doi:10.1016/j.jmb.2007.09.051 *J. Mol. Biol.* (2007) **374**, 837–863

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Exploring the Mitochondrial Proteome of the Ciliate Protozoon *Tetrahymena thermophila*: Direct Analysis by Tandem Mass Spectrometry

Daryl G. S. Smith[1]†, Ryan M. R. Gawryluk[2]†, David F. Spencer[2] Ronald E. Pearlman[3], K. W. Michael Siu[1] and Michael W. Gray[2]*

[1]*Department of Chemistry and Centre for Research in Mass Spectrometry, York University Toronto, Ontario, Canada*

[2]*Department of Biochemistry and Molecular Biology Dalhousie University, Halifax Nova Scotia, Canada*

[3]*Department of Biology and Centre for Research in Mass Spectrometry, York University Toronto, Ontario, Canada*

Received 10 July 2007; received in revised form 18 September 2007; accepted 19 September 2007
Available online
22 September 2007

To date, direct analysis of mitochondrial proteomes has largely been limited to animals, fungi and plants. To broaden our knowledge of mitochondrial structure and function, and to provide additional insight into the evolution of this key eukaryotic organelle, we have undertaken the first comprehensive analysis of the mitochondrial proteome of a protist. Highly purified mitochondria from *Tetrahymena thermophila*, a ciliated protozoon, were digested exhaustively with trypsin and the resulting peptides subjected to tandem liquid chromatography-tandem mass spectrometry (LC/LC-MS/MS). In this way, we directly identified a total of 573 mitochondrial proteins, 545 of which are encoded by the nuclear genome and 28 by the mitochondrial genome. The latter number includes a novel, 44 residue protein (which we designate Ymf78) that had not been recognized during annotation of the *T. thermophila* mtDNA sequence. The corresponding gene, *ymf78*, is highly conserved in genomic position, size and sequence within the genus *Tetrahymena*. Our analysis has provided broad coverage of both membrane-bound and soluble proteins from the various submitochondrial compartments, with prominent representatives including components of the tricarboxylic acid cycle, Complexes I–IV of the electron transport chain and Complex V (ATP synthase), the mitochondrial transcription and translation machinery, the TOM and TIM protein translocases, various mitochondrial transporters, chaperonins (Cpn60, Hsp70, Hsp90), at least four FtsH family ATP-dependent metalloproteases implicated in *m*-AAA and *i*-AAA protease function, and enzymes involved in lipid, amino acid and coenzyme metabolism, as well as iron–sulfur cluster formation. Unexpectedly, six of the ten enzymes of glycolysis were found by MS analysis of purified *T. thermophila* mitochondria, whereas no hits were seen to any cytosolic ribosomal proteins. At least one of the glycolytic proteins, enolase, has an evident N-terminal extension that exhibits characteristics of a typical mitochondrial targeting peptide. As in other organisms, phylogenetic analysis of functionally annotated mitochondrial proteins demonstrates that <20% can be traced confidently to the α-proteobacterial lineage of Bacteria, emphasizing the chimeric evolutionary nature of the mitochondrial proteome. Notably, about 45% of the proteins identified in our analysis have no known function, and most of these do not have obvious homologs outside of the ciliate lineage. About two-thirds of these ORFan proteins have putative homologs in another ciliate, *Paramecium tetraurelia*, whereas the remainder appear to be *Tetrahymena*-specific. These results emphasize the power and importance of direct MS-based analysis of mitochondria in revealing novel mitochondrial proteins in different

**Corresponding author.* E-mail address: m.w.gray@dal.ca.

† D.G.S.S. and R.M.R.G. contributed equally to this work.

Abbreviations used: LC, liquid chromatography; MS, mass spectrometry; TGD, *Tetrahymena* Genome Database; ORF, open reading frame; MPP, Mitochondrial processing peptidase; IMP, intermediate processing peptidase.

eukaryotic lineages. Our observations reinforce an emerging view of the mitochondrion as an evolutionarily flexible organelle, with novel proteins (and presumably functions) being added in a lineage-specific fashion to an ancient, highly conserved functional core, much of which was contributed by the presumptive α-proteobacterial symbiont from which the mitochondrial genome was derived.

Edited by J. Karn

Keywords: mitochondria; proteome; mass spectrometry; protist; evolution

Introduction

Although the origin and evolution of mitochondria have been investigated intensively from the perspective of the mitochondrial genome,[1,2] we still have a rudimentary and narrowly focused knowledge of the proteins that comprise the mitochondrial proteome, the vast majority of which are encoded by the nuclear genome, synthesized in the cytosol and targeted to this key eukaryotic organelle.[3] Using genomic data, several bioinformatics tools are available to predict mitochondrion-directed gene products,[4,5] by screening for putative mitochondrial targeting sequences. Such predictions suggest that functional mitochondria could harbor as few as several hundred proteins in *Plasmodium falciparum*, the malaria parasite, to >3000 in vertebrate animals.[6] Conspicuously absent from these surveys are non-parasitic protists, in large part because relatively few complete nuclear genome sequences are available for these mostly unicellular eukaryotes, even though protists constitute the bulk of the phylogenetic diversity within the eukaryotic domain (Eucarya). Although reconstruction of proto-mitochondrial metabolism suggests that a core of mitochondrial proteins is common to all eukaryotes containing this organelle,[7] we have little comprehension on a phylogenetically broad scale of how variable the mitochondrial proteome may actually be, and what the evolutionary origins were and the current functions are of the additional proteins in larger mitochondrial proteomes.

In a bioinformatics approach based on identification of N-terminal targeting sequences, a major limitation is that the accuracy of the algorithms used is variable,[6,8] and may have limited sensitivity in cases of highly divergent mitochondrial protein sequences. In addition, it is not clear that the "rules" for mitochondrial protein targeting on which these algorithms are based are broadly applicable throughout Eucarya. As well, these methods obviously fail in those cases where mitochondrial proteins lack N-terminal targeting sequences,[3,9] although approaches based on other distinguishing features of mitochondrial proteins have been developed; e.g. MITOPRED.[10] With all of these *in silico* programs, there is a trade-off between sensitivity and specificity in the predictions they make. Accordingly, even with complete genome sequences, direct analyses of mitochondria are ultimately required to establish unequivocally and comprehensively the composition of the mitochondrial proteome in different organisms.

Subcellular proteomics[11,12] combines direct analysis by mass spectrometry (MS)[13–17] of isolated organelles such as mitochondria,[18] with bioinformatics and subcellular localization techniques. This combined approach brings unprecedented sensitivity to the definition of complete organellar proteomes,[19] including the identification of novel proteins of unknown function. In the case of the budding yeast *Saccharomyces cerevisiae*, recent analyses estimate the number of mitochondrial proteins to range from ~400 to ~800,[20–24] or between ~7% and ~13% of the ~6100 proteins that comprise the total proteome.

About 50–60% of the yeast mitochondrial proteome is classified as prokaryote-specific, comprising proteins having homologs in Bacteria and Archaea, but only a small fraction of this category can be traced with confidence to the α-proteobacterial ancestor of mitochondria.[25] Another portion (~20–30%) of the yeast mitochondrial proteome is eukaryote-specific, i.e. proteins that have identifiable homologs only in other eukaryotes. About 20% of yeast mitochondrial proteins are unique. These data indicate that the mitochondrial proteome has multiple evolutionary origins, and a complex evolutionary history.[26,27]

To date, MS-based mitochondrial proteomics projects have largely focused on multicellular organisms for which complete genome information is available.[22,28–34] In the study reported here, we undertook a direct mass spectrometric analysis of the mitochondrial proteome of *Tetrahymena thermophila*, a ciliated protozoon. Both mitochondrial[35] and nuclear[36] genome sequences for this organism have been determined, a large EST data set is available for comparison, and highly purified mitochondria may be obtained in quantity from cultures grown axenically. Moreover, *T. thermophila* is amenable to a variety of biochemical and genetic techniques, including gene knockout and gene replacement,[37,38] that offer the possibility of confirming subcellular localization and examining gene function. Finally, the genome sequence of a ciliate relative of *T. thermophila*, *Paramecium tetraurelia*, is available,[39] which provides a further framework for exploring the phylogenetic distribution of novel mitochondrial proteins from *T. thermophila*. Here, we present the results of the most comprehensive analysis to date of the mitochondrial proteome of a protist, and we

discuss the implications of these results for our further understanding of mitochondrial evolution, biogenesis and function.

Results and Discussion

Mass spectrometric identification of *Tetrahymena thermophila* mitochondrial proteins

Nucleus-encoded proteins

The peptide data derived from MS analysis of isolated mitochondria allowed us to identify with high confidence sequences that mapped to 568 preliminary gene predictions (PGPs, designated PreTt at TGD and in accompanying tables; see below) for the macronuclear genome of *T. thermophila* (see *Tetrahymena* Genome Database, TGD)‡. We divided these nucleus-encoded entries (Supplementary Data Table 1) into two categories: (A) 385 identified by more than one peptide, and (B) 183 identified by one exceptional peptide (i.e. one whose fragmentation pattern in the MS/MS spectrum is especially well categorized, as confirmed by manual inspection). We have estimated the probability of false positives in these categories to be 0% and ~1%, respectively, and we therefore consider these MS identifications to be definitive. These estimates were made by searching the MS/MS spectra against a concatenated database comprising the predicted protein sequences and a decoy of the reverse sequences.[40,41] Because these estimates likely overstate the extent of false-positive assignments, the very low false-positive rates we obtained for the A and B categories are confirmation of the robustness of our analyses.

An additional 121 hits that we classify as tentative (category C) are represented in the analysis by one acceptable peptide: i.e. a peptide that exceeds Mascot thresholds and shows some significant (≥3 consecutive) y-ion coverage, but is otherwise not exceptionally convincing. However, because we estimated the probability of false positives in this category to be very high (searching against the concatenated target/decoy database returned a high false positive estimate of ~70%), these identifications are not robust and so are not considered further here, with the exception of eight entries that are very likely mitochondrial proteins on the basis of BLAST analysis and annotation.

A total of 17 nucleus-encoded proteins in categories A and B, and two in category C are represented by two consecutively numbered PGPs that proved to comprise a single open reading frame (ORF). In 13 cases, both constituent PGPs were identified independently in the MS analysis, which reduces the actual number of identified proteins by 13 (to the 555 nucleus-encoded entries listed in Supplementary Data Table 1 in categories A and B, plus an additional eight in category C, for a total of 563).

No peptide corresponding to any *Tetrahymena* cytosolic ribosomal proteins was identified in this analysis. Mitochondrion-bound cytosolic ribosomes have been demonstrated in some organisms,[42,43] and ribosomes bound to fragments of endoplasmic reticulum (rough ER) usually contaminate crude mitochondrial fractions. Moreover, cytosolic ribosomal proteins are among the most abundant and highly conserved of cellular proteins; so, if present in the MS analysis, they should be readily recognized as contaminants, especially as we detected a number of *bona fide Tetrahymena* mitochondrial ribosomal proteins (see below). Failure to detect any cytosolic ribosomal proteins at all argues strongly that contamination with cytosolic or rough ER-associated polysomes is minimal in our mitochondrial preparations.

On the other hand, we did find 18 proteins previously identified in proteomic analyses of isolated cilia[44] and/or phagosomes[45] from *T. thermophila* (Supplementary Data Table 2). Although some of these cilia and phagosome proteins might actually be authentic mitochondrial proteins arising from mitochondrial contamination of the organelle preparations analyzed in these previous studies, as was seen in the case of dense core granule preparations from *T. thermophila*,[46] our finding is consistent with trace amounts of cilia fragments evident in electron micrographs of isolated *Tetrahymena* mitochondria (Figure 1). Excluding these presumptive cilia and phagosome proteins from the list of identified mitochondrial proteins gives a total

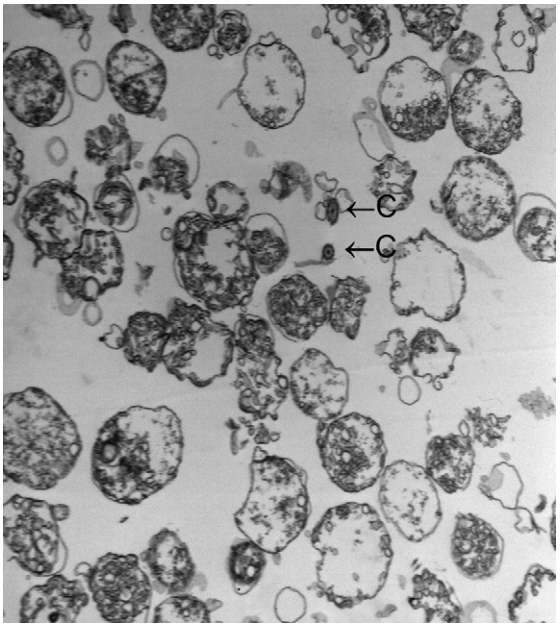

Figure 1. Thin-section electron micrograph of a highly purified *T. thermophila* mitochondrial preparation. Arrows point to contaminating fragments of cilia (C) in the final mitochondrial fraction.

‡ http://www.ciliate.org/

of 545 (363 in category A, 174 in category B, 8 in category C).

Mitochondrion-encoded proteins

The mitochondrial genome of *T. thermophila* is predicted to encode 44 proteins,[35] of which 25 have been assigned putative functions; separate N-terminal and C-terminal portions of Nad1 (Nad1_a and Nad1_b, respectively), which are independently encoded and expressed in characterized ciliate mitochondrial genomes,[47,48] are considered here as a single protein. The remaining 19 ORFs are of unknown function. An earlier analysis of the sequenced mtDNA of *T. pyriformis* had identified 22 ORFs (designated *ymf56* to *ymf77*) of unknown function, 13 of which are found also in the mtDNA of *P. tetraurelia*,[49] and were designated ciliate-specific.[47] All of the genes identified in *T. pyriformis* mtDNA are present also in the *T. thermophila* mitochondrial genome, with the study reported by Brunk *et al.* providing evidence that *ymf58*, *ymf60* and *ymf62* correspond to *nad4L*, *rpl6* and *nad6*, respectively.[35] Of the remaining 19 unidentified Ymf proteins, analysis of their predicted physicochemical properties suggested that five are likely highly diverged ribosomal proteins, whereas 14 are probably non-ribosomal.[35]

In the present study, we identified 15 of the 25 functionally annotated mtDNA-encoded proteins (60%); these include subunits of Complex I (Nad1_a,4,5,6,7,9,10), Complex III (Cob) and Complex IV (Cox1,2), as well as small subunit (S12,13) and large subunit (L2,14,16) ribosomal proteins. In addition, we found 12 of the 19 unidentified Ymf proteins (63%), including four of the five that were tentatively assigned to the mitochondrial ribosome. Overall, we recovered 27 out of a possible 44 mtDNA-encoded proteins (61%) in this proteomic analysis (Table 1).

One MS peptide mapped to a region between genes encoding Ymf76 and Ymf66. Inspection of this intergenic region revealed a previously unannotated 44 residue ORF, in the same transcriptional orientation as *ymf76* and *ymf66* (Figure 2). This ORF, which we designate *ymf78*, is highly conserved in position, length and sequence in the mtDNA of various species of the genus *Tetrahymena*. A 48 codon ORF encoding the homologous protein is present in *P. tetraurelia* mtDNA, but in this case the corresponding *ymf78* gene is flanked by *cox1* upstream and by *nad1_a* downstream. Figure 2 presents an alignment of the predicted amino acid sequence of this novel mitochondrial protein, which to date we have not identified beyond ciliates.

