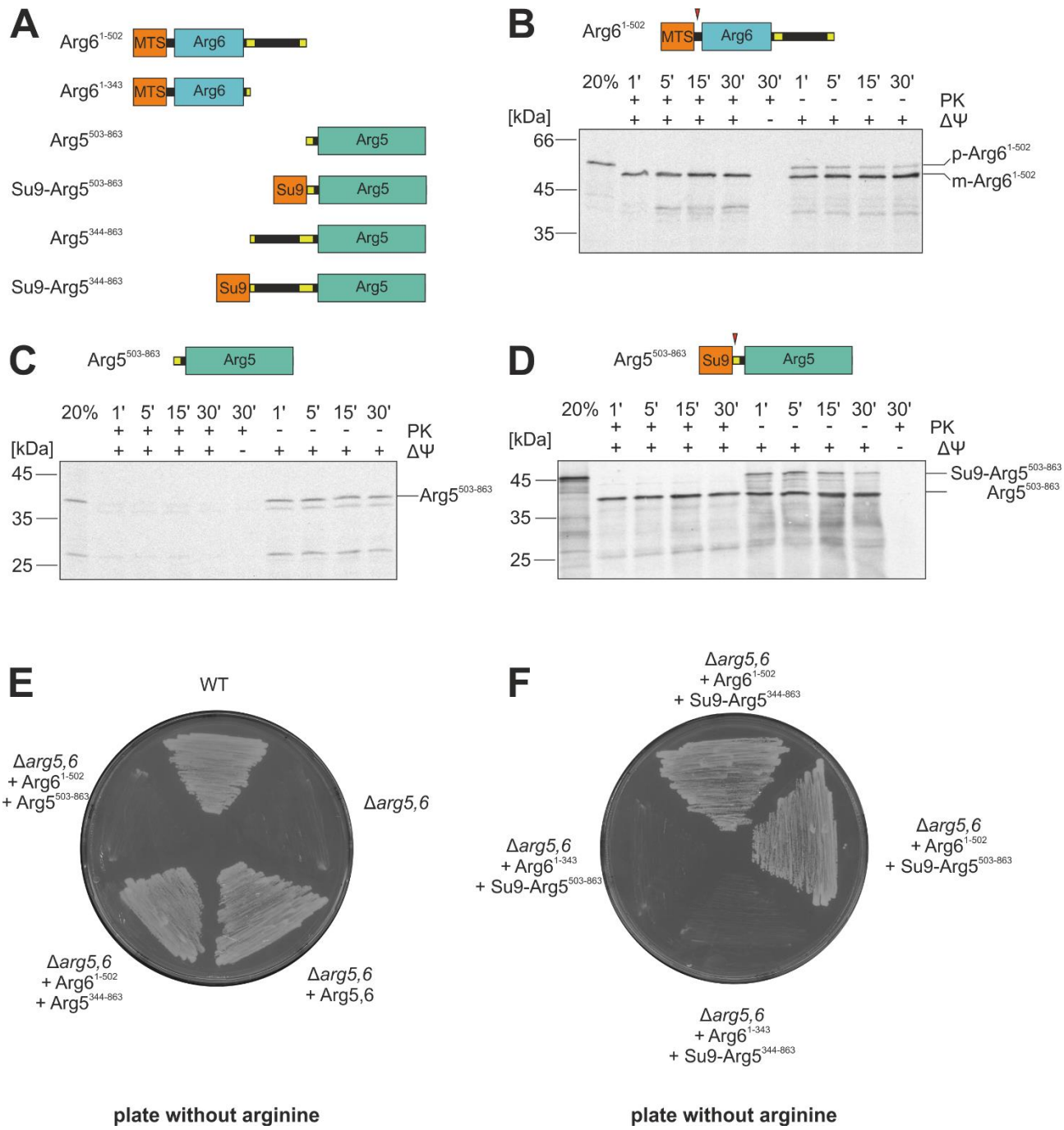


Figure 2. Arg5 and Arg6 can be imported separately *in vitro* and *in vivo*. A, Overview of truncated Arg5,6 variants used in this study. Su9, presequence of *N. crassa* subunit 9. **B-D,** Radiolabeled precursor proteins of $Arg6^{1-502}$, $Arg5^{503-862}$ and $Su9\text{-}Arg5^{503-862}$ were incubated with isolated mitochondria for the indicated times and analyzed by SDS-PAGE and autoradiography. Non-imported material was digested with proteinase K (left half). 20% of the total lysate used per import lane is loaded for control. The membrane potential (Δψ) was depleted with VAO. Red arrowheads indicate processing sites. p, precursor, m, mature. **E-F,** Yeast cells that lack endogenous Arg5,6

605 (Δ*arg5,6*) were transformed with plasmids for expression of the indicated Arg5,6 variants and
606 streaked out on plates containing minimal growth medium without arginine.

607

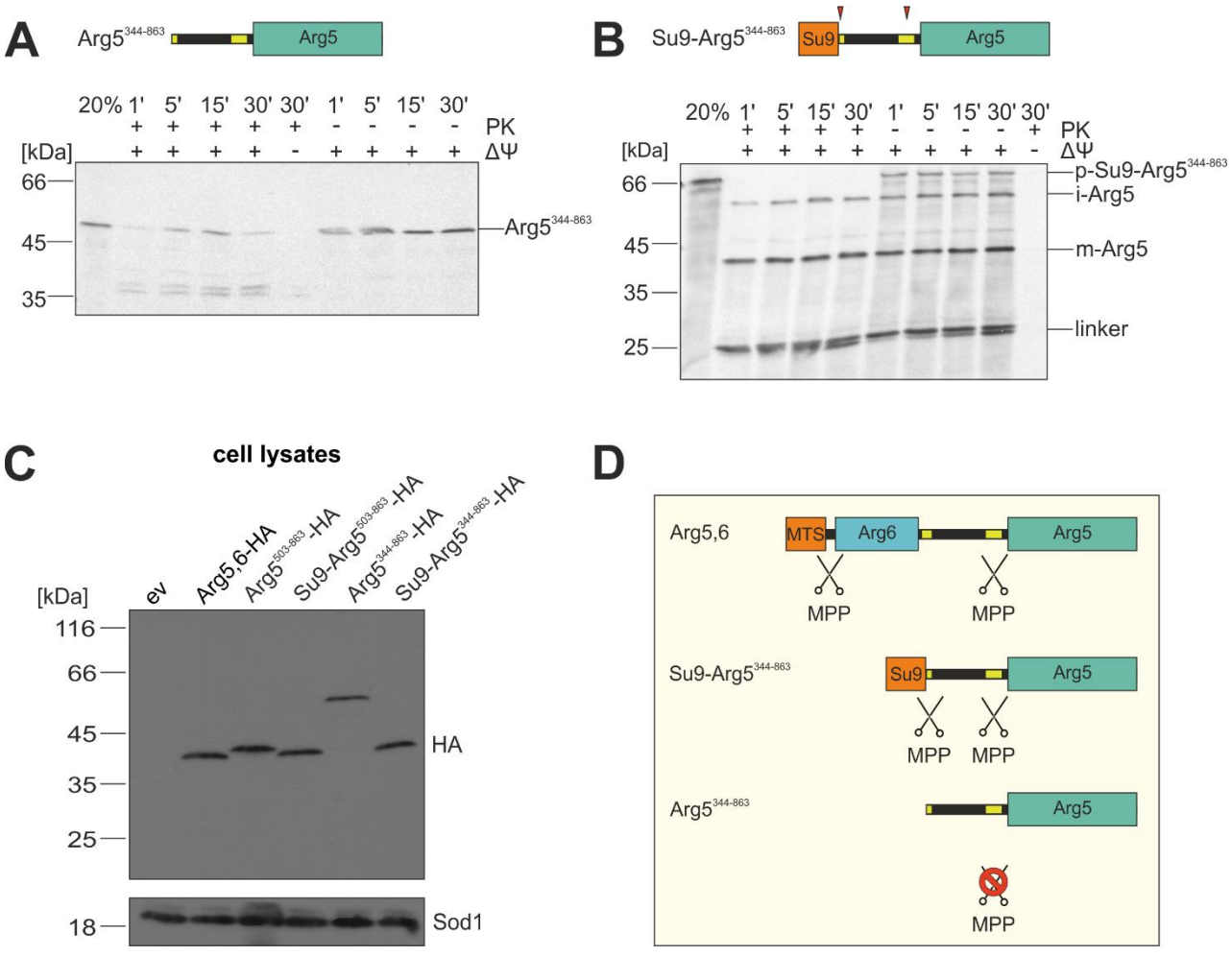


Figure 3. MPP requires a strong N-terminal MTS for internal processing of precursor proteins. **A-B,** Radiolabeled precursor proteins of $Arg5^{344\text{-}862}$ and $Su9\text{-}Arg5^{344\text{-}862}$ were incubated with isolated mitochondria for the indicated times and analyzed by SDS-PAGE and autoradiography. Non-imported material is digested with proteinase K (left half). 20% of the total lysate used per import lane is loaded for control. The membrane potential ($\Delta\psi$) was depleted with VAO. Red arrowheads indicate processing sites. p, precursor, i, intermediate, m, mature. **C,** Yeast cells expressing indicated variants of Arg5,6, all carrying a C-terminal HA tag, were lysed and protein extracts were analyzed by SDS-PAGE and immunoblotting directed against the HA epitope or Sod1 as a loading control. ev, empty vector. **D,** MPP cleaves the Arg5,6 precursors at its internal processing site only if they carry a *bona fide* N-terminal presequence.

