COMPARATIVE PROTEOMICS: STUDIES ON THE COMPOSITION AND EVOLUTION OF THE MITOCHONDRIAL PROTEOME IN EUKARYOTIC MICROBES (PROTISTS)

by

Ryan M.R. Gawryluk

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
August 2011

DALHOUSIE UNIVERSITY

DEPARTMENT OF BIOCHEMISTRY & MOLECULAR BIOLOGY

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "COMPARATIVE PROTEOMICS: STUDIES ON THE COMPOSITION AND EVOLUTION OF THE MITOCHONDRIAL PROTEOME IN EUKARYOTIC MICROBES (PROTISTS)" by Ryan M.R. Gawryluk in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Dated: August 11, 2011

External Examiner: _____

Research Supervisor: _____

Examining Committee: _____

_____

_____

Departmental Representative: _____

DALHOUSIE UNIVERSITY

DATE: August 11, 2011

AUTHOR: Ryan M.R. Gawryluk

TITLE: COMPARATIVE PROTEOMICS: STUDIES ON THE COMPOSITION AND EVOLUTION OF THE MITOCHONDRIAL PROTEOME IN EUKARYOTIC MICROBES (PROTISTS)

DEPARTMENT OR SCHOOL: Department of Biochemistry & Molecular Biology

DEGREE: Ph.D. CONVOCATION: October YEAR: 2011

_____

Signature of Author

To my Jodie

Table of Contents

List of Tables

List of Figures

Abstract

Mitochondria are eukaryotic organelles derived in evolution from within the α subdivision of Proteobacteria. Although mitochondria are structurally and metabolically complex, modern-day mitochondrial genomes (mtDNA) encode only a small number of RNAs and proteins predominantly involved in adenosine triphosphate (ATP) formation through electron transport coupled to oxidative phosphorylation, as well as translation of mtDNA-encoded proteins. In humans, only 13 of the >1000 polypeptides that constitute the complete mitochondrial protein complement (proteome) are encoded in mtDNA; the remainder is encoded by nuclear DNA (nuDNA). It is therefore imperative to comprehensively catalog nuDNA-encoded mitochondrial proteins in order to understand holistically the evolution of mitochondria.

Mitochondrial proteome investigations of animals, fungi and land plants have dramatically altered our conception of mitochondrial evolution; in contrast to mtDNA-encoded proteins, few nuDNA-encoded mitochondrial proteins are demonstrably derived from the eubacterial progenitor of mitochondria, and many are found only in eukaryotes. Notably, however, little is known about the mitochondria of eukaryotic microbes (protists), which constitute the bulk of biochemical and genetic diversity within the domain Eucarya. The proteomic characterization of protist mitochondria is therefore crucial to fully elucidating mitochondrial function and evolution.

Employing tandem mass spectrometry (MS/MS), I have analyzed highly purified mitochondria from *Acanthamoeba castellanii* (Amoebozoa). In combination, nearly 750 nuDNA- and mtDNA-encoded proteins were identified. These data were used to catalog metabolic pathways and protein complexes, and to infer functional and evolutionary profiles of *A. castellanii* mitochondria. My analyses suggest that while *A. castellanii* mitochondria have many features in common with other eukaryotes, they possess several novel attributes and pronounced metabolic versatility.

An analysis of the *A. castellanii* electron transport chain (ETC) was also performed, utilizing a combination of blue native polyacrylamide gel electrophoresis (BN-PAGE), MS/MS and bioinformatic queries. A significant proportion of *A. castellanii* ETC proteins was identified, yielding several insights into ETC evolution in eukaryotes.

Lastly, I present two unusual cases of 'split' mitochondrial proteins: the iron-sulfur subunit SdhB of succinate:ubiquinone oxidoreductase (Complex II) in the phylum Euglenozoa and Cox1 of cytochrome *c*:O_2 oxidoreductase (Complex IV) in various eukaryotes, including *A. castellanii*. Functional and evolutionary implications of these findings are discussed.

List Of Abbreviations Used

γCA	γ-type carbonic anhydrase
$\Delta\psi_m$	electrical membrane potential
AAR	α-aminoadipate reductase
AB	ammonium bicarbonate
ACN	acetonitrile
AOX	cyanide-insensitive alternative oxidase
ASCT	acetate:succinate CoA transferase
ATP	adenosine triphosphate
bis-tris	bis-(2-hydroxyethyl)-amino-tris(hydroxymethyl)-methane
BLAST	basic local alignment search tool
BN-PAGE	blue native polyacrylamide gel electrophoresis
BSA	bovine serum albumin
bya	billion years ago
CI	Complex I; NADH:ubiquinone oxidoreductase
CII	Complex II; succinate:ubiquinone oxidoreductase
CIII	Complex III; ubiquinol:cytochrome *c* oxidoreductase
CIV	Complex IV; cytochrome *c*:O_2 oxidoreductase
CV	Complex V; F_oF_1 ATP synthase
DDM	*n*-dodecyl-β-D-maltoside
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EGT	endosymbiotic gene transfer
ER	endoplasmic reticulum
ERMES	ER-mitochondria encounter structure
EST	expressed sequence tag
EtBr	ethidium bromide

ETC	electron transport chain
$FADH_2$	flavin adenine dinucleotide
Fe-S	iron-sulfur
FMN	flavin mononucleotide
ICL	isocitrate lyase
IMM	inner mitochondrial membrane
IMP	inner membrane peptidase
IMS	intermembrane space
kDa	kilodalton
LSU	large subunit (ribosomal)
MDa	megadalton
MDH	malate dehydrogenase
MS	malate synthase
ML	maximum likelihood
MOPS	4-morpholinepropanesulfonic acid
MPP	matrix processing peptidase
MRO	mitochondrion-related organelle
MS/MS	tandem mass spectrometry
mtDNA	mitochondrial DNA
mTP	mitochondrial targeting peptide
NADH	nicotinamide adenine dinucleotide
nuDNA	nuclear DNA
OMM	outer mitochondrial membrane
ORF	open reading frame
PDH	pyruvate dehydrogenase
PEPCK	phosphoenolpyruvate carboxykinase
PFL	pyruvate formate-lyase

PFO	pyruvate:ferredoxin oxidoreductase
PMSF	phenylmethylsulfonyl fluoride
PPR	pentatricopeptide repeat
RCF	relative centrifugal force
RNR	ribonucleotide reductase
ROS	reactive oxygen species
RP-HPLC	reversed-phase high-performance liquid chromatography
rpm	revolutions per minute
SAM	sorting and assembly machinery
SCX-HPLC	strong cation-exchange high-performance liquid chromatography
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SL	spliced leader
SSU	small subunit (ribosomal)
TIM	translocase of the inner mitochondrial membrane
TOM	translocase of the outer mitochondrial membrane
Tris	tris(hydroxymethyl)-aminomethane

Acknowledgements

I would like to thank my supervisor, Mike Gray, for agreeing to take on one last student, even if it meant working on a thesis well into his retirement years. I was afforded quite a bit of intellectual freedom in terms of the projects that I decided to pursue, but Mike was always more than willing to give his invaluable input when I asked for it. I feel very fortunate to have worked with Mike, who is a great in mitochondrial research and a genuinely nice person.

When I first arrived, the Gray Lab was made up of a fairly large number of people; shortly thereafter, everyone left, except Dave Spencer and Murray Schnare. Dave and Murray have taught me almost everything that I know how to do in the lab (although they might not want to acknowledge it in public). They are, as others have noted many times, the most knowledgeable molecular biologists around. I have learned so much from them and consider them very good friends. Their acerbic wit and curmudgeonly ways have been a real highlight of my time here. I also thank former members of the Gray Lab, especially Tony Russell and Amanda Lohan.

I would like to thank the members of my Supervisory Committee: John Archibald, Andrew Roger and Christian Blouin. I greatly appreciate all of the effort and input they have given over the course of my Ph.D. studies.

I am indebted to various collaborators who have helped me with tandem mass spectrometry experiments. Firstly, I thank the York University Mass Spectrometry Group, including Daryl Smith, Ron Pearlman and Michael Siu, for their contributions to the *Tetrahymena thermophila* mitochondrial proteome project (Smith *et al*. 2007; not

presented here). More recently, I have had the opportunity to work with Dev Pinto and Ken Chisholm at the National Research Council of Canada Institute for Marine Biosciences. Ken has always been very willing to process samples for me and has taught me a lot about the workings of mass spectrometry. Most of the work presented in this thesis was possible only because Dev and Ken agreed to collaborate with me.

I am very grateful for the funding support that I have received over the past 6 years. I thank NSERC and the Killam Trusts for making graduate school much less stressful from a financial perspective.

My parents have always been very supportive of me. They are kind, generous people and their advice, encouragement and friendship continues to mean a lot.

I sincerely thank my wife, Jodie, who has been my best friend for 15 years and my wife for 6. We have experienced the ups and downs of junior high school, high school, undergrad and now grad school/parenthood together. I can't think of a better person to spend my life with. Lastly, I thank my 15-month-old son Reed, whose curious and playful personality makes me look forward to each day.

Chapter 1 Introduction

1.1 Mitochondrial Evolution

1.1.1 Genomics versus Proteomics

Mitochondria – metabolically and structurally diverse organelles derived in evolution via a symbiotic relationship between a eubacterium and a host of unresolved provenance [1] – are found in all aerobic eukaryotes studied to date. A number of eukaryotes (typically anaerobic and frequently parasitic) lack classical mitochondria; on the basis of small subunit ribosomal RNA (SSU rRNA) phylogenies, these amitochondrial eukaryotes were considered to branch early on the eukaryotic evolutionary tree [2]. Initially, it was proposed that these 'early-branching' lineages (collectively referred to as Archezoa [3]) had diverged from other eukaryotes prior to the acquisition of the protomitochondrial endosymbiont. However, more recent phylogenetic reconstructions dispute the early-branching positions of putative archezoan lineages [4], and the discovery of classical, nuclear DNA (nuDNA)-encoded mitochondrial-type proteins [5-7] that function within double-membraned organelles (hydrogenosomes [8-11] and mitosomes [12; 13]) suggests that primitively amitochondriate eukaryotes may never have existed. These observations have led several investigators to suggest that the establishment of mitochondria was perhaps *the* seminal event in the emergence of the eukaryotic lineage [14-16].

Mitochondria play vital roles in a wide variety of eukaryotic cellular processes, including energy generation (i.e., the electron transport chain; ETC), iron-sulfur (Fe-S) cluster biosynthesis, cell signaling and differentiation, programmed cell death

(apoptosis), amino and fatty acid metabolism, ion homeostasis and vitamin biosynthesis. Consistent with their functional significance in myriad cellular functions, the complete protein repertoire of diverse mitochondria (i.e., the mitochondrial proteome) consists of several hundred to >1000 proteins [17-26]. Analyses of the protein components of the mitochondrial proteomes of various eukaryotes (discussed below) indicate a bewilderingly complex and multifaceted evolutionary origin, with genetic contributions from nearly every source imaginable [27-29].

In stark contrast to the evolutionary complexity and lability of the mitochondrial proteome, the vestigial genome retained by mitochondria (i.e., mtDNA) portrays a much simpler picture [1], although the structure and expression of mtDNA is anything but uncomplicated [15; 30]. Genomic sequencing efforts spanning a broad range of eukaryotic taxa have provided strong evidence that mitochondria are descendants of α-proteobacteria [1]. Specifically, phylogenetic reconstructions of mtDNA-encoded rRNAs and proteins have suggested that the closest evolutionary relatives of mitochondria may be affiliated with either the *Rickettsiales* subdivision [31-34] or SAR11 clade [35] of α-Proteobacteria. Studies of the content, expression and organization of mitochondrial genomes across the eukaryotic domain have provided further significant insight into the origin of mitochondria [30; 36]. For instance, nearly all genes with identifiable homologs encoded by mitochondria are a subset of those encoded by the most bacteria-like mitochondrial genome, that of the heterotrophic freshwater jakobid flagellate, *Reclinomonas americana* [1; 37]. These genes specify RNAs and proteins associated with a limited range of functions, including coupled oxidative phosphorylation, translation of mtDNA-encoded proteins and, in some minimally derived mtDNAs, protein maturation and translocation

and gene expression. The most parsimonious interpretation of these observations, together with shared ribosomal protein operon organizations (and mtDNA-specific deletions within these operons [1; 38]) and the position of mitochondrial rRNA and protein sequences in phylogenetic reconstructions [1; 31; 34], is that mitochondria are monophyletic – that is, they arose only once in evolution.

1.1.2 Limitations of the Genomic Approach

Although comparative genomic analyses have provided important and novel insights into the evolutionary origin of mitochondria, this perspective is necessarily constrained by the small number of functionally biased genes encoded by mitochondria. For instance, the number of conserved proteins encoded by mitochondria ranges from 3 (all respiratory proteins) in *Plasmodium falciparum* [39], the causative agent of malaria, to 67 in *R. americana* [37]; in contrast, in *Rickettsia prowazekii*, an obligate intracellular α-proteobacterial parasite, the highly reduced genome codes for 834 proteins [40], whereas the free-living α-proteobacterium *Bradyrhizobium japonicum* encodes >8000 polypeptides [41]. Clearly then, the limited information present in mtDNA cannot account for the metabolic versatility of the endosymbiont, whatever its coding capacity initially was. Furthermore, mitochondrial genomics cannot address another crucial aspect of mitochondrial evolution: mitochondria are integrated organelles (as opposed to endosymbiotic bacteria) acquired > 1 bya and, as such, must perform many novel, eukaryote-specific roles, such as ATP export and protein import [28; 42; 43]. In sum, mitochondrial genomics provides only limited information about the protomitochondrial genome [44], the nature of the symbiosis that led to the establishment of mitochondria, and the proteome of modern-day mitochondria. From a functional perspective, mitochondrial

genomics ultimately tells us only that the ancestor of mitochondria had the capacity to perform coupled oxidative phosphorylation and express/translate the associated genes. In order to achieve a more holistic representation of mitochondrial evolution, it is imperative to identify and characterize nuDNA-encoded mitochondrial proteins as well.

1.1.3 On Becoming an Organelle

Attempts to understand the nature of the protomitochondrial genome by searching for genes of α-proteobacterial ancestry in eukaryotic nuclear genomes have provided important insights into the metabolic capacities of the endosymbiont, revealing 630 [44] or 842 [45] orthologous groups of putative protomitochondrial origin. Essentially, the rationale for these studies is that genes transferred from the protomitochondrial to the nuclear genome should bear an α-proteobacterial signature (uncovered using phylogenetic reconstructions), whether or not the proteins are localized in mitochondria in modern eukaryotes; accordingly, the functional classification of these proteins provides insight into protomitochondrial metabolism. Although this approach fails to account for the probability that certain protomitochondrion-derived proteins do not exhibit an α-proteobacterial signature [46] it has provided a valuable minimal representation of the protomitochondrion as a (probably facultatively) aerobic organism capable of electron transport, catabolizing lipids, glycerol and amino acids provided by the host cytosol, fructose/mannose metabolism and the synthesis of heme, biotin and Fe-S clusters, among others [44]. Notably, however, relatively few of the protomitochondrion-derived proteins encoded by eukaryotes are present in modern mitochondria [44], indicating that the biochemistry of modern mitochondria is substantially different than that of their evolutionary antecedents and that acquisition of the endosymbiont had other important

impacts on eukaryotic cellular function.

Although the precise number and nature of genes encoded by the protomitochondrion is not known (many, like those coding for proteins associated with cell wall biosynthesis, were lost from eukaryotes wholesale), what is apparent is that the protomitochondrial genome underwent extensive gene loss during its transformation into a mitochondrial genome, and that most of this transformation likely happened very early in eukaryotic evolution [1]. Genes that disappeared from the protomitochondrial genome have apparently done so for a variety of reasons [27]. For instance, many of these genes have likely been completely lost (i.e., deleted via mutational degradation) due to functional redundancy; that is, the functions of certain gene products encoded by the protomitochondrial genome were either superfluous (e.g., cell wall biosynthesis) or could be substituted by proteins encoded by the host (e.g., glycolysis). Conversely, another large proportion of the protomitochondrial coding repertoire appears to have been transferred to the host genome (endosymbiotic gene transfer; EGT), as detailed below. In some cases, the products of transferred genes provided novel functions to the host; for instance, it has been demonstrated that 13-18% of the proteins that constitute the peroxisomal proteome are most closely affiliated with α-proteobacterial homologs [47]. Peroxisomes are thought to be autogenously derived organelles [47; 48]. Importantly, however, another collection of transferred genes encodes proteins that have retained their ancestral function in mitochondria. Because functional mitochondrion-to-nucleus EGT still occurs in certain eukaryotic lineages (plants [49]), the process of EGT that led to the relocation of protomitochondrial genes is relatively well understood.