Table 1. MS/MS-identified mitochondrial proteins encoded by *T. thermophila* mtDNA

No.	Protein	GenBank identifier	Ion score sum	Unique/total peptides
D-1[a]				
564	Nad7 (NADH dehydrogenase subunit 7)	gi|15011477|gb|AAK77572.1|	548	11/23
565	Cob (apocytochrome *b*)	gi|15011497|gb|AAK77592.1|	87	2/5
566	Cox1 (cytochrome *c* oxidase subunit 1)	gi|15011504|gb|AAK77599.1|	78	3/3
567	Ymf77	gi|15027660|ref|NP_149394.1|	75	3/3
568	Cox2 (cytochrome *c* oxidase subunit 2)	gi|15027663|ref|NP_149397.1|	164	4/4
569	Ymf68	gi|15011502|gb|AAK77597.1|	151	5/6
570	Nad5 (NADH dehydrogenase subunit 5)	gi|15011498|gb|AAK77593.1|	158	2/8
571	Ymf67	gi|15027665|ref|NP_149399.1|	72	2/2
572	RpS13 (ribosomal protein S13)	gi|15027633|ref|NP_149367.1|	61	2/2
573	Ymf63	gi|15027648|ref|NP_149382.1|	38	2/2
574	RpL2 (ribosomal protein L2	gi|15011472|gb|AAK77567.1|	82	2/2
575	Ymf76	gi|15011468|gb|AAK77563.1|	82	3/3
576	Nad9 (NADH dehydrogenase subunit 9)[b]	gi|15011494|gb|AAK77589.1|	212	4/7
577	Nad10 (NADH dehydrogenase subunit 10)	gi|15027638|ref|NP_149372.1|	105	2/2
578	Nad6 (NADH dehydrogenase subunit 6)[c]	gi|15011506|gb|AAK77601.1|	99	2/2
579	Ymf73	gi|15011509|gb|AAK77604.1|	63	2/2
580	Ymf57	gi|15027630|ref|NP_149364.1|	171	3/4
581	–	–	57	2/2
D-2[a]				
582	Ymf66	gi|15027631|ref|NP_149365.1|	18	1/1
583	Nad4 (NADH dehydrogenase subunit 4)	gi|15011508|gb|AAK77603.1|	37	1/1
584	Nad1_a (NADH dehydrogenase subunit 1a)	gi|15027669|ref|NP_149403.1|	55	1/1
585	Ymf65	gi|15011487|gb|AAK77582.1|	41	1/1
586	Ymf61	gi|15027654|ref|NP_149388.1|	36	1/1
587	RpS12 (ribosomal protein S12)	gi|15027639|ref|NP_149373.1|	33	1/1
588	RpL14 (ribosomal protein L14)	gi|15027671|ref|NP_149405.1|	26	1/1
589	RpL16 (ribosomal protein L16)	gi|15027652|ref|NP_149386.1|	21	1/1
590	Ymf59	gi|15027650|ref|NP_149385.1|	15	1/1
591	Ymf56	gi|15027664|ref|NP_149398.1|	37	1/1

[a] Two or more unique tryptic peptides identify proteins in category D-1, whereas a single exceptional peptide identifies proteins in category D-2.
[b] Nad9 is encoded by duplicate genes in the *T. thermophila* mitochondrial genome.[35]
[c] Ymf62 has been identified as NADH dehydrogenase subunit 6.[35]

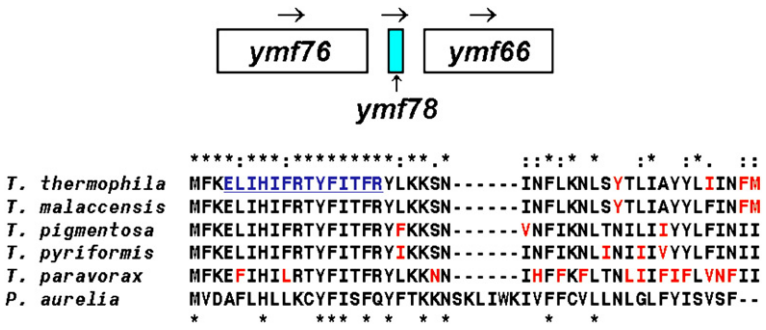

Figure 2. Organization and transcriptional orientation (horizontal arrows) of the *ymf78* gene (turquoise rectangle) in the mitochondrial genome of the listed *Tetrahymena* species and CLUSTALW alignment of Ymf78 protein sequences from *Tetrahymena* species and a related ciliate, *Paramecium aurelia*. The degree of conservation for the five *Tetrahymena* species is shown above the alignment (*, identical residue in all sequences; :, highly conserved column; ., weakly conserved column). Residues that differ from the consensus for that position are highlighted in red. The diagnostic MS peptide is depicted in blue and underlined. Asterisks below the alignment mark positions that have the same amino acid in all five *Tetrahymena* species and *P. aurelia*.

Altogether, we identified 545 nucleus-encoded proteins (excluding presumptive cilia and/or phagosome contaminants; see above) and 28 mitochondrion-encoded proteins, for a combined total of 573. If one assumes, based on the recovery of expected mitochondrial proteins, that we have identified ~60% of the proteins that comprise the *T. thermophila* mitochondrion, we can estimate that the mitochondrial proteome in this organism probably consists of ~900–1000 proteins. This number should be considered a minimal estimate, in view of the likelihood that the nucleus-encoded fraction of the proteome contains substantially more low-abundance proteins than the fraction encoded by mtDNA.

Analysis and correction of predicted protein sequences

Before further analysis, we screened PGPs corresponding to the proteins identified here against available EST data§. This analysis completely (Y) or partially (N) confirmed 40% of the PGPs (225 sequences; Supplementary Data Table 1). For 36% of the initial PGPs (203 in total), the predicted sequence was found to be incorrect due to misassignment of 5′ and/or 3′ intron-exon splice junctions, leading to intron retention, creation of spurious in-frame insertions, and/or exclusion of exons. We were able to infer correct, complete sequences for 134 proteins (Y* entries) and to make partial corrections for the remaining 71 (N*; in the latter category, the EST data did not encompass the entire sequence). For 133 sequences (X* and X entries in Supplementary Data Table 1), insufficient information is available to allow independent authentication of the current PGP. In most cases, correction of the PGP data did not change the associated annotation deduced by BLAST analysis.

Subsequently, we re-screened the corrected protein sequences against the peptide data, which yielded additional informative peptide hits. Supplementary Data Table 3 lists the confirmed and corrected mitochondrial protein sequences identified in our MS/MS analysis, with the initial tryptic peptide hits highlighted in red, and additional peptides recovered by the re-screening procedure denoted in blue.

§ http://tbestdb.bcm.umontreal.ca/searches/login.php; http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=t_thermophila

Functional categories

Of the 545 nucleus-encoded mitochondrial proteins identified here, functional assignments (based mainly on BLAST results) could be made for 297 (55%); the remaining 248 (45%) are unannotated. A few of the latter proteins are conserved in other eukaryotes, but the vast majority (90%) are either unique to *T. thermophila* or have homologs only in other ciliates, in which case we deem them "ciliate-specific" (categories A-2 and B-2 in Supplementary Data Table 1).

Figure 3 shows that the majority of the functionally annotated proteins are associated with energy cycle, catalysis/cofactors, and protein binding/transport, with these three categories together accounting for >80% of the annotated entries. Most of the identified proteins listed in Supplementary Data Table 1 are known to be mitochondrial in

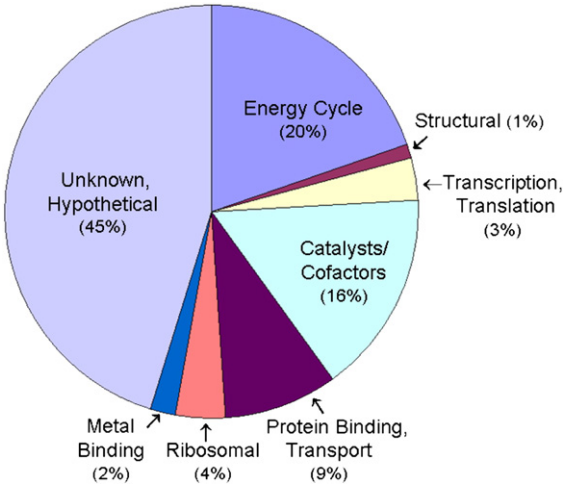

Figure 3. Proportion of individual functional groups within the total nucleus-encoded, MS-identified mitochondrial proteome (545 proteins) of *T. thermophila*.

localization and function in other organisms. Prominent among these are components of the tricarboxylic acid cycle; proteins of the electron transport chain (Complexes I–IV) and ATP synthase (Complex V) that carry out coupled oxidative phosphorylation; components of the mitochondrial transcription and translation machinery, including many mitoribosomal proteins; subunits of the TOM and TIM translocases that mediate protein import into mitochondria; mitochondrial carrier proteins; chaperonins (Cpn60, Hsp70, Hsp90); at least four FtsH family ATP-dependent metalloproteases; and enzymes involved in lipid, amino acid and coenzyme metabolism, as well as iron–sulfur cluster formation. Both soluble (matrix and intermembrane) and membrane-bound proteins are well represented in the list of identified mitochondrial proteins.

In the following sections, we consider in more detail the protein components of several key mitochondrial biosynthetic pathways and functional categories.

Mitochondrial carrier proteins

Our MS analysis identified 23 proteins annotated as mitochondrial carrier proteins; genes encoding an additional 30 proteins of this class were identified by bioinformatics analysis of the genome sequence. The total number (53) approximates the reported number (45) of mitochondrial carrier proteins in *Arabidopsis thaliana*,[50] although we detected a substantially larger number of mitochondrial carrier proteins in our MS analysis than did the *Arabidopsis* study. A similar number (48) of mitochondrial carrier proteins is likely encoded by the human genome.[51] Because the diversity of mitochondrial carrier proteins is thought to reflect the metabolic capacity and diversity of mitochondria,[51] it is notable that in a eukaryotic microbe such as *Tetrahymena*, the number and variety of such transporters is comparable to what is seen in multicellular eukaryotes.

Putative transport specificities, given in Table 2, include a range of metabolites; in particular, at least 11 adenine nucleotide transporters (ADP/ATP) are encoded by the *T. thermophila* nuclear genome. Two of these, PreTt01834 and PreTt25833, are identical in sequence and presumably represent duplicated genes, although PreTt01834 is truncated at the N terminus (127 amino acid residues shorter than PreTt25833) and the corresponding genes are not closely linked in the genome. On the other hand, PreTt03158 and PreTt03159 are highly similar (89% amino acid identity) phosphate transporters encoded by adjacent genes ~2 kb apart: again, a situation presumably reflective of a gene duplication event.

Notably, five mitochondrial carrier proteins, four of them identified in the MS analysis, display the highest level of amino acid identity with H^+ transporters (uncoupling proteins, UCPs) from animals and plants. UCPs are involved in thermogenesis and in the maintenance of redox balance and reduction

Table 2. Mitochondrial carrier proteins identified in *T. thermophila*

Metabolite transported (predicted)	Number (Supplementary Data Table 1)	PreTt
A. *Identified by proteomic analysis (MS/MS)*		
ADP/ATP	78	11411
Phosphate	99	03159
ADP/ATP	100	13437
ADP/ATP	101	14498
Phosphate	103	03158
ADP/ATP	113	23795
H^+ (uncoupling protein)	115	22438
Citrate	126	22715
Oxoglutarate/malate	145	16898
ADP/ATP	166	16838
Phosphate	170	18231
Oxoglutarate/malate	173	05157
ADP/ATP	188	09321
H^+ (uncoupling protein)	222	26420
Oxoglutarate/malate	247	09391
Deoxynucleotide	381	03141
H^+ (uncoupling protein)	438	20788
Carnitine/acylcarnitine	439	27752
Phosphate	447	22402
H^+ (uncoupling protein)	556a/b	04299/ 04300
Oxoglutarate/malate	557	18472
Deoxynucleotide	558	18721
Carnitine/acylcarnitine	560a/b	24642/ 24643
B. *Identified by genomic analysis (BLASTp)*		
ADP/ATP		07780
ADP/ATP		01834
ADP/ATP		19774
ADP/ATP		25833
ADP/ATP		18528
Carnitine/acylcarnitine		07127
Carnitine/acylcarnitine		24644
Folate[a]		23573
Folate		27014
Folate		14396
Iron		10054
NAD^+		02258
Oxoglutarate/malate		29306
Phosphate		28710
Phosphate[b]		28932
Phosphate		02267
H^+ (uncoupling protein)		21383
No consensus		15796
No consensus		04736
No consensus		03141
No consensus: *S*-AdoMet or phosphate		07707
No consensus: *S*-AdoMet or succinate/fumarate		04585
No consensus		03023
No consensus		24246
No consensus: *S*-AdoMet or phosphate		26061
No consensus		09934
No consensus		24169
No consensus: NAD^+ or folate		09235
No consensus: *S*-AdoMet or aspartate/glutamate		08597
No consensus		21910

[a] Possibly peroxisomal.
[b] Similar to Grave's Disease autoantigen.

in the generation of reactive oxygen species.[52] Other than Amoebozoa,[53] evidence for the presence of UCPs has not been reported in protists.

Iron–sulfur-protein biogenesis

Iron–sulfur (Fe–S) cluster formation[54] is increasingly recognized as an indispensable function of mitochondria,[55] with evidence of enzymes involved in Fe–S-protein biogenesis even in hydrogenosomes and mitosomes,[56,57] mitochondrion-related organelles that lack DNA and typical mitochondrial functions such as oxidative phosphorylation. In mitochondria, Fe–S protein biogenesis is carried out by the Fe–S cluster assembly pathway comprising 11 proteins.[54] We identified homologs of six of these proteins in our MS analysis of *T. thermophila* mitochondria, with homologs of the remaining five yeast proteins identified by genomic analysis (Table 3).

In yeast, maturation of extra-mitochondrial Fe–S-proteins requires the Fe–S cluster export machinery, comprising Atm1 (an ABC transporter protein) and Erv1.[54] Somewhat surprisingly, only a single ABC transporter protein was identified in our MS analysis (Table 3) out of 147 annotated ABC transporter family proteins among the *T. thermophila* PGPs. Two homologs of Erv1 were identified by genomic analysis (Table 3), neither of which was detected by MS.

AAA+ family proteins

The widespread and functionally diverse AAA+ family (ATPases associated with various cellular activities) comprises proteins that are able to induce conformational changes in a broad range of substrate proteins.[58] Among the most important and ubiquitously distributed members of this class are mitochondrial inner membrane AAA+ proteases that recognize and degrade non-assembled or excess membrane proteins: the *m*-AAA+ protease, active within the matrix and comprising (in *S. cerevisiae*) multiple copies of homologous ATP-dependent

Table 3. *T. thermophila* homologs of yeast proteins involved in Fe–S cluster formation[a]

Yeast homolog	PreTt	No.[a]	TGD automatic annotation	E-value (BLASTp)
A. *ISC-assembly machinery*				
Nfs1 (cysteine desulfurase)	16358		Aminotransferase, class V family protein	1e-26
	13618		Aminotransferase, class V family protein	1e-85
Isu1/Isu2	02137	433	NifU-like N-terminal domain-containing protein	3e-49
Isa1	28026		Iron–sulfur cluster assembly accessory protein	5e-27
Yah1 (ferredoxin)	28574		Ferredoxin, 2Fe-2S, putative	1e-22
Arh1 (ferredoxin reductase)	22600[b]		Conserved hypothetical protein[a]	4e-56
Yfh1 (frataxin)	07836	464	Hypothetical protein TTHERM_00989460	3e-06
Nfu1	26839		NifU-like domain-containing protein	3e-24
Grx5 (glutaredoxin)	09308		Glutaredoxin-related protein	3e-12
	03016	379	Glutaredoxin-related protein	1e-11
	29301		Glutaredoxin-related protein	3e-11
Ssq1 (Hsp70)[c]	04604	1a	Hypothetical protein TTHERM_01014740	1e-140
	04605	1b	DnaK protein	
Jac1	11816		Co-chaperone Hsc20 family protein	6e-08
	08674		Hypothetical protein	3e-04
	21083	21	DnaJ C-terminal region family protein	6e-04
Mge1	16425	169	Co-chaperone GrpE family protein	6e-24
B. *ISC-export machinery*				
Atm1 (ABC transporter)	18156[d]		ABC transporter family protein	6e-76
	05951[e]	17	ABC transporter family protein	7e-37
Erv1[f]	08946		Erv1/Alr family protein	2e-24
	04264		Erv1/Alr family protein	8e-13

Yeast (*S. cerevisiae*) proteins were used as queries in a BLASTp search of *T. thermophila* PGPs. Reciprocal BLASTp as well as the presence of diagnostic conserved domains confirmed the assignments.

[a] See Supplementary Data Table 1.

[b] Chimeric PGP: only the C-terminal 491 residues of this 1040 residue ORF represent an Arh1 homolog (see gi_146143186; gb_EDK31272.1).

[c] Ssq1 is an Hsp70 isoform,[114] the result of a fungal-specific Hsp70 gene duplication.[115] One additional *T. thermophila* Hsp70 (PreTt13040) may be localized in mitochondria, out of a total of 13 annotated Hsp70/DnaK homologs in TGD.

[d] *T. thermophila* protein exhibiting highest BLASTp similarity to yeast Atm1 among *T. thermophila* ABC transporters annotated in TGD; likely mitochondrial in localization. Among conserved domains retrieved using ABC transporters as query sequences in a BLASTp search at NCBI, PreTt18156 uniquely displayed highest E-values with cd03253 ('ABCC_ATM1_transporter') and COG5265 ('ATM1').

[e] Sole ABC transporter protein identified by MS/MS analysis of isolated *T. thermophila* mitochondria.

[f] In a BLASTp search, yeast Erv1 (mitochondrial) and Erv2 (endoplasmic reticulum) both retrieve the same two *T. thermophila* PGPs, PreTt08946 and PreTt04264, with the Erv1 E-values slightly lower compared to Erv2 in the case of PreTt08946 (2e-24 *versus* 5.8e-17) but slightly higher in the case of PreTt08946 (8e-13 *versus* 2.0e-16). PreTt08946 and PreTt04264 are both seen to contain the same conserved domain (pfam04777.8, Erv1_Alr), suggesting they are mitochondrial intermembrane proteins. PreTt08946 (187 residues) is about the same size as yeast Erv1 (189) and Erv2 (196), whereas the much larger PreTt04264 (659 residues) contains an additional (N-terminal) cNMP-binding domain (pfam00027). The two ciliate proteins are highly divergent in sequence, hence phylogenetic analysis is not informative. Lacking direct proteomic evidence of mitochondrial localization and in the absence of definitive bioinformatic data, we cannot infer with certainty the subcellular location of either *Tetrahymena* Erv1 homolog.

metalloproteases Yta10 and Yta12; and the *i*-AAA+ protease, working in the intermembrane space and constituted by Yme1, in turn a homolog of Yta10 and Yta12.[59]

Our MS analysis revealed four Yta10/Yta12/Yme1 homologs, annotated as ATP-dependent metalloprotease FtsH family proteins. From BLASTp analysis, we tentatively identify PreTt19244 and PreTt13597 as likely Yta10/Yta12 orthologs, and PreTt18351 and PreTt24473 as probable Yme1 orthologs. A fifth, more distant FtsH metalloprotease homolog encoded by the *T. thermophila* macronuclear genome (PreTt17441), but not detected in our MS analysis, displays BLAST similarity with eubacterial proteins (peptidase M41), with highest BLAST hits to cyanobacterial and other Gram-positive eubacterial sequences.

Prohibitins (Phb1 and Phb2 in *S. cerevisiae*) are highly conserved eukaryotic proteins, localized in mitochondria and implicated in regulation of the replicative life-span of cells and in maintenance of mitochondrial morphology. Prohibitins have been shown to physically interact with and regulate membrane protein degradation by the *m*-AAA protease.[60] Our MS analysis identified two closely related prohibitin homologs in *T. thermophila* mitochondria), both annotated as SPFH domain / Band 7 family proteins: PreTt16871, a putative Phb1 ortholog, and PreTt24859, putatively a Phb2 ortholog.

A final AAA+ family protein of interest in the context of our mitochondrial analysis is PreTt15615, which appears to be a homolog of mouse ATAD3 (human Tob3). The mouse protein has recently been reported to have displacement loop-binding properties, and to be involved in mitochondrial nucleoid organization.[61]

Energy cycle proteins

Glycolysis

Among the proteins identified in this study are components of six of the ten enzymes of glycolysis (Table 4), the exceptions being hexokinase, phosphoglucose isomerase, fructose bisphosphate aldolase, and pyruvate kinase. (Curiously, although sequences encoding homologs of the last three enzymes are readily identifiable in the *T. thermophila* genome, a classical hexokinase sequence could not be recovered by BLAST searches of either genomic or EST data: nor does a hexokinase gene appear to be present in the *Paramecium* genome). Our finding of mitochondrion-associated glycolytic enzymes is consistent with earlier reports that several glycolytic enzymes are bound to mitochondria in *T. pyriformis*,[62] amounting in the case of phosphofructokinase to 75% of the total cellular activity.[63]

Our observation is not without precedent, as the entire glycolytic pathway has recently been shown to be functionally associated with mitochondria in a plant, *A. thaliana*,[64] and in the budding yeast, *S. cerevisiae*.[65] In the former instance, protease accessibility experiments and lack of N-terminal targeting sequences suggested that the glycolytic enzymes are outside mitochondria *per se*, perhaps attached to the cytosolic face of the outer mitochondrial membrane. In *T. thermophila*, phosphoglycerate mutase and enolase both possess an evident N-terminal extension, which in the case of enolase (Table 4), strongly resembles a mitochondrial targeting peptide (see below). Furthermore, although triosephosphate isomerase, glyceraldehyde phosphate dehydrogenase, and phosphoglycerate kinase do not exhibit N-terminal extensions, their N-terminal regions appear compositionally similar to *Tetrahymena* targeting peptides: in fact, to an extreme extent in the case of phosphoglycerate kinase. These observations suggest that some components of the glycolytic pathway may well be localized to, and perhaps even function within,

Table 4. Enzymes of glycolysis identified by proteomic analysis of purified *T. thermophila* mitochondrial fraction

Name	No.[a]	PreTt	BLAST annotation
Hexokinase[b]			
Phosphoglucose isomerase			
Phosphofructokinase[c]	73	29743	Phosphofructokinase family protein
	84	06068	Phosphofructokinase family protein
Fructose bisphosphate aldolase			
Triosephosphate isomerase	134	29845	Triosephosphate isomerase
Glyceraldehyde phosphate dehydrogenase[d]	46	29220	Glyceraldehyde-3-phosphate dehydrogenase, type I family protein
Phosphoglycerate kinase[e]	48	17357	Phosphoglycerate kinase family protein <GTP>
Phosphoglycerate mutase[f]	123	10592	Phosphoglycerate mutase 1 family protein
	198	29896	Phosphoglycerate mutase family protein
Enolase[g]	91	11127	Enolase family protein
Pyruvate kinase			

[a] See Supplementary Data Table 1.
[b] Neither genomic nor EST data provide evidence of a classical hexokinase sequence in the *T. thermophila* genome.
[c] Three phosphofructokinase family proteins are encoded in the *T. thermophila* genome, two of which were recovered here (the third is PreTt06087).
[d] Also reported in the phagosome proteome (see Supplementary Data Table 2). A paralogous sequence (PreTt29226) was not identified here.
[e] Single gene.
[f] At least seven family members are encoded in the *T. thermophila* genome. Very little similarity is evident between the two proteins identified here, even though both are annotated as phosphoglycerate mutase family proteins and both exhibit a GmpA conserved domain. Pret29896 appears to be a bacteria-like enzyme, whereas PreTt10592 is highly similar to phosphoglycerate mutases in other eukaryotes.
[g] PreTt11127 and a second gene, PreTt02295, encode highly similar protein sequences, with the one found here (PreTt11127) displaying an evident N-terminal mitochondrial targeting motif:

```
11127  MLSKLTKNLYMNKPTFGFSSIQNVFAREILDSRGNPTIEAEVVTSKG-AFRAAVPSG..
02295                    MATIKDIKAREILDSRGNPTVEVDLTVDNGQVFRAAVPSG..
                         :::*::: ************:*.::...:* .********
```

mitochondria in *Tetrahymena*. Genes for several of the glycolytic enzymes occur in multiple copies in the *Tetrahymena* genome, raising the possibility that certain of the multiple gene products are specifically targeted to mitochondria as appears to be the case for enolase (Table 4).

Notably, we found no evidence in our MS data for any of the ~30 annotated glucose or other sugar transporters encoded by the *T. thermophila* genome; hence, what the function might be of any imported glycolytic enzymes in *Tetrahymena* mitochondria is unclear. Further work will be necessary to demonstrate actual import into mitochondria as opposed to simple targeting to the outer mitochondrial membrane. An alternative, intra-mitochondrial function unrelated to the normal enzymatic role of these enzymes in glycolysis is a distinct possibility. In yeast, for example, enolase is recruited as a cofactor to assist in targeting a cytosolic tRNA to mitochondria for subsequent import into the organelle.[66]

Tricarboxylic acid (TCA) cycle

MS analysis identified all expected components of the pyruvate dehydrogenase complex (E1α, E1β, E2, E3), as well as components of all eight enzymatic activities of the tricarboxylic acid cycle (Table 5).

Citrate synthase. Citrate synthase functions as a homodimeric enzyme. We detected two paralogs in *Tetrahymena* mitochondria, which phylogenetic analyses indicate are the typical eukaryotic type. Mukai and Endoh recently identified a bacteria-like citrate synthase in *T. thermophila*, possibly acquired through lateral gene transfer.[67] Neither this protein (encoded by PreTt11286) nor related bacteria-like citrate synthase family proteins (PreTt03659, PreTt10419) were detected in our analysis, consistent with the absence of an evident N-terminal mitochondrial matrix targeting peptide in these proteins.

Aconitase (aconitate hydratase). Most eukaryotic cells contain two types of aconitase, one cytosolic and another targeted to the mitochondrial matrix. We detected one of *Tetrahymena*'s two aconitase paralogs, PreTt14482, which shares 64% identity with PreTt08830. Interestingly, the aconitase that is present in *Tetrahymena* mitochondria does not appear to be a typical mitochondrial aconitase, belonging instead to the iron regulatory protein

Table 5. Pyruvate dehydrogenase and TCA cycle enzymes

Name	No.[a]	PreTt	BLAST annotation
Pyruvate dehydrogenase			
PDH E1 α	183	06606	Pyruvate dehydrogenase E1 component
PDH E1 β	63	08277	Transketolase, pyridine-binding-domain-containing protein; pyruvate dehydrogenase E1b (acetyl-transferring) (*Arabidopsis thaliana*)
	66	25561	Transketolase, C-terminal-domain-containing protein, pyruvate dehydrogenase E1 β subunit (*Dictyostelium discoideum* AX4)
PDH E2	77	04738	Pyruvate dehydrogenase complex, dihydrolipoamide acetyltransferase family protein
PDH E3	52	11064	Dihydrolipoamide dehydrogenase family protein
Citrate synthase	2	04648	Citrate synthase family protein
	53	21058	Citrate synthase
Aconitase	14	14482	Aconitate hydratase 1 family protein
Isocitrate dehydrogenase (NAD^+)	76	18099	Dehydrogenase, isocitrate/isopropylmalate family protein (α subunit), NAD^+-dependent
	95	26182	Dehydrogenase, isocitrate/isopropylmalate family protein (β subunit), NAD^+-dependent
Isocitrate dehydrogenase ($NADP^+$)	47	06805	Isocitrate dehydrogenase, NADP-dependent family protein
2-Oxoglutarate dehydrogenase			
E1 α	32	05014	2-Oxoglutarate dehydrogenase, E1 component family protein
	4	06647	2-Oxoglutarate dehydrogenase, E1 component family protein
	11	09842	2-Oxoglutarate dehydrogenase, E1 component family protein
E1 β	559	20176[a]	2-Oxoisovalerate dehydrogenase β subunit, mitochondrial precursor, putative
E2	92	23045	2-Oxoglutarate dehydrogenase, E2 component, dihydrolipoamide succinyltransferase family
E3	52	11064	Dihydrolipoamide dehydrogenase family protein
Succinyl-CoA synthetase			
α subunit	85	13203	Succinyl-CoA synthetase, α subunit family protein
β subunit	87	05416	Succinyl-CoA synthetase, β subunit family protein
	88	10199	Succinyl-CoA synthetase, β subunit family protein
	104	03111	Succinyl-CoA synthetase, β subunit family protein
Succinate dehydrogenase			
SdhA	13	11187	Succinate dehydrogenase, flavoprotein subunit
SdhB	225	18907/18908[b]	Succinate dehydrogenase and fumarate reductase iron–sulfur protein
Fumarase	41	02578	Fumarate hydratase, class II family protein
Malate dehydrogenase	425	11406	Malate dehydrogenase, cytoplasmic, putative
	75	14250	Malate dehydrogenase family protein

[a] See Supplementary Data Table 1.
[b] Single gene initially predicted as two adjacent ORFs.

class (often annotated as cytosolic aconitase enzymes in plants). The other aconitase paralog in *Tetrahymena* is of the same type. It therefore appears that the sole form of aconitase in *Tetrahymena* is the "cytosolic" isoform, with at least one of the two paralogs apparently having acquired mitochondrial targeting information.

Isocitrate dehydrogenase(NAD$^+$). Human isocitrate dehydrogenase (NAD$^+$) is a heterotetramer consisting of two α subunits, one β subunit and one γ subunit, all very similar in sequence. *S. cerevisiae* isocitrate dehydrogenase is a complex consisting of two different subunits, IDH1 and IDH2, again very similar in sequence.

Our MS analysis identified two NAD$^+$ isocitrate dehydrogenase proteins, PreTt26182 and PreTt18099. As additional subunits of this complex do not appear to be encoded by the *T. thermophila* genome, the subunit composition of the *Tetrahymena* enzyme appears more similar to that of its yeast rather than its human counterpart. Superficially, BLASTp analyses suggest that PreTt26182 may correspond to the IDH1 from yeast, whereas PreTt18099 would correspond to IDH2. However, the subunits are quite similar (~43% identical), so it is difficult to be certain.

Isocitrate dehydrogenase (NADP$^+$). Two isoforms of the NADP$^+$ isocitrate dehydrogenase are present in most eukaryotes, one in the cytosol and one in the mitochondrial matrix. The single enzyme detected by our MS analysis (PreTt06805) corresponds, as expected, to the mitochondrial NADP$^+$ isocitrate dehydrogenase, whereas the enzyme encoded by PreTt21410, which was not detected in our MS analysis, is presumably a cytosolic version of the enzyme.

2-Oxoglutarate dehydrogenase. Typically, the 2-oxoglutarate dehydrogenase complex comprises three different components: E1 (oxoglutarate dehydrogenase, which has α and β subunits); E2 (dihydrolipoyllysine-residue succinyltransferase); and E3 (dihydrolipoyl dehydrogenase, which is also the E3 component of the pyruvate dehydrogenase complex). We identified three different E1-α paralogs, one E1-β subunit, one E2 and one E3 (Table 5).

Succinyl-CoA synthetase. Succinyl-CoA synthetase is a heterotetramer composed of α and β subunits, with mammals having ATP-specific and GTP-specific isoforms.

Our MS analysis identified one α subunit protein, the only such subunit encoded in the *Tetrahymena* genome and likely GTP-specific according to BLASTp analysis. We identified three different β subunits; in this case, BLASTp suggests that two isoforms, PreTt05416 and PreTt10199, correspond to ATP-specific β isoforms, whereas the third, PreTt03111, is GTP-specific. It is unclear what function the ATP-specific β isoforms may be performing in the absence of ATP-specific α subunits.

Succinate dehydrogenase. TCA cycle succinate dehydrogenase comprises two subunits, SdhA (flavoprotein subunit) and SdhB (iron-sulfur subunit), both of which were identified in the MS analysis (Table 5).

Fumarase. Fumarase, which is active as a homotetramer, was detected (Table 5).

Malate dehydrogenase. Our MS analysis detected two (Table 5) of four paralogous malate dehydrogenases encoded in the *Tetrahymena* genome (i.e. PreTt11406 and PreTt14250); the remaining two are specified by PreTt18794 and PreTt10826. All of these paralogs are most similar to a cytosolic type rather than a mitochondrial type of malate dehydrogenase. Interestingly, BLASTp and maximum likelihood reconstructions suggest that PreTt14250 (Table 5) is more similar to a proteobacterial, although not specifically α-proteobacterial, homolog than to any eukaryotic homolog, with the exception of the *Paramecium* one. This observation suggests the possibility that the *Tetrahymena* enzyme originated *via* lateral transfer of a malate dehydrogenase gene from a proteobacterium to an ancestral ciliate.

Electron transport chain (ETC) and ATP synthase

Complex I (NADH:ubiquinone oxidoreductase). Complex I is the largest and most structurally complicated of the four membrane-bound assemblies of the mitochondrial electron transport chain (Brandt, 2006),[68] containing a set of 14 central subunits that are equivalent to the minimal form of the enzyme that exists in bacteria. Variable sub-sets of these subunits are encoded by mtDNA (ten in the case of *T. thermophila*), with the remainder encoded by the nuclear genome, although a few eukaryotes lack Complex I altogether.[68–70] In our MS analysis, we identified 11 of these central subunits: seven of the mtDNA-encoded subunits as well as all four of the nuclear DNA-encoded subunits (Table 6A).

As well as these central subunits, Complex I contains a large number of so-called accessory subunits, of which 18 (Table 6B) have evident homologs that are phylogenetically widespread throughout eukaryotes.[68,70] Our proteomic analysis has identified nine of these accessory subunits in *T. thermophila* mitochondria. Together with three other proteins, namely NUXM and two assembly proteins designated CI30 and CI84,[70–72] these accessory proteins have been postulated to comprise an ancestral eukaryotic core that was added early in mitochondrial evolution to the 14 subunits that comprise the central ("bacterial") core. Six of the proteins in this ancestral eukaryotic core have bacterial homologs; five of these proteins are among the nine proteins we have identified in this group (Table 6B).

In addition to the ancestral eukaryotic core proteins, a phylogenetically more restricted set of

Table 6. Identified subunits of mitochondrial proton translocating NADH:ubiquinone oxidoreductase (Complex I)

Mitochondrial	Bovine	SwissProt	Human	Eubacterial	Genome[a]	Proteome[b]	PreTt
A. *Core ("central") subunits*							
Nad1	ND1	NU1M	ND1	NuoH	M	584	
Nad2	ND2	NU2M	ND2	NuoN	M		
Nad3	ND3	NU3M	ND3	NuoA	M		
Nad4	ND4	NU4M	ND4	NuoM	M	583	
Nad4L	ND4L	NULM	ND4L	NuoK	M (Ymf58)		
Nad5	ND5	NU5M	ND5	NuoL	M	570	
Nad6	ND6	NU6M	ND6	NuoJ	M (Ymf62)	578	
Nad7	49 kDa	NUCM	NDUFS2	NuoD	M	564	
Nad8	TYKY	NUIM	NDUFS8	NuoI	N	136	15574
Nad9	30 kDa	NUGM	NDUFS3	NuoC	M	576	
Nad10	PSST	NUKM	NDUFS7	NuoB	M	577	
Nad11	75 kDa	NUAM	NDUFS1	NuoG	N	6	06710
	24 kDa	NUHM	NDUFV2	NuoE	N	105	06008
	51 kDa	NUBM	NDUFV1	NuoF	N	5	06675
Bovine	SwissProt	Human	Proteome	PreTt			
B. *Accessory proteins (nucleus-encoded)*[c]							
39 kDa	NUEM	NDUFA9	153	07874[d]			
15 kDa	NIPM	NDUFS5	461	09517			
13 kDa	NUMM	NDUFS6	217	06922[d]			
AQDQ	NUYM	NDUFS4	209	06431			
ESSS	NESM	NDUFB11					
MWFE	NIMM	NDUFA1					
PDSW	NIDM	NDUFB10					
PGIV	NUPM	NDUFA8					
SDAP	ACPM	NDUFAB1	214	21241[d]			
B22	NI2M	NDUFB9					
B18	NB8M	NDUFB7					
B17.2	N7BM	DAP13	207	06733[d]			
B16.6	NB6M	GRIM19					
B14.7	NUJM	NDUFA11					
B14	NB4M	NDUFA6	466	13063[e]			
B13	NUFM	NDUFA5	246	26354			
B12	NB2M	NDUFB3					
B8	NI8M	NDUFA2	205	02192[d]			

Acronyms (bovine, SwissProt, human, eubacterial) for the various subunits are as described.[68]

[a] M, mitochondrial; N, nuclear.

[b] See Table 1 (M) and Supplementary Data Table 1 (N) for listing of individual entries.

[c] Accessory subunits having eubacterial homologs.

[d] Three other accessory subunits (CI30, CI84, NUXM) assigned to ancestral eukaryotic core[70] were not detected by BLASTp analysis.

[e] Currently annotated as a hypothetical protein.

Complex I subunits has been identified in metazoans,[68] and in fungi and plants and/or algae.[70,73,74] BLASTp analysis provided no evidence for any of these lineage-specific proteins, with the exception of three cases (PreTt24369, PreTt27618, PreTt14992) that are homologous to plant/algal-specific proteins designated Plant1, Plant2 and Plant3.[70] The corresponding proteins in *Chlamydomonas reinhardtii* (accession numbers Q6S7R7, Q6QIV7, and Q6QIV6) are annotated as Complex I subunits,[74] and they and their homologs in *A. thaliana* and *Oryza sativa* (rice) as well as the *Tetrahymena* sequences all contain a highly conserved domain that places them in the carbonic anhydrases/acetyltransferases, isoleucine patch superfamily (COG06663.2; PaaY). In *Arabidopsis*, disruption of the nuclear gene for one of these γ-type carbonic anhydrases results in a reduction in the level of Complex I and alters mitochondrial physiology.[75]

Complex II (succinate:ubiquinone oxidoreductase). Complex II comprises the TCA cycle enzyme succinate dehydrogenase (SDH, composed of SdhA and SdhB subunits) plus two integral membrane proteins that anchor SDH to the inner mitochondrial membrane, apocytochrome b_{560} (SdhC) and a hydrophobic 13 kDa polypeptide (SdhD). In most eukaryotes, these proteins are encoded in the nuclear genome, although in a few organisms several of the SDH subunits are mtDNA-encoded.[1,76] As discussed above, MS analysis identified the two nucleus-encoded subunits of SDH (Table 7); however, we were unable to detect a gene encoding either SdhC or SdhD in either the macronuclear or mitochondrial genomes of *Tetrahymena* or *Paramecium*. Presumably SdhC and SdhD, wherever they are encoded, have diverged in sequence too extensively in ciliates to allow their ready recognition by BLAST analysis.

Complex III (ubiquinol: cytochrome c oxidoreductase). In human mitochondria, this complex, cytochrome *c* reductase, consists of ten subunits, three of

Table 7. Structural and assembly components of respiratory complexes II-V identified by MS/MS analysis

Name	No.[a]	PreTt
A. *Complex II*		
Succinate dehydrogenase		
SdhA	13	11187
SdhB	225	18907/18908[b]
B. *Complex III*		
Apocytochrome *b* (Cob)	565	M
Cytochrome c_1	187	07044/07045[b]
Rieske iron–sulfur protein	175	15618
C. *Complex IV (cytochrome* c *oxidase)*		
Subunit 1 (Cox1)	566	M
subunit 2 (Cox2)	568	M
Subunit 5b (Cox5b)	272	10316/10317[b]
Assembly protein (Cox15)	127	06878
Assembly protein (Cox19)	462	20651
Sco1/Sco2 family protein	37	07889
Sco1/Sco2 family protein	405	05778
D. *Complex V (F_1F_O ATP synthase)*		
α Subunit (Atp1)	12	10516
β Subunit (Atp2)	42	04132
γ Subunit (Atp3)	160	27180
OSCP (Atp5)	192	18192/18913[b]
Assembly protein (Atp12)	452	28163

[a] See Supplementary Data Table 1.
[b] Single gene initially predicted as two adjacent ORFs.

which we detected here: apocytochrome *b* (Cob, encoded in the mtDNA); cytochrome c_1; and the Rieske iron–sulfur protein (Table 7). Using human Complex III sequences as BLASTp queries, human core protein I (CORI; P31930) gives significant hits to three *Tetrahymena* mitochondrial proteins detected by MS analysis: PreTt08077, PreTt01072 and PreTt17781. All three of these proteins are annotated, as is the human query sequence, as members of the peptidase M16 family. In fact, PreTt01072 is identified as a Mas1 homolog of the mitochondrial processing peptidase complex (see below). Homologs of the remaining five proteins (QCR2, 6, 7, 8, 9, 10) could not be detected by BLAST analysis in the *T. thermophila* genome sequence.

Complex IV (cytochrome c:O_2 oxidoreductase). The human version of this oligomeric complex consists of 13 polypeptides, three of which, Cox1, Cox2 and Cox3, are encoded in the human mitochondrial genome. Although a *cox1* gene is universally found in mtDNA, in some organisms genes for Cox2 and Cox3 have evidently been moved to the nuclear genome, from which they are now expressed through the cytosolic translation system for import into mitochondria.[1] *T. thermophila* mtDNA encodes Cox1 and Cox2, both of which were detected here (Table 7). Curiously, we have not been able to identify a Cox3-encoding sequence in either the mtDNA or the nuclear DNA of either *T. thermophila* or *P. tetraurelia*, even though Cox3 is a reasonably well conserved protein and Cox3 sequences have been reported in the mtDNA of other alveolates, including apicomplexans[77] and dinoflagellates.[78,79] Cox3 does not contain any prosthetic groups and is not involved directly in proton translocation.[80] Moreover, in the α-proteobacterium *Paracoccus dentrificans*, a two-subunit cytochrome oxidase (aa_3) evidently functions with only Cox1 and Cox2,[81] suggesting that Cox3 may be dispensable in some systems. Alternatively, Cox3 may have diverged too extensively in sequence in ciliates to be detectable by standard BLAST searches, or it may have been functionally substituted in ciliates by a different protein.

Of the ten nucleus-encoded cytochrome oxidase subunits in humans, a gene for only one, Cox5b, is evident in the *Tetrahymena* genome, and the corresponding protein was identified here (Table 7). The genome does encode homologs of human cytochrome oxidase assembly proteins Cox10 (PreTt28597), Cox11 (PreTt19744), Cox15 (PreTt06878) and Cox17 (PreTt19470) as well as Sco1/Sco2 (PreTt05778, PreTt07889). Three of these proteins, Cox15 and the two Sco1/Sco2 homologs, were identified here (Table 7) as well as a protein homologous to Cox19, a putative metallochaperone required for the assembly of cytochrome oxidase in yeast (Table 7).

Complex V (F_1F_O ATP Synthase). Of the 15 proteins that comprise bovine Complex V,[82] three subunits of the F_1 portion (α, Atp1; β, Atp2; and γ, Atp3) were identified here, as well as F_O δ subunit (Atp5; also called oligomycin sensitivity conferring protein (OSCP)) (Table 7). A single ATP synthase subunit, Atp9, is encoded by the *T. thermophila* mitochondrial genome,[35] but a corresponding protein was not seen in our MS analysis, nor were proteins corresponding to Atp6 or Atp8, both commonly encoded in other mtDNAs but often difficult to recognize,[1,83] or other subunits that comprise Complex V. Genes for these additional subunits have not been annotated in the *Tetrahymena* nuclear genome sequence. Genes for two Complex V assembly proteins have been reported for *T. thermophila*,[84] and one of these proteins, Atp12, was identified in our analysis (Table 7).

Protein import systems in *T. thermophila* mitochondria

Reductive mitochondrial genome evolution characterized by massive gene loss and transfer of genes from mtDNA to the nuclear genome, coupled with recruitment of novel proteins to the mitochondrion, has necessitated the largely *de novo* appearance of a mitochondrial protein import machinery,[85,86] although a subset of proteins involved in mitochondrial import comprises conserved bacterial proteins.[87] Functional analyses of these protein import machines have been largely limited to *S. cerevisiae*, although homologs have been identified in many animals and plants. Here, we utilized a combined bioinformatics (BLASTp)/proteomics approach in

identifying *T. thermophila* homologs of *S. cerevisiae* mitochondrial protein import components.

Translocase of the outer mitochondrial membrane (TOM)

The TOM complex is a multi-protein agglomeration responsible for the translocation of all proteins destined to move across the outer mitochondrial membrane. In general, the TOM complex appears to be poorly conserved in *T. thermophila*, with only three of a total of seven subunits identified by bioinformatics and/or proteomics methods. Tom40, the so-called general insertion pore, and Tom7 were identified through both approaches, whereas Tom22 was identified only *via* BLASTp similarity searches (Table 8). The other components of the TOM complex (Tom5, 6, 70 and 20) were not identified. However, considering that Tom5 and Tom6 are very short proteins and that their sequences are poorly conserved, we cannot exclude the possibility that homologous proteins exist in *T. thermophila.*

SAM complex

The SAM complex, which functions to insert β-barrel proteins such as Tom40 into the outer mitochondrial membrane, contains five subunits: Sam35, Sam37, Sam50, Mim1 and Mdm10. None of these proteins was identified in *T. thermophila*. Failure to find a Sam50 homolog (the core subunit of this complex) is particularly surprising, considering that Dolezal *et al*.[87] have suggested that Sam50 is part of the eukaryotic core protein import machinery.

Translocases of the inner mitochondrial membrane (TIM)

TIM22. The TIM22 complex is responsible for the insertion of a variety of proteins into the inner mitochondrial membrane. Typically, these proteins lack a canonical targeting peptide. In the present analysis, one of the four subunits (Tim22) was identified, although not by the MS analysis (Table 8). The remaining three proteins, Tim18, Tim54 and Tim12, were not found by MS or bioinformatics analysis.

TIM23. The TIM23 complex mediates the translocation of matrix-targeted proteins across the inner mitochondrial membrane. This appears to be a well conserved protein import pathway, with seven out of eight subunits identified in this analysis: Tim16, Tim17, Tim23, Tim50 and Hsp70 by proteomics and Pam18 and Tim44 by bioinformatics analysis of genome sequence data (Table 8). A Tim21 homolog was not identified by either approach.

Tiny Tims. A group of TIM proteins called the Tiny Tims functions in the transport of proteins to the TIM22 or SAM complexes. Four different Tiny Tims have been described in *S. cerevisiae*, of which we identified three in *Tetrahymena*, although they are generally not annotated as such: Tim10 in the proteomics analysis, Tim8 and Tim9 by BLASTp (Table 8). We have not identified a Tim13 homolog.

OXA complex. The OXA complex, which functions in the assembly of proteins in the inner mitochondrial membrane, consists of four subunits, two of which were detected here, Oxa1 in the MS analysis and Mdm38 by bioinformatics (Table 8). Mba1 and Y1h47 were not identified. OXA subunits other than Oxa1 are involved in docking of mitochondrial ribosomes at the inner membrane.

Mitochondrial processing peptidase (MPP) complex. The MPP complex in *S. cerevisiae* is made up of two subunits (Mas1, Mas2), both of which were identified in the proteomics analysis for *Tetrahymena* (Table 8). These peptidases are responsible for cleaving N-terminal matrix-targeting peptides.

Intermediate processing peptidase (IMP) complex. Of the three subunits in the IMP category (Imp1, Imp2, Som1), only Imp1 was found in this analysis (Table 8).

In summary, 36 proteins are known to be involved in mitochondrial protein import in *S.* cerevisiae,[87] of which 19 appear to have homologs in *Tetrahymena* (Table 9). As reported previously, the mitochondrial protein import machinery is a mosaic of proteins that have apparently been recruited from the endosymbiont, including Hsp70, signal peptidases and Oxa1, and proteins that have evolved within the eukaryotic lineage, notably TOM and TIM proteins. It appears that the proteins of bacterial descent are better conserved at a sequence level than those that are strictly eukaryotic. Of the 19 protein import subunits found here, 11 were identified in the MS analysis, whereas eight were identified through bioinformatics analysis of genome data alone. Caution must be exercised in assigning homology for some of these proteins, as BLASTp E values were quite high in several cases. For instance, BLASTp using the putative *Tetrahymena* Tim23 as query does not pick up anything in GenBank, whereas the *Paramecium* Tim23 homolog yields discernible BLASTp hits. Thus, a more detailed approach has been taken in defining some of these proteins as TIM/TOM subunits. In general, the *Paramecium* TIM/TOM sequences appear less divergent than those of *Tetrahymena*, and the protein predictions are more accurate; thus, the *Paramecium* protein predictions appear to have more value in defining the evolution of the ciliate TIM/TOM complex. On the basis of this analysis, an emerging generalization is that "core" proteins (i.e. Tom40, Tom22, Tom7, Tim23 and Tim17, Tim22) are more apt to be present than are subunits that serve largely receptor/accessory functions (i.e. Tom70 and Tom20, Tim12, Tim21, Tim54). Also, subunits of the Tim23 complex appear to be the most highly conserved. We

Table 8. Proteins of the mitochondrial import machinery in *T. thermophila*

Name	Subcomplex	No.[a]	PreTt	MS[b]	Bioinf[c]
TOM complex					
Tom40	Core translocase	30	03459	+	
Tom22[d]	Core translocase	–	19603		+
Tom7[e]	Core translocase	342	01966	+	
Tom6	Small subunit	–	×		
Tom5	Small subunit	–	×		
Tom70[f]	Receptor	–	×		
Tom20	Receptor	–	×		
SAM Complex					
Sam50	Core translocase	–	×		
Sam35	Metaxin	–	×		
Sam37	Metaxin	–	×		
Mdm10	Mdm10	–	×		
Tiny Tims					
Tim9[g]	Core complexes	–	05154		+
Tim8[h]	Core complexes	486	16268	+	
Tim10	Core complexes	–	22455		+
Tim13	Core complexes	–	×		
TIM22 Complex					
Tim22	Core translocase	–	12801		+
Tim12	Peripheral Tim	–	×		
Tim22	Accessory subunit	–	×		
Tim18	Accessory subunit	–	×		
TIM23 Complex					
Tim23	Core translocase	349	03540	+	
Tim17	Core translocase	167	06293	+	
Tim50[i]	Tim50	178	04277	+	
Pam18[j]	Pam complex	–	03698		+
Pam16	Pam complex	260	01262	+	
Tim44[k]	Pam complex	–	20198		+
mtHsp70	Pam complex	1	04604/04605[l]	+	
Tim21	Tim21	–	×		
Oxa Complex					
Oxa1[m]	Core chaperone	226	30328	+	
Mba1	Ribosome receptor	–	×		
Mdm38[n]	Ribosome receptor	–	18547		+
Ylh47	Ribosome receptor	–	×		
Imp complex					
Imp1	Core peptidase	–	08934		+
Imp2	Core peptidase	–	×		
MPP complex					
Mas1[o]	Core peptidase	441	01072	+	
Mas2[o]	Core peptidase	384	25767	+	

[a] See Supplementary Data Table 1.
[b] A plus sign (+) means detected by mass spectrometry (MS).
[c] A plus sign (+) means detected by BLASTp search (Bioinf).
[d] The Tom22 proteins in both *Paramecium* and *Tetrahymena* are quite divergent relative to their *Saccharomyces* and animal counterparts. BLAST scores alone are not convincing; however, small conserved regions are evident in the alignment.
[e] Tom7 represents the extreme C-terminal portion of a rather large, incorrect PGP (PreTt01966), an inference supported by the fact that the two identified MS peptides are both at the extreme C terminus of PreTt01966. E-values for Tom7 are extremely poor, but the alignment demonstrates convincingly that this protein is Tom7.
[f] Yeast Tom70 is composed of numerous tetratricopeptide repeat (TPR) domains; hence, similarity with a number of *Tetrahymena* protein sequences is presumably only on the basis of shared domains rather than a reflection of true homology.
[g] When the *Tetrahymena* Tim9 sequence is used as a BLASTp query, there is little to suggest that it is a Tim protein; however, its *Paramecium* ortholog convincingly recovers Tim9 in a BLASTp search.
[h] Annotated as a Tim10/DDP family zinc-finger-containing protein. The *Tetrahymena* Tim8 homolog is a very short protein; in consequence, E-values are poor. However, the alignment is persuasive and, like all suggested Tiny Tim proteins, it exhibits a characteristic CXXXC–(~15 residues)–CXXXC pattern.
[i] Numerous *Tetrahymena* proteins align with Tim50 from other eukaryotes (often annotated as NLI phosphatases), but of the top eight with BLASTp E-values of <1e-20 relative to the yeast Tim50 protein sequence, only PreTt04277 was identified in the MS analysis.
[j] A protein found by MS (PreTt03699), identified tentatively as a Tim14/Pam18 protein, proved on further analysis not to be a true Tim14. It exhibits good similarity to Tim14 proteins (better, in fact, than the actual Tim14); however, it has a long N-terminal extension (supported by EST data) that suggests it is not truly Tim14. Interestingly, genes encoding this protein and the authentic Tim14 are adjacent in the *Tetrahymena* genome, suggesting a possible recent duplication event. We conclude that there is a *Tetrahymena* homolog of Tim14, but that it was not identified in our MS analysis.
[k] *Paramecium* has a clear homolog of *S. cerevisiae* Tim44; however, the similarity is not as obvious in the *Tetrahymena* case.
[l] Single gene initially predicted as two adjacent ORFs.
[m] Corrected PGP sequence significantly improves alignment to homologous *S. cerevisiae* and *Paramecium* proteins.
[n] The *S. cerevisiae* sequences for Mdm38 and Ylh47 both hit PreTt18547 equally well. On the basis of BLASTp values, we suggest that Mdm38 is the true ortholog (note that the N-terminal region of the predicted protein has been revised on the basis of EST data).
[o] Mas1 and Mas2 protein sequences have been corrected based on consideration of corresponding EST data. Two additional related sequences detected in the MS analysis appear to be MPP paralogs: PreTt08077 (putative Mas1 ortholog) and PreTt17781 (putative Mas2 ortholog).

emphasize that many of the proteins involved in these complexes are quite small, and fairly divergent. For this reason, bioinformatics-based identifications cannot rely solely on BLAST E values but also require careful scrutiny of the accompanying protein alignments.

Proteins of the mitochondrial transcription and translation systems

In the MS analysis, we found two proteins of the mitochondrial transcription system: PreTt14793, corresponding to the phage T7-like RNA polymerase previously described in *T. pyriformis*,[88] and PreTt12339, which appears to be a homolog of mitochondrial transcription termination factor mTERF.[89] We also found a homolog of Pet127p (PreTt17447), a yeast protein involved in mitochondrial RNA turnover.[90,91] Additionally, we identified a number of components of the mitochondrial translation apparatus, including translation initiation and elongation factors (but no translation termination factor) and several aminoacyl-tRNA synthases (Table 9A).

Table 9 lists small subunit (SSU) and large subunit (LSU) mitochondrial ribosomal proteins (MRPs) found in our analysis. Relative to their eubacterial evolutionary antecedents, mitochondrial ribosomes (mitoribosomes) are highly divergent in both their RNA and protein constituents.[92] In *T. pyriformis*, Curgy *et al.*[93] have characterized 80 S mitoribosomes that dissociate into subunits that have the same sedimentation coefficient (55 S), the same electrophoretic mobility, and a similar morphology. Only mammalian (e.g. human and rat) and yeast (*S. cerevisiae*) mitoribosomes have been thoroughly characterized with respect to their protein constituents, by direct proteomic analysis of isolated mitoribosomes and their small and large subunits in the case of mammals. The results indicate that homologs of certain eubacterial (e.g. *E. coli*) ribosomal proteins can be identified in mammalian and yeast mitoribosomes, but others cannot (i.e. they appear to have been lost in evolution or have become so divergent in sequence that they can no longer be readily recognized). On the other hand, novel proteins have been added, particularly in the case of mammalian mitoribosomes. For example, the small subunit of human mitoribosomes, comprising 30 proteins, lacks *E. coli* homologs S1, S3, S4, S8, S13, S19 and S20 but has 15 additional proteins (MRPS22-S36), six of which appear to have homologs in yeast.[92] The large subunit of human mitoribosomes, containing 48 proteins, lacks *E. coli* homologs L5, L6, L25, L29 and L31 but has 20 additional proteins (MRPL37-L56), nine of which appear to have homologs in yeast. All mitoribosomal proteins are encoded in the nuclear genome in animals, and all but one in yeast. In *T. thermophila*, at least nine ribosomal protein genes are encoded by the mtDNA.[35]

To test how many mitoribosomal proteins could be identified *via* a bioinformatics approach and how many of these were found also in our MS analysis, we used the set of ribosomal proteins encoded by *Rickettsia prowazekii*, a specific α-proteobacterial relative of mitochondria, as BLASTp queries against *T. thermophila* gene predictions. In addition, we used the unique human MRP proteins as BLASTp queries in a separate search.

In this way, we identified eight nucleus-encoded SSU MRPs, five independently by both MS and BLASTp and three by MS alone (Table 9B). None of these SSU MRPs was confidently identified by the *in silico* approach alone (due to high BLASTp E-values). Two of the five SSU MRPs encoded by *Tetrahymena* mtDNA (S12 and S13) were also found in the MS analysis. Unique human SSU MRP sequences gave two BLASTp hits, one of which (S29) appears to be authentic and has a corresponding MS entry (Table 9B). This protein is one of the additional human SSU MRPs with a yeast homolog.

We identified 19 nucleus-encoded LSU MRPs by a combination of approaches: 14 by both MS and BLASTp, one by MS alone (L9 and L28 did give legitimate BLASTp hits, but at a very high E value that we would have discounted in the absence of MS data), and an additional four by BLASTp alone (Table 9C). BLASTp using unique human LSU MRP sequences gave three strong candidates (L43, L47, and L49), all of which have yeast as well as human homologs. Three of the four LSU MRPs encoded by *Tetrahymena* mtDNA were identified by MS analysis.

In total, using a combination of bioinformatics-based sequence comparisons and direct MS analysis, we identified with confidence 10 SSU MRP sequences, all with a corresponding MS entry, and 22 LSU MRP sequences, 18 with an MS entry. By extension, many of the remaining, unidentified *Tetrahymena* MRPs are likely to have a corresponding MS entry, but one that is currently unannotated. Clearly, the only way that the full complement of MRPs will be identified in *Tetrahymena*, or any other organism, is by direct proteomics analysis of isolated mitoribosomes and/or their separated subunits.

Whether encoded by nuclear DNA or mtDNA, virtually all of the identified *Tetrahymena* MRPs have homologs in human and/or yeast mitoribosomes. An exception is L6, which is mtDNA-encoded in *Tetrahymena* but absent from both mammalian and yeast mitoribosomes, although we do not know whether *rpl6* (*ymf60*[35]) is actually functional in *Tetrahymena* mitochondria. Notably, *Tetrahymena* seems to have homologs of several of the proteins that are so far unique to mammalian and/or yeast mitoribosomes, i.e. not present in eubacterial ribosomes. In all cases, the *Tetrahymena* candidates are found in both human and yeast mitoribosomes. This result indicates that these particular "eukaryote-specific" MRPs must have originated well before the emergence of the last common ancestor of the animal-fungal lineage (opisthokonts).

Recently, Smits *et al.* published a comprehensive bioinformatics survey of MRP sequences encoded in a range of eukaryotic genomes, including that of *T. thermophila*.[94] Our analysis of *Tetrahymena* MRPs

(Table 9) compares very favorably with that reported by Smits *et al.*,[94] with virtually the same set of nucleus-encoded MRPs (eight SSU, 18 LSU) identified in the two independent studies. Our results differ from those of Smits *et al.* as follows: (1) our bioinformatics screen did not retrieve eight proteins reported by Smits *et al.*: S34, S35, S36, yeast Rsm22 homolog, L19, L23, L36, and L46 (see Table 9B and C). Detection of these additional proteins presumably reflects the more sensitive

Table 9. Mitochondrial translation proteins in *T. thermophila*

Name	No.[a]	PreTt			
A. *Protein synthesis factors and aminoacyl-tRNA synthases*					
Protein synthesis factors					
Initiation					
IF-2	383	14786			
IF-3	221	09256			
Elongation					
EF-G	8	07503			
EF-Tu	112	20839/20840			
EF-Ts	161	26771			
Aminoacyl-tRNA synthases					
Asparagine	55	03027			
Phenylalanine	385	07589			
Threonine	375	06749			
Tryptophan	410	24986			

Name	No.[a]	PreTt	EST[b]	MS	Bioinf[c]
B. *Small subunit ribosomal proteins*					
S3					M
S5	463	06911	–	+	–
S6	256	03140	–	+	–
S9	193	01041[d]	+	+	+
S12	587		–	+	M
S13	572		–	+	M
S14					M
S15	458	28371	–	+	–
S16	206	03801	–	+	+
S17	563	04131[e]	+	+	+
S18	258	06737[f]	–	+	(+)
S19					M
S29	165	11814	–	+	+
S34[g]	469	04187	–	+	–
S35[g]		06755	+ (Y)	–	–
S36[g]	322	13073	+	+	–
Rsm22[g,h]		02642	–	–	–

Name	No.[a]	PreTt	EST[b]	MS	Bioinf[c]
C. *Large subunit ribosomal proteins*					
L1	156	19816	+	+	+*
L2	574	–		+	M
L3	129	19743	+	+	+
L4	195	08966	+	+	+
L6 (Ymf60[35])					M
L7/L12	62	07046[i]	+	+	+
L9	102	17535	–	+	(+)
L11		04140[j]	–	–	+*
L13	44	09119[k]	+	+	+
L14	588	–		+	M
L15	186	24349	–	+	+
L16	589	–		+	M
L17	442[l]	[i]	+	+	+
L19[g]	264	21511	+	+	–
L20		22682	+	–	+*
L21	414	04837	–	+	+
L22	488	11431	–	+	–
L23[g]		07787	+ (Y)	–	–
L24	456	17534	+	+	+
L27	50	07115	–	+	*
L28	218	26603	+	+	(+)
L33		15127	–	–	+
L36[g]		03923	+ (Y)	–	–
L43	455	11934	–	+	+
L46[g]	253	07951	+	+	–
L47	215	28488	–	+	+
L49		19187	–	–	+

bioinformatics approach used by Smits *et al*.[94] Notably, our MS analysis provides direct support for a mitochondrial location for S34, S36, L19 and L46 (Table 9B and C). (2) Conversely, we identified MRP L17 (see Table 9C), not reported by Smits *et al*.[94] (3) Finally, three of the *Tetrahymena* entries listed by Smits *et al.* appear to represent cytosolic (cy), not mitochondrial, ribosomal proteins. Specifically, we conclude on the basis of reciprocal BLAST analyses that S8 in the work by Smits *et al*.[94] is cyS15a; S11 is cyS14; and S17 is cyS17 (see Table 9B for an alternative MRP S17 candidate).

Ciliate-specific ORFan proteins

Many (~45%) of the *T. thermophila* mitochondrial proteins identified in our MS analysis are without functional annotation at this time, and exhaustive *in silico* analysis has failed to suggest a possible function or indicate (a) conserved domain(s) in these cases. In Supplementary Data Table 1, >200 such proteins are designated hypothetical, their genes sharing no significant sequence similarity with any ORFs outside of the ciliate lineage. Some of these hypothetical proteins appear to be specific to *T. thermophila* (representing "singleton" ORFans),[95] but about two-thirds have putative homologs in *P. tetraurelia* ("homologous" ORFans).

Of the ciliate-specific ORFans 22 are <100 amino acid residues in length, and most (20) of these small ORFan sequences are confirmed by EST data. Of these small ORFan proteins, 12 have putative homologs in *P. tetraurelia*; alignments for three of these are shown in Figure 4. In these particular cases, and more generally as well, the ORFan coding sequences are single-copy in *Tetrahymena* but in multiple copies in *Paramecium*, consistent with the inferred serial duplications that underlie genome evolution in the latter ciliate.[39]

We used the algorithm TargetP[96] to compare the proportion of hypothetical proteins inferred to be localized in *Tetrahymena* mitochondria (see below) relative to a set of functionally annotated (TCA cycle and ETC) mitochondrial proteins. In a set of 214 hypothetical proteins, only 39 (18%) are predicted to be targeted to mitochondria at the 60% confidence level, compared to ~35% for the TCA cycle + ETC set.

Unexpectedly, at the 60% confidence level, 16 of the hypothetical proteins are predicted to be secreted, whereas at the 80% confidence level ten are predicted to be localized specifically to the endoplasmic reticulum. We tested these ten proteins with another program, Predotar.[97] At the 80% confidence level, seven of them are again predicted to be ER proteins, with an additional two predicted at a lower confidence level (69%). Because two of the ten ER-predicted proteins are likely cilia and/or phagosome contaminants, as discussed earlier, it is possible that some of these "ER-directed" proteins are additional components of unknown function contributed by contaminating cilia and/or phagosomes.

Phylogenetic analysis

Based on BLAST searches, almost half (~46%) of the mitochondrial proteins that we identified in our MS analysis have eubacterial homologs. However, only 11% of the identified proteome represents nucleus-encoded proteins that in phylogenetic reconstructions cluster robustly (i.e. with >75% bootstrap support) with α-Proteobacteria. A further 5% of the MS-identified mitochondrial proteome represents proteins that are encoded in the *T. thermophila* mitochondrion. We consider that these proteins are also of α-proteobacterial origin, even though not all of them are clearly identifiable as such through sequence similarity searches. Altogether,

Notes to Table 9:

[a] See Table 1 and Supplementary Data Table 1 for listing of individual entries.

[b] A plus sign (+) means that supporting EST data (TTL) are available through TBestDB (http://amoebidia.bcm.umontreal.ca/pepdb/searches/login.php?bye=true).

[c] M, encoded in mtDNA; +, sequence retrieved using *Rickettsia prowazeckii* protein query; *, sequence retrieved using *Reclinomonas americana* mitochondrial (mtDNA-encoded) protein query; +*, sequence retrieved using both *R. prowazeckii* and *R. americana* mitochondrial protein queries. Parentheses enclose associated E-values greater than the cutoff value of e-05.

[d] PreTt01041 is 753 residues long; however, only the final 280 residues correspond to S9. All three MS peptides associated with this PGP are upstream of the S9-homologous portion; however, this upstream region does not recover any obvious homologs in a BLASTp search in GenBank. Thus, the MS data for PreTt01041 provide evidence for a different (unidentified) mitochondrial protein, not S9.

[e] In Supplementary Data 1 of ref.[94], the entry for S17 (PreTt17947) is to cytosolic, not mitochondrial, S17.

[f] BLASTp values are very poor, but residues ~450–525 of PreTt06737 consistently recover residues ~10–85 of eubacterial S18.

[g] *T. thermophila* MRP not retrieved in our bioinformatic screen but reported recently.[94] For those entries not listed in Supplementary Data Table 1, (Y) indicates that available EST data confirm the corresponding PGP in the TGD.

[h] Yeast nomenclature (the human homolog is designated Q9H7HD/Q9BZH2).[94]

[i] The corresponding TBestDB EST cluster (TTL00000109) has 51 ESTs, an unusually high number for a mitoribosomal protein (usually, mitoribosomal proteins are represented by a single or, at most, a few ESTs). However, the encoded L7/L12 protein definitely appears to be mitochondrial, not cytosolic.

[j] The inferred sequence (1347 residues) is abnormally long for a ribosomal protein. PreTt04140 is almost certainly chimeric: the L11-homologous region is represented by the C-terminal ~175 residues, with the region upstream homologous to an EST (TTL00007811) annotated as unassigned protein.

[k] A corresponding EST cluster (TTL00000434) again has an unusually high number of ESTs (37) for a mitoribosomal protein. The PGP sequence (PreTt09119; 779 residues) is chimeric: the first 303 residues are represented in EST cluster TTL00003010, annotated as Cdc2 protein kinase-related cluster, whereas the final 419 residues correspond to the mitoribosomal L13 sequence, as inferred from EST data (TTL00000434).

[l] EST cluster TTL00005099 predicts an ORF of at least 237 residues (7.m04851 is the corresponding entry in TGD).

Figure 4. CLUSTALW alignment of three small ORFan proteins from *T. thermophila* (T.t.) mitochondria with duplicated homologs from *Paramecium tetraurelia* (P.t.). The degree of conservation of individual positions is as in Figure 2. The diagnostic MS peptides from the *T. thermophila* analysis are shown in bold and underlined.

therefore, we can infer with confidence that a minimum of ~16% of the MS-identified mitochondrial proteome in *T. thermophila* was inherited directly from the proto-mitochondrial endosymbiont. Another 30% of the *T. thermophila* mitochondrial proteins either group with α-Proteobacteria, but with <75% bootstrap support; branch within Proteobacteria but not specifically with α-Proteobacteria; or display prokaryotic but not specifically proteobacterial affinity.

We find, consistent with the observations of Karlberg *et al*.[25] regarding mitochondrial proteins that robustly group with α-Proteobacteria, that many of the *Tetrahymena* proteins in this category are involved in core bioenergetic processes. In particular, many of the known subunits of the pyruvate dehydrogenase multi-enzyme complex, the tricarboxylic acid cycle, the electron transport chain (Complexes I–IV) and the ATP synthase (Complex V) branch convincingly with α-Proteobacteria in phylogenetic reconstructions (data not shown). For instance, the E1-α (PreTt06606), E1-β (PreTt08277, PreTt25561), and E2 (PreTt04738) subunits of the pyruvate dehydrogenase multi-enzyme complex group confidently with α-Proteobacteria. Among other TCA cycle enzymes, fumarase (PreTt02578) and the NADP-specific isoform of isocitrate dehydrogenase (Pre06805) are clearly α-Proteobacterial in origin as well. Additional support for the α-proteobacterial ancestry of mitochondria derives from the robust clustering of mitochondrial proteins involved in translation (several ribosomal proteins, IF-2 and an EF-Tu-associated protein) and protein maturation/metabolism (Hsp70, Cpn60, Mge1p, glutamate dehydrogenase) with α-proteobacterial homologs.

Nevertheless, several proteins involved in the foregoing mitochondrial processes either do not cluster robustly with α-Proteobacteria, or group non-specifically with other proteobacterial groups. For instance, the mitochondrial elongation factors EF-Tu, EF-Ts and EF-G, along with several ribosomal proteins, cluster with their α-proteobacterial homologs to the exclusion of other groups, but with bootstrap confidence values below 75%. Similar results were obtained for the E3 subunit of pyruvate dehydrogenase (PreTt11064), two succinyl-CoA synthetase β paralogs (PreTt03111 and PreTt05416), the 24-kDa subunit of NADH dehydrogenase (PreTt06008), SdhA (PreTt11187) and the δ subunit of ATP synthase (PreTt18192/18193). Although inadequate bootstrap support leads us to assign these mitochondrial proteins to the non-specific bacterial category, a good case can be made for their having been inherited from the α-proteobacterial endosymbiont. Indeed, several more focused studies have demonstrated that most of these proteins are α-proteobacterial in origin in other species.[98,99] Furthermore, the fact that most of these proteins function in multi-enzyme complexes that contain protein partners of undisputed α-proteobacterial origin argues in favor of their α-proteobacterial antecedence.

Additionally, in the case of identified *Tetrahymena* mitochondrial proteins that we designate simply as prokaryotic, a substantial proportion cluster non-

specifically within Proteobacteria. Many of these proteins are involved in fatty acid and protein metabolism, including acyl-CoA dehydrogenase and acetyl-CoA carboxylase (PreTt24442 and PreTt16455, respectively). Also, some components of the TCA cycle (2-oxoglutarate dehydrogenase E2 and malate dehydrogenase) group non-specifically within Proteobacteria. As we argue above, it is quite likely that many of these proteins are derived from the α-proteobacterial endosymbiont. Factors that confound standard tree reconstruction methods, namely small proteins, sparse phylogenetic sampling, and relatively rapid rates of sequence evolution, have, at least to some extent, probably obscured the true origins of proteins in this group.

A sizeable contingent of mitochondrial proteins exhibits a more complex evolutionary history. Among this group are several well conserved proteins, such as superoxide dismutase and Hsp90, along with the TCA cycle enzymes aconitase and isocitrate dehydrogenase. The fact that these TCA cycle enzymes branch non-specifically within the eubacterial domain is not completely unexpected.[98] However, the observation does raise important questions about the phylogenetic affiliations of the gene content of the eubacterial progenitor of the mitochondrion and the processes of lateral gene transfer and sequence divergence after endosymbiotic gene transfer.

Characteristics of *T. thermophila* mitochondrial targeting peptides

Expressed sequence tag (EST) data allowed us to predict N termini for 368 of the 545 nucleus-encoded mitochondrial proteins identified here. As reported for other species,[3] two distinct classes of mitochondrial proteins were evident: those with cleavable N-terminal targeting peptides and those containing internal or cryptic (non-cleaved) signals. The proteins with apparent canonical targeting (as deduced from N-terminal sequences) constitute the majority of mitochondrial proteins, although the other class represents a substantial proportion (~1/3) of the total.

In *T. thermophila*, canonical mitochondrial targeting signals are similar to those defined in other species. In particular, studies performed in other model organisms suggest that mitochondrial targeting peptides are enriched in basic (Lys and Arg) and hydroxylated (Ser) residues, and nearly devoid of acidic (Asp and Glu) amino acids.[100,101] Mitochondrial targeting peptides in *Tetrahymena* and *Paramecium* generally appear to follow these trends, with an elevated level of particular basic and hydroxylated residues but with the two acidic amino acids being quite rare. This feature is demonstrated by a comparison of the frequency of amino acids in the entire predicted cellular proteome to amino acid frequency in the first 15 and 30 residues in mitochondrial proteins with canonical targeting sequences. In the case of basic residues, Lys constitutes 8.8% of amino acids in the entire proteome, whereas it represents 9.8% and 10.8%, of the first 30 and 15 residues of targeting peptides, respectively. Arg is present at higher levels in targeting peptides relative to the entire proteome, constituting 2.7% of the predicted cellular proteome but 5.7% of the first 30 residues of target peptides and 6.4% of the first 15 residues. The fact that Lys appears to be more common than Arg may reflect constraints associated with the high A+T content of the *T. thermophila* genome. A similar but even more dramatic trend is observed for Arg residues in targeting peptides in *P. tetraurelia*, which also has a high-A+T genome (72%). Whereas Arg is rare in the total cellular proteome (3.5%), it is represented abundantly in mitochondrial target peptides (7.0% and 9.2% for the first 30 and 15 residues, respectively). These data point to a selective pressure that has favored the accumulation of positively charged amino acids in mitochondrial targeting peptides of *Paramecium* and *Tetrahymena*.

The prevalence of Ser in the entire proteome and targeting peptides of both *Tetrahymena* and *Paramecium* follows a pattern very similar to that observed for Lys in *Tetrahymena*; however the abundance of Thr is essentially equal in targeting peptides and the entire proteome of both species. Interestingly, Lys and Ser are part of the most prominent motif, MLSK, observed in *Tetrahymena* targeting peptides (Figure 5). In contrast, neither Lys nor Ser is as prominent in the sequence logo generated for *Paramecium* (not shown), in keeping with the slightly reduced Lys content in *Paramecium* targeting peptides compared to the situation in *Tetrahymena*. In the case of *Paramecium*, it appears that there is a requirement for positively charged and hydrophobic residues at the extreme N terminus, although no motif is present that is as prominent as the MLSK signature from *Tetrahymena*.

Perhaps the most striking compositional feature of the canonical targeting peptides in *Tetrahymena* and *Paramecium* is the scarcity of acidic residues relative to the entire predicted proteome, consistent with observations made in other systems.[100,101] Both Asp and Glu comprise between 5% and 7% of the total predicted proteome, while making up ~2% of the first 30 residues and 0.6% or less of the first 15 residues of canonical mitochondrial targeting peptides.

Another feature that canonical *T. thermophila* target peptides have in common with those in other organisms is the ability to assume an α-helical secondary structure. Typically, characterized mitochondrial targeting peptides are amphiphilic, positively charged on one face, and hydrophobic and uncharged on the other, features thought to be important in interactions between the targeting signals and the associated mitochondrial import machinery.[86,100] In *T. thermophila*, helical wheel projections of EST-verified N termini usually conform to this rule. Additionally, this requirement helps to explain the increased prevalence of hydrophobic residues such as Ala, Phe and Leu in targeting peptides. This apparent consistency in targeting peptide structure and composition among organisms

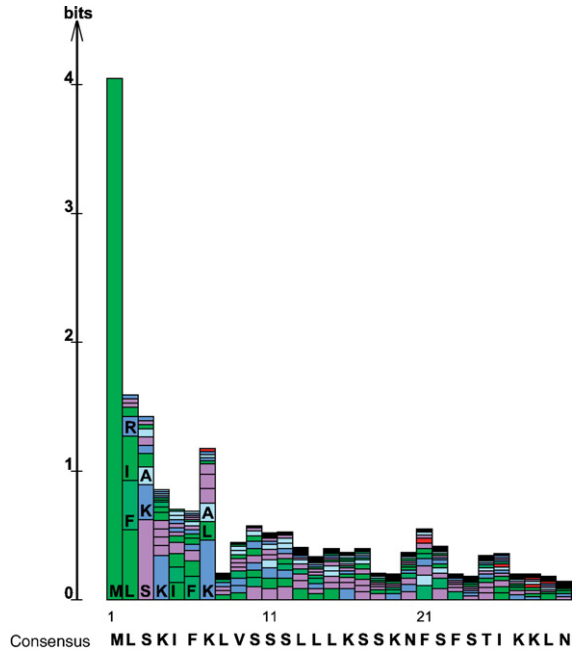

Figure 5. Sequence logo for the first 30 residues of 273 canonical mitochondrial targeting peptides from *T. thermophila*.

in diverse eukaryotic groups lends support to the notion of a common mitochondrial protein import mechanism, a view further reinforced by our identification of several known components of the TIM/TOM complexes, as discussed earlier.

Processing of mitochondrial targeting signals

Alignments of *T. thermophila* mitochondrial proteins, along with their most similar counterparts from *Paramecium*, to highly conserved α-proteobacterial proteins allow us to infer the length of the N-terminal sequences that are required for import into mitochondria. By extension, we can deduce that the site of peptidase cleavages occur at the C-terminal ends of targeting peptides. For the purposes of this investigation, only the mitochondrial proteins that are known to localize to the matrix were examined, as they are cleaved by a conserved MPP, whereas proteins that are localized to other compartments often require further proteolytic processing. In both *Tetrahymena* (Figure 6) and *Paramecium* (not shown), alignments indicate a well-conserved Phe or Tyr residue at or near the proteolytic processing site. This observation is consistent with the demonstration that purified rat and bovine MPP have a preference for aromatic or other bulky hydrophobic residues at the +1 position relative to the processing site.[102] Furthermore, a conserved Phe residue (F94), shown in *S. cerevisiae* MPP to accommodate bulky hydrophobic or aromatic groups present at the +1 position,[103] is conserved in the homologous *T. thermophila* protein. Examination of mitochondrial targeting peptides from other species has demonstrated a requirement for Arg at positions −10, −3 and −2 from the processing site. Although basic amino acids are frequently located near the putative processing sites in *Tetrahymena* mitochondrial targeting peptides, no specific motif is evident. Direct experimental approaches will be required to verify these processing sites, and to pinpoint other required residues in a more systematic way.

Miscellaneous mitochondrial proteins detected by MS

- PreTt13388 (metallo-β-lactamase superfamily protein). It is not clear why *Tetrahymena* (and *Paramecium*) would have an enzyme to degrade β-lactams. Additionally, PreTt20857 (hypothetical protein TTHERM_00151990) bears some domain similarity to isopenicillin *N*-synthase, the final enzyme in penicillin production.
- PreTt19260 and PreTt07358 (pyridine nucleotide-disulphide oxidoreductase family proteins; similar to apoptosis-inducing factor, AIF). Detection of these mitochondrion-localized, apoptosis-related proteins may be relevant to the proposed apoptosis-like cell death of macronuclei in *T. thermophila*.[104]
- PreTt20799 (ATP:guanidophosphotransferase, C-terminal catalytic domain-containing protein; arginine kinase). The *Tetrahymena* enzyme is very similar to that seen in marine animals (even more similar than to its *Paramecium* homolog).

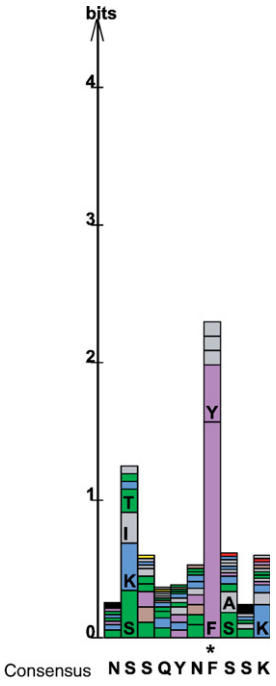

Figure 6. Sequence logo of the region immediately flanking the matrix-processing peptidase (MPP) site for 18 matrix-bound mitochondrial proteins. The asterisk denotes the presumptive site of cleavage by the MPP.

- PreTt10091 (mitochondrial glycoprotein). This protein is similar to *S. cerevisiae* acidic protein of the mitochondrial matrix, which is involved in oxidative phosphorylation and which, in turn, is related to the human complement receptor gC1q-R.

Concluding Remarks

The study reported here provides a comprehensive overview of mitochondrial structure and function in the ciliated protozoon, *T. thermophila*, the first such detailed, MS-based analysis of the proteome of a protist. By analyzing the peptides generated by exhaustive tryptic digestion of isolated mitochondria, we obtained excellent coverage of both soluble and membrane-associated proteins from all four submitochondrial compartments. Judging from the recovery of expected mitochondrial proteins, our MS analysis has likely detected ~60% of the total mitochondrial proteome, which we estimate totals minimally 900–1000 proteins. Analysis of selected, key biochemical pathways and complexes indicates that *T. thermophila* mitochondria carry out a range of functions approximating those described for the mitochondria of multicellular organisms.

Of particular interest is the finding that ~45% of the proteins identified in our analysis have no known function, and most have no demonstrable sequence homologs outside of the ciliate lineage. The conclusion that these proteins are functional is supported by our direct demonstration of their presence in mitochondria, and by supporting EST data in many cases, and by the fact that putative homologs for about two-thirds of them are found in another ciliate, *P. tetraurelia*. In general, we find that mitochondrial targeting algorithms such as TargetP have a low probability of predicting a mitochondrial localization for the proteins directly identified here by MS, and this is particularly true with regard to the set of ciliate-specific ORFans. Lack of sequence homologs of identified function outside ciliates coupled with the relatively low predictive capacity of mitochondrial targeting algorithms emphasizes both the power and the necessity of direct, MS-based proteomic analyses for the identification of novel mitochondrial proteins.

The advantage of an MS approach in complementing genomic analysis is further illustrated by our identification of a novel small protein, Ymf78, encoded by the mtDNA of various *Tetrahymena* species. The 44 codon ORF for this protein was overlooked during annotation of the mitochondrial genomes of both *T. pyriformis*[47] and *T. thermophila*,[35] as well as *P. aurelia*,[49] the latter mtDNA encoding a 48 codon homolog (Figure 2). ORFs of <50 codons are generally discounted during genome annotation, but peptide data provide direct evidence for the existence of the corresponding proteins. In our analysis, tryptic peptides were screened against a sixfold translation of the entire mtDNA sequence rather than against the set of previously inferred mtDNA-encoded protein sequences, an approach that was instrumental in the discovery of Ymf78.

In our MS analysis, we identified a number of proteins corresponding to "ciliate-specific ORFs" of unknown function, found initially in the *T. pyriformis* and *P. aurelia* mitochondrial genomes (Table 1).[47] We regard these genes as most likely representing highly divergent homologs of standard mtDNA-encoded genes found in other organisms. Indeed, on the basis of subsequent analyses, Brunk *et al.* suggested functional identities for three of these ciliate-specific ORFs, also found in *T. thermophila* mtDNA.[35] There may be exceptions to this general inference of evolutionary origin as, e.g. Ymf78, which could well be a ciliate-specific invention.

A surprising finding in our analysis was the detection of components of six of the ten enzymes of glycolysis in isolated mitochondria. The complete absence of MS hits to any cytosolic ribosomal proteins or to other abundant cytosolic components (e.g. translation factors) argues against simple contamination as an explanation for this observation. Association of a glycolytic enzyme complex with the outer mitochondrial membrane, as reported for *Arabidopsis*,[64] is understandable from the perspective of metabolic efficiency. However, our further observation that at least some of the *Tetrahymena* glycolytic enzymes (in particular enolase) appear to contain N-terminal matrix-targeting information raises the possibility that these proteins are actually imported into and function within mitochondria. In that case, it may be that these enzymes perform an alternative function to their usual role in glycolysis.

Our results provide further insights into the origin and evolution of the mitochondrial proteome. As in *S. cerevisiae*,[25,26] the mitochondrial proteome in *Tetrahymena* is clearly an evolutionary chimera, with the ancestry of only a minority (~16%) of the proteins we identified being robustly traceable to an α-proteobacterial progenitor. An additional fraction (~30%) are generally "prokaryotic" in phylogenetic affinity but cannot be affiliated confidently with α-Proteobacteria. About 13% of the MS-identified proteins are eukaryote-specific, having evident homologs in other eukaryotes but not in Bacteria or Archaea. Finally, ~45% of the MS-identified proteins are of unknown function, and most of these (~41%) appear to be unique either to the ciliate lineage (29%) or specifically to *Tetrahymena* (12%). Although some of these hypothetical proteins may be highly diverged versions of mitochondrial proteins widely present in other organisms, it appears that a relatively large proportion of the mitochondrial proteome may have emerged within particular eukaryotic lineages, where these proteins either substitute for known functions (i.e. they are functional analogs rather than homologs) or contribute entirely new functions. In the case of respiratory Complex I, for example, *T. thermophila* has all 14 subunits that comprise the central ("bacterial") core

as well as most (9/14) of the components of the "ancestral eukaryotic core" that has been proposed to have been added early in eukaryote evolution.[70] However, we have not detected, either by proteomic or genomic analysis, any of the additional accessory subunits that contribute to Complex I assembly and function in metazoan animals.[68,70] Presumably, other proteins have been recruited in ciliates to serve similar roles. Our data support an emerging picture of mitochondrial proteome evolution in which a surprisingly large and variable set of lineage-specific proteins has been added to a conserved functional core, much of which was inherited from an α-proteobacterial ancestor.[7]

Additional approaches will be required to more fully define the components of various functional complexes and pathways operating in ciliate mitochondria. Further proteomic analysis of isolated submitochondrial fractions (e.g. inner and outer membrane) would undoubtedly reveal additional mitochondrial proteins as well as further localize proteins identified in the analysis reported here. Although we were successful in identifying a large number of subunits of oxidative phosphorylation Complexes I–V, proteomic analysis of the individual complexes isolated, e.g. by means of blue-native polyacrylamide gel electrophoresis,[105,106] has the potential to substantially augment the existing data set by revealing novel proteins specifically associated with each of these complexes. Similarly, although we have catalogued a number of mitoribosomal proteins, including several that had previously been reported in only animals and fungi, proteomic analysis of isolated mitoribosomes[93] and their subunits will ultimately be required to complete this inventory.

Finally, it is notable that, although many of the novel mitochondrial proteins reported here are encoded by single-copy genes in the macronuclear genome of *T. thermophila*, the corresponding genes are present in multiple copies in the genome of *P. tetraurelia*. Thus, *T. thermophila* is obviously a preferred model ciliate system for functional analysis by means of the gene knockout and gene replacement techniques that have been developed for this protist.[37,38]

Materials and Methods

Chemicals

Reagent-grade chemicals were purchased from Sigma Aldrich (Oakville, ON, Canada), or Fisher Scientific (Nepean, ON).

Preparation of mitochondria

T. thermophila strain CU428 was cultured in SPP medium[107] with the sequestrine replaced with DTPA-Fe (III) (Iron Chelate 7%, Plant Products Co. Ltd., Brampton, ON).

For isolation of mitochondria, a total of 3 l of filter-sterilized SPP medium was prepared and distributed among 15 750 ml vented tissue-culture flasks. Each flask was seeded with 1 ml of fresh early to mid-log phase cells and the flasks shaken slowly on an orbital shaker at room temperature (~25 °C) for 1.5 to two days to about mid-log phase. Cultures were distributed among four 750 ml centrifuge bottles, chilled on ice for 15 min, and the cells collected at 800*g* for 5 min in an IEC CR-6000 preparative, swinging-bucket centrifuge. The cell pellets were suspended in residual medium, transferred to four 50 ml conical centrifuge tubes and the above centrifugation step repeated. The final cell pellets were suspended gently in 150 ml of lysis buffer (250 mM sorbitol, 0.2% (w/v) BSA, and 5 mM iodoacetamide, in the EM buffer of *Sickmann et al.*).[22] Cells were lysed by three passes through a metal hand-held emulsifier (Fisher Scientific) with the control nut backed off 2.5 turns. The lysate was centrifuged for 5 min at 1000*g* (IEC CR-6000), the supernatant decanted into six 30 ml Corex centrifuge tubes, followed by centrifugation at 8000*g* for 5 min in a #875 fixed-angle rotor using an IEC B22M preparative centrifuge. The uppermost supernatant material, comprising both liquid and a viscous, semi-solid material, was decanted . Each crude mitochondrial pellet was resuspended gently in 1 ml of SEM buffer (EM buffer containing 250 mM sucrose), then combined, and an additional 6 ml of SEM buffer was added. The mitochondria were purified on discontinuous sucrose gradients with steps consisting of 1.55 M (3 ml) and 1.15 M (6 ml) sucrose in SEM buffer in Beckman SW41Ti centrifuge tubes. Two ml of a suspension of crude mitochondria were layered onto six such gradients, and the tubes centrifuged at 22,500 rpm for 1 h at 4 °C using a Beckman SW41Ti rotor and a Beckman L8-70M ultracentrifuge. The mitochondrial bands were carefully removed from the 1.15/1.55 M sucrose interface using a large-bore disposable plastic pipette, the recovered material was distributed between two 30 ml Corex centrifuge tubes, and the mitochondrial suspensions were diluted slowly by dropwise addition at 4 °C of two volumes of SEM buffer. The mitochondria were collected by centrifugation at 8000*g* for 10 min (IEC B22M), the supernatants discarded, and the mitochondrial pellets suspended in a total of 12 ml of SEM buffer. The discontinuous sucrose gradient step was repeated as described and the final purified mitochondrial pellets were stored at −20 °C.

Sample preparation

Intact mitochondria (~1 mg) were suspended in 100 μl of 50 mM ammonium bicarbonate and lysed by vortex mixing for 1 min both before and after heating at 70 °C for 5 min. Proteins in solution were reduced by the addition of dithiothreitol (DTT) to a final concentration of 5 mM and incubation at 56 °C for 1 h, then alkylated at cysteine residues by the addition of iodoacetamide to final concentration of 10 mM followed by incubation in the dark at room temperature for 1 h. Proteolytic digestion was achieved by the addition of 50 μg of porcine trypsin (Promega) and incubation at 37 °C overnight. The insoluble fraction was removed by centrifugation in a microfuge for 30 min at 16,000*g* and the remaining soluble peptides were acidified by mixing with eluent A (see below) to a total volume of 2.0 ml and then filtered through a 0.4 μm pore size filter (Millipore) to prepare for offline separation by strong cation-exchange (SCX) HPLC.

SCX HPLC parameters

The prepared sample (1.5 ml) was injected into an HP1050 LC system (Agilent, Palo Alto, CA) with a 1.5 ml injection loop and a PolyLC Polysulfoethyl A column (2.1 mm internal diameter (ID)×100 mm length) packed with 5 μm beads with 300 Å pores (The Nest Group, Southborough, MA). Fractionation was effected by a binary mobile-phase gradient at a total flow-rate of 0.2 ml/min using eluent A (10 mM KH_2PO_4 in 25% (v/v) acetonitrile, 75% (v/v) deionized water) and eluent B (10 mM KH_2PO_4 and 350 mM KCl in 25% acetonitrile, 75% deionized water), each acidified to pH 3.0 with phosphoric acid. Chromatography was initiated with 100% eluent A, then from the second to the 58th minute the proportion of eluent B was changed linearly from 0 to 100% and remained at 100% until the run was terminated at the 60th minute. A total of 30 fractions (0.4 ml, 2 min each) were collected using an SF-2120 Super Fraction Collector (Advantec MFS, Dublin, CA). Each fraction was vacuum-dried to 10 μl total volume and 1 μl was analyzed by nanobore reverse phase (RP) HPLC over a 120 min gradient.

RP HPLC parameters

The nanobore LC system was from LC Packings (Amsterdam, The Netherlands) and consisted of a Famos autosampler and an Ultimate Nano LC system. It was interfaced to an API QSTAR Pulsar QqTOF mass spectrometer (Applied Biosystems/MDS Sciex, Foster City, CA) using a Protana NanoES ion source (Protana Engineering A/S, Odense, Denmark). The spray capillary was a PicoTip SilicaTip emitter with a 10 μm ID tip (New Objective, Woburn, MA) and the nanobore LC column was a 75 μm ID×150 mm length RP PepMap C18 nano capillary column (LC Packings, Amsterdam, The Netherlands), or in-house equivalent, packed with 3 μm beads with 100 Å pores. From each sample, 1 μl was injected *via* the full-loop mode and separation was performed using a binary mobile-phase gradient at a total flow-rate of 200 nl/min. Eluent C consisted of 94.9% deionized water, 5.0% acetonitrile and 0.1% formic acid (pH ~3). Eluent D consisted of 5.0% deionized water, 94.9% acetonitrile and 0.1% formic acid.

The following binary gradient (time (min), % eluent D) was used for the RP HPLC analysis: (0, 5; 8, 5; 10, 15; 20, 20; 70, 40; 80, 60; 90, 80; 102, 80; 105, 5; 120, stop).

Resolved peptides were analyzed online by nanospray MS/MS using the QStar Pulsar mass spectrometer with the following instrumental settings: curtain gas, 20; ionspray voltage, 1800–3000 V (optimized daily); declustering potential (DP), 65 V; focussing potential (FP), 265 V; collision gas, nitrogen; data acquisition software, Analyst QS SP8 with Bioanalyst Extension 1.1 (Applied Biosystems/MDS Sciex) using the information-dependent acquisition (IDA) mode. IDA software allowed for the collection of cycles of data: 1 s MS survey scan (m/z=400–1500, z=2–4), then 4×1 s MS/MS scans (m/z=80–2000) on the most intense MS peaks. Redundant analysis was limited by 30 s rolling exclusion lists set to reject previously analyzed masses within a ±6.0 Th range. Every fourth cycle was dedicated to analyzing low-intensity peaks (closest to, but exceeding 10 cps threshold) to aid identification of low-abundance peptides. Tandem MS peak lists from all 30 SCX fractions were concatenated using basic text editing software and submitted to Matrix Science's Mascot software platform. This analysis was used to compare MS/MS data to the preliminary gene predictions (PGP) database (year 2004) released by The Institute for Genomic Research (TIGR)‖, as well as to a six-reading-frames translation of the *T. thermophila* (TETTH) mitochondrial genome (GenBank accession number NC_003029). Five separate samples of mitochondria were lysed, separated and analyzed.

Peptide identifications exceeding Mascot search program's 95% significance threshold (usually scoring between 20 and 100) and passing manual inspection were accepted along with their corresponding PGPs and categorized in the following way: PGPs identified by two or more tryptic peptides from any of the repeat analyses, or by a single peptide at least eight residues in length and showing extensive (>5) sequential y-ion peaks or other extremely convincing spectral features, were classified as confident identifications. PGPs identified by a single peptide, passing Mascot's 95% significance threshold and showing some but not extensive y-ion coverage, were classified as tentative identifications. The BLASTp tool at the NCBI website¶ was then used to compare both confident and tentative PGP amino acid sequences for similarity to those of all known proteins in the NCBI database. BLAST scores are given as a probability of random match, or E-value; PGPs showing similarity to a known protein with an E-value ≤e-05 were considered significant for the purposes of this study. The underlying assumption is that significant sequence similarity implies functional similarity.

The robustness of the protein identifications was estimated by searching the MS/MS spectra against a concatenated database comprising the predicted protein sequences (target) and a decoy of the same sequences in reverse.[40,41] The false-positive rate (%) was calculated as 100×(2× number of decoy hits) / total number of hits. Proteins that were identified by two or more tryptic peptides, by one tryptic peptide with extensive sequence coverage, and by one tryptic peptide with less than extensive coverage were evaluated separately to determine the reliability of the identifications in each grouping. False-positive rates calculated as such are likely to be high estimates for real searches that are conducted only with the predicted protein sequences.

Phylogenetic methods

The BLASTp program[108] was used to query GenBank for bacterial and eukaryotic homologs of MS-identified *Tetrahymena* mitochondrial proteins. Where possible, homologs were identified in a broad selection of eukaryotes and several major bacterial groups including α, β, γ, δ and ε Proteobacteria, Cyanobacteria, Bacteroidetes, Chlorobi, Chlamydiae, Actinobacteria, Deinococcus-Thermus, Spirochaetes, Aquificae and Firmicutes. Additionally, tBLASTn was used to query TBestDB[109] for eukaryotic homologs not present in GenBank. Proteins corresponding to full-length EST clusters were collected. A fairly liberal E-value cutoff threshold of e-05 was used for inclusion of homologous sequences in the alignments, although most proteins considered for phylogenetic analysis scored significantly lower E-values. Homologs of identified *Tetrahymena* mitochondrial protein sequences were aligned with Clustal X[110] and edited

‖ ftp://ftp.tigr.org/pub/data/Eukaryotic_Projects/t_thermophila/Gene_Predictions/

¶ http://www.ncbi.nlm.nih.gov/BLAST

manually with BIOEDIT. Manual editing of alignments was performed for regions of the alignment where it was unclear whether Clustal X had performed adequately, which in turn prevented the comparison of non-homologous sequences. Edited alignments were then used for the reconstruction of maximum likelihood phylogenies with IQPNNI version 3-0.1.[111] IQPNNI trees were created using the WAG amino acid substitution model,[112] and no stopping rule, with γ-distributed rates, no invariable sites, an estimated γ parameter and six categories of substitution rates. Bootstrap values were obtained from PHYML[113] by generating 100 non-parametric bootstrap replicates using a BIONJ input tree, the WAG substitution model, no invariable sites, six substitution rate categories and an estimated γ distribution parameter. When the phylogenetic positions of *Tetrahymena* mitochondrial proteins among the bacterial groups were established, they were assigned to one of two groups: those that are derived from α-proteobacteria, and those with non-specific prokaryotic affinities. In order for a protein to be classified as α-proteobacterial in origin, it was required to group with α-Proteobacteria to the exclusion of other bacterial groups, with a bootstrap support of at least 75%.

Defining bacterial homologs was not possible for all proteins, as some genes have very narrow phylogenetic distributions, whereas others have no bacterial homologs. Trees were not reconstructed for proteins for which no eubacterial homologs were evident; these proteins were annotated as eukaryotic in origin. However, all proteins with no bacterial relatives were examined in detail, and classified according to whether they had specific relatives in other eukaryotic supergroups, other ciliates (*P. tetraurelia* and/or *Anophryoides haemophila*), or exist only in *Tetrahymena*. Mitochondrial proteins were assigned as general eukaryotic proteins if they scored BLASTp E-values ≤e-05. Conversely, *Tetrahymena* mitochondrial proteins were designated as ciliate-specific if they received an ungapped BLASTp score of 100 or higher when compared to the predicted *Paramecium* proteome. An ungapped BLASTp was deemed necessary in order to reduce the effect that general compositional similarities between *Tetrahymena* and *Paramecium* have on the assignment of homology. If no BLAST scores met this threshold for a given protein, it was ascribed a *Tetrahymena*-specific origin.

Acknowledgements

This work was supported by funding to M.W.G. and to R.E.P. from the Canadian Institutes of Health Research (grants MOP-4124 and MOP-13347, respectively) and to K.W.M.S. and R.E.P. from the Natural Sciences and Engineering Research Council (NSERC) and Eli Lilly Canada's Collaborative Research and Development Grants. Infrastructure support to K.W.M.S. from the Ontario Research and Development Challenge Fund and MDS SCIEX is gratefully acknowledged. R.M.R.G. was supported by an NSERC PGS-A award and a Predoctoral Scholarship from the Killam Trusts. M.W.G. is pleased to acknowledge salary support from the Canada Research Chairs Program and the Canadian Institute for Advanced Research. The research reported here was carried out while M.W.G. was a Fellow and R.E.P. was an Associate of the Canadian Institute for Advanced Research (Program in Evolutionary Biology).

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2007.09.051

References

1. Gray, M. W., Lang, B. F., Cedergren, R., Golding, G. B., Lemieux, C., Sankoff, D. *et al.* (1998). Genome structure and gene content in protist mitochondrial DNAs. *Nucl. Acids Res.* **26**, 865–878.
2. Gray, M. W., Burger, G. & Lang, B. F. (1999). Mitochondrial evolution. *Science*, **283**, 1476–1481.
3. Pfanner, N. & Wiedemann, N. (2002). Mitochondrial protein import: two membranes, three translocases. *Curr. Opin. Cell Biol.* **14**, 400–411.
4. Nakai, K. (2000). Protein sorting signals and prediction of subcellular localization. *Advan. Protein Chem.* **54**, 277–344.
5. Emanuelsson, O. & von Heijne, G. (2001). Prediction of organellar targeting signals. *Biochim. Biophys. Acta*, **1541**, 114–119.
6. Richly, E., Chinnery, P. F. & Leister, D. (2003). Evolutionary diversification of mitochondrial proteomes: implications for human disease. *Trends Genet.* **19**, 356–362.
7. Gabaldón, T. & Huynen, M. A. (2003). Reconstruction of the proto-mitochondrial metabolism. *Science*, **301**, 609.
8. Heazlewood, J. L., Tonti-Filippini, J. S., Gout, A. M., Day, D. A., Whelan, J. & Millar, A. H. (2004). Experimental analysis of the *Arabidopsis* mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins. *Plant Cell*, **16**, 241–256.
9. Rehling, P., Pfanner, N. & Meisinger, C. (2003). Insertion of hydrophobic membrane proteins into the inner mitochondrial membrane—a guided tour. *J. Mol. Biol.* **326**, 639–657.
10. Guda, C., Fahy, E. & Subramanian, S. (2004). MITOPRED: a genome-scale method for prediction of nucleus-encoded mitochondrial proteins. *Bioinformatics*, **20**, 1785–1794.
11. Dreger, M. (2003). Proteome analysis at the level of subcellular structures. *Eur. J. Biochem.* **270**, 589–599.
12. Taylor, S. W., Fahy, E. & Ghosh, S. S. (2003). Global organellar proteomics. *Trends Biotechnol.* **21**, 82–88.
13. Griffin, T. J. & Aebersold, R. (2001). Advances in proteome analysis by mass spectrometry. *J. Biol. Chem.* **276**, 45497–45500.
14. Griffin, T. J., Goodlett, D. R. & Aebersold, R. (2001). Advances in proteome analysis by mass spectrometry. *Curr. Opin. Biotech.* **12**, 607–612.
15. Aebersold, R. & Mann, M. (2003). Mass spectrometry-based proteomics. *Nature*, **422**, 198–207.
16. Bauer, A. & Kuster, B. (2003). Affinity purification-mass spectrometry. Powerful tools for the characterization of protein complexes. *Eur. J. Biochem.* **270**, 570–578.
17. Dreger, M. (2003). Emerging strategies in mass-spectrometry based proteomics. *Eur. J. Biochem.* **270**, 569.

18. Pflieger, D., Le Caer, J.-P., Lemaire, C., Bernard, B. A., Dujardin, G. & Rossier, J. (2002). Systematic identification of mitochondrial proteins by LC-MS/MS. *Anal. Chem.* **74**, 2400–2406.
19. Reichert, A. S. & Neupert, W. (2004). Mitochondriomics or what makes us breathe. *Trends Genet.* **20**, 555–562.
20. Marcotte, E. M., Xenarios, I., van der Blik, A. M. & Eisenberg, D. (2000). Localizing proteins in the cell from their phylogenetic profiles. *Proc. Natl Acad. Sci. USA*, **97**, 12115–12120.
21. Kumar, A., Agarwal, S., Heyman, J. A., Matson, S., Heidtman, M., Piccirillo, S. *et al.* (2002). Subcellular localization of the yeast proteome. *Genes Dev.* **16**, 707–719.
22. Sickmann, A., Reinders, J., Wagner, Y., Joppich, C., Zahedi, R., Meyer, H. E. *et al.* (2003). The proteome of *Saccharomyces cerevisiae* mitochondria. *Proc. Natl Acad. Sci. USA*, **100**, 13207–13212.
23. Prokisch, H., Scharfe, C., Camp, D. G., II, Xiao, W., David, L., Andreoli, C. *et al.* (2004). Integrative analysis of the mitochondrial proteome in yeast. *PLoS Biol.* **2**, 795–804.
24. Reinders, J., Zahedi, R. P., Pfanner, N., Meisinger, C. & Sickmann, A. (2006). Toward the complete yeast mitochondrial proteome: multidimensional separation techniques for mitochondrial proteomics. *J. Proteome Res.* **5**, 1543–1554.
25. Karlberg, O., Canbäck, B., Kurland, C. G. & Andersson, S. G. E. (2000). The dual origin of the yeast mitochondrial proteome. *Yeast*, **17**, 170–187.
26. Kurland, C. G. & Andersson, S. G. E. (2000). Origin and evolution of the mitochondrial proteome. *Microbiol. Mol. Biol. Rev.* **64**, 786–820.
27. Gray, M. W., Burger, G. & Lang, B. F. (2001). The origin and early evolution of mitochondria. *Genome Biol.* **2**, reviews 1018.1–1018.5.
28. Kruft, V., Eubel, H., Jänsch, L., Werhahn, W. & Braun, H.-P. (2001). Proteomic approach to identify novel mitochondrial proteins in Arabidopsis. *Plant Physiol.* **127**, 1694–1710.
29. Millar, A. H., Sweetlove, L. J., Giegé, P. & Leaver, C. J. (2001). Analysis of the Arabidopsis mitochondrial proteome. *Plant Physiol.* **127**, 1711–1727.
30. Mootha, V. K., Bunkenborg, J., Olsen, J. V., Hjerrid, M., Wisniewski, J. R., Stahl, E. *et al.* (2003). Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell*, **115**, 629–640.
31. Taylor, S. W., Fahy, E., Zhang, B., Glenn, G. M., Warnock, D. E., Wiley, S. *et al.* (2003). Characterization of the human heart mitochondrial proteome. *Nature Biotech.* **21**, 281–286.
32. Heazlewood, J. L., Howell, K. A., Whelan, J. & Millar, A. H. (2003). Towards an analysis of the rice mitochondrial proteome. *Plant Physiology*, **132**, 230–242.
33. Alonso, J., Robriguez, J. M., Baena-López, L. A. & Santarén, J. F. (2005). Characterization of the *Drosophila melanogaster* mitochondrial proteome. *J. Proteome Res.* **4**, 1636–1645.
34. Millar, A. H., Heazlewood, J. L., Kristensen, B. K., Braun, H.-P. & Møller, I. M. (2005). The plant mitochondrial proteome. *Trends Plant Sci.* **10**, 1360–1385.
35. Brunk, C. F., Lee, L. C., Tran, A. B. & Li, J. (2003). Complete sequence of the mitochondrial genome of *Tetrahymena thermophila* and comparative methods for identifying highly divergent genes. *Nucl. Acids Res.* **31**, 1673–1682.
36. Eisen, J. A., Coyne, R. S., Wu, M., Wu, D., Thiagarajan, M., Wortman, J. R. *et al.* (2006). Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol.* **4**, e286.
37. Turkewitz, A. P., Orias, E. & Kapler, G. (2002). Functional genomics: the coming of age for *Tetrahymena thermophila*. *Trends Genet.* **18**, 35–40.
38. Collins, K. & Gorovsky, M. A. (2005). *Tetrahymena thermophila*. *Curr. Biol.* **15**, R317–R318.
39. Aury, J.-M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B. M. *et al.* (2006). Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature*, **444**, 171–178.
40. Elias, J. E., Haas, W., Faherty, B. K. & Gygi, S. P. (2005). Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nature Methods*, **2**, 667–675.
41. Elias, J. E. & Gygi, S. P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods*, **4**, 207–214.
42. Suissa, M. & Schatz, G. (1982). Import of protein into mitochondria. Translatable mRNAs for imported mitochondrial proteins are present in free as well as mitochondria-bound cytoplasmic polysomes. *J. Biol. Chem.* **257**, 13048–13055.
43. Marc, P., Margeot, A., Devaux, F., Blugeon, C., Corral-Debrinski, M. & Jacq, C. (2002). Genome-wide analysis of mRNAs targeted to yeast mitochondria. *EMBO Rep.* **3**, 159–164.
44. Smith, J. C., Northey, J. G. B., Garg, J., Pearlman, R. E. & Siu, K. W. M. (2005). Robust method for proteome analysis by MS/MS using an entire translated genome: demonstration on the ciliome of *Tetrahymena thermophila*. *J. Proteome Res.* **4**, 909–919.
45. Jacobs, M. E., Desouza, L. V., Samaranayake, H., Pearlman, R. E., Siu, K. W. M. & Klobutcher, L. A. (2006). The *Tetrahymena thermophila* phagosome proteome. *Eukaryot. Cell*, **5**, 1990–2000.
46. Bowman, G. R., Smith, D. G. S., Siu, K. W. M., Pearlman, R. E. & Turkewitz, A. P. (2005). Genomic and proteomic evidence for a second family of sense core granule cargo proteins in *Tetrahymena thermophila*. *J. Eukaryot. Microbiol.* **52**, 291–297.
47. Burger, G., Zhu, Y., Littlejohn, T. G., Greenwood, S. J., Schnare, M. N., Lang, B. F. & Gray, M. W. (2000). Complete sequence of the mitochondrial genome of *Tetrahymena pyriformis* and comparison with *Paramecium aurelia* mitochondrial DNA. *J. Mol. Biol.* **297**, 365–380.
48. Edqvist, J., Burger, G. & Gray, M. W. (2000). Expression of mitochondrial protein-coding genes in *Tetrahymena pyriformis*. *J. Mol .Biol.* **297**, 381–393.
49. Pritchard, A. E., Seilhamer, J. J., Mahalingam, R., Sable, C. L., Venuti, S. E. & Cummings, D. J. (1990). Nucleotide sequence of the mitochondrial genome of *Paramecium*. *Nucl. Acids Res.* **11**, 173–180.
50. Millar, A. H. & Heazlewood, J. L. (2003). Genomic and proteomic analysis of mitochondrial carrier proteins in Arabidopsis. *Plant Physiol.* **131**, 443–453.
51. Kunji, E. R. S. (2004). The role and structure of mitochondrial carriers. *FEBS Letters*, **564**, 239–244.
52. Ledesma, A., de Lacoba, M. G. & Rial, E. (2002). The mitochondrial uncoupling proteins. *Genome Biol.* **3**, reviews 3015.1–3015.9.
53. Sluse, F. E. & Jarmuszkiewicz, W. (2002). Uncoupling proteins outside the animal and plant kingdoms: functional and evolutionary aspects. *FEBS Letters*, **510**, 117–120.
54. Lill, R. & Mühlenhoff, U. (2005). Iron-sulfur-protein

biogenesis in eukaryotes. *Trends Biochem. Sci.* **30**, 133–141.
55. Henze, K. & Martin, W. (2003). Essence of mitochondria. *Nature*, **426**, 127–128.
56. Tovar, J., León-Avila, G., Sánchez, L. B., Sutak, R., Tachezy, J., van der Giezen, M. *et al.* (2003). Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation. *Nature*, **426**, 172–176.
57. Sutak, R., Dolezal, P., Fiumera, H. L., Hrdy, I., Dancis, A., Delgadillo-Correa, M. *et al.* (2004). Mitochondrial-type assembly of FeS centers in the hydrogenosomes of the amitochondriate eukaryote *Trichomonas vaginalis*. *Proc. Natl Acad. Sci. USA*, **101**, 10368–10373.
58. Hanson, P. I. & Whiteheart, S. W. (2005). AAA+ proteins: have engine, will work. *Nature Rev. Mol. Cell Biol.* **6**, 519–529.
59. Leonhard, K., Herrmann, J. M., Stuart, R. A., Mannhaupt, G., Neupert, W. & Langer, T. (1996). AAA proteases with catalytic sites on opposite membrane surfaces comprise a proteolytic system for ATP-dependent degradation of inner membrane proteins in mitochondria. *EMBO J.* **15**, 4218–4229.
60. Steglich, G., Neupert, W. & Langer, T. (1999). Prohibitins regulate membrane protein degradation by the *m*-AAA protease in mitochondria. *Mol. Cell. Biol.* **19**, 3435–3442.
61. He, J., Mao, C.-C., Reyes, A., Sembongi, H., Di Re, M., Granycome, C. *et al.* (2007). The AAA^+ protein ATAD3 has displacement loop binding properties and is involved in mitochondrial nucleoid organization. *J. Cell Biol.* **15**, 141–146.
62. Srere, P. A. (1987). Complexes of sequential metabolic enzymes. *Annu. Rev. Biochem.* **56**, 89–124.
63. Eldan, M. & Blum, J. J. (1973). Localization of phosphofructokinase on the mitochondria of *Tetrahymena pyriformis*. *J. Biol. Chem.* **248**, 7445–7448.
64. Giegé, P., Heazlewood, J. L., Roessner-Tunali, U., Millar, A. H., Fernie, A. R., Leaver, C. J. & Sweetlove, L. J. (2003). Enzymes of glycolysis are functionally associated with the mitochondrion in Arabidopsis cells. *Plant Cell*, **15**, 2140–2151.
65. Brandina, I., Graham, J., Lemaitre-Guillier, C., Entelis, N., Krasheninnikov, I., Sweetlove, L. *et al.* (2006). Enolase takes part in a macromolecular complex associated to mitochondria in yeast. *Biochim. Biophys. Acta*, **1757**, 1217–1228.
66. Entelis, N., Brandina, I., Kamenski, P., Krasheninnikov, I. A., Martin, R. P. & Tarassov, I. (2006). A glycolytic enzyme, enolase, is recruited as a cofactor of tRNA targeting toward mitochondria in *Saccharomyces cerevisiae*. *Genes Dev.* **15**, 1609–1620.
67. Mukai, A. & Endoh, H. (2004). Presence of a bacterial-like citrate synthase gene in *Tetrahymena thermophila*: recent lateral gene transfers (LGT) or multiple gene losses subsequent to a single ancient LGT? *J. Mol. Evol.* **58**, 540–549.
68. Brandt, U. (2006). Energy converting NADH:quinone oxidoreductase (Complex I). *Annu. Rev. Biochem.* **75**, 69–92.38.
69. Ohnishi, T., Kawaguchi, K. & Hagihara, B. (1966). Preparation and some properties of yeast mitochondria. *J. Biol. Chem.* **241**, 1797–1806.
70. Gabaldón, T., Rainey, D. & Huynen, M. A. (2005). Tracing the evolution of a large protein complex in the eukaryotes, NADH:ubiquinone oxidoreductase (Complex I). *J. Mol. Biol.* **348**, 857–870.
71. Küffner, R., Rohr, A., Schmiede, A. & Krülle, U. (1998). Involvement of two novel chaperones in the assembly of mitochondrial NADH:ubiquinone oxidoreductase (Complex I). *J. Mol. Biol.* **283**, 409–417.
72. Schulte, U. (2001). Biogenesis of respiratory complex I. *J. Bioenerg. Biomembr.* **33**, 205–212.
73. Heazlewood, J. L., Howell, K. A. & Millar, A. H. (2003). Mitochondrial complex I from Arabidopsis and rice: orthologs of mammalian and fungal components coupled with plant-specific subunits. *Biochim. Biophys. Acta*, **1604**, 159–169.
74. van Lis, R., Atteia, A., Mendoza-Hernández, G. & González-Halphen, D. (2003). Identification of novel mitochondrial protein components of *Chlamydomonas reinhardtii*. A proteomic approach. *Plant Physiol.* **132**, 318–330.
75. Perales, M., Eubel, H., Heinemeyer, J., Colaneri, A., Zabaleta, E. & Braun, H.-P. (2005). Disruption of a nuclear gene encoding a mitochondrial gamma carbonic anhydrase reduces complex I and supercomplex $I+III_2$ levels and alters mitochondrial physiology in Arabidopsis. *J. Mol. Biol.* **350**, 263–277.
76. Burger, G., Lang, B. F., Reith, M. & Gray, M. W. (1996). Genes encoding the same three subunits of respiratory complex II are present in the mitochondrial DNA of two phylogenetically distant eukaryotes. *Proc. Natl Acad. Sci. USA*, **93**, 2328–2332.
77. Feagin, J. E. (1992). The 6-kb element of *Plasmodium falciparum* encodes mitochondrial cytochrome genes. *Mol. Biochem. Parasitol.* **52**, 145–148.
78. Jackson, C. J., Norman, J. E., Schnare, M. N., Gray, M. W., Keeling, P. J. & Waller, R. F. (2007). Broad genomic and transcriptional analysis reveals a highly derived genome in dinoflagellate mitochondria. *BMC Biol.* **5**, 41.
79. Chaput, H., Wang, Y. & Morse, D. (2002). Polyadenylated transcripts containing random gene fragments are expressed in dinoflagellate mitochondria. *Protist*, **15**, 111–122.
80. Stiburek, L., Hansikova, H., Tesarova, M., Cerna, L. & Zeman, J. (2006). Biogenesis of eukaryotic cytochrome c oxidase. *Physiol. Res.* **55**, S27–S41.
81. Ludwig, B. & Schatz, G. (1980). A two-subunit cytochrome *c* oxidase (cytochrome aa_3) from *Paracoccus dentrificans*. *Proc. Natl Acad. Sci. USA*, **77**, 196–200.
82. Walker, J. E. & Dickson, V. K. (2006). The peripheral stalk of the mitochondrial ATP synthase. *Biochim. Biophys. Acta*, **1757**, 286–296.
83. Sabar, M., Gagliardi, D., Balk, J. & Leaver, C. J. (2003). ORFB is a subunit of F_1F_O-ATP synthase: insight into the basis of cytoplasmic male sterility in sunflower. *EMBO Rep.* **4**, 381–386.
84. Pícková, A., Potocký, M. & Houštěk, J. (2005). Assembly factors of F_1F_O-ATP synthase across genomes. *Proteins: Struct. Funct. Genet.* **59**, 393–402.
85. Bohnert, M., Pfanner, N. & van der Laan, M. (2007). A dynamic machinery for import of mitochondrial precursor proteins. *FEBS Letters*, **581**, 2802–2810.
86. Neupert, W. & Herrmann, J. M. (2007). Translocation of proteins into mitochondria. *Annu. Rev. Biochem.* **76**, 723–749.
87. Dolezal, P., Likic, V., Tachezy, J. & Lithgow, T. (2006). Evolution of the molecular machines for protein import into mitochondria. *Science*, **313**, 314–318.
88. Cermakian, N., Ikeda, T. M., Cedergren, R. & Gray, M. W. (1996). Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage. *Nucl. Acids Res.* **15**, 648–654.
89. Daga, A., Micol, V., Hess, D., Aeversold, R. & Attardi, G. (1993). Molecular characterization of the

transcription termination factor from human mitochondria. *J. Biol. Chem.* **268**, 8123–8130.
90. Wiesenberger, G. & Fox, T. D. (1997). Pet127p, a membrane-associated protein involved in stavility and processing in *Saccharomyces cerevisiae* mitochondrial RNAs. *Mol. Cell. Biol.* **17**, 2816–2824.
91. Wiesenberger, G., Speer, F., Haller, G., Bonnefoy, N., Schleiffer, A. & Schafer, B. (2007). RNA degradation in fission yeast mitochondria is stimulated by a member of a new family of proteins that are conserved in lower eukaryotes. *J. Mol. Biol.* **367**, 681–691.
92. O'Brien, T. W. (2002). Evolution of a protein-rich mitochondrial ribosome: implications for human genetic disease. *Gene*, **286**, 73–79.
93. Curgy, J. J., Ledoigt, G., Stevens, B. J. & André, J. (1974). Mitochondrial and cytoplasmic ribosomes from *Tetrahymena pyriformis*. Correlative analysis by gel electrophoresis and electron microscopy. *J. Cell Biol.* **60**, 628–640.
94. Smits, P., Smeitink, J. A. M., van den Heuvel, L. P., Huynen, M. A. & Ettema, T. J. G. (2007). Reconstructing the evolution of the mitochondrial ribosomal proteins. *Nucl.Acids Res.* **35**, 4686–4703.
95. Siew, N. & Fischer, D. (2004). Structural biology sheds light on the puzzle of genomic ORFans. *J. Mol. Biol.* **342**, 369–373.
96. Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005–1016.
97. Small, I., Peeters, N., Legeai, F. & Lurin, C. (2004). Predotar: A tool for rapidly screening proteomes for *N*-terminal targeting sequences. *Proteomics*, **4**, 1581–1590.
98. Schnarrenberger, C. & Martin, W. (2002). Evolution of the enzymes of the citric acid cycle and the glyoxylate cycle of higher plants. A case study of endosymbiotic gene transfer. *Eur. J. Biochem.* **269**, 868–883.
99. Emelyanov, V. V. (2003). Common evolutionary origin of mitochondrial and rickettsial respiratory chains. *Arch. Biochem. Biophys.* **420**, 130–141.
100. von Heijne, G. (1986). Mitochondrial targeting sequences may form amphiphilic helices. *EMBO J.* **5**, 1335–1342.
101. Hammen, P. K. & Weiner, H. (1998). Mitochondrial leader sequences: structural similarities and sequence differences. *J. Expt. Zool.* **282**, 280–283.
102. Ogishima, T., Niidome, T., Shimokata, K., Kitada, S. & Ito, A. (1995). Analysis of elements in the substrate required for processing by mitochondrial processing peptidase. *J. Biol. Chem.* **270**, 30322–30326.
103. Taylor, A. B., Smith, B. S., Kitada, S., Kojima, K., Miyaura, H., Otwinowksi, Z. *et al.* (2001). Crystal structures of mitochondrial processing peptidase reveal the mode for specific cleavage of import signal sequences. *Structure*, **9**, 615–625.
104. Kobayashi, T. & Endoh, H. (2005). A possible role of mitochondria in the apoptotic-like programmed nuclear death of *Tetrahymena thermophila*. *FEBS J.* **272**, 5378–5387.
105. Schäger, H. (1995). Native electrophoresis for isolation of mitochondrial oxidative phosphorylation protein complexes. *Methods Enzymol.* **260**, 190–203.
106. Brookes, P. S., Pinner, A., Ramachandran, A., Coward, L., Barnes, S., Kim, H. & Darley-Usmar, V. M. (2002). High throughput two-dimensional blue-native electrophoresis: A tool for functional proteomics of mitochondria and signaling complexes. *Proteomics*, **2**, 969–977.
107. Orias, E., Hamilton, E. P. & Orias, J. D. (2000). *Tetrahymena* as a laboratory organism: useful strains, cell culture, and cell line maintenance. *Methods Cell Biol.* **62**, 189–211.
108. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389–3402.
109. O'Brien, E. A., Koski, L. B., Zhang, Y., Yang, L., Wang, E., Gray, M. W. *et al.* (2007). TBestDB: a taxonomically broad database of expressed sequence tags (ESTs). *Nucl. Acids Res.* **35**, D445–D451.
110. Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. & Gibson, T. J. (1998). Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* **23**, 403–405.
111. Vinh, L. S. & von Haeseler, A. (2004). IQPNNI: moving fast through tree space and stopping in time. *Mol. Biol. Evol.* **21**, 1565–1571.
112. Whelan, S. & Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **18**, 691–699.
113. Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704.
114. Dutkiewicz, R., Schilke, B., Knieszner, H., Walter, W., Craig, E. A. & Marszalek, J. (2003). Ssq1, a mitochondrial Hsp70 involved in iron-sulfur (Fe/S) center biogenesis. Similarities to and differences from its bacterial counterpart. *J. Biol. Chem.* **278**, 29719–29727.
115. Schilke, B., Williams, B., Knieszner, H., Pukszta, S., D'Silva, P., Craig, E. A. & Marszalek, J. (2006). Evolution of mitochondrial chaperones utilized in Fe–S cluster biogenesis. *Curr. Biol.* **16**, 1660–1665.