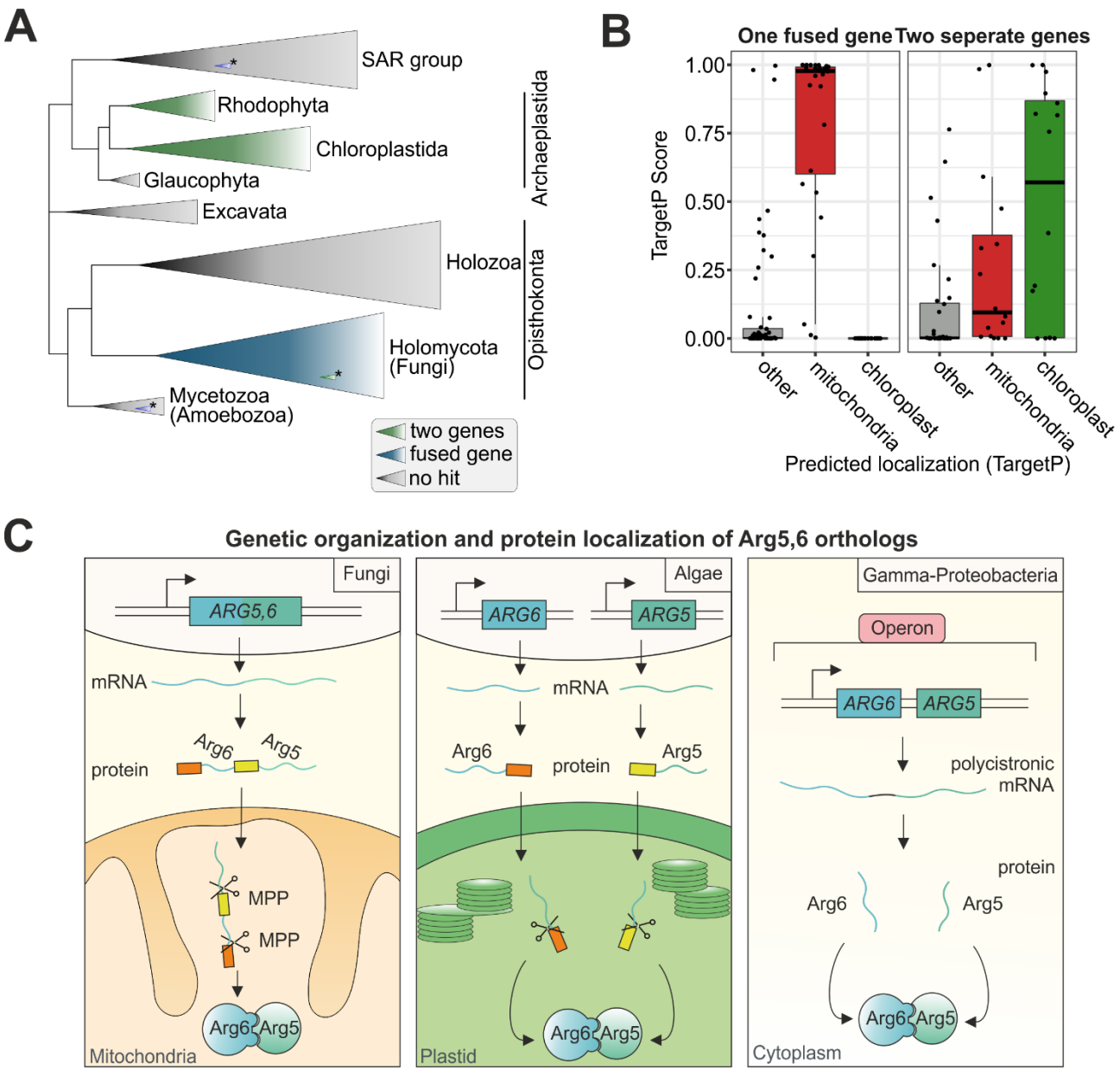


Figure 4. Tandem organization and mitochondrial localization of Arg5,6 is conserved in fungi, whereas algae synthesize two separate proteins that localize to their plastids. A, A database of 150 eukaryotes was searched for homologs of full length Arg5,6 and the separate Arg5 and Arg6 proteins of *S. cerevisiae*. Clades in which Arg5,6 homologs were encoded as a single fusion protein are coloured in blue, clades in which two separate Arg6 and Arg5 genes were found are coloured in green. Grey, no Arg5,6 homologs were found in these clades. **B,** In species that encode Arg5,6 as fusion protein, TargetP predicts the protein to be localized to mitochondria. In species with separate genes for Arg5 and Arg6, localization is predicted to be plastidal. **C,** In fungi, Arg5,6 is a fusion protein which is matured into two separate enzymes by MPP in the mitochondrial matrix. In contrast,

631 algae encode two separate proteins that localize to their plastids. Gamma-proteobacteria encode Arg5

632 and Arg6 polycistronically.

633

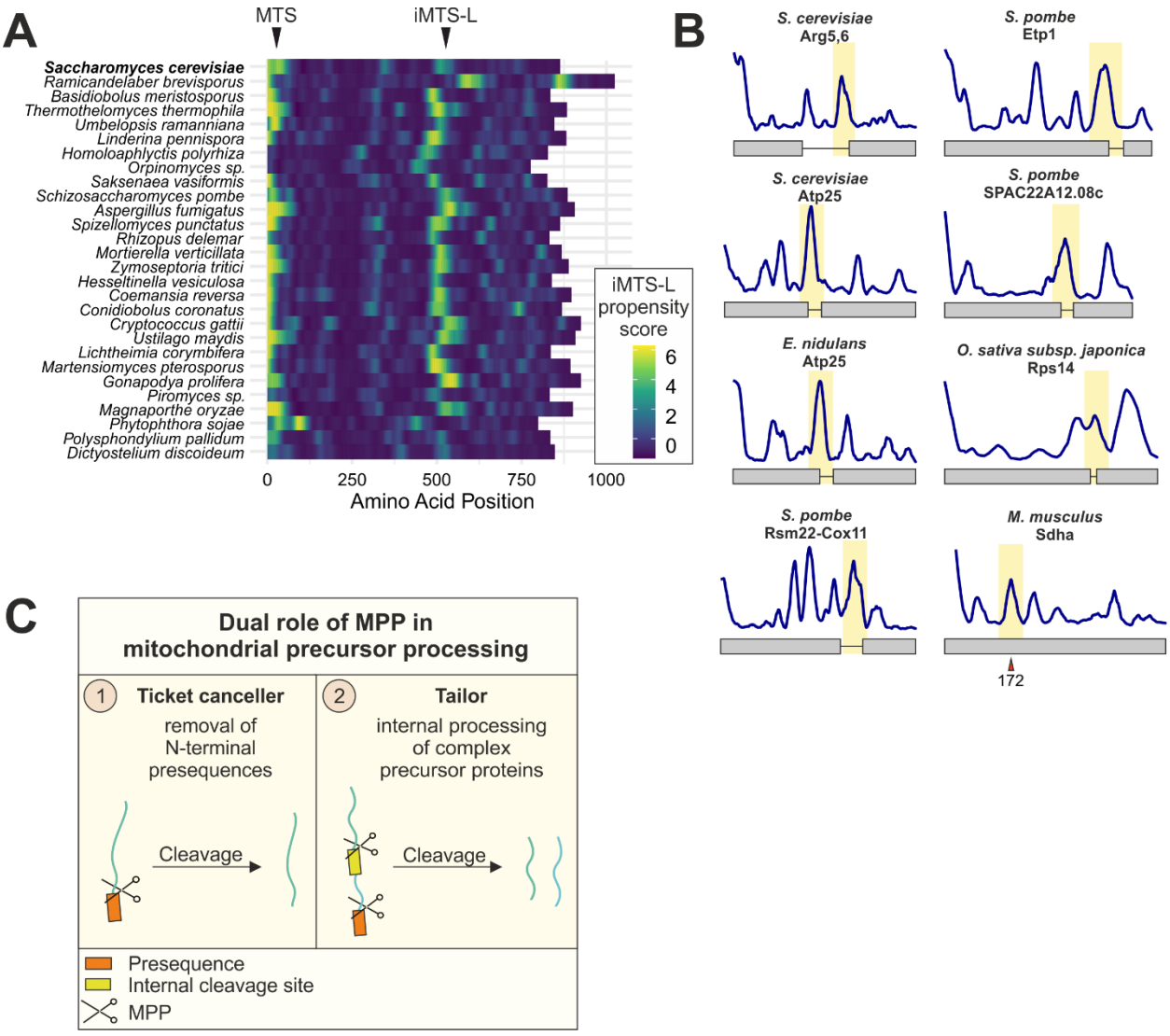
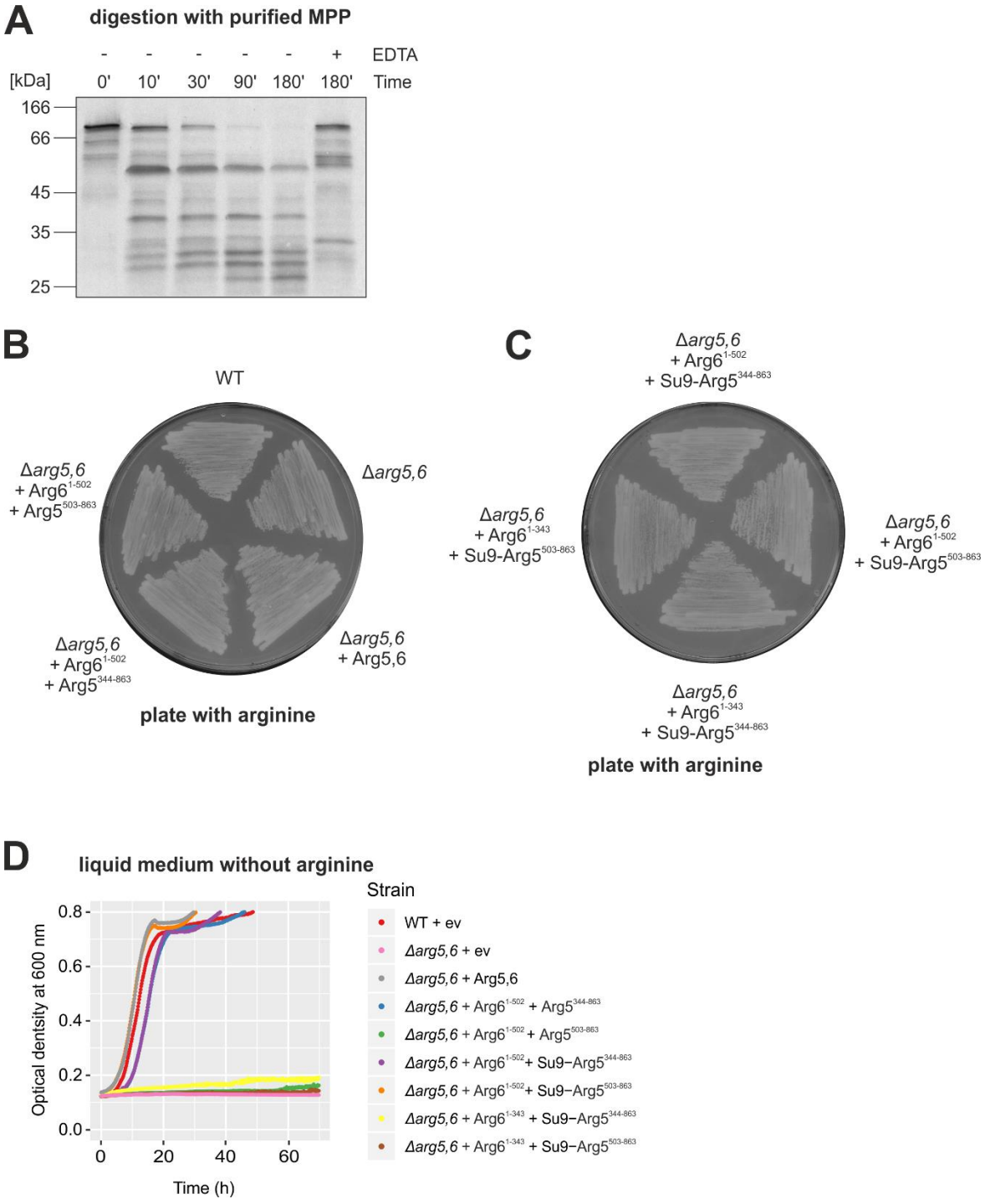


Figure 5. MPP functions not only as presequence peptidase, but also has a conserved processing activity at internal cleavage sites of composite precursor proteins. A, The iMTS-L at which MPP cleaves the Arg5,6 precursor in *S. cerevisiae* is conserved among species that encode Arg5,6 as fusion protein. Shown are iMTS-L propensity profiles along the sequence of Arg5,6 fusion protein homologs. **B,** Previously described mitochondrial polyproteins from different species harbor an iMTS-L at the position of the junction between the fused polypeptides, indicating potential MPP cleavage sites. Sdha is no polyprotein, but an alternative N-terminus (red arrowhead) was recently identified, presumably the result of a posttranslational cleavage (Calvo et al., 2017). **C,** In addition to its canonical role in presequence removal from mitochondrial precursor proteins, MPP has a second, conserved function. It recognizes internal cleavage sites and processes complex precursor proteins into separate polypeptides.

646 **Supplementary Figures**

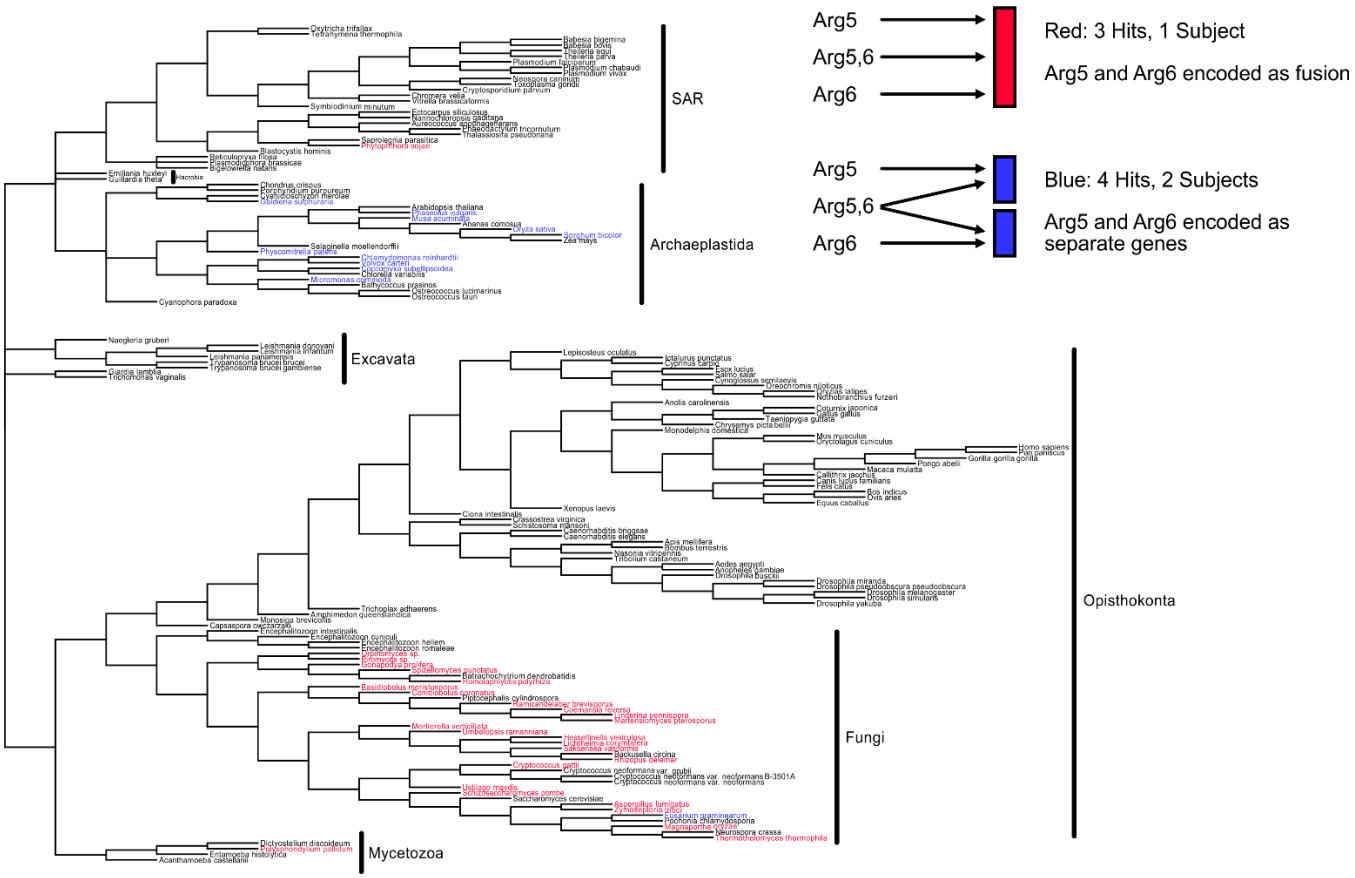


647

648 **Supplementary Figure 1. Arg5,6 is internally processed by MPP *in vitro* and separate expression**
649 **of Arg5 and Arg6 *in vivo* is non-toxic and can restore arginine prototrophy. A,** Radiolabeled
650 Arg5,6 precursor is incubated with purified MPP for indicated times. EDTA is added as a control to
651 inhibit the enzymatic activity of MPP. Reactions were analyzed by SDS-PAGE, Western blotting and
652 autoradiography. **B-C,** The indicated Arg5,6 variants are expressed in yeast cells lacking *ARG5,6* and

653 cells are streaked out on plates with minimal medium containing arginine. **D,** Cells are grown in liquid
654 minimal medium without arginine for 72 hours at 30°C. The optical density at 600 nm is measured
655 every 10 min.

656



Supplementary Figure 2. Tandem organization of Arg5,6 is conserved in fungi, whereas algae synthesize two separate proteins. A database of 150 eukaryotes was searched for homologs of full length Arg5,6 and the separate Arg5 and Arg6 proteins of *S. cerevisiae*. Red, species in which the Arg5,6 homolog is encoded as fusion protein. Blue, Arg6 and Arg5 homologs are encoded as separate proteins. Black, no homolog of Arg5,6 found or unclear pattern.

Supplementary Tables

Supplementary Table 1. Evolutionary conservation of the *ARG5,6* gene structure

Supplementary Table 2. Yeast strains and plasmids used in this study.