Successful EGT has several pre-requisites [50], including a gene duplication

generated via the escape of genetic material from mitochondria and integration into nuDNA. Mitochondrial release of genetic material can be attributed to the digestion of mitochondria by autophagous vacuoles [51] and the fusion and fission events characteristic of mitochondria. Subsequent to the stable integration of either mitochondrial genomic DNA or cDNA intermediates into the nuclear genome, probably via microhomology-mediated recombination [52; 53], the duplicated nuclear copy must acquire sequence elements (e.g., transcriptional promoters, polyadenylation signals) that permit expression in the nucleus and translation by cytosolic ribosomes. Moreover, the nuclear copy must encode information that targets the protein to the proper subcellular compartment; in the case of proteins targeted back to mitochondria, this signal is typically an N-terminal extension (mTP) or internal targeting signal [43] (i.e., within the protein sequence). Once the nuclear copy has acquired the necessary expression and targeting information, either the nuclear or mitochondrial copy of the gene may be lost, presumably via neutral mechanisms. An assumption underlying EGT is that functional nuclear and mitochondrial genes must co-exist for a period of time prior to the complete replacement of a mitochondrial gene by its nuclear counterpart.

In mitochondria, N-terminal mTPs are poorly conserved at the sequence level, but are enriched in basic, hydroxylated and certain hydrophobic residues and depleted in acidic amino acids [54]. In general, mTPs have in common the ability to form amphipathic α-helices [55] (with basic and hydrophobic faces) that specifically interact with protein translocation complexes [56] of the outer and inner mitochondrial membranes (TOM and TIM23 complexes, respectively). N-terminal mTPs, which typically direct proteins to the mitochondrial matrix or inner membrane and may contain multiple targeting signals, are

cleaved off after import by specific processing peptidases to yield a mature protein sequence. Conversely, other nuDNA-encoded mitochondrial proteins do not contain cleavable leader sequences; in these cases, the signals that direct proteins to mitochondria are less well understood. However, it is believed that internal structural determinants are necessary and sufficient for interaction with the relevant complexes (including TOM and TIM22). The ultimate evolutionary origins of mTPs and the protein complexes that direct the vectorial import of nuDNA-encoded mitochondrial proteins are still somewhat unclear, as protein import is not required in prokaryotes. Randomly generated peptides consisting only of Met, Arg, Ser and Leu are able to effect import *in vivo* [57], suggesting that mTPs are not especially complex. Moreover, certain bacterial [58] and mtDNA-encoded [59] proteins are predisposed to mitochondrial import. The complexes that direct protein import are mostly eukaryotic in origin [60; 61], although some complexes (e.g., SAM complex), and certain subunits of other complexes (e.g., Hsp70) have bacterial homologs [60; 61] that apparently serve alternate purposes.

The transfer of genes encoding essential proteins to the nucleus, enabled by the co-evolution of the mTP and protein import systems that permit the corresponding protein products to function in mitochondria, was a key step in the transition from protomitochondrion (endosymbiont) to mitochondrion [28; 62] as the fitness of the endosymbiont became tied to that of the host. These events were likely key to molding the modern-day mitochondrial proteome, as they not only allowed the retargeting of endosymbiont-derived proteins back to mitochondria, but also allowed targeting of proteins of other origins (e.g., host-derived) to mitochondria [27; 28; 45; 63]. For instance, once protein targeting/import mechanisms evolved, the host was able to tap the vast

energetic resources produced by the endosymbiont via coupled oxidative phosphorylation (ATP) through elaboration of a multipass inner membrane protein – the ADP/ATP translocator – that exports ATP from mitochondria. Obviously, ATP export would be detrimental or lethal to the free-living eubacterial progenitor of the mitochondrion. Thus, in a sense, the system designed to re-target and import endosymbiont-derived proteins has been 'hijacked' by the host in order to shape the mitochondrial proteome according to its own metabolic requirements [45].

1.1.4 Evolution of the Mitochondrial Proteome: Model Species

Mitochondria from animals [18; 20], fungi [19; 22] and, to a lesser extent, land plants [21; 64] were the subject of the earliest mitochondrial proteomic investigations. Perhaps unsurprisingly, the mitochondrial proteome of the yeast *Saccharomyces cerevisiae* remains the best defined, with over 800 authentic proteins identified thus far (estimated coverage of >85%) [19; 22; 65], although recent integrative analyses including data from MS/MS and GFP studies, along with literature curation, have significantly expanded our understanding of mammalian mitochondria [24]. Yeast mitochondrial proteins were the first to be subjected to a rigorous evolutionary analysis, the results of which demonstrated a surprising phylogenetic profile [27; 63]. Somewhat naïvely, on the basis of mitochondrial genome ancestry alone, one might have expected most of the mitochondrial protein complement to be of α-proteobacterial provenance; however, this is not the case. Although a large proportion of proteins involved in mitochondrial translation, protein maturation and energy transduction are most closely affiliated with α-proteobacterial homologs, the percentage of yeast mitochondrial proteins that could be confidently assigned as α-proteobacterial by means of phylogenetic inference was ~ 10% [63]. A

similar proportion of the nuDNA-encoded human mitochondrial proteome (15%) is derived from α-proteobacteria [29]. These proteins represent the α-proteobacterial 'core' of the mitochondrial proteome; because this subset of proteins affiliate with α-proteobacteria in phylogenetic reconstructions, it is generally assumed that the genes encoding these proteins were transferred from the protomitochondrial genome to the nucleus [1; 28; 62]. However, while this is the most parsimonious explanation, only proteins that are mtDNA-encoded in other lineages (e.g., the ATP synthase α subunit and SdhB) are unequivocally of protomitochondrial origin, whereas others (e.g., cytochrome *c*, Cpn60, Hsp70 and pyruvate dehydrogenase subunits) are always nuDNA-encoded and their proposed α-proteobacterial provenance is necessarily inferential [1]. In general, α-proteobacterial 'core' proteins are involved in bioenergetic processes (e.g., pyruvate dehydrogenase [63] and certain TCA cycle enzymes [66]), respiration and translation of mtDNA-encoded proteins (e.g., endosymbiont-derived components of the ETC and mitochondrial ribosome), Fe-S cluster biosynthesis and protein maturation/folding (e.g., chaperonins) [29; 63].

Another large proportion of yeast and human mitochondrial proteins are 'prokaryote-like': that is, eukaryotic homologs affiliate with a wide variety of prokaryotic groups, but not with α-Proteobacteria specifically. Importantly, however, no other specific prokaryotic groups are overrepresented, as α-Proteobacteria is [29]. The genes encoding these 'prokaryote-like' proteins could have been acquired in a variety of ways. For instance, some of the genes likely were, in fact, inherited from the protomitochondrion but have subsequently lost the phylogenetic signature that would allow us to confidently assign them as α-proteobacterial homologs (in essence, this would

be a misclassification based on the decay of evolutionary signal). Approximately 28% of human mitochondrial proteins are most closely affiliated with homologs from other proteobacteria [29]; it is tempting to speculate that these proteins, too, were derived from the protomitochondrial endosymbiont. Similarly, it may be that these genes were inherited from the protomitochondrion, but never possessed an α-proteobacterial phylogenetic signal [46]. Lateral gene transfer (LGT), by which genetic information is exchanged between different species, is rampant among modern-day bacteria [67]. Assuming that this process was frequent in bacteria at the time that the protomitochondrion was acquired, it follows that a proportion of the proteins encoded by the protomitochondrial genome could affiliate with the prokaryotic groups from which they were originally derived [46]. Alternatively, prokaryote-like genes may have been transferred directly to nuDNA via LGT at some other point in eukaryotic evolution by way of transient endosymbioses [68], or they may have been ancestral components of the host genome that acquired mitochondrial targeting information. Host-derived mitochondrial proteins may have provided a novel function in mitochondria or have functionally substituted for endosymbiont-derived proteins. For example, certain enzymes involved in mitochondrial tRNA metabolism, such as tRNA synthetases and tRNA maturation enzymes, are putatively host-derived [28; 29; 63] and replace the function of endosymbiont-derived homologs. The mitochondrial localization of these proteins can be rationalized by gene duplication events that target one copy to mitochondria [63], or by dual localization to multiple subcellular locations (i.e., certain tRNA synthetases participate in both cytosolic and mitochondrial translation [69; 70]). Other replacements of endosymbiont-derived genes, though 'viral' (T-odd phage) and not prokaryotic in origin, include

components of the mitochondrial DNA replication and transcription machinery, in particular the replicative DNA polymerase and helicase (Twinkle [71]) and RNA polymerase (substituting the multisubunit RNA polymerase still encoded by jakobid mtDNA [37; 72]). Many other proteins of non-α-proteobacterial origin are involved in mitochondrial metabolism, including various TCA cycle enzymes [66]. Intriguingly, it has been suggested that replacement of endosymbiont-derived genes is correlated with the presence or absence of permanent protein interaction partners [29] (as in the Complexity Hypothesis [73] – the proposal that the lateral transferability of a gene is more limited if its product interacts in large, complex networks). For example, host-derived proteins infrequently replace mitoribosomal and ETC components, although it does occur [74]. Taken together, the lack of phylogenetic affiliation with α-Proteobacteria does not necessarily negate the possibility that some of these genes were inherited from the endosymbiont by transfer to the nuDNA; however, it does suggest that the prokaryote-related portion of the proteome likely has multiple and diverse origins.

More surprisingly, it was reported that nearly half of all proteins composing the yeast mitochondrial proteome do not have a detectable prokaryotic homolog [63]. Most proteins in this cohort have homologs in animals and plants, suggesting that they were acquired early in mitochondrial evolution, before the diversification of the major eukaryotic lineages [62]. Among the widespread eukaryote-specific additions are many membrane proteins [29], including those involved in sensing cellular conditions, metabolite transport, protein import and mitochondrial fission/fusion. Many of these proteins perform functions not required of (or deleterious to) endosymbionts and reflect the full transformation of mitochondria into organelles [28]. Additional widespread eukaryote-

specific proteins are new subunits of ancient endosymbiont-derived complexes, including the respiratory chain enzyme complexes [75; 76] (discussed below) and mitochondrial ribosomes [77].

Lastly, approximately 17% of yeast mitochondrial proteins examined do not have homologs outside of Fungi [63]; a similar proportion of proteins with unidentified function was reported for the *Arabidopsis thaliana* (20%) [21] and human (19%) [18] proteomes . Certain of these proteins, referred to as lineage- or species-specific, have presumably evolved fairly recently in the nuclear genome; conversely, others likely have homologs in other mitochondrial systems that are too divergent to permit confident identification using standard bioinformatic techniques.

Although similar trends shape the human and yeast mitochondrial proteomes, they differ considerably in content. It was reported that only 42% of yeast mitochondrial proteins have orthologs in human mitochondria [28] (incomplete proteome coverage appears to have only a minor effect on this result), indicating that mitochondrial proteomes can be quite variable. Gene loss accounts for a certain proportion of these differences; for instance, ETC Complex I (NADH:ubiquinone oxidoreductase), comprising 45 subunits in mammals, is missing altogether in yeast. Similarly, duplication of genes encoding either mitochondrial or non-mitochondrial proteins, followed by divergence and neo-functionalization, provides raw material for novel mitochondrial function [78].

Although rather strict expect values (E-value; the expected number of hits from a database of a given size with the observed alignment score or greater) were employed in

defining orthologs of yeast mitochondrial proteins (a cutoff of 10^{-10}) [63], meaning that prokaryotic and/or eukaryotic homologs of short and highly divergent proteins may have been misclassified, these findings have nonetheless radically altered our conception of mitochondrial evolution. On top of the relatively modest, albeit essential, contribution of the endosymbiont, the recruitment of novel proteins from the host, both early in eukaryotic evolution and within established lines, and potentially from other prokaryotes, suggests an intricate and mosaic history for the mitochondrial proteome. Thus, it has become very apparent that additional proteomic investigations of mitochondria are required if we wish to understand the forces driving the evolution of mitochondria in concert with nuclear genomes throughout the eukaryotic radiation. While these findings in no way refute inferences drawn from mitochondrial genome sequencing efforts (i.e., that mitochondria are a monophyletic assemblage ultimately derived from within α-Proteobacteria), they demonstrate that mitochondria have been altered significantly relative to their eubacterial progenitors; indeed, one wonders whether the same conclusions about mitochondrial origins would have been made on the basis of nuDNA-encoded mitochondrial proteins (i.e., in the absence of mtDNA). However, it is important to stress that these proteome-based conclusions about mitochondrial evolution and function are based predominantly on data drawn from analyses of fungal, animal and land plant mitochondria; consequently, it is necessary to study more diverse taxa in order to make more robust generalizations about the protein content and evolution of mitochondria. This relative paucity of data inherently limits and biases the conclusions drawn from these studies because the predominantly unicellular protists, which encompass the bulk of eukaryotic cellular, ecological and evolutionary diversity, had,

until recently, been ignored.

1.1.5 Mitochondrial Proteomes of Protists (So Far)

Recent forays into characterizing the mitochondrial proteomes of eukaryotic microbes (protists) have been facilitated by the increased feasibility of sequencing nuclear and mitochondrial genomes. In 2007, our group performed the first comprehensive MS/MS-based analysis of a protist mitochondrial proteome (*Tetrahymena thermophila*) [23]; since then, MS/MS-based examinations of mitochondria from the green alga *Chlamydomonas reinhardtii* [25] and the causative agent of African sleeping sickness, *Trypanosoma brucei* [79], have been completed. Furthermore, recent proteomic investigations of mitosomes [80; 81] (degenerate mitochondrion-related organelles of parasitic anaerobes) have provided novel insights into these organelles.

From the perspective of general evolutionary trends, analyses of protist mitochondria (*T. thermophila* and *C. reinhardtii*, in particular) largely support conclusions drawn from human and yeast mitochondria. Approximately 11% and 12%, respectively, of *T. thermophila* [23] and *C. reinhardtii* [25] proteins group specifically with α-proteobacterial homologs (with acceptable bootstrap support), confirming a relatively small contribution from the protomitochondrial endosymbiont. As well, the functional profile of endosymbiont-derived proteins mirrors that reported in other organisms [23; 25]. Similarly, a large contingent of prokaryote-like proteins was identified, including in *T. thermophila* proteins such as superoxide dismutase, Hsp90 and various TCA cycle enzymes. Intriguingly, while a large percentage of the *T. thermophila* and *C. reinhardtii* mitochondrial proteomes comprises eukaryote-specific proteins, nearly half (45%) of the *T. thermophila* proteins lack homologs outside of ciliates, compared to 17% in *C.*

reinhardtii and roughly similar proportions in yeast and animals. While some of these lineage-specific proteins are probably extremely divergent proteins with homologs in other species, rendered unrecognizable due to fast protein evolution in ciliates, it is also likely that many of these proteins are novel, ciliate-specific additions to *T. thermophila* mitochondria [23].

Characterization of protist mitochondria has also furthered our understanding of proteome variability and versatility. Notably, in agreement with the notion that duplication of otherwise non-mitochondrial proteins contributes to mitochondrial proteome expansion [63; 78], a mitochondrion-targeted isoform of enolase (a glycolytic enzyme) with an mTP was reported in *T. thermophila* [23], indicating that at least some glycolytic steps may occur in mitochondria. Alternatively, mitochondrial enolase may serve a separate function [82; 83]. In *C. reinhardtii* mitochondria [25; 84], pyruvate:formate lyase, acetate kinase and phosphotransacetylase provide a metabolic route for anaerobic ATP generation not present in other mitochondrial proteomes described thus far. Further examinations of protist mitochondrial proteomes will be useful in answering a broad array of questions about mitochondrial evolution, including how similar mitochondrial functions are across the eukaryotes and whether the evolutionary trends observed in yeast and other model organisms hold in general.

1.1.6 Defining a Mitochondrial Proteome

As described above, $\geq$ 95% of proteins that function in mitochondria are encoded in nuDNA [19; 22; 43; 85]. Consequently, if we wish to achieve a comprehensive understanding of the biology and evolution of mitochondria, it is imperative that we characterize the complete proteome and not a limited number of proteins involved in a

metabolically narrow range of functions.

Numerous techniques have been utilized in an attempt to define the protein complement of mitochondria in a high-throughput manner. Several different bioinformatic approaches have been developed [86], including attempts to infer likely mitochondrial proteins via phylogenetic profiling [87], the assignment of mitochondrial proteins by physicochemical properties [88; 89], and recognition of cleavable mTPs by machine-learning programs [90-92]. Although useful in analyzing the likelihood of mitochondrial localization of a given protein when limitations are taken into account, strictly bioinformatic approaches are inappropriate tools for comprehensive proteomic investigations of mitochondria (and other subcellular fractions), as they suffer from a number of inherent systematic errors that result in large numbers of false positives and negatives [21; 25]. For instance, phylogenetic profiling [87] (assigning proteins as being specific to an organelle based on phylogenetic distribution of homologs) has limited use in the identification of 'novel' mitochondrial proteins that do not have recognizable homologs in other mitochondria-possessing cells; moreover, this approach incorrectly assigns re-targeted endosymbiont-derived proteins as authentic mitochondrial proteins. Similarly, algorithms that recognize cleavable mTPs are limited in that they are incapable of recognizing mitochondrial proteins with non-cleavable mTPs, which may constitute more than 40% of mitochondrial proteins [19; 65; 93]. Moreover, in analyzing mitochondrial proteins with cleavable mTPs, these algorithms are variably accurate; reported sensitivity and specificity values are discouraging (specificity is improved when multiple predictors are employed, but at the expense of sensitivity [17; 21]). In addition, the predicted mitochondrial proteome size scales with the total number of proteins encoded by an

organism, leading to unreasonable mitochondrial proteome size predictions in eukaryotes with large nuclear genomes [17].

In contrast, biochemical and/or molecular biological methods, which can be complemented by bioinformatic interrogations, have proven more useful in providing an accurate and comprehensive mitochondrial protein inventory. Although techniques such as global analysis of epitope- [94; 95] and GFP-tagged [96] proteins have been employed to help define the mitochondrial proteome of yeast, the most popular method relies on the efficient purification of mitochondria via density gradient centrifugation coupled to tandem mass spectrometric (MS/MS) analysis of peptides derived from exhaustive proteolytic digestion of proteins [97-99] from whole mitochondria [18-22], suborganellar compartments [100-102] or protein complexes [103-108]. This method, made possible through recent advances in peptide ionization technologies, is advantageous as it requires only that relatively small quantities of highly pure mitochondria be isolated (i.e., the organism does not need to be a genetically tractable system, as in the GFP- tagging experiments [96]) and that mitochondrial and nuclear genome sequences (and/or deeply-sequenced transcript libraries) be available for querying of tandem mass spectra. Results vary somewhat based on the complexity and composition of the proteome, instrumentation used, and the interpretation of the data, although several hundred different proteins are routinely identified in one experiment. There is, however, a small proportion of proteins refractory to this high-throughput approach; in particular, very low abundance proteins, hydrophobic proteins and proteins with few suitable tryptic peptides are often difficult to detect. MS/MS-based approaches also identify abundant non-mitochondrial proteins associated with a purified organellar fraction [18; 19; 23; 25]; as a result, careful scrutiny of

results is required in order to avoid the inclusion of false positives. Additionally, this approach cannot address dual targeting of proteins [109] (i.e., identified mitochondrial proteins may have important non-mitochondrial functions as well); recent estimates suggest that up to 25% of yeast mitochondrial proteins are dual targeted [110]. Moreover, proteins that are scored as contaminants of purified mitochondria may in fact be *bona fide* mitochondrial components. Indeed, the complete elucidation of the mitochondrial proteome in any organism requires combined large- and small-scale approaches, along with the examination of the proteome under various metabolic conditions and, in multicellular organisms, diverse tissue types and developmental stages [20; 24]. Nonetheless, MS/MS analysis of whole mitochondria provides an excellent first approximation of the complexity, size, metabolic function and evolution of mitochondria from a particular species.

I have performed a combined MS/MS and bioinformatic analysis of mitochondria from the amoeboid protist, *Acanthamoeba castellanii* - a member of the supergroup Amoebozoa, the sister lineage to Opisthokonta (animals + fungi) [111]. I have identified in excess of 700 mitochondrial proteins from whole mitochondria, membrane-enriched fractions and soluble protein-enriched fractions via MS/MS analysis and have performed various bioinformatic interrogations in order to annotate and characterize proteins and metabolic pathways. My results suggest that *A. castellanii* mitochondria contain many of the proteins characteristic of previously characterized mitochondria, but also exhibit a number of metabolic and evolutionary novelties.

1.2 Annotation of the Mitochondrial Proteome

1.2.1 Annotating Mitochondrial Proteins of Unknown Function

Functional assignments are lacking for a significant proportion of all mitochondrial proteins [85], even in well-studied systems such as human [24] and yeast [22]. Among these poorly understood mitochondrial proteins are those bearing conserved domains that provide tantalizing hints to protein function but do not elucidate the precise molecular role (e.g., pentatricopeptide repeat (PPR) proteins are often involved in RNA binding and metabolism, but can participate in diverse functions, ranging from intron splicing to translational activation and RNA editing [112]), as well as proteins apparently lacking conserved domain features and homologs in other species: often referred to as 'hypothetical' mitochondrial proteins. This problem is even more pronounced in the study of protist mitochondria; for instance, in *T. thermophila*, nearly one-half of all MS/MS-identified mitochondrial proteins lack homologs outside of ciliates [23]. As a result, a major challenge to a full understanding of the mitochondrial proteome is the functional annotation of its constituent proteins [85]. This may be achieved using more targeted approaches, in particular those that provide insight into suborganellar localization (i.e., which mitochondrial compartment is a protein associated with), protein-protein interactions, and the consequences of functional impairment.

Numerous mitochondrial proteins reside in multiprotein complexes, including, most prominently, the ETC and mitochondrial ribosomes (mitoribosomes). Intriguingly, as noted earlier, many mitochondrial multiprotein complexes have α-proteobacterial antecedents, although their subunit complexity has expanded markedly in the eukaryotic line [75-77] with the addition of supernumerary subunits. Many of the latter appear to have

been 'invented' within the eukaryotic domain (i.e., they lack detectable prokaryotic homologs); a large proportion has a phylogenetically broad distribution – generally interpreted as an early eukaryotic feature – while others appear to be more recent additions in particular eukaryotic lineages [75; 77]. Conversely, certain other supernumerary subunits have bacterial homologs, although these particular proteins were apparently not components of the ancestral protomitochondrial complex [75; 113-115]. Thus, the proteomes of these multiprotein enzyme complexes can be viewed as a microcosm of the mitochondrial proteome as a whole, encompassing a number of core subunits that can be traced to the α-proteobacterial ancestor of mitochondria and a larger set of proteins that has evolved subsequent to the establishment of mitochondria.

Biochemical cataloging of the individual proteins that constitute these multiprotein complexes is a valuable tool in functional annotation, as it provides information on their suborganellar localization and functional context and, by extension, a rudimentary understanding of their putative role on the basis of 'guilt by association' with proteins of known function. Moreover, the characterization of multiprotein complexes in diverse eukaryotes permits important insights into the evolutionary processes that shape the mitochondrial proteome; for instance, it will help in the understanding of the 'types' of proteins added to complexes and permit the mapping of gene/protein gain and loss patterns.

1.2.2 The Mitochondrial Electron Transport Chain (ETC)

In aerobic eukaryotes, coupled oxidative phosphorylation via the mitochondrial ETC is responsible for generating the bulk of cellular energy in the form of adenosine triphosphate (ATP). The ETC is composed of four multisubunit oxidoreductase

complexes known as Complex I (CI; NADH:ubiquinone oxidoreductase), Complex II (CII; succinate:ubiquinone oxidoreductase), Complex III (CIII; ubiquinol:cytochrome *c* oxidoreductase) and Complex IV (CIV; cytochrome *c*:O_2 oxidoreductase); in general, the naming of ETC complexes reflects their sequential order within the respiratory chain (although the naming of CI versus CII is arbitrary in this regard as they represent separate primary electron inputs). These complexes, with the exception of CII, couple the energy derived via the transfer of electrons from NADH and $FADH_2$, ultimately to O_2, to the translocation of protons from the mitochondrial matrix to the intermembrane space. The vectorial transport of protons across the inner mitochondrial membrane results in a proton gradient that is harnessed by Complex V (CV; F_oF_1 ATP synthase), which employs a rotary catalytic mechanism in order to generate ATP from ADP and P_i. Although the ETC complexes have traditionally been considered as distinct enzyme entities, evidence for the stoichiometric association of certain individual ETC complexes with each other (termed respiratory supercomplexes) has increased dramatically in recent years [116-120].

In comparison with bacteria, two general trends are apparent in the mitochondrial ETC [121]. Firstly, mitochondria have become streamlined with respect to the number of distinct complexes or pathways by which electron transport can occur. Specifically, free-living bacteria employ a highly branched and redundant ETC capable of metabolizing a broader spectrum of compounds and, as such, require a greater diversity of oxidoreductases in order to accommodate this variability [121]. Like mitochondria, reduced metabolic versatility is also observed in endoparasitic and endosymbiotic bacteria [122], suggesting convergent evolutionary trajectories, possibly due to similarly nutrient-rich intracellular conditions. Notably, however, certain of these observations are biased

towards a streamlined, mammal-centric view of mitochondrial electron transport and do not take into consideration the branched nature of electron transport in fungal, plant and protist mitochondria [123; 124].

Secondly, and in stark contrast to a reduction in metabolic flexibility, the intricacy of the complexes that constitute the classical mitochondrial ETC has increased considerably in comparison to bacterial counterparts [75; 76; 121; 125]. Whereas mitochondrial ETC complexes have retained all of the 'core' subunits present in the bacterial complexes (many are encoded in mtDNA and are considered to be vertically inherited from the ancestor of mitochondria [1]), a comparison of the proteins that comprise each ETC complex in mitochondria and bacteria demonstrates the evolution of myriad 'novel' subunits in eukaryotes. This expansion in complexity relative to bacteria is most dramatically observed in CI [75; 76], although it is also observed to some extent in all ETC complexes.

In contrast to the 14 (Nad1-Nad11 and Nad4L along with 24-kDa and 51-kDa subunits) and 17 subunits (also including B17.2, AQDQ and 13 kDa subunits) that comprise CI in *E. coli* and *P. denitrificans* [126], respectively, mammalian and plant CI are composed of 45 [107] and 49 proteins [108], respectively. These comparisons highlight the fact that the number of CI subunits has roughly tripled in relation to the homologous bacterial complex. Comparative genomic analyses of animals, fungi, land plants and green algae demonstrate that the majority of novel CI proteins are broadly distributed across eukaryotes [75]; in relation to the 17-subunit α-proteobacterial complex [126], 16 proteins are believed to be ancestral eukaryote-specific CI subunits, yielding a predicted ancestral mitochondrial enzyme of 33 subunits. Although some novel eukaryotic CI proteins have

prokaryotic non-CI homologs, including e.g., subunit SDAP, the acyl-carrier protein [113], most do not have bacterial homologs and appear to be genuine eukaryotic inventions. The final number of CI subunits in a particular lineage (for example, 49 in land plants, in contrast to the 33-subunit eukaryotic core) is thought to be a result of lineage-specific gains and losses. For instance, *A. thaliana* and *C. reinhardtii* CI contain 5 [108] and 3 [104] subunits, respectively, related to γ-type carbonic anhydrases. In *A. thaliana*, these subunits form a novel domain of unknown function [115]. No γ-type carbonic anhydrases are encoded by animals or fungi; as a result, it was initially concluded that these proteins were added to CI after the divergence of Opisthokonta and Plantae, but prior to the divergence of green algae and land plants [75; 127].

As is the case with the mitochondrial proteome as a whole, the vast majority of mitochondrial ETC research has been carried out in a small number of species, most of which are animals [107; 128], fungi [103; 129; 130], and land plants [105; 108; 131; 132]. Consequently, proteomic investigations of the ETC from a greater diversity of organisms are required in order to fully comprehend the evolution and function of these complexes. For example, it is not yet clear to what extent the expansion of ETC complexes (in comparison to bacteria) is a general feature of all eukaryotes versus multicellular lineages in particular. Notably, only 23 of the 33 ancestral CI subunits were reported in a combined MS/MS and bioinformatic analysis of the ciliated protozoon, *T. thermophila* [23], although the highly divergent nature of *T. thermophila* proteins likely accounts for much of this apparent discrepancy.

Relatively few direct biochemical investigations of protist ETC complexes have been performed. Examinations thus far have provided several intriguing glimpses of

structurally and compositionally conservative [133; 134] and unusual [106; 135-139] ETC complexes. It is, however, unclear how representative certain results are, especially in regard to highly unusual complexes, since much attention has been paid to ETC complexes from divergent parasitic eukaryotes, such as *T. brucei*.

A phylogenetically broader understanding of ETC composition will provide additional details about the evolutionary trajectory of particular ETC complexes and subunits. In some cases, as noted above, comparative analyses of ETC proteins considering only animals, fungi and land plants have highlighted apparently lineage-specific additions to ETC complexes [75; 115]; however, the conclusion of lineage-specific gains with such a narrow phylogenetic sampling neglects the possibility of lineage-specific losses as an alternative hypothesis and highlights the requirement for more data. Furthermore, the direct characterization of protist ETC complexes will potentially allow the identification of both novel protein subunits and highly divergent, but phylogenetically conserved, subunits. In particular, using strictly bioinformatic techniques, it is sometimes not possible to confidently discern classical subunits (often, but not always serving structural roles). For example, in an analysis of whole mitochondria from the ciliate *T. thermophila*, we failed to identify homologs of a number of conserved CV proteins [23], whereas subsequent MS/MS analysis of purified *T. thermophila* CV uncovered putative homologs for several of these 'missing' subunits, including Atpa (Atp6), an essential and catalytically important component [139]. Effectively, the characterization of individual ETC complexes provides a much smaller dataset in which to search for 'missing' subunits and may increase the probability of identifying them. Thus, the elucidation of protist ETC complex composition will be

important to our understanding of the composition and evolution of these complexes across the eukaryotic domain and the annotation of novel mitochondrial proteins.

1.2.3 Blue Native Polyacrylamide Gel Electrophoresis (BN-PAGE)

In order to analyze the composition and ultimately reconstruct the evolution of respiratory chain complexes, it is necessary to employ efficient and high-resolution protein purification techniques. There are a variety of methods by which respiratory chain complexes (or any protein complex) can be isolated, including (combinations of) immunoprecipitation with highly specific monoclonal or polyclonal antibodies, affinity chromatography, size-exclusion chromatography and others [103; 126; 128; 129]. Although these methods are ideal for the detailed characterization of a particular protein complex, their usefulness for a large-scale characterization of all five respiratory complexes is limited in several ways, especially for work in organisms that are not well studied or genetically tractable, like many protists. For instance, chromatographic methods typically require large quantities of starting material (often hundreds of milligrams) for the characterization of a single complex - a prohibitive amount in many non-model systems. Similarly, a broad ETC characterization using immunological methods would require several specific antibodies. Although antibodies that recognize mammalian and fungal ETC complex subunits are commercially available, heterologous antibodies are unlikely to be specific enough for many protistan systems, especially for fast-evolving lineages. Consequently, the production of multiple antibodies (at least one per complex) would be both expensive and inefficient.

BN-PAGE is a native electrophoretic technique that was developed as a means to

achieve a relatively efficient and inexpensive analysis of mitochondrial ETC complexes [140; 141]. Based on the gentle solubilization of protein complexes with non-ionic detergents and migration of these in the presence of an electric field due to a charge shift imparted by the binding of the anionic dye Coomassie Brilliant Blue G-250 to proteins, BN-PAGE is a high-resolution method capable of efficiently separating all ETC complexes in a single experiment. Little material is required (~ 50-100 μg per lane), approximate estimates of protein complex size and relative abundance can be made, and few specialized instruments are required. Furthermore, BN-PAGE is compatible with a variety of other valuable downstream techniques, including in-gel enzyme activity assays for ETC complexes [142], 2D BN/SDS-PAGE for visualization of protein profiles [141], and sensitive MS/MS analyses. In sum, BN-PAGE is a valuable technique for examination of the ETC, especially for the large-scale characterizations of multiple complexes in non-model organisms not amenable to other approaches.

I have used BN-PAGE, in combination with in-gel enzyme assays, 2D BN/SDS-PAGE, MS/MS and sensitive bioinformatic searches, to elucidate the structure, protein composition and evolution of the *A. castellanii* mitochondrial respiratory chain. I present evidence for a relatively conservative ETC along with several unexpected results, such as γ-type carbonic anhydrases as components of *A. castellanii* CI (and possibly CI in most eukaryotes) and a stable dimeric, relatively divergent CV. My findings highlight the importance of a phylogenetically broad comparative approach to understanding the evolution of these complexes and how direct characterizations of protein complexes can aid in the annotation of conserved, but divergent, subunits that cannot be confidently identified on the basis of bioinformatic searches alone.

1.3 The Occurrence of 'Split' Mitochondrial Proteins

1.3.1 Split Genes Encoding Mitochondrial Proteins

Increased efforts to determine nuclear and mitochondrial genome sequences have provided the raw material for many novel insights into the evolution of eukaryotes. Among these insights are several recent discoveries of 'fissioned' genes that encode 'split' mitochondrial proteins [135; 143-147]. Specifically, in these cases genes are split (and usually rearranged) such that they encode separate N- and C-terminal halves of the classical protein sequence. Importantly, and in contrast to other phenomena, such as *trans*-splicing [148] or split-intein splicing [149], the split proteins remain covalently discontinuous: i.e., separate, mature halves of the classical protein presumably interact in *trans* in order to fulfill the function of the singular ancestral protein. In several ways, this situation is reminiscent of independently transcribed, sub-genic units of fragmented ribosomal RNA species [150; 151], which interact in *trans* via complementary base-pairing interactions. Perhaps more saliently from a functional perspective, the phenomenon of α-complementation [152], employed in blue-white screening, relies upon the *trans* interaction of two (artificially) split LacZ peptides.

1.3.2 Examples of Split Genes

Fissioned genes encoding split mitochondrial proteins are arranged in three different ways with respect to the genome(s) in which they are encoded. 1) In certain cases, both halves of the split protein are encoded in mtDNA. For instance, Nad1 – a component of CI – is specified in the mtDNA of the ciliates *Tetrahymena pyrifomis* and *Paramecium aurelia* by two distinct genes (*nad1_a* and *nad1_b*) located on opposite

strands [143]. Analysis of *nad1_a* and *nad1_b* transcripts from *T. pyriformis* revealed that they are not *trans*-spliced into a single mRNA unit [144], suggesting that mature Nad1_a (284 aa) and Nad1_b (59 aa) are distinct proteins in CI, although a post-translational fusion cannot be formally discounted [30]. Similarly, whereas a component of the cytochrome *c* biogenesis system is encoded by one gene in the mtDNA of certain angiosperms, it is encoded by two separately transcribed mtDNA loci in others, including *A. thaliana* [153; 154]. 2) In other instances, both halves of fissioned genes specifying mitochondrial proteins are in nuDNA. In particular, it has been demonstrated that Cox2 (a hydrophobic component of CIV that is mtDNA-encoded in most eukaryotes) is encoded by two distinct nuclear genes (*cox2a* and *cox2b*) in the green algae *C. reinhardtii* and *Polytomella* sp. [135], as well as in apicomplexans and dinoflagellates [146; 147]. Notably, biochemical characterization of *C. reinhardtii* CIV has confirmed that it contains distinct COX IIa and COX IIb subunits, indicating that the halves are independently transcribed, translated and subsequently imported into mitochondria, where they are incorporated into CIV [135]. 3) In the final scenario, the distinct portions of a split protein are encoded by two genomes; that is, one half is mtDNA-encoded whereas the other is nuDNA-encoded. In contrast to *C. reinhardtii*, COX IIa from the related green alga *Scenedesmus obliquus* has been retained in mtDNA [155], while COX IIb is nuDNA-encoded [146]. This suggests that the fission of *cox2* had already occurred in the common ancestor of *C. reinhardtii* and *S. obliquus* and that differential effects of endosymbiotic gene transfer account for the current locations of *cox2a* and *cox2b*. The deviant genetic code of *S. obliquus* mitochondria [155] may represent a barrier to the functional transfer of *cox2a*. A similar situation was reported for *rpl2* (encoding ribosomal protein L2) in eudicot mitochondria

[145]. In certain eudicots the 5′ portion of classical *rpl2* is mtDNA-encoded, whereas it is nuDNA-encoded in other eudicots; the 3′ portion appears invariably to be nuDNA-encoded. Lastly, *nad11* (a CI protein) has split, possibly multiple times, in various stramenopiles. Two adjacent genes in *Phaeodactylum tricornutum* mtDNA encode classical Nad11; conversely, the mtDNAs of *Ectocarpus siliculosus*, *Pylaiella littoralis and Cafeteria roenbergensis* encode only a small, N-terminal portion of Nad11 [156-158]. A separate, C-terminal portion of Nad11 is encoded in the *E. siliculosus* nuclear genome [158] – presumably, a similar situation exists in the unsequenced *P. littoralis* and *C. roenbergensis* nuclear genomes. To the best of my knowledge, these examples represent all reported cases of split mitochondrial proteins, aside from those I will detail herein.

1.3.3 Process of Gene Fission

The processes that result in split mitochondrial proteins are essentially a special case of those associated with EGT, as detailed above. Specifically, it is important to note that (complete or partial) gene duplication events, with both copies initially specifying the original function of the protein portion they encode, are necessary in order for gene fission to occur. The gene duplication could result in two mtDNA-encoded copies, or one mtDNA- and one nuDNA-encoded copy. For instance, the fission of ciliate *nad1* obviously occurred in mtDNA [143; 144], whereas Adams *et al*. have proposed that the split of eudicot *rpl2* was enabled by the acquisition of a stop codon in the 5′ portion of classical *rpl2* subsequent to the partial duplication and functional relocation of the 3′ portion to the nucleus [145]. When relocation to nuDNA occurs, the nuDNA-encoded duplicate must acquire necessary expression, translation and mitochondrial targeting information before the original copy can be partially lost. Subsequently, if the protein

encoded by a duplicated sequence becomes fixed as a partial source of the original function in question, presumably by neutral processes, the constraints on the non-required coding portion of the original copy are relaxed.

1.3.4 Potential Insights Provided by Split Proteins

While at first glance, split mitochondrial proteins may appear to be nothing more than an evolutionary oddity, their existence has the potential to inform biological research on a number of fronts. Several investigators have argued that rare genomic changes, including gene fusion and fission events, should be considered as highly improbable (unique) molecular events that can be used as evolutionary markers independent of the many assumptions associated with phylogenetic models [146; 159-161]. In particular, phylogenetic reconstructions based on comparison of protein or nucleic acid sequences that aim to untangle deep eukaryotic relationships often suffer from artifacts resulting from the vast evolutionary distances separating major eukaryotic groups [162]. If gene fission events are indeed unique, the co-occurrence of the same split mitochondrial gene in distinct evolutionary lineages would provide concrete evidence for shared ancestry, potentially alleviating the problems associated with sequence comparisons. To this end, the discovery of fissioned genes encoding mitochondrial proteins has already provided intriguing and highly controversial results. Specifically, the co-occurrence of split *cox2* in green algae and plastid-bearing alveolates (apicomplexans and dinoflagellates but not ciliates) has been used to argue that the ancestor of the eukaryotic endosymbiont that gave rise to the alveolate plastid (apicoplast) was a green alga [146]. The proposed scenario for this endosymbiotic gene replacement is that both of the split *cox2* genes, already present in the nuDNA of a *Chlamydomonas*-like organism, were transferred to the

nucleus of the apicomplexan/dinoflagellate ancestor, eventually replacing the classical *cox2* encoded in mtDNA or nuDNA. This finding is in direct contrast with other strong evidence that the ancestor of the apicoplast was derived from an association with red algae [163; 164]. Interestingly, Waller and Keeling have provided phylogenetic evidence indicating that the split apicomplexan/dinflagellate *cox2* is more closely related to the very divergent, mtDNA-encoded ciliate *cox2*, which contains a large insertion specifically where *cox2* is split in other alveolates [165]. This finding makes sense from a phylogenetic perspective (apicomplexans, dinoflagellates and ciliates are all alveolates) [111], but suggests that the fission of *cox2* has occurred in *precisely* the same way more than once. Further examples of (and careful analyses of) split mitochondrial genes/proteins are required in order to resolve whether they truly are unique molecular events capable of pruning and refining the eukaryotic tree or stunning examples of convergent evolution that offer false hope of true phylogenetic insight.

Aside from phylogenetic implications, the presence of split proteins may provide novel perspectives on the forces guiding the evolution, structure and function of said proteins and the complexes into which they integrate. Most obviously, but surprisingly, split proteins demonstrate that a given protein does not have to remain a single continuous entity in order to execute its function, so long as the parts of that protein are able to interact properly. It is not clear how split proteins achieve proper association with each other, or how they integrate into protein complexes. In some cases, the existence of N- and C-terminal protein extensions (specifically, a C-terminal extension of the N-terminal half and an N-terminal extension of the C-terminal half) is suggested to facilitate this interaction [135], although direct support for this proposal is lacking. Intriguingly, it

appears that certain proteins are split *between* conserved protein folding domains. For instance, Nad11 consists of a short N-terminal Fe-S cluster binding domain (fer2) and a larger C-terminal domain with a putative molybdopterin-binding domain; in stramenopiles, the N-terminal domain is mtDNA-encoded while the C-terminal domain is nuDNA-encoded [156-158]. It may be that gene fissions in between regions encoding functional domains are more easily permitted than fissions within functional domains, as this may perturb protein function excessively. Lastly, in regards to genome annotation and protein function, a general awareness of the phenomenon of split genes/proteins is important. In particular, investigators may be misled to believe that the identification of a truncated version of a gene in a given genome suggests that the remainder of the protein (the 'missing' part, possibly encoded in another genome in the same cell) is less functionally important, as it is not conserved across all species [156]. Quite the opposite is true, however, if the other portion of the protein is in fact retained: i.e., the retention of a gene portion after a fission event argues strongly that the corresponding protein region serves an important role.

Below I present an analysis of two gene fission events that I discovered (somewhat serendipitously) during my studies. The first is a fission of the Fe-S subunit of mitochondrial succinate dehydrogenase (CII) in *Euglena gracilis* and trypanosomatids (members of the phylum Euglenozoa) that resulted in the presence of two nuDNA-encoded halves, termed SdhB-n and SdhB-c. I also describe a fission of mitochondrial *cox1* that has resulted in the relocation to the nucleus of a small *cox1* fragment (encoding a hydrophilic region of intact, ancestral Cox1, termed Cox1-c) of this otherwise hydrophobic, universally mtDNA-encoded protein. Notably, it was found that numerous

eukaryotes encode a Cox1-c in nuDNA and a C-terminally truncated Cox1 in mtDNA that lacks the portion corresponding to Cox1-c (termed Cox1(-)). These two gene fissions provide further examples of split mitochondrial proteins, yielding several structural and evolutionary insights.

Chapter 2 A Comprehensive Proteomic Investigation of *Acanthamoeba castellanii* Mitochondria

This chapter includes unpublished work (in prep.) by **Gawryluk, R.M.R**, Chisholm, K.A., Pinto, D.M. & Gray M.W.

2.1 Abstract

Most mitochondrial proteomic analyses to date have focused on animals, fungi and plants. Recently, progress in elucidating the protein content of mitochondria from eukaryotic microbes (protists) has been made; however, sampling remains limited and much remains to be learned about the proteins that constitute protist mitochondrial proteomes in a broad sense. I present a combined tandem mass spectrometry (MS/MS) and bioinformatic investigation of mitochondria from the amoeboid protist *Acanthamoeba castellanii* (supergroup Amoebozoa). In sum, MS/MS analyses of whole mitochondria, along with analyses of fractions enriched in either membrane or soluble proteins, have led to the identification of peptides from 716 nucleus-encoded and 32 mitochondrion-encoded proteins (748 in total). Bioinformatic interrogations of *A. castellanii* sequence databases have uncovered a further 113 putative mitochondrial proteins (861 in total). Among those identified by MS/MS are hydrophobic and hydrophilic proteins representing all mitochondrial subcompartments. The results are discussed in the context of a number of important mitochondrial complexes and metabolic pathways. In addition, a characterization of mitochondrial electron transport chain (ETC) complexes was carried out, employing blue native polyacrylamide gel electrophoresis (BN-PAGE) coupled to MS/MS, as well as sensitive bioinformatic

searches. I describe the structure and composition of each of the five ETC complexes, highlight putative evolutionary relationships between several divergent ETC subunits and provide evidence for a relatively unusual, dimeric F_oF_1 ATP synthase.

2.2 Introduction

Mitochondria are organelles involved in a broad array of eukaryotic cellular processes, including energy generation, Fe-S cluster biosynthesis, apoptosis and the metabolism of amino and fatty acids. Although mitochondria retain a distinct genome (mtDNA) that encodes a limited number of proteins ($\leq$67) and has been used to infer the α-proteobacterial origin of the mitochondrial genome [1; 15], 95-99% of mitochondrial proteins are encoded in the nucleus and imported post-translationally [43]. As a result, the information encoded in mtDNA is necessarily restricted (although invaluable), so that insight into the overall function and evolution of the organelle requires the systematic identification and characterization of mitochondrial proteins encoded in nuclear DNA (nuDNA).

Although a number of bioinformatic techniques – often based on detecting the presence of N-terminal mitochondrial targeting peptides (mTPs) – can provide valuable auxiliary information relating to subcellular localization, the algorithms employed in these techniques are insufficiently robust for the comprehensive characterization of mitochondrial proteomes on their own. In particular, false negatives and positives are common in cases where mTPs are not similar to those used to train the prediction algorithms, where N-terminal sequence data are incomplete, or where mitochondrial proteins lack N-terminal mTPs [17; 21; 23]. As a result, methods that allow direct determination of the mitochondrial protein complement (i.e., via the characterization of

purified mitochondria and/or submitochondrial compartments and complexes) are integral to the elucidation of mitochondrial function and evolution. Methods that rely on peptide identification via tandem mass spectrometry (MS/MS) and the subsequent characterization of the proteins from which the peptides are derived by bioinformatic techniques [97; 98] have been especially valuable and popular.

To date, the overwhelming majority of MS/MS-based mitochondrial proteomic studies has been carried out on mitochondria derived from animals [18; 20; 24; 26], fungi [19; 22; 65; 101; 166] and land plants [21; 64; 167; 168]. While these studies have provided important insights into mitochondrial evolution [27; 44; 45; 63; 169] and mitochondria-associated diseases [170], a more thorough understanding of the proteomes and the evolutionary histories of individual proteins from mitochondria of free-living single-celled eukaryotes (protists), which constitute the bulk of biodiversity within the eukaryotic lineage (domain Eucarya), is required to understand the evolutionary origin of the mitochondrial proteome and how similar it is among different eukaryotic groups. To this end, MS/MS analyses of mitochondria from the ciliated protozoon *Tetrahymena thermophila* [23] and the green alga *Chlamydomonas reinhardtii* [25] (along with recent proteomic investigations of mitochondria and mitochondrion-related organelles (MROs) from parasitic eukaryotes [79-81]) have provided the first glimpses into the evolution of protist mitochondrial proteomes. These studies have reinforced an emerging concept of the mitochondrial proteome as an evolutionarily labile entity; a relatively small core set of proteins (~15%), largely associated with respiration and translation, is inferred to be α-proteobacterial in origin, whereas the rest of the proteome is encoded by genes of unresolved prokaryotic origin, conserved genes 'invented' early in eukaryotic evolution, and novel, lineage-specific

additions. Nonetheless, proteomic analysis of protist mitochondria is still very much in its infancy and much remains to be learned within a phylogenetically broad context about the metabolic capacities of these organelles.

In order to expand the understanding of mitochondrial proteome evolution amongst free-living protozoa, I have undertaken an MS/MS-based investigation of the mitochondria and mitochondrial respiratory chain complexes of the amoeboid protozoon, *A. castellanii*, a cyst-forming soil and fresh water amoeba that feeds on fungi, other protists and bacteria [171]. Recently, species of the genus *Acanthamoeba* have received increasing attention due to their biomedical relevance as opportunistic human pathogens responsible for amoebic keratitis and granulomatous amoebic encephalitis [172; 173] and as important reservoirs for bacterial pathogens [174; 175]. Furthermore, *A. castellanii* occupies an interesting evolutionary position, as it is a member of the Amoebozoa supergroup [111] that is sister to Opisthokonta (animals + fungi). Analyses of its mitochondria have already uncovered several evolutionary and biochemical novelties [176; 177].

A. castellanii mitochondria are a suitable target for an MS/MS investigation as the mtDNA has been sequenced [178] and a high-coverage expressed sequence tag dataset is available, along with a draft nuclear genome sequence (http://www.hgsc.bcm.tmc.edu/microbial-detail.xsp?project_id=163). Furthermore, techniques for isolating mitochondria from axenically grown cultures have been described [179] and the mitochondrial and nuclear genomes of other free-living amoebozoan species, *Dictyostelium discoideum* [180] and *Polysphondylium pallidum* (GenBank: ADBJ01000010.1), are publicly available for comparison.

Below I present my MS/MS analysis of the *A. castellanii* mitochondrial proteome, identifying a large number of authentic mitochondrial proteins and uncovering novel metabolic and evolutionary features of the organelle.

2.3 Materials and Methods

2.3.1 Growth of A. castellanii and Purification of Mitochondria

Two 500-ml *A. castellanii* (strain Neff) cultures were shaken at ~60 rpm in 2.8-L Fernbach flasks at 30°C until they reached an OD_{580} of approximately 1.0 in modified Neff's medium [179]. For preparation of mitochondria, cells were sedimented at 4°C in two 750-ml bottles for 5 min at 900 x *g* (2000 rpm) in an IEC-CR 6000 preparative centrifuge with a swinging bucket rotor. Most of the medium was aspirated; cells were subsequently resuspended in residual medium and sedimented in two conical 50-ml tubes, as above. Cells were combined and washed once with 40 ml of homogenization buffer (10 mM tris(hydroxymethyl)-aminomethane (Tris)·HCl, pH 7.6, 0.25 M sucrose, 1 mM $ZnCl_2$, 1 mM β-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride (PMSF) and 0.1% (w:v) BSA) and sedimented, as above. *A. castellanii* cells were resuspended in 40 ml homogenization buffer and disrupted in a 40-ml Dounce homogenizer (Wheaton) using 35 strokes with the loose-fitting pestle and 5 with the tight-fitting pestle. The lysate was centrifuged for 10 min at 1000 x *g* (2200 rpm) in order to sediment unbroken cells and nuclei, following which the supernatant was centrifuged for 15 min at 9000 x *g* in a #875 fixed-angle rotor using an IEC B22M preparative centrifuge, yielding a caramel-coloured pellet. This crude mitochondrial pellet was resuspended in 25 ml mitochondrial wash buffer (10 mM Tris·HCl, pH 7.6, 0.35 M sucrose, 1 mM β-mercaptoethanol, 10 mM

Na_2EDTA and 0.1% (w:v) BSA). Mitochondria were sedimented as above and the wash process was repeated one more time.

The final crude mitochondrial pellet was resuspended to a final volume of 4 ml with pre-sucrose-gradient buffer (50 mM Tris·HCl, pH 7.6, 0.25 M sucrose, 1 mM β-mercaptoethanol, 3 mM Na_2EDTA) and layered on top of a two-step sucrose gradient consisting of 1.3 M (22 ml) and 1.55 M (11 ml) sucrose solutions made in pre-sucrose-gradient buffer. The sucrose gradients were centrifuged at 4°C for 60 min in an SW27 Ti swinging bucket rotor at 22,000 rpm (RCF_{avg} of ~64,000 x *g*). The clot-like mitochondrial band that formed at the interface of the 1.3 and 1.55 M sucrose steps was recovered using a large bore disposable pipette and diluted dropwise with 2.5 vol. pre-sucrose-gradient buffer on ice over a period of 30 min. Mitochondria were sedimented via centrifugation at 12,000 x *g* for 20 min, resuspended to a final volume of 4 ml with pre-sucrose-gradient buffer and banded again on a step sucrose gradient, as described above. Purified mitochondria were diluted and sedimented as previously described, except post-sucrose-gradient/storage buffer (20 mM Tris·HCl, pH 7.6, 0.25 M sucrose, 1 mM β-mercaptoethanol, 0.5 mM Na_2EDTA, 1 mM PMSF and 15% glycerol) was used. Protein concentration was assayed with the BioRad D_C Protein Assay kit, indicating a yield of roughly 45 mg mitochondrial protein per 1 L of cell culture. Purified mitochondria were divided into 1 mg aliquots, sedimented and stored as frozen pellets at -70°C.

2.3.2 Lysis of Mitochondria and Enrichment for Soluble and Membrane Fractions

Two milligrams of whole mitochondria purified on two-step sucrose gradients were resuspended to a final volume of 2 ml in a solution containing 50 mM Tris·HCl, pH

7.6, 1 mM Na_2EDTA and 1 mM PMSF. Mitochondria were passed twice through a French Press at 10,000 psi and the lysate was centrifuged at 33,000 rpm (RCF_{avg} of ~ 70,000 x *g*) for 1 hr at 4°C in a Beckman Type 75 Ti rotor. The resulting pellet (membrane-enriched fraction) was washed with buffer, re-sedimented and stored at -70°C. The supernatant (dilute soluble-protein-enriched fraction) was supplemented with a cold solution containing 10% (w:v) trichloroacetic acid and 1 mM β-mercaptoethanol dissolved in acetone to a final volume of 18 ml (approximately 10-fold the original supernatant volume) and proteins were precipitated at -20°C for 1.5 hr. The solution was centrifuged for 10 min at 10,000 x *g* at 4°C and the supernatant was decanted. Subsequently, the precipitated proteins were washed three times for 2 hr each time in 18 ml of an acetone solution containing 1 mM β-mercaptoethanol at -20°C in order to remove as much of the TCA as possible. The protein pellet was vacuum-dried for 1 hr at room temperature. Proteins were dissolved in 1x SDS-PAGE loading buffer without dye (62.5 mM Tris·HCl, pH 6.8, 2% SDS, 10% glycerol, 100 mM β-mercaptoethanol) at 90°C for 10 min. The protein concentration was assayed with the BioRad D_C Protein Assay kit.

2.3.3 SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE) and Excision of Bands

SDS-PAGE experiments were carried out according to Laemmli [181]. Proteins were incubated for 10 min at 90°C in SDS-PAGE loading buffer containing 0.01% bromophenol blue and briefly centrifuged afterwards. Approximately 40 μg of protein were loaded onto an 8 x 7.3 cm gel (Bio-Rad Mini-Protean II system) with an ~0.5 cm 4% stacking gel and a 9-15% linear acrylamide gradient resolving gel. Electrophoresis was carried out at 15 mA constant current until the dye front was within ~0.5 cm of the

bottom of the gel. Gels were stained overnight in a solution containing 40% (v:v) ethanol, 10% (v:v) acetic acid and 0.025% (w:v) Coomassie Blue R250, destained for ~4 hr in the same solution without Coomassie Blue R250 and finally destained overnight in 7% (v:v) acetic acid. The lanes were excised into approximately equal-sized bands (19 bands from the soluble-enriched fraction, 22 from the membrane-enriched fraction and 24 from the whole mitochondrial fraction) for subsequent MS/MS analysis.

2.3.4 Blue Native Polyacrylamide Gel Electrophoresis (BN-PAGE)

BN-PAGE of whole *A. castellanii* mitochondria was carried out as described previously [182] on either 3.5-12%, 4-12% or 4-15% polyacrylamide gels.

2.3.5 Centrifugation of Solubilized Complexes in Linear Sucrose Gradients

Whole mitochondria, representing approximately 20 mg of protein, were resuspended to a final volume of 800 µl with BN-PAGE isolation buffer (750 mM 6-aminocaproic acid/50 mM bis-tris [bis-(2-hydroxyethyl)-amino-tris(hydroxymethyl)-methane], pH 7.0). Subsequently, mitochondria were solubilized on ice by the addition of 200 µl of 20% (w:v) *n*-dodecyl-β-D-maltoside (DDM) dissolved in BN-PAGE isolation buffer in 20 µl aliquots over a 30-min period (yielding a detergent:protein ratio of 2:1). The mixture was vortexed briefly after the addition of each aliquot. The mitochondrial lysate was centrifuged for 30 min at 18,000 x *g* at 4°C in order to sediment insoluble material. The supernatant (approximately 1 ml) was then layered on top of an 11-ml linear sucrose gradient containing 75 mM 6-aminocaproic acid, 50 mM bis-tris (pH 7.0), 1 mM PMSF, 0.1 mg/ml DDM and gradients of 10-40% (w:v) sucrose. The linear sucrose gradients were centrifuged for 19 hr at 4°C in a Beckman SW41 Ti rotor at 35,000 rpm (RCF_{avg} = ~150,000 x *g*) and ~500-µl fractions were collected from bottom to

top using a fine cannula attached to a peristaltic pump. Complexes were stable at 4°C for at least several months. For BN-PAGE analysis, small volumes (usually ranging from 10-30 µl) of each fraction were supplemented with 1.5 µl 5% Coomassie Brilliant Blue G250 in 750 mM 6-aminocaproic acid/50 mM bis-tris and 5.0 µl 50% (v:v) glycerol and loaded directly onto BN gels. BN-PAGE experiments were run as described above.

2.3.6 In-Gel Enzyme Activity Assays

All in-gel enzyme activity assays were carried out overnight, at room temperature, immediately after BN-PAGE experiments, essentially according to Zerbetto et al. [142]. For the NADH dehydrogenase assay, a gel strip was incubated in 50 mM MOPS·NaOH, pH 7.4 buffer containing 0.1 mg/ml reduced β-NADH and 1 mg/ml nitroblue tetrazolium overnight. Active CI was identified by the formation of a purple-blue formazan precipitate at approximately 940 kDa. The succinate dehydrogenase assay was carried out in 50 mM MOPS·NaOH, pH 7.4 buffer containing 84 mM succinate dibasic hexahydrate, 0.2 mM phenazine methosulfate and 1 mg/ml nitroblue tetrazolium. A formazan precipitate was detected at ~130 kDa. Cytochrome oxidase activity was assayed in 50 mM MOPS·NaOH, pH 7.4 containing 1 mg/ml cytochrome *c* and 0.5 mg/ml 3,3′-diaminobenzidine tetrahydrochloride. ATPase activity was assayed in a 35 mM Tris, 270 mM glycine, pH 7.8 solution containing 14 mM $MgSO_4$, 8 mM ATP and 0.2% $Pb(OAc)_2$.

2.3.7 Assessment of Mitochondrial Purity

2.3.7.1 Extraction of Total Cellular RNA and Total Mitochondrial RNA from *A. castellanii*

All steps were carried out between 0 and 4°C. *A. castellanii* cells were chilled on ice for 30 min and sedimented by centrifugation at 1,000 x *g* for 10 min at 4°C in an IEC-

CR 6000 preparative centrifuge with a swinging bucket rotor. Cells were washed twice in a 0.85% (w:v) NaCl solution and once in 50 mM Tris·HCl, pH 7.6. Cells were resuspended to a final volume of 10 ml with 50 mM Tris·HCl, pH 7.6, homogenized using a Teflon plunger and lysed by shaking 10 min at 4°C after combination with 10 ml 2x detergent mix (2% (w:v) sodium sarkosyl, 0.1 M NaCl) and 20 ml of phenol-cresol solution (70 ml of *m*-cresol and 0.5 g of 8-hydroxyquinoline dissolved in 500 ml of liquefied phenol and equilibrated to pH 7.6 with Tris·HCl). The solution was centrifuged for 10 min at 3,000 rpm and the aqueous phase was recovered. Solid NaCl (0.6 g) was added to the 20 ml aqueous phase and dissolved by vortex mixing. The phenol-cresol extraction was repeated twice more and RNA was precipitated from the final aqueous phase by incubation at -20°C overnight in 2 vol. anhydrous ethanol. Subsequently, the RNA was sedimented by centrifugation for 10 min at 3,000 x *g* and washed with 5 ml of 80% ethanol. RNA was dissolved in 500 µl dH_2O, transferred to a microcentrifuge tube, and extracted three times with 500 µl phenol-cresol and 50 µl of 3.0 M NaOAc. RNA was precipitated by addition of 1.0 ml anhydrous ethanol and incubation at -70°C for 30 min. Precipitated RNA was dissolved in 400 µl of dH_2O and precipitated again with 40 µl 3.0 M NaOAc and 1.0 ml anhydrous ethanol at -70°C for 30 min. RNA was washed in 80% ethanol and the dried pellet was dissolved in 400 µl dH_2O. The concentration of RNA was estimated by determining the absorbance at 260 nm; approximately 15 mg of RNA was extracted from 1 L of cells.

Total *A. castellanii* mitochondrial RNA was obtained and quantified from a fraction of purified *A. castellanii* mitochondria corresponding to 10 mg of mitochondrial

protein, using the same method as above (but with smaller volumes). Approximately 280 μg of RNA were obtained from the mitochondria-enriched fraction.

2.3.7.2 Electrophoresis of RNA in Denaturing (7 M Urea) Polyacrylamide Gels

Denaturing 7M urea polyacrylamide gels [42 g urea, 20 ml 5x TBE (250 mM Tris, 250 mM boric acid, 5 mM Na_2EDTA) per 100 ml; pH ~8.3], 20 x 20 cm, 1.5 mm thick, were used to compare the rRNA profiles of *A. castellanii* total cellular RNA and RNA isolated from fractions enriched in mitochondria. Gels containing 6% (w:v) polyacrylamide (5.7 g acrylamide, 0.3 g bis-acrylamide per 100 ml) were used to assess the large rRNAs, which are not resolved on higher percentage acrylamide gels. Conversely, 10% (w:v) polyacrylamide gels (9.5 g acrylamide, 0.5 g bisacrylamide per 100 ml) were preferred for the separation and visualization of small rRNAs (i.e., 5.8S and 5S rRNA) and tRNAs because the cytosolic and mitochondrial 5S rRNAs are not resolved in 6% gels [183]. Gels were pre-run for approximately 1.5 hr at 350 V. Samples were dissolved in loading buffer, electrophoresed at the same voltage for 3 hr, stained in 40 μl of a 10 mg/ml ethidium bromide (EtBr) stock dissolved in 400 ml of dH_2O for 30 min and visualized at 302 nm. In contrast to a previous report [183], the mitochondrial 5S rRNA was visible in purified mitochondrial fractions after 30 min of EtBr staining and was even detectable in total cellular RNA preparations.

2.3.7.3 Chemical Sequencing of RNA

A faint band corresponding in size to cytosolic 5.8S rRNA was detected in purified mitochondrial fractions separated on 10% acrylamide gels (although no contaminating cytosolic 5S rRNA was detected, nor were cytosolic 18S or 28S rRNA

detected in 6% gels). In order to determine the identity of this RNA species, the band was excised and passively eluted by homogenization in 1 ml 0.5 M NH_4OAc, 10 mM $Mg(OAc)_2$, 1 mM EDTA and 1 ml phenol-cresol by shaking overnight at 4°C. The solution was centrifuged 10 min at 10,000 x *g*, the aqueous phase was re-extracted with phenol-cresol and precipitated twice with 2 vol. ethanol at -70°C.

The RNA species was 3′-end labeled with [5′-32P]pCp by RNA ligase and purified on a 6% acrylamide sequencing gel (33 x 40 x 0.05 cm) that was pre-run in TBE buffer at 1300 V for ~1.5 hr and run for 4 hr at 1700 V. The 3′-end-labeled RNA species was detected with autoradiography, then excised, eluted and precipitated as above, before being subjected to base-specific chemical cleavage [184]. Fragments were separated on a 20% (w:v) acrylamide, 7 M urea, 33 x 40 x 0.05 cm gel for 2 hr 20 min at 2100 V (pre-run 30 min at 1700 V) to determine the 3′-terminal sequence of the RNA species. X-ray film with an intensifying screen was exposed at -70°C.

2.3.8 Mass Spectrometry

2.3.8.1 In-Gel Protein Digest

Excised protein bands (either SDS-PAGE or BN-PAGE) were digested manually or with a Genomic Solutions Progest automated system. Briefly, gel slices were washed with 0.1 M NH_4HCO_3 (AB), pH 8.0, followed by a wash with acetonitrile (ACN). Samples were reduced with 10 mM dithiothreitol (DTT) in 0.1 M AB at 56°C for 30 min, then alkylated with 100 mM iodoacetamide for 30 min at room temperature. Gel slices were washed with 50% acetonitrile in 0.1 M AB for 1 hr, then 100% ACN for 10 min. Solvent was removed and 12.5 ng/µl trypsin (Promega V5111) in 0.1 M AB was added in

a minimum volume to cover the gel slices. Samples were incubated overnight at 37°C. Peptides were extracted from the gel with 50% ACN, 0.1% formic acid. Samples were taken to near dryness and re-dissolved in 5% ACN, 0.1% formic acid in preparation for MS/MS.

2.3.8.2 In-Solution Protein Digest

Approximately 500 µg of mitochondrial protein was diluted to a concentration of 1 mg/ml in 50 mM AB (pH 8.0) with protease inhibitor (Set III, Calbiochem 539134). An acid-labile surfactant (Rapigest, Waters 186001861) was added at a concentration of 0.1% (w/v) to improve solubilization and protein digestion. The sample was reduced with 5 mM DTT at 56°C for 60 min, then alkylated with 15 mM iodoacetamide for 30 min at room temperature. Proteins were digested with trypsin (Promega V5111) at a ratio of 50:1 protein:trypsin (w:w) and incubated overnight at 37°C. Following digestion, the pH of the resulting peptide solution was decreased to <2 with trifluoroacetic acid in order to precipitate the surfactant. The precipitate was removed using a 0.45 µm filter. The sample was lyophilized then re-dissolved in 10 mM ammonium formate, 25% acetonitrile, pH 3.0, in preparation for strong cation-exchange high-performance liquid chromatography (SCX-HPLC).

2.3.8.3 SCX-HPLC of Tryptic Peptides

Five hundred µg of tryptic peptides were fractionated by SCX-HPLC into 45 fractions using an Agilent 1100 LC system equipped with a 100 x 2.1 mm polysulfoethyl A column (PolyLC 102SE0502, Columbia, MD). The separation was carried out using the following gradient at a constant flow rate of 0.2 ml/min: A - 10 mM ammonium

formate, 25% ACN (pH 3.0); B – 600 mM ammonium formate, 25% ACN (pH 3.0). The following binary gradient (time (min), % eluent B) was used for SCX-HPLC separation (0, 0; 5, 0; 20, 10; 30, 20; 40, 60; 45, 60; 50, 0; 60, 0).

Fractions were lyophilized, then resuspended in 5% ACN, 0.1% formic acid for MS/MS analysis.

2.3.8.4 MS/MS AND PROTEIN IDENTIFICATION

All samples were analyzed by reversed-phase LC-MS/MS (RP-HPLC) using an Agilent 1100 HPLC system equipped with a 15 cm x 100 mm Onyx Monolithic C_{18} column (Phenomenex, Torrance, California). The separation was carried out using the following gradient: 2% B for 3 min, increasing to 25% B over 45 min and 95% B over 10 min (A: 0.1% formic acid in water, B: 0.1% formic acid in ACN) at 2 ml/min. The HPLC was interfaced to an AB/MDS-SCIEX QTrap 4000 mass spectrometer via a nanoflow source. Data were acquired in the information-dependent acquisition mode, i.e., the m/z values of the tryptic peptides were measured using an MS scan, followed by enhanced resolution scans of the three most intense peaks and finally three tandem MS scans. The tandem mass spectra were submitted to the database search program MASCOT (Matrix Science Ltd., England) in order to identify peptides. Data files were searched against a 6-frame translation of a CAP3-clustered EST database, consisting of expressed sequence tags (ESTs) generated by random 454 sequencing and through TBestDB [185] via dideoxynucleotide chain termination sequencing of random cDNA libraries; a set of previously predicted *A. castellanii* mtDNA-encoded proteins [178]; a 6-frame translation of the complete *A. castellanii* mitochondrial genome sequence [178]; and an in-house predicted set of nuDNA-encoded proteins (preliminary gene predictions;

PGP), generated by Augustus [186] from a draft version of the *A. castellanii* nuclear genome sequence. For nuDNA-encoded proteins identified in the whole proteome analysis (either in EST or PGP datasets), only those containing one or more peptides with ion scores $\geq$ 38 (exceeding the MASCOT 95% significance threshold) were considered. Typically, proteins from the PGP dataset were only considered if they were poorly represented in the EST dataset (in order to offset the biases associated with using an expression-based database alone). In the case of mtDNA-encoded proteins or proteins identified by MS/MS after BN-PAGE, certain peptides below this threshold were allowed after manual inspection.

2.3.8.5 Summary of Various Fractions Analyzed by MS/MS

In sum, four different strategies were used in the MS/MS-based analysis of *A. castellanii* mitochondria (aside from analysis of complexes isolated via BN-PAGE): 1) LC/LC-MS/MS analysis of whole mitochondria (WM), 2) 1D SDS-PAGE/LC-MS/MS analysis of whole mitochondria (SWM), 3) 1D SDS-PAGE/LC-MS/MS analysis of soluble-protein-enriched fractions (SPE) and 4) 1D SDS-PAGE/LC-MS/MS analysis of membrane-protein-enriched fractions (MPE).

2.3.9 Sequence Analysis And Prediction Of Gene Structures

2.3.9.1 Prediction of Gene Structures

As described above, a 6-frame translation of CAP3-clustered *A. castellanii* ESTs served as the primary database from which nuDNA-encoded mitochondrial proteins were identified by MASCOT, as predicted protein sequences for *A. castellanii* are currently unavailable. As such, many of the translated clusters represent incomplete protein sequences; the correction of these sequences (i.e., inferring the proper gene structure

using a combination of EST and genomic data, along with homology analyses) is important in cataloging the number and putative localization of proteins identified by MS/MS. In order to augment the information in the CAP3 EST clusters, I also considered unclustered 454 ESTs, unclustered TBestDB ESTs [185], ESTs clustered with MIRA [187], homology-based inference from the nuclear genome, and a set of in-house-predicted proteins generated with Augustus [186]. Completeness of protein sequences was judged manually.

2.3.9.2 Methods for Detection of Homology

Annotation of proteins identified by MS/MS analysis was performed by querying a variety of databases, including NCBI nr, *Saccharomyces* Genome Database, TBestDB [185] and DictyBase [188] using BLAST [189] (BLASTp, tBLASTn and PSI-BLAST algorithms). Alignments were inspected manually.

In order to identify certain proteins expected to be found in *A. castellanii* but not easily uncovered by standard homology searches, or to assign tentative orthology to identified components of ETC complexes without obvious matches in sequence databases, profile hidden Markov searches were performed. Briefly, convincing orthologs of the protein of interest were collected and aligned with MUSCLE v.3.6 [190]. HMMER 3.0 [191] was used to interrogate a 6-frame translation of the *A. castellanii* 454 EST dataset or the mitochondrial proteome dataset. Results of HMMER 3.0 searches were inspected manually. In combination, I considered 1) the E-value and rank of the putative homolog when queried with a profile HMM (all putative homologs ranked first when the mitochondrial proteome dataset was queried with the profile HMM; expect (E-) values were relatively high, though most were ≤ 0.05), 2) whether the highest ranking

result was identified via MS/MS as a component of a protein complex (e.g., if a CI subunit was being searched for, was the top HMMER match a component of CI?), 3) the size and physicochemical properties of the protein in question in relation to authentic homologs.

The overlap between my *A. castellanii* mitochondrial proteome data and the mitochondrial proteome datasets from yeast (851 proteins) [22], human (1098 proteins) [24], *Arabidopsis thaliana* (414 proteins) [21], *T. thermophila* (573 proteins) [23] and *C. reinhardtii* (349 proteins) [25] was assessed using BLASTp; comparisons to bacterial proteins (corresponding to all bacterial proteins in the NCBI nr database) were done in the same way. Briefly, proteins were assigned as putative homologs if E-values were lower than 1e-10, although exceptions were made for certain small and divergent proteins (e.g., small respiratory, mitoribosomal or import-associated proteins). Due to the limitations associated with our EST database (and lack of a predicted protein dataset), reciprocal best BLAST hit comparisons were not feasible; as a result, my analyses report putative homologs (not necessarily orthologs) and therefore represent an overestimate of the orthology relationships between *A. castellanii* mitochondrial proteins and the proteins encoded in the specified databases.

2.3.9.3 Analysis of Protein Targeting Information

As described above, protein sequences were inferred on the basis of EST sequence data. Putative mitochondrial proteins considered as complete at their respective N-termini (generally on the basis of BLAST alignments) were collected and TargetP [90] was used to test their likelihood of mitochondrial localization (probability $\geq$ 75% used as a criterion for mitochondrial localization).

Assessment of classical mTP properties was performed on a manually curated set of 299 high-confidence *A. castellanii* mitochondrial proteins likely bearing N-terminal extensions. LogoBar [192] was used to create sequence logo depictions on a de-gapped alignment of the first 30 amino acid residues from these proteins. Characterization of the relative amino acid content of classical mTPs was achieved via comparison of the frequencies of amino acids in the first 15 and 30 residues vs. the entire protein sequences.

2.3.9.4 Detection of Contaminating Proteins

Subcellular localization prediction and homology detection algorithms both played important roles in defining putative contaminating (e.g., secreted and peroxisomal) proteins identified by MS/MS analysis. Briefly, both TargetP [90] and Predotar [91] were used to identify likely contaminating proteins bearing signal peptides. Putative peroxisomal proteins were screened using the Target Signal Predictor tool (http://www.peroxisomedb.org/Target_signal.php) capable of predicting PTS1, PTS2 and Pex19BS binding sites. Other classes of contaminating proteins (e.g., cytosolic) cannot be positively identified on the basis of targeting peptides; BLASTp analyses, along with consideration of protein function, were used to identify other potential contaminating proteins. BLAST query results were also considered in defining putative secreted and peroxisomal proteins.

2.3.10 Phylogeny of MS-ICL Fusion Protein

Two separate phylogenetic trees were reconstructed for the malate synthase-isocitrate lyase fusion protein (below): one tree for the portion that corresponds to malate synthase and another that corresponds to isocitrate lyase. Homologs were collected from

various eukaryotes and bacteria by querying the NCBI nr database with BLASTp. Alignments were performed using MUSCLE v.3.6 [190] and edited manually. Maximum likelihood trees were reconstructed with RAxML-HPC [193] under the WAG + Γ model (using the PROTGAMMAWAGF option) with 25 categories of substitution rate variation. One hundred bootstrap replicates were performed as a measure of statistical support for inferred nodes.

2.4 Results and Discussion

2.4.1 Mitochondrial Proteins Encoded by Mitochondrial DNA (mtDNA)

The *A. castellanii* mitochondrial genome [178] is currently annotated as specifying 41 proteins, all of which are encoded on the 'positive' strand (although there are only 40 predicted open reading frames (ORFs), the single continuous Cox1/2 ORF encodes two distinct proteins [194]). Of these 41 proteins, most have been assigned functions associated with respiration (17) and translation (16). Conversely, 3 intron-encoded ORFs (*orf142*, *orf168*, *orf164*) are predicted to specify LAGLI-DADG homing endonucleases, 2 ORFs (*orfB* and *orf25*) are hypothesized to encode subunits of the F_o sector of ATP synthase [36; 134] and 3 ORFs (*orf83*, *orf115*, *orf349*) appear unique to *A. castellanii*.

Our proteomic analysis identified peptides from 31 proteins (76%), 21 of which were identified by more than one significant peptide (Table 2.1). I detected 8/10 subunits of Complex I (but not subunits 2 and 4L); the sole Complex III subunit (Cob); 3/3 Complex IV subunits (Cox1, Cox2, Cox3); 3/3 Complex V subunits (Atp1, Atp6, Atp9), along with hypothetical Complex V subunits OrfB and Orf25; 5/6 LSU ribosomal proteins (but not L6) and 9/10 SSU ribosomal proteins (but not S12). Together, these results suggest a high coverage of both hydrophobic and hydrophilic mtDNA-encoded

proteins. Neither intron ORF-encoded proteins nor proteins corresponding to the three *A. castellanii*-specific ORFs of putative but unknown function were confidently identified in this analysis.

Examination of the ions scores (a number reflective of the probability that a peptide match between experimentally determined and theoretical tandem mass spectra is due to chance) for mtDNA-encoded proteins identified by MS/MS (Table 2.1) highlights several noteworthy trends. Firstly, the LC/LC-MS/MS approach (i.e., WM fraction) almost always generated the highest ion scores for abundant mtDNA-encoded proteins (compare, e.g., the ion scores for Atp1 in Table 2.1). Secondly, these results demonstrate the effectiveness of the suborganellar enrichment strategies; the soluble-protein-enriched fraction lacked peptides derived from a number of hydrophobic membrane proteins, including Cox3, Cob and Nad1, while membrane protein enrichment yielded the only peptide data for the hydrophobic Nad6 and Atp6 proteins.

Interestingly, peptide data corresponding to two directly adjacent, previously unannotated small ORFs were identified here via MS/MS by searching against a 6-frame translation of the mitochondrial genome. These two ORFs – 92 and 99 amino acid residues in length – are encoded between the *atp9* and *trnL2* genes, and represent the only proteins encoded on the negative strand of *A. castellanii* mtDNA. I argue below that these proteins are homologs of ribosomal proteins S16 and L19, respectively, in which case the *A. castellanii* mitochondrial genome encodes at least 43 proteins.

Table 2.1 *A. castellanii* mtDNA-encoded proteins identified my MS/MS. Cumulative peptide ion scores for proteins identified in each analysis are presented. The highest ion score for each protein among the four conditions is bolded. Identified proteins are available in Supplemental File 2.1. SWM, 1D SDS-PAGE/LC-MS/MS analysis of whole mitochondria; WM, LC/LC-MS/MS analysis of whole mitochondria; SPE, 1D SDS-PAGE/LC-MS/MS analysis of soluble-protein-enriched fractions; MPE, 1D SDS-PAGE/LC-MS/MS analysis of membrane-protein-enriched fractions.

Protein	SWM	WM	SPE	MPE	>1 peptide
		Ion Score			
Atp1	2560	**10960**	838	3910	+
Cox1/2	1419	**2638**	124	1096	+
Cox3	775	**1491**	-	494	+
Nad11	674	**1128**	1089	474	+
Atp9	589	108	48	**1090**	+
Nad7	508	**1529**	138	885	+
Nad9	456	**808**	307	153	+
OrfB	423	**855**	-	248	+
Rpl11	233	**276**	35	-	+
Orf25	201	**914**	146	147	+
Cob	161	**483**	-	206	+
Rpl14	140	**211**	-	-	+
Rpl2	97	**148**	55	-	+
Rpl5	**85**	36	81	-	+
Rps2	80	**128**	78	-	+
Rps8	64	**106**	-	46	+
Rps7	34	**95**	-	-	+
Rps11	29	**83**	-	-	+
Rpl16	29	**274**	53	-	+
Rps4		**189**	-	-	+
Nad1	86	51	-	**98**	-
Rps19	54	77	**114**	37	-

Protein	SWM	WM	SPE	MPE	>1 peptide
			Ion score		
Rps14	**26**	-	-	-	-
Rps13	-	-	33	**71**	-
Nad3	-	**127**	-	-	-
Nad5	-	**101**	-	-	-
Nad4	-	33	-	**57**	-
Rps3	-	**28**	-	-	-
Nad6	-	-	-	**77**	-
Atp6	-	-	-	**33**	-
Novel protein					
Rpl19	-	31	-	**42**	-
ORF92/Rps16	-	47	**55**	33	+

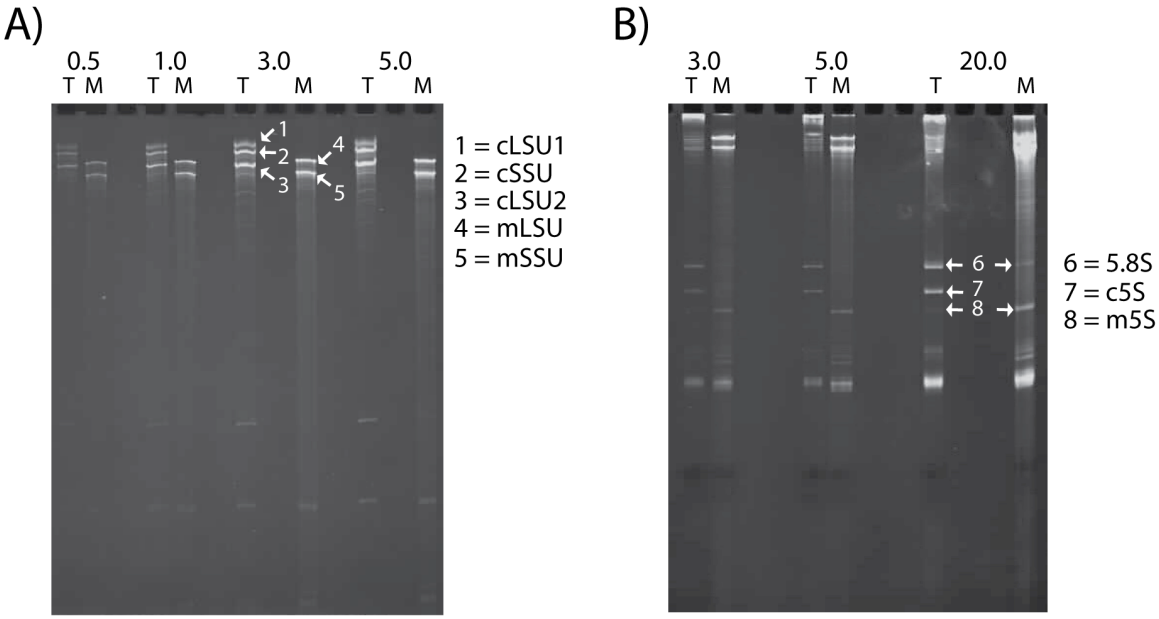
2.4.2 Mitochondrial Proteins Encoded by Nuclear DNA (nuDNA)

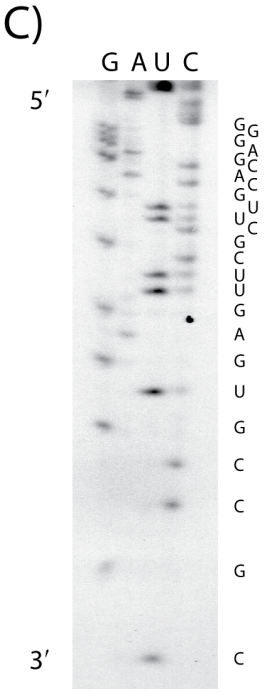
2.4.2.1 Assessment of Mitochondrial Purity

In order to assess the purity of isolated mitochondrial fractions prior to MS/MS analysis, I compared the EtBr-stained profiles of total RNA isolated from *A. castellanii* cells and RNA isolated from a portion of the whole mitochondrial fraction used for MS/MS analysis (Figure 2.1). I chose this approach because the availability of antibodies specific for appropriate *A. castellanii* subcellular markers is very limited whereas cytosolic ribosomes (containing rRNA species) represent abundant cellular components that frequently contaminate mitochondrial preparations. Consistent with a high degree of mitochondrial enrichment, neither large nor small subunit cytosolic rRNA species were

detectable in the total mitochondrial RNA fraction on 6% acrylamide gels, nor was cytosolic 5S rRNA visible on 10% acrylamide gels. Conversely, a mitochondrial 5S rRNA was clearly seen after EtBr staining (the cytosolic and mitochondrial 5S species migrate at the same rate in 6% acrylamide gels, but their mobilities differ in 10% acrylamide gels), and another small RNA the size of cytosolic 5.8S rRNA was visible when large amounts of mitochondrial RNA were loaded on a 10% acrylamide gel. Direct RNA sequencing demonstrated that the latter species was identical to *A. castellanii* 5.8S rRNA; however, I do not interpret this as evidence of significant contamination by cytosolic ribosomes due to the absence of the other cytosolic rRNA species (which, in the case of large rRNAs, should stain much more intensely than 5.8S rRNA if present in equimolar amounts). Interestingly, the selective presence of cytosolic 5.8S rRNA has been previously reported in human mitochondria [195], so it may be that the 5.8S rRNA is actively imported into *A. castellanii* mitochondria.

Figure 2.1 Assessment of mitochondrial purity by comparing rRNA profiles of total cellular RNA and total mitochondrial RNA from *A. castellanii*. A) An EtBr-stained 6% (w:v) acrylamide, 7 M urea gel loaded with varying amounts of total cellular (T) and total mitochondrial (M) RNA. Numbers are amounts of RNA (in μg) applied in each lane. The profiles of large rRNA species in T and M are distinct, suggesting that the mitochondrial fraction is relatively free of contaminating cytosolic rRNA (and presumably cytosolic ribosomal proteins). *A. castellanii* cytosolic (c) and mitochondrial (m) LSU and SSU species are identified (note that the cLSU is split). B) 10% acrylamide gel as in A). Note that a mitochondrial 5S rRNA (m5S) is visualized in the M lane and, to a much lower extent, in the T lane. A band corresponding in size to cytosolic 5.8S rRNA is visible in an overloaded M lane; however, no cytosolic 5S (c5S) is detectable. C) Direct chemical sequencing of 3′-end-labeled RNA (band "6" from 20.0 μg "M" lane in B) that comigrates with cytosolic 5.8S rRNA in a 20% (w:v) acrylamide, 7 M urea gel. The RNA sequence determined (bolded letters) mapped onto the sequence of *A. castellanii* 5.8S rRNA (GenBank: K00471.1), confirms the identity of this RNA species as 5.8S rRNA.





5′ AACUCCUAACAACGGAUAUCUUGGUUCUCGCGAGGAUGA
AGAACGCAGCGAAAUGCGAUACGUAGUGUGAAUCGCAGG
GAUCAGUGAAUCAUCGAAUCUUUGAACGCAAGUUGCGCU
CUCGUGGUUUAACCCCCC**GGGAGCACGUUCGCUUGAGUG**
CCGCUU 3′

In general, MS/MS analyses showed the preparation to be highly enriched in mitochondria; however, a number of contaminating proteins (158) from various subcellular compartments – most notably the endoplasmic reticulum (ER) (68), cytosol (10), cytoskeleton (12), cytosolic ribosomes (21) and peroxisomes (30) – were detected (identified contaminants are available in Supplemental File 2.2). Contamination by certain cytosolic proteins, in particular cytoskeletal elements, reflects an association critical for mitochondrial function and motility [196]. The presence of cytoribosomal proteins (21 in total, all components of the large ribosomal subunit) may reflect the presence of ribosomes that are often associated with the outer mitochondrial membrane (OMM; the small ribosomal subunit would be expect to be dissociated during routine washing of the mitochondria fraction with EDTA-containing buffer during isolation), as has been reported in yeast [197; 198] and mammals [199]. Conversely, contamination by ER proteins leaves open the possibility that the detection of cytoribosomal proteins simply reflects the contamination of mitochondrial preparations by rough ER, which is quite commonly seen. Direct protein associations between the ER and mitochondrial compartments [200] may partially account for these results. Finally, it is possible that the unusual clot-forming behaviour of *A. castellanii* mitochondria on step sucrose gradients [179] may have contributed to the spurious trapping of other cellular components during purification of mitochondria. Importantly, however, many of the suspected contaminating sequences identified by MS/MS analysis appear to be present in relatively low abundance (i.e., many were identified by only one exceptional peptide). Attempts to perform a third step sucrose gradient were made; however, the integrity of *A. castellanii* mitochondria was lost at this additional step, giving very poor yields.

Predicted contaminating proteins were removed from the putative mitochondrial proteome dataset. Encouragingly, only five of the predicted contaminating proteins with discernible targeting information are 'hypothetical' proteins; in general, contaminating proteins tend to represent abundant cellular proteins, suggesting that many of the novel proteins identified in my analysis are likely to be authentic mitochondrial proteins.

2.4.2.2 Overlap Between WM, SWM, SPE and MPE Datasets

After removal of likely non-mitochondrial contaminant sequences from consideration, a set of 716 putative nuDNA-encoded mitochondrial proteins remained (available in Supplemental File 2.3). In combination with mtDNA-encoded proteins, a total of 748 mitochondrial proteins were confidently identified in this study. I estimate that 77% (553/716) of the inferred protein sequences are complete or likely complete, based on clustered EST sequences, and that 624/716 proteins (87%) probably have complete N-termini (targeting information being located within the N-terminal ~30 residues for many mitochondrial proteins). A further 113 putative mitochondrial proteins were identified bioinformatically in the course of characterizing protein complexes and metabolic pathways (discussed below; available in Supplemental File 2.4).

I identified proteins from all submitochondrial compartments, including the outer membrane, intermembrane space, inner membrane and matrix, suggesting that mitochondrial integrity was maintained in the purification procedure. As described above (see Table 2.1), four distinct strategies were employed in order to maximize coverage of the mitochondrial proteome. Of the 716 proteins identified by MS/MS, 148 (~21%) were detected in all four conditions and a further 178 (25%) were found in at least three

conditions (see Figure 2.2). BN-PAGE analyses revealed five proteins not identified by other analyses (not shown in Figure 2.2).

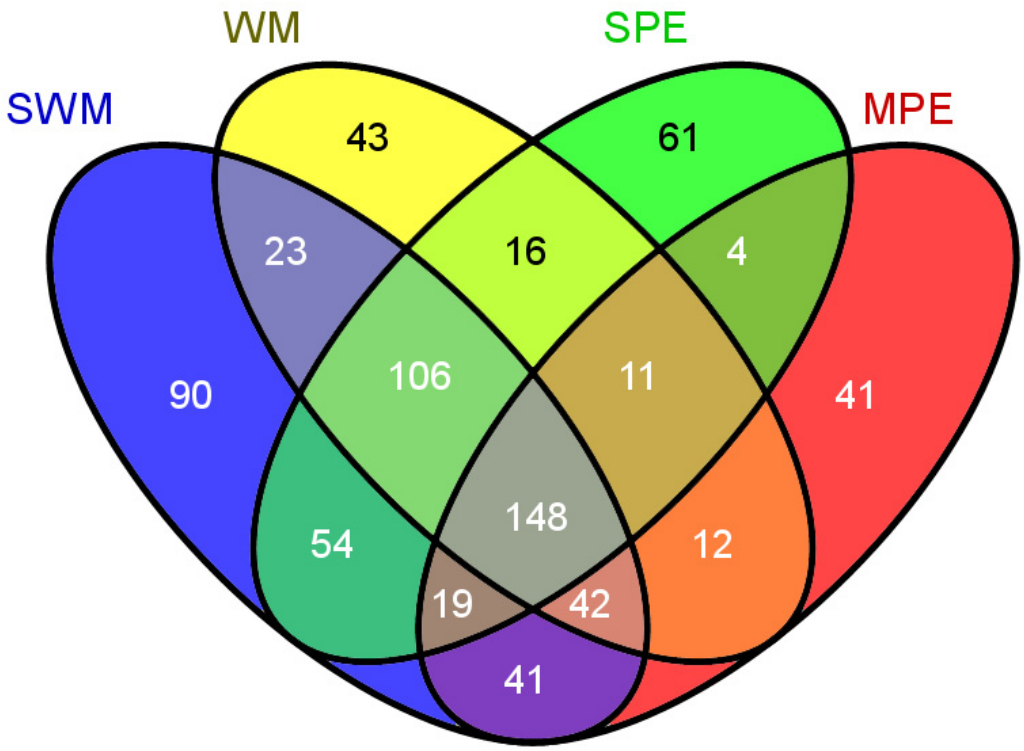


Figure 2.2 Overlap of nuDNA-encoded proteins among four separate MS/MS analyses. A Venn diagram depicts the number of shared and unique proteins identified in each of: WM (whole mitochondrial tryptic peptide sample separated by SCX-HPLC into 45 fractions and analyzed by RP-HPLC-MS/MS); SWM (whole mitochondrial protein sample separated via 9-15% acrylamide SDS-PAGE and analyzed by RP-HPLC-MS/MS); SPE (soluble-protein-enriched sample); and MPE (membrane-protein-enriched sample). SPE and MPE samples were each separated and analyzed as in SWM.

2.4.2.3 Functional Categories

A. castellanii nuDNA-encoded proteins were categorized manually according to their inferred function (Figure 2.3) including: 1) Metabolism (amino acid, fatty acid and nucleotide metabolism, among other processes), 2) Pyruvate metabolism, TCA cycle and

respiration, 3) Replication and transcription (including proteins directly involved in these processes or known to be associated with nucleic acids), 4) Translation and ribosomes (mitochondrial ribosomal proteins and various translation factors), 5) Transport and import (metabolite carrier proteins and protein import machinery), 6) Protein folding and degradation (chaperones and proteases), 7) Structural, 8) Reactive oxygen species (ROS) defense and signaling, 9) Metal binding (enzymes involved in Fe-S cluster and heme biosynthesis) and 10) Unknown (primarily proteins that are specific to *A. castellanii*, but also including other proteins of unknown function with homologs in other organisms).

As observed in other systems, proteins involved in energy generation, metabolism, transport processes and translation constitute the bulk (>80%) of the MS-identified mitochondrial proteome. The proportion of all proteins of unknown function (~23%) is consistent with observations in the case of certain other mitochondrial proteomes [19; 24], but significantly lower than is seen in more derived protists, like *Tetrahymena* [23]. Most *A. castellanii* mitochondrial proteins lacking homologs in non-amoebozoan groups do not have apparent homologs in other amoebozoan genomes either (i.e., these tend to be *A. castellanii*-specific rather than amoebozoan-specific proteins).

Below I describe in finer detail my characterizations of the most novel and interesting aspects of the proteins and pathways that comprise these functional categories.

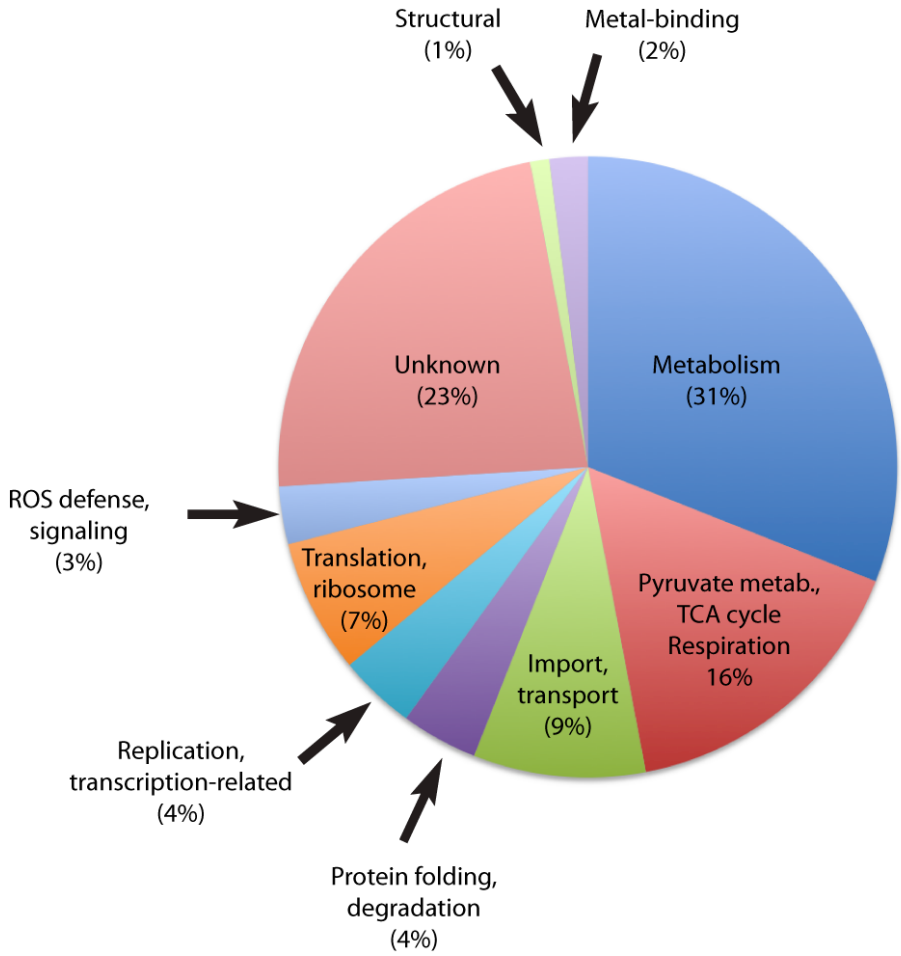


Figure 2.3 Functional categorization of nuDNA-encoded *A. castellanii* mitochondrial proteins. Relative proportions of each functional class of the nuDNA-encoded proteins (716) identified in this study by MS/MS analyses of *A. castellanii* mitochondria.

2.4.3 Pyruvate Metabolism

In classical aerobic mitochondria, the pyruvate dehydrogenase (PDH) multi-enzyme complex effects the NAD^+-dependent oxidative decarboxylation of pyruvate, yielding acetyl-CoA, NADH and CO_2. In *A. castellanii*, I detected multiple unique peptides from all of the PDH subunits, including E1 α and β, E2 (2 isoforms) and E3 (Table 2.2). With MS/MS, peptides from PDH kinase (present in EST and genome data), which phosphorylates and inhibits the E1 (catalytic) subunit of PDH, were not detected;

however, peptides were identified from both the catalytic and regulatory subunits of PDH phosphatase, which reverses the effects of PDH kinase and thereby activates the enzyme.

Interestingly, I have also confidently detected pyruvate:ferredoxin oxidoreductase (PFO), an O_2-sensitive, Fe-S cluster-containing enzyme frequently present in anaerobic bacteria and in the MROs (or cytoplasm [13; 201] and possibly chloroplast [202]) of anaerobic or microaerophilic eukaryotes (Table 2.2). Like PDH, PFO catalyzes the decarboxylation of pyruvate, yielding acetyl-CoA and CO_2; however, instead of reducing NAD^+, PFO donates electrons to mitochondrial ferredoxin. Subsequently, electrons from ferredoxin are transferred to protons by [FeFe]-hydrogenase, producing molecular hydrogen. The CoA moiety of acetyl-CoA is transferred to succinate by acetate:succinate CoA transferase (ASCT), yielding succinyl-CoA and acetate; substrate level phosphorylation of P_i by succinate thiokinase (also known as succinyl-CoA synthetase) then yields ATP and succinate.

The existence of a PFO gene in *A. castellanii*, along with the genes encoding [FeFe]-hydrogenase and two of three assembly proteins reported from other organisms, has been described previously [203]. PFO, which in *A. castellanii* is predicted to be >130 kDa in size and contains an obvious mTP (probability of 0.958 with TargetP), was identified in the WM, SWM and SPE fractions, but not the MPE fraction, suggesting that it is likely a soluble, mitochondrial matrix protein. The total ion scores for PFO were relatively high, suggesting that PFO is not a low-abundance protein in *A. castellanii* mitochondria. Neither [FeFe]-hydrogenase (which also has a strong predicted mTP) nor the [FeFe]-hydrogenase assembly factors HydE or HydG were detected in any of the fractions; however, I detected ferredoxin, a putative homolog of ASCT (most similar to

the type employed by *Fasciola hepatica* [204]) and one peptide from the [FeFe]-hydrogenase assembly factor, HydF (not reported by Hug *et al.* [203]).

The identification of both PDH and PFO in mitochondria from cells grown under aerobic conditions is somewhat surprising, given the supposed O_2 sensitivity of PFO. However, this situation is reminiscent of *C. reinhardtii* mitochondria, in which PDH and a different O_2-sensitive pyruvate-catabolizing enzyme, pyruvate formate-lyase (PFL), are both present in abundance [25; 84]. It might be argued that the fluctuating environmental conditions with which *A. castellanii* is presented (i.e., low O_2 concentrations in soil) would favor a low-level constitutive expression of enzymes of 'anaerobic' energy generation pathways under aerobic conditions such that the transition to an anaerobic metabolism would be expedited. In this case, it is possible that while PFO/[FeFe]-hydrogenase system proteins are present under aerobic conditions, they are not enzymatically active. Indeed, mechanisms for reversible and irreversible deactivation of PFL under aerobic conditions have been described [205; 206]. Conversely, previous studies have demonstrated that *A. castellanii* is strictly dependent on oxygen for growth, and encystation is complete within several hours of exposure to anoxia, although these conclusions were reached under laboratory conditions and may not reflect the true behavior of *A. castellanii* in nature. As a result, the regulation and precise functions of PFO and the [FeFe]-hydrogenase system in *A. castellanii* mitochondria remain enigmatic.

Although present in *A. castellanii*, neither PFO nor [FeFe]-hydrogenase is encoded by the nuDNA of the other mitochondriate amoebozoans (*D. discoideum*, *P. pallidum* and *Physarum polycephalum*) for which high-coverage nuclear genome sequence assemblies are available, although TBestDB EST data [185] do suggest the

presence of a [FeFe]-hydrogenase in another amoebozoan relative, *Hartmannella vermiformis*. As described elsewhere, the phylogenetic distribution of PFO and [FeFe]-hydrogenase is patchy within eukaryotes and these enzymes tend to be found in anaerobic/microaerophilic eukaryotes (*Trichomonas vaginalis*, *Entamoeba histolytica*, *Giardia intestinalis*, *Blastocystis hominis*) or species that can be expected to encounter both ambient and low O_2 conditions (*A. castellanii*, *Naegleria gruberi*, *C. reinhardtii*) [203; 207]. In other words, their distribution appears to better reflect common ecology than common descent. Recent analyses support the idea that the genes encoding [FeFe]-hydrogenase and, to a lesser extent, PFO, have been acquired from multiple sources by eukaryotes [203], although it is formally possible that genes encoding certain isozymes were inherited from the protomitochondrial genome [14; 46], while others were acquired via lateral transfer from other sources. Thus, it may be that the acquisition of genes encoding hallmark anaerobic organellar proteins by *A. castellanii* has enabled adaptation to environments that fluctuate in O_2 abundance.

2.4.4 TCA Cycle

MS/MS analysis revealed all subunits associated with the eight enzymatic activities of the TCA cycle (Table 2.2).

I identified two distinct isoforms of citrate synthase; however, BLAST analyses and the presence of a predicted N-terminal PTS2 peroxisome binding site on one isoform led me to exclude this protein as a contaminant.

Two isoforms of aconitase (64% sequence identity) were found, and both of these proteins appear to be typical mitochondrial aconitase proteins, unlike the case of *T. thermophila* mitochondria, in which a typically cytosolic variant is the sole aconitase protein [23]. Peptides from one isoform (Contig17472) were detected to a much greater extent than the other (Contig806; Table 2.2, Supplemental File 2.5), strongly suggesting that they are not present in the same abundance in *A. castellanii* mitochondria.

Isocitrate dehydrogenase (NAD^+-dependent) is composed of two distinct but related proteins, the α and β subunits. Both subunits were detected by MS/MS. Interestingly, BN-PAGE-MS/MS demonstrated that these two subunits co-migrate with respiratory CIII in a band >500 kDa in size (below). The other isocitrate dehydrogenase isoform typically present in mitochondria ($NADP^+$-dependent) was also identified.

The 2-oxoglutarate dehydrogenase complex consists of three enzyme activities: oxoglutarate dehydrogenase (thiamine pyrophosphate-dependent; E1α and E1β subunits), dihydrolipoyl succinyltransferase (lipoic acid cofactor; E2) and dihydrolipoyl dehydrogenase (FAD-dependent; E3). I identified two isoforms of E1α and one each of E1β, E2 and E3. E3 is also a component of the mitochondrial glycine cleavage system, along with the PDH and branched-chain α-keto acid dehydrogenase multienzyme complexes. Subunits E1α and E2 were identified as minor contaminants of a putative tetrameric ATP synthase (Complex V) isolated via BN-PAGE (below), consistent with previous reports of a 2-oxoglutarate dehydrogenase complex molecular weight of several MDa in other organisms [208].

Succinyl-CoA synthetase consists of α and β subunits. I identified one α and two β subunit variants via MS/MS analysis; these appear to be the only succinyl-CoA synthetase proteins encoded by *A. castellanii*.

All subunits of the succinate dehydrogenase complex were detected by MS/MS (see Section 2.4.5.2 for further discussion).

Two distinct and unrelated forms of fumarase - Class I and Class II - are known to exist [66]. Class II enzymes have a broad phylogenetic distribution in archaea, bacteria and eukaryotes and are the class present in the mitochondria of animals, fungi, land plants and ciliates [19; 21; 23; 24]. Intriguingly, as in the case of the green alga *C. reinhardtii* [25], only a Class I fumarase enzyme was identified via MS/MS in *A. castellanii* mitochondria. Until recently, Class I fumarase enzymes were thought to be limited to prokaryotes [66]; however, expanded sequencing of eukaryotic nuclear genomes demonstrates the presence of Class I fumarase enzymes in a wide variety of eukaryotes, including diatoms, heterolobosean, euglenozoans, green algae, alveolates, stramenopiles, chlorarachniophytes, cryptophytes, choanoflagellates and some animals. Moreover, many of the eukaryotic Class I fumarase proteins have predicted mTPs, suggesting that Class I fumarase proteins may be the mitochondrial fumarase in many lineages. The *A. castellanii* nuclear genome does encode a Class II fumarase; however, this protein does not bear an N-terminal mTP and it appears unlikely that it is a mitochondrial protein. Somewhat surprisingly, the distribution of fumarase proteins within the slime molds is quite complicated: *D. discoideum* and *P. pallidum* encode only Class II fumarase proteins, whereas *Physarum polycephalum* encodes both Class I and Class II enzymes.

BLAST analyses indicate that *A. castellanii* nuDNA encodes three distinct malate dehydrogenase (MDH) proteins, two of which were detected in the analysis. Interestingly, neither of the putative mitochondrial MDH proteins encoded by *A. castellanii* are the 'typical' mitochondrial type: one is related to the eukaryotic cytosolic MDH (Contig26422) whereas the other (Contig10193; Table 2.2, Supplemental File 2.5) appears to be closely related to a bacterial type MDH (the third, non-mitochondrial MDH of *A. castellanii* is a cytosolic type isozyme as well). Both of the mitochondrial MDH proteins have N-terminal extensions relative to the putative non-mitochondrial *A. castellanii* MDH and cytosolic MDH proteins from other eukaryotes, supporting the conclusion that they both represent authentic mitochondrial proteins. In fact, the bacteria-like MDH bears an unusually long N-terminal extension relative to cytosolic homologs (~100 amino acids), suggesting it may retain an N-terminal extension even after removal of the N-terminal mTP. Interestingly, BN-PAGE analyses demonstrated that this protein (but not the other mitochondrial MDH) may interact with the mitochondrial ATP synthase of *A. castellanii* (see Section 2.4.5.5.2.3).

Table 2.2 *A. castellanii* mitochondrial proteins associated with pyruvate metabolism and the TCA cycle. Proteins identified via MS/MS analyses (MS) are noted by "+", whereas proteins identified only by bioinformatic searches are marked as "-". "Identifier" refers to a unique identifier (from clustered EST or PGP datasets) that can be used to retrieve the relevant entry from Supplemental File 2.5.

Protein	MS	Identifier
Pyruvate metabolism		
Pyruvate dehydrogenase		
PDH E1 α	+	15930
PDH E1 β	+	19193
PDH E2	+	7200
	+	10054
PDH E3	+	588
Pyruvate dehydrogenase-associated		
Pyruvate carboxylase	+	23294
Pyruvate dehydrogenase kinase		5370/2871
Pyruvate dehydrogenase phosphatase regulatory subunit	+	3418
Pyruvate dehydrogenase phosphatase catalytic subunit	+	27417
Pyruvate dehydrogenase kinase	-	2871
Anaerobic pyruvate metabolism		
Pyruvate:ferredoxin oxidoreductase	+	16068
Acetate:succinate CoA transferase	+	19900
Ferredoxin	+	18230
HydF	+	g4978.t1
Tricarboxylic acid cycle		
Citrate synthase	+	19816
Aconitase	+	17472

Protein	MS	Identifier
	+	806
Isocitrate dehydrogenase (NAD^+), α	+	8357
Isocitrate dehydrogenase (NAD^+), β	+	22692
Isocitrate dehydrogenase ($NADP^+$)	+	11724
2-Oxoglutarate dehydrogenase		
E1 α	+	813
	+	20262
E1 β	+	5667
E2	+	8827
E3	+	588
Succinyl-CoA synthetase		
α subunit	+	9628
β subunit	+	2792
	+	20053
Succinate dehydrogenase		
SdhA	+	11902
SdhB	+	16868
SdhC	+	5019
SdhD	+	22329
Fumarase, class I	+	5809
MDH, bacterial type	+	10193
MDH, cytosolic type	+	26422

2.4.5 Respiration

In spite of the many diverse cellular functions associated with mitochondria, they are best known for the vital role they play in cellular energy generation, primarily through coupled oxidative phosphorylation [209]. While the assembly, composition and function of the respiratory chain is understood relatively well in certain prokaryotic and eukaryotic model species, significantly less is known about these complexes in protists. Below I present a synopsis of the *A. castellanii* electron transport chain (ETC) that combines direct MS/MS analyses of individual respiratory chain complexes separated via BN-PAGE along with bioinformatic interrogations of the *A. castellanii* mitochondrial proteome and EST datasets.

2.4.5.1 Complex I (NADH:Ubiquinone Oxidoreductase)

Complex I (CI) of the mitochondrial respiratory chain is a large, multi-subunit molecular machine that couples the energy derived from the transfer of electrons from NADH to ubiquinone via flavin mononucleotide (FMN) and 8 Fe-S clusters to the active translocation of protons across the inner mitochondrial membrane (IMM). Structurally, CI is an L-shaped complex with a membrane-integrated domain and an NADH-oxidizing, peripheral domain projecting into the mitochondrial matrix [76; 210]. In plants, another domain, made up of γ-type carbonic anhydrases, projects into the mitochondrial matrix from the middle of the membrane arm [115; 117].

E. coli CI consists of 14 subunits (Nad1-Nad11 and Nad4L along with 24-kDa and 51-kDa subunits); however, the corresponding α-proteobacterial complex was recently shown [126] to have 17 subunits, including three additional proteins (B17.2, AQDQ and 13 kDa) that were previously believed to be specific to mitochondrial CI. In order to

simplify comparisons with other publications, I will continue to refer here to the 'traditional' 14-subunit bacterial core. In contrast to the bacterial enzyme, the mitochondrial complex has accrued a large number of 'supernumerary' subunits [75; 76]. For instance, bovine and *A. thaliana* CI comprise 45 [107] and 49 [108] components, suggesting a gain of 31 and 35 subunits, respectively. At least 19 of the supernumerary subunits were acquired early in eukaryotic evolution, although certain others appear to have evolved within particular eukaryotic lineages [75].

Many of the novel eukaryotic additions to CI are small proteins, often with single transmembrane helices. In general, little is known about their specific function(s), although it is assumed that many may play a role in the assembly and stability of CI [76], which unlike its bacterial counterpart, is under dual genetic control. Conversely, certain other supernumerary subunits confer novel properties on mitochondrial CI; for instance, the SDAP subunit (also known as acyl-carrier protein), which is a CI subunit in animals, fungi and green algae [25; 103; 107], plays a role in fatty acid metabolism while subunit B16.6 (known also as GRIM-19) is an apoptosis-inducing factor in animals that is dual-localized to CI and the nucleus [211; 212].

2.4.5.1.1 Size of CI

Enzymatically active CI isolated via BN-PAGE was detected as an ~940-kDa complex, as demonstrated by in-gel enzyme activity assay (Figure 2.4), while an inactive ~820-kDa isoform was identified by 2D BN/SDS-PAGE profile similarity to active CI [182]. The estimated size for monomeric CI is consistent with the size reported for CI from a variety of other eukaryotes, including a prior investigation of *A. castellanii* CI [213]. After staining with Coomassie Blue R250, the band representing active CI was very faint

relative to neighboring bands, such as dimeric Complex V (discussed below), whereas the 820-kDa isoform co-migrates with other complexes, including a breakdown product of CV (see below). As a result, comprehensive MS/MS detection of subunits from isolated CI has proven difficult, although all prokaryotic and eukaryotic core subunits were identified via a combination of BN-PAGE with detailed bioinformatic searches of proteins identified in my whole mitochondrial proteome analysis and/or relevant *A. castellