Evolutionary and genetic analyses of mitochondrial translation initiation factors identify the missing mitochondrial IF3 in S. cerevisiae

Gemma C. Atkinson1,*, Anton Kuzmenko1,2, Piotr Kamenski2, Mikhail Y. Vysokikh3,4, Valentina Lakunina2, Stoyan Tankov1, Ekaterina Smirnova2, Aksel Soosaar1, Tanel Tenson1 and Vasili Hauryliuk1,*

1University of Tartu, Institute of Technology, Tartu, Estonia, 2Department of Molecular Biology, Faculty of Biology, Moscow State University, Moscow, Russia, 3UMR No. 7156, Université de Strasbourg - CNRS, “Molecular Genetics, Genomics, Microbiology”, Strasbourg, France and 4Laboratory of the Molecular Mechanisms of Aging, A.N. Belozersky Physico-Chemical Biology Institute, Moscow State University, Moscow, Russia

Received February 23, 2012; Revised March 9, 2012; Accepted March 13, 2012

ABSTRACT

Mitochondrial translation is essentially bacteria-like, reflecting the bacterial endosymbiotic ancestry of the eukaryotic organelle. However, unlike the translation system of its bacterial ancestors, mitochondrial translation is limited to just a few mRNAs, mainly coding for components of the respiratory complex. The classical bacterial initiation factors (IFs) IF1, IF2 and IF3 are universal in bacteria, but only IF2 is universal in mitochondria (mIF2). We analyse the distribution of mitochondrial translation initiation factors and their sequence features, given two well-propagated claims: first, a sequence insertion in mitochondrial IF2 (mIF2) compensates for the universal lack of IF1 in mitochondria, and secondly, no homologue of mitochondrial IF3 (mIF3) is identifiable in Saccharomyces cerevisiae. Our comparative sequence analysis shows that, in fact, the mIF2 insertion is highly variable and restricted in length and primary sequence conservation to vertebrates, while phylogenetic and in vivo complementation analyses reveal that an uncharacterized S. cerevisiae mitochondrial protein currently named Aim23p is a bona fide evolutionary and functional orthologue of mIF3. Our results highlight the lineage-specific nature of mitochondrial translation and emphasise that comparative analyses among diverse taxa are essential for understanding whether generalizations from model organisms can be made across eukaryotes.

INTRODUCTION

Mitochondria are multifunctional organelles of virtually all eukaryotic cells and were probably present in the last common ancestor of all extant eukaryotes (1). They take part in production of energy, fatty acid metabolism, apoptosis and many other cellular processes. According to the endosymbiotic hypothesis, mitochondria are of bacterial origin (2), which explains why they contain their own genome and are competent in transcription and translation of their genetic material.

Translation initiation in bacteria is facilitated by three universal and essential initiation factors (IFs), IF1, IF2 and IF3. IF2 in the GTP-bound form promotes binding of aminoacylated and formylated initiator tRNA (fMet-tRNAi) to the small ribosomal subunit and subsequent docking of the large subunit to this pre-initiation complex (3–5). IF1 and IF3 together contribute to selection of the initiator codon and fMet-tRNAi through a delicate kinetic mechanism (6–9), and have equally important roles in ribosomal recycling (10–12).

One important difference between the mitochondrial and bacterial translational systems is that the former deals with a very limited set of different mRNAs coding for a handful of proteins, mostly components of the respiratory complex. Most of what we know about translational control in mitochondria comes from the model organism Saccharomyces cerevisiae. In this yeast, the translational machinery is extremely specialized for translating these mRNAs (13). Sequence-specific
translational activators that interact with specific elements in the mRNAs link translation of the specific mRNA to their localization, as well as activate translation (14,15). Thus, the classical IF set in mitochondria in this yeast is less involved in initiation codon and nascent tRNA selection: selection of the initiation region is performed by activators, and the need for fMet-tRNAi is relaxed, allowing initiation with elongator Met-tRNA (16).

The mitochondrial translation machinery has a modified set of classical IFs in comparison to bacteria. IF1 is absent altogether, and mitochondrial IF2 (mIF2), which is universal in mitochondrial eukaryotes, has been suggested to be the functional equivalent of both bacterial IF1 and IF2. A short insertion between domains V and VI has been identified as the element responsible for IF1-like function using complementation assays in Escherichia coli (17). Indeed, cryo-electron microscopy of mIF2 in complex with initiator tRNA and the bacterial ribosome suggests the insertion occupies the same binding site on the ribosome that would be occupied by IF1 (18).

However, no detailed comparative sequence analysis of the insertion across a broad distribution of eukaryotes has been carried out to confirm or reject a relationship with IF1 loss.

Orthologues of IF3 (mIF3) are present in a number of eukaryotes including the fission yeast Schizosaccharomyces pombe, but a clear homologue of mIF3 has not been identified in budding yeast S. cerevisiae (19,20). This raises questions regarding the mechanism of translation initiation and recycling in this organism. Recently, S. cerevisiae mitochondrial translational activators Aep3p and Rsm28p have been shown to interact genetically and physically with mIF2 and initiator tRNA, thus being directly involved in selection of formation of the pre-initiation complex (21–23). These observations raise the possibility that Aep3p and Rsm28p may perform analogous functions of mIF1 and/or mIF3 in this organism (21–23). Again, however, the distributions of Aep3p, Rsm28p and other translational activators and their potential relationship to mIF3 distribution have not previously been addressed.

Here, present a systematic in silico analysis of mitochondrial IF2, IF3 and translational activators. We identify S. cerevisiae Aim23p as the hitherto unidentified mIF3 in Saccharomycetales. By means of in vivo complementation assays, we show that S. pombe mIF3 effectively complements a genomic disruption of S. cerevisiae AIM23, verifying that Aim23p is a bona fide mitochondrial IF3.

### MATERIALS AND METHODS

#### Sequence retrieval and phylogenetic analysis

Sequences homologous to mIF2, mIF3 and 17 S. cerevisiae translational activators were retrieved by BlatP and PSI-Blast searches at the NCBI. Sequences were aligned using MAFFT (24), and maximum likelihood (ML) and Bayesian inference (BI) phylogenetic analyses were carried out using RAxML v7.0.4 (25) and MrBayes v3.1.2 (26). Full methods for sequence analysis are presented in Supplementary Text S1: SI Methods.

Amino acid composition, subcellular targeting and conservation analyses

The amino acid composition of peptides was calculated using the Expasy ProtParam tool (27). Mitochondrial and plastid targeting peptides were predicted using TargetP (28), MitoProt (II) (29), PATS (30) and Plasmid (31). Consensus sequences were calculated using the Consensus Finder Python script (32).

#### Aim23p in vivo complementation experiments

To investigate whether Aim23p is a functional orthologue of mIF3, a strain of S. cerevisiae lacking Aim23p was first obtained. The heterozygous AIM23 knockout diploid strain Y21294, which carries the chromosomal AIM23 gene disrupted by a geneticin (G418) resistance cassette was purchased from EUROSCARF, and sporulation was induced to obtain the haploid AIM23::kanMX4 strain (referred to here as AIM23Δ). This was complemented by a plasmid expressing the gene for S. pombe mIF3 (gene name SPBC18E5.13, henceforth referred to as S. pombe MIF3), and growth was assayed in fermentable and non-fermentable media. Mitochondrial oxygen uptake was measured by Clark electrode and mitochondrial membrane potential was assayed by Rhodamine 123 and DiOC6 labelling. Full methods for complementation analyses are presented in Supplementary Text S1: SI Methods, and the full list of strains and plasmids used in this study is found in Supplementary Table S1.

#### RESULTS

The conserved insertion within mIF2 is only found in vertebrates

IF2 is a translational GTPase (trGTPase) and is the only IF universal in bacteria, archaea, eukaryotes and both the eukaryotic organelles capable of translation: plastids (chloroplasts and apicoplasts derived from secondary endosymbiosis) and mitochondria. We therefore use IF2 as a framework for mapping the presence and absence of other IFs. Phylogenetic analysis of the ‘full IF2 data set’, that is IF2 homologues across bacteria, eukaryotes and archaea, rooted with archaeal and eukaryotic orthologues a/eIF5B shows a lack of resolution in the backbone of the bacterial and organellar IF2 part of the tree (Supplementary Figure S1). This is not surprising given the vast evolutionary distances covered in the whole tree (dating back to the last common ancestor of all life) and the differences in evolutionary rate among organelar sequences (as seen by the long branches for organellar IF2, Supplementary Figure S1). The longest branched protist organellar IF2 sequences are mostly found at the base of the IF2 part of the tree, but relationships among groups have no statistical support in this part of the tree, and possibly represent long-branch attraction (LBA) to the out-group. Scans for organellar transit peptides enable grouping of sequences into putative organelle-specific groups (Supplementary Figure S1). A surprising feature of the tree is a long-branched clade referred to here as mIF2-2, which is a second copy of mIF2 in members of...

There has been a particularly large insertion in bacteria and cpIF2, both of which carry IF1 (Figure 1). Many independent insertions and deletions, as ambiguous homology, the boundaries of the insertion are 1.05%, respectively; Supplementary Table S3). Due to am- over-represented in glutamate and lysine (by 1.54 and bias is not seen in out the function of IF1 in mitochondria, such dramatic loss is responsible for the inability to accommodate quite large variations in size and sequence seemingly without perturbing protein function.

Aim23p is the orthologue of mitochondrial initiation factor mIF3

BlastP searches alone failed to identify mIF3 homologues from distantly related eukaryotes. However homologues from across the eukaryotic tree of life were identified with the more sensitive PSI-Blast (Supplementary Table S2). The PSI-Blast hits included previously identified mIF3s from human and S. pombe, chloroplast IF3s (cpIF3s), and a Saccharomyces protein called Aim23p/ YJL131C in S. cerevisiae (henceforth referred to as Aim23p). Aim23p has been identified in the S. cerevisiae mitochondrial proteome, and mutation of Aim23p results severe respiratory defect suggesting its importance for mitochondrial functionality (37–39). However, no specific function has previously been assigned or predicted for Aim23p.

Phylogenetic analysis of the full data set comprising IF3/mIF3/cpIF3 and Aim23p shows that the Aim23p sequences group with other fungi (Supplementary Figure S2). Although this has only weak support in the full IF3 tree (51% MLBP), phylogenetic analysis of a cut-down IF3 data set with the longest branched protist sequences included has very strong support for the monophyly of Aim23p and fungal mIF3s (1.0 BIPP, 98% MLBP; Figure 3). Thus, although the position of Aim23p within fungi has no significant support, it clearly groups with fungi, and the most parsimonious explanation is that Aim23p is in fact the previously unidentified mIF3 orthologue. The inability to find significant mIF3 hits across fungi using BlastP alone is probably due to the short length of IF3 and its highly biased amino acid composition. IF3 is enriched in charged amino acids, particularly lysine; E. coli IF3, human mIF3 and S. cerevisiae Aim23p contain 11.7, 10.4 and 14.9% lysine, respectively (Supplementary Table S3). Schizosaccharomyces pombe mIF3 is also surprisingly enriched in serine (12.4%). The bias in lysine composition in IF3 and the IF2 insertions may reflect the role of these proteins in binding negatively charged RNA; enrichment of lysine and other charged amino acids is a feature of ribosomal proteins (40).

Although we have identified the missing mIF3 orthologue in S. cerevisiae, and we find mIF3 is repre- sented in all major groups of mitochondriate eukaryotes, we could not identify mIF3 in a handful of species: alveolate Toxoplasma gondii, fungus Yarrowia lipolytica, platyhelminth Schistosoma mansoni and chromodorean nematodes Caenorhabditis briggsae, Caenorhabditis elegans and Brugia malayi (although a homologue was found in Enoplean nematode Trichinella spiralis; Figure 1). Kinetoplastida, on the other hand, have experienced a duplication of mIF3 (NCBI GI numbers in Supplementary Table S2). These paralogues are more similar to each other than to any other mIF3 orthologue, indicating that they are in-paralogues, i.e. the duplication occurred in the lineage to kineoplastida. The PSI-Blast
Figure 1. Phylogenetic tree of mitochondrial IF2 (mIF2), bacterial IF2 and chloroplast IF2 (cpIF2). The tree is a MrBayes consensus tree, generated from 405 aligned amino acids. The standard deviation of split frequencies at the end of the MrBayes run was 0.018. Bayesian inference posterior probability (BIPP) and maximum likelihood bootstrap percentage (MLBP) support are indicated on branches. Only branches with >0.70 BIPP support are labelled with BIPP and MLPP values. The scale bar below the tree shows the evolutionary distance expressed as substitutions per site. Numbers in taxon names are NCBI GI numbers. Vertical blocks show the distribution of IF1, mitochondrial IF1 (mIF1), chloroplast IF1 (cpIF1), IF3, mitochondrial IF3 (mIF3), Aim23p and chloroplast IF3 (cpIF3). Blocks in the same column indicate orthologous proteins. The black bracket shows the taxonomic boundary of the full-length conserved animal IF2 insertion.
Figure 3. Phosphoglycerin tree of mitochondrial IF3 (mIF3), Aim23p and bacterial IF3. The tree is a MrBayes consensus tree, generated from 156 aligned amino acids. The standard deviation of split frequencies at the end of the MrBayes run was 0.015. Branch support, GI numbers and substitutions per site are indicated as per Figure 1. The Aim23p clade is indicated with a dashed box.

Figure 2. Sequence alignment of the mIF2 insertion region. Example sequences from across the IF2 tree are aligned, with their major taxonomic groupings indicated. The domain structure of human mIF2 is shown above the alignment, with dotted lines indicating the location of the insertion.
results suggest kinetoplastida mIF3-1 and mIF3-2 are genuine mIF3 homologues, and TargetP predicts mitochondrial transit peptides. However, the sequences are too divergent to be reliably aligned and included in either the full or reduced mIF3 phylogenies.

IF3 is comprised of two domains joined by a flexible linker (41). NMR has shown that the ribosome interacting sites of bacterial IF3 are mainly found in the C-terminal domain (CTD), with fewer interacting sites in the N-terminal domain (NTD) (42). Sequence alignment shows that Aim23p, mIF3 and bacterial IF3 can be aligned and contain conserved positions across both the NTD and CTD (yellow columns, Figure 4A–C). A striking clustering of functionally important sites is found in and around helix H3 and sheet S6 of the CTD, regions shown with NMR and X-ray crystallography to interact with the ribosome (42,43). Several sites are conserved in Aim23p and other mIF3s in these regions (Figure 4A). The linker between the N and C domains is extended by an insertion of up to 19 amino acids in Aim23p (alignment positions 216–234; Figure 4A), and secondary structure prediction of S. cerevisiae Aim23p with PSIPred (44) suggests the insertion in this region extends the H2 helical region of the linker (Figure 4A). There are low complexity regions at both the extreme N and C termini of mIF3; N and C-terminal extensions have been previously reported in human mtIF3, relative to bacterial IF3 (Figure 4A). The extensions are poorly conserved in Aim23p and other mIF3s in these regions (Figure 4A). The linker between the N and C domains is extended by an insertion of up to 19 amino acids in Aim23p (alignment positions 216–234; Figure 4A), and secondary structure prediction of S. cerevisiae Aim23p with PSIPred (44) suggests the insertion in this region extends the H2 helical region of the linker (Figure 4A). There are low complexity regions at both the extreme N and C termini of mIF3; N and C-terminal extensions have been previously reported in human mtIF3, relative to bacterial IF3 (Figure 4A). The extensions are poorly conserved in primary sequence, but are rich in charged amino acids, particularly glutamate, glutamine, asparagine and lysine, similarly to the insertion in mIF2 (Figure 4A).

Plotting the sites that are conserved in Aim23p and mIF3 or bacterial IF3 onto the 3D structures of the NTDs and CTDs shows that these sites are largely buried, with the side chains oriented into the center of each domain (Figure 4B and C). This is particularly apparent in the CTD, where the conserved residues may be involved in intramolecular interactions to maintain the bundle shape.

Residues H170, D171 and K175 of human mIF3 (positions 253, 254 and 259, respectively in Figure 4A) have been found to be particularly important for the dissociation of mitochondrial 55S ribosomes (46). D254 is universally conserved in Aim23p and IF3, H253 is unconserved in Aim23p, and K259 is universally conserved as K in Aim23, with a conservative substitution to R in some mIF3s (Figure 4A). The positions D254 (D106 in E. coli) and K259 (K110 in E. coli) have also been implicated in ribosome interactions in bacteria (42,47,48). The functional importance of all arginine residues in the E. coli IF3 CTD has been determined (49). However, none of these arginines are conserved outside of bacteria, with the possible of exceptions of R112 (alignment position 262), which has the chemically similar lysine in this position in mIF3 (but not Aim23p), and R131, which has lysine at this position in both mIF3 and Aim23p (Figure 4A). Mutation of R112 on the surface of helix H3 severely disrupts ribosome binding and subunit dissociation in E. coli (49). R131 is part of an exposed loop (Figure 4A and C) and is proposed to contact the mRNA (49), as is supported by its involvement in start codon discrimination (50). Another critical residue of E. coli IF3, Y75 (51) is unconserved outside of bacteria (Figure 4A).

Validation of Aim23p as a bona fide mitochondrial IF3

To test whether Aim23p is a functional mIF3 in S. cerevisiae, we employed an in vivo complementation strategy. We obtained a haploid S. cerevisiae strain with the chromosomal AIM23 gene disrupted by a genetin (G418) resistance cassette by inducing sporulation in a heterozygous AIM23 knockout diploid yeast strain Y21294 (EUROSCARF). By assessing growth of the wild-type and AIM23A strains on fermentable (i.e. not requiring mitochondrial functionality) and non-fermentable (i.e. requiring mitochondrial functionality) media, we confirm the AIM23A respiration deficient phenotype reported in (37–39) (Figure 5A and B). This defect in growth on non-fermentable media is reversed by complementation with a plasmid encoding the gene for S. pombe mIF3 (S.p.MIF3) with S. cerevisiae 5' and 3' flanking regions driving its expression (Figure 5A and B).

To demonstrate that the AIM23 knockout indeed specifically affects mitochondrial functionality and that these effects are rescued by S. pombe mIF3, we analysed mitochondrial membrane potential of live yeast cells both by Rhodamine 123 staining followed by fluorescence microscopy and by DiOC6 staining followed by flow cytometry. In the AIM23A strain, unlike wild-type, Rhodamine 123 fluorescence signal is almost indetectable, indicating failure in the electron transport chain and resulting lack of membrane potential. Fluorescence signal is restored in the mIF3 complementation strain (Figure 5C). Similarly, mean DiOC6 fluorescence decreases in the AIM23A strain by around an order of magnitude in comparison to wild-type and over half an order of magnitude in comparison to the S.p.MIF3 complementation strain (Figure 5D). To assess effective oxygen uptake by all strains, oxygen consumption was followed using a Clark-type oxygen electrode. FCCP-induced oxygen consumption is severely affected in the AIM23A strain as compared to wild-type, and, again, this defect is largely (although not entirely) rescued by complementation with S.p.MIF3 (Figure 5E).

Importantly, oxygen uptake by the knockout strain was insensitive to inhibition of cytochrome oxidase by 0.5 mM potassium cyanide, unlike the wild-type and complemented strains (data not shown), suggesting that only the latter two display bona fide mitochondrial respiration. Finally, we tested binding of Aim23p in vitro using E. coli ribosomes and showed that, indeed, purified Aim23p forms a tight complex with the E. coli 30S ribosome, suggestive of at least partial function in this heterologous system (Supplementary Figure S3A). It was, however, not able to split the E. coli 70S ribosome (Supplementary Figure S3B).

Distribution of yeast mitochondrial translational activators

The S. cerevisiae translational activator Aep3p has been observed to interact with mIF2 genetically and physically,
suggesting it is a mitochondrial translation initiation accessory factor (21). Similarly, genetic screens suggest the Rsm28p protein of the same organism may have overlapping roles with mIF2, and is possibly performing IF1 or IF3-like functions in translation initiation (22,23).

**Figure 4.** IF3 NTD and CTD structures showing patterns of sequences conservation. IF3 subgroup consensus and example sequences are aligned in (A), with consensus sequences calculated at the 60% level using the Python script Consensus Finder (32). Yellow columns indicate those sites that are conserved in Aim23p, mIF3 and/or bacterial IF3, exposed sites of which are underlined in black. Green columns are those sites limited in conservation to Aim23p. The secondary structure of *E. coli* IF3 (41) is shown in red below the alignment, with ‘closing parenthesis’ representing helices and ‘greater than’ symbol representing sheets. Secondary structure in blue shows the extension of the linker helical region in *S. cerevisiae* Aim23p, as predicted with PSI-Pred symbols above the alignment show sites predicted to affect IF3 function. Dark blue squares: (42); red squares: (49); yellow triangles: (51); turquoise triangles: (48); green circles: (50); light blue circles: (47); pink circles: (46). The structure of (B) *Bacillus stearothermophilus* IF3 NTD [PDB entry 1TIF (41)], and (C) *Mus musculus* mIF3 CTD (NMR structure, PDB entry 2CRQ). Helix H3 is indicated, along with the universally conserved D106 (*E. coli* numbering), which is found at alignment position 254 in (A). As in (A), yellow residues indicate those sites that are conserved in Aim23p, mIF3 and/or bacterial IF3. The side-chains of these residues are also shown. Green residues are those limited in conservation to Aim23p. Blue residues show the location of insertions in Aim23p.
many initiation-associated activators that have been discovered using *S. cerevisiae* as a model system [Aep1p (52), Aep2p (53), Atp22p (54,55) Cbp1p (56), Cbp6p, Cbs1p, Cbs2p (57), Cbt1p (58,59), Mss51p (60), Mtf2p (61), Nca2p (62), Nca3p (63), Pet111p (64), Pet54p, Pet112p and Pet494p (65), Pet122p (66), Pet309p (67) and Rmd9p (23)]. As the distributions of these activators across eukaryotes have not been systematically addressed, we searched with PSI-Blast for these 21 activators to uncover their patterns of presence and absence (Supplementary Table S2).

Activators are often members of multi-protein families sharing a common domain (particularly the pentatricopeptide repeat (PPR) domain). Therefore, in some cases, the sequence searching identified distant paralogues in other groups of organisms. However, such paralogues were not included, in order to record only putative functional orthologues that form a distinct clade in their phylogenetic tree. Using these criteria, putative orthologues were only identified in fungi (Supplementary Table S2), but with differences in distributions within fungi in each protein. We find that Rsm28p, Pet309p, Pet494p, Pet122p, Cbs1p, Cbs2p and Cbt1p are limited to the Saccharomycetaceae family. Aep1p, Aep3p, Atp22p, Rmd9p, Pet54p and Pet111p have a wider distribution, being specific to the order *Saccharomycetes*. Rmd9p has duplicates in *Candida glabrata*, *S. cerevisiae* and *Vanderwaltozyma polyspora*, possibly as a result of the whole genome duplication in Saccharomycete evolution (68). Aep2p and Cbp1p have representatives across the phylum Ascomycota, although *Aspergillus* Cbp1ps are very different in sequence. Finally, Cbp6p, Nca2p, Mtf2p and Mss51p are found across the fungal kingdom (Supplementary Table S2). Nca3p is also

---

Figure 5. *S. pombe* mIF3 rescues a respiratory defect in an *S. cerevisiae* strain lacking the genomic copy of *AIM23*. (A and B) Ten-fold serial dilutions (made on the basis of measurement of OD600 of wild-type (‘wt’, Y21294a), *AIM23Δ* and *AIM23Δ + S.p.MIF3* strains were spotted on YPGly (A, non-fermentable) or YPD (B, fermentable) plates and incubated at 30°C for 4 and 1 days, respectively. (C) Differential interference contrast (DIC, left column) and fluorescence (right column) images of the yeast cells stained with mitochondrial membrane potential marker Rhodamine 123. (D) Flow cytometry analysis of DOOr6-stained wt (black trace), *AIM23Δ* (red trace) and *AIM23Δ + S.p.MIF3* (blue trace) yeast cells with the FITC-A detection channel. (E) Measurements of oxygen consumption using a Clark-type oxygen electrode. Oxygen consumption values (in pmoles O2 per minute per 10^6 cells) prior to FCCP treatment /after FCCP treatment /after treatment with 0.5 mM KCN are 49/92/2 for wt (black trace), 26/29/18 for *AIM23Δ* (red trace) and 29/58/2 for *AIM23Δ + S.p.MIF3* (blue trace).
found in many species of fungi, but this gene has experienced multiple duplications in fungi, and therefore was not included in the table.

Pet309, a member of the PPR protein family (67) was earlier identified to have a human homolog [LRPPRC, (69,70)]. However, sequence alignment and phylogenetic analysis suggest that while they are homologous, they are distant relatives, and may be paralogues. As such, the mammalian homologues are not included in Supplementary Table S2. Pet111p also has homologues in animals and may be a highly diverged orthologue of the nucleus-localized human RNA polymerase-associated-protein CTR9 homologue. Rsm28p is targeted to mitochondria in *Saccharomycetales*, while its closest homologue outside of this subset (aromatic amino acid aminotransferase Ar08/YGL202Wp) is seemingly cytoplasmic (38,71).

**DISCUSSION**

**IF2 functional evolution and loss of IF1**

Formation of the 70S initiation complex (IC) in bacteria requires the initiation factors IF1, IF2 and IF3. Together with IF3, IF1 inhibits the formation of initiator tRNA-less 70S ribosomes in *E. coli* by accelerating the binding of initiator tRNA to the 30S subunit and inhibiting joining of 50S subunits with initiator tRNA-less 30S subunits, thereby minimizing the abundance of initiator-tRNA lacking 70S ribosomes (7). IF1 plays an additional role in translation fidelity, increasing the accuracy of initiator tRNA selection over elongator aa-tRNAs (6). In addition to their role in initiation, IF1 and IF3 are also involved in ribosomal recycling (10–12).

Given all its crucial functions, it is unsurprising that IF1 is universal in bacteria, archaea (aIF1A) and the eukaryotic cytoplasm (eIF1A), as well as eukaryotic plastids (Figure 1) (72,73). However, it is not present in mitochondria, where mIF2 and mIF3 are able to (Figure 1) (72,73). In the most parsimonious scenario, the supposed secondarily serve as an additional anchoring point for mIF2, which could be the reason why in the *E. coli* system, bovine mtIF2 rescues the IF1-knock out strain (17).

The lack of IF1 raises the question of whether the mitochondrial translational system has a reduced requirement for classical IFs. Even in the absence of universal initiation factor mIF2, yeast still can translate a subset of mRNAs (16). We propose that this phenomenon may be the combined result of a reduction in the set of translated mRNAs, along with lineage-specific evolution of activators, specific for just one or two mRNAs. Such dedication of initiation-associated factors to specific mRNAs would not be feasible for a genome with a large protein-coding component. When the activators Aep3p and Rsm28p were found to affect mIF2 binding and dependence on initiator tRNA formylation for initiation (21,23), this raised the question of whether these proteins could be IF1 and/or IF3 functional analogues. We analysed the distribution of Aep3p, Rsm28p and other activator proteins and found that these particular means of regulating of initiation appear to be peculiar to distinct lineages. While it is likely that other mRNA-specific activators are also found in other organisms, the *S. cerevisiae* activators we searched for are fungi-specific, and mostly limited to *Saccharomycetales*.

**Identification and verification of the missing *Saccharomycetale* mIF3 (Aim23p)**

Through PSI-Blast, we have identified Aim23p as the previously undiscovered *S. cerevisiae* mIF3 orthologue. An earlier attempt to find the missing mIF3 gene identified Ygl159wp and not Aim23p as a possible homologue (19). Ygl159wp does not possess a mitochondrial transit peptide and has not been identified in mitochondrial proteomes. We find no sequence homology between Ygl159wp and mIF3, and searching Ygl159wp against the Pfam domain database suggests it, in fact, belongs to the Ornithine cyclodeaminase family, unrelated to IF3 (33).

While no specific function has previously been assigned to Aim23p, it has been identified in the *S. cerevisiae*
mitochondrial proteome, and deletion of Aim23p in high-throughput screening analyses results in abnormal mitochondrial genome maintenance and the absence of respiratory growth (37–39). Through in vivo complementation analyses, we show that S. pombe mIF3 can compensate for a deletion of Aim23p in S. cerevisiae (Figure 5), confirming that Aim23p is the functional, as well as evolutionary orthologue of mIF3. S. cerevisiae is a promising model system for the study of human mitochondrial disorders, as has recently been exemplified in a screen for compounds able to suppress the respiratory growth disorder caused by a defect in an ATP synthase assembly protein (79). A polymorphism in human mIF3 that results in reduced levels of the protein is associated with onset of Parkinson’s disease (80–82). With the identification of Aim23p as yeast mIF3, a similar screen could be carried out to find compounds that suppress the phenotypic effect of defects in Aim23p. However, while S. cerevisiae is a very promising model system, it is important to keep in mind the significant differences between the mammalian and yeast mitochondrial systems, such as the contribution of translational activators in yeast that are absent in animal lineages.

Lineage-specific duplications of classical IFs

Other branches of the eukaryotic tree of life have experienced their own remodeling of the set of classical IFs. Our sequence searching has identified duplications of mIF2 in alveolates and kinetoplastida (mIF2-2; Figure 1) and mIF3 in kinetoplastida (Supplementary Table S2). It is not clear why these protists would require multiple mitochondrial initiation factors. However, their mitochondrial translational apparatuses have several distinctive features. Many alveolates and kinetoplastida translate as few as three proteins, have fragmented rRNA genes that are assembled post-transcriptionally, and import all or most tRNAs from the cytoplasm, possibly in their aminoacylated form in vivo (9S and 12S ribosomal RNAs in the case of Trypanosoma brucei) and have experienced a dramatic expansion in ribosomal protein content (84,85). Recent investigations of gene duplications of mitochondrial EF-Tu (86) and EF-G (32,87) have shed light on the subfunctionalization (i.e. functional divergence of two copies of a duplicated multifunctional protein) of these core components of translational machinery, and similar investigations of mIF2 and mIF3 would be very interesting. Indeed, both of these proteins are multifunctional; in addition to their roles in initiator tRNA recruitment to the ribosome, IF2 promotes subunit docking to the pre-initiation complex, while IF3 promotes premature subunit association in the recycling step of protein synthesis (3–11).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables S1–S3, Supplementary Figures S1–S3, Supplementary Text S1: SI Methods and Supplementary References [19,24–26,29,88–94].

ACKNOWLEDGEMENTS

Many thanks to Ivan Tarassov for providing yeast strain Y21294, Fedor Severin for the opportunity to use the micromanipulator for yeast tetrad dissection and Arvi Jõers and Antonio Caballero Reyes for advice in flow cytometry. V.H. and G.C.A. conceived of this study and GCA carried out sequence searching and phylogenetic analyses. G.C.A. and V.H. drafted the article with input from A.K. and T.T. T.T. helped draft the article and participated in the coordination of this study. A.S. performed in vitro 30S binding and 70S splitting experiments. A.K., P.K., V.L., M.Y.V., S.T. and E.S. performed in vivo complementation experiments. All authors read and approved the final article.

FUNDING

The European Social Fund (MJD99, Mobilitas grant to G.C.A.); Estonian Science Foundation (ETF7616 and ETF9012 to V.H., ETF9020 to G.C.A.); the European Regional Development Fund through the Center of Excellence in Chemical Biology (code RLOTITIPP to V.H. and T.T.); Russian Foundation for Basic Research, Ministry of Education and Science of Russia, the President of Russian Federation Grant for Young Researchers and funds from International Associated Laboratory ‘RNA-mitocure’ (to P.K., A.K., V.L. and E.S.); SA Archimedes foundation, European Social Fund and DoRa program (to A.K. and S.T.); U.M.N.I.K. program (E.S. and A.K.); ANR (Agence Nationale de la Recherche to M.V.). Funding for open access charge: Estonian Science Foundation grant of G.C.A. and the Center of Excellence in Chemical Biology.

Conflict of interest statement. None declared.

REFERENCES

from the mitochondrial COX1 locus of Saccharomyces cerevisiae. EMBO J., 14, 4031–4043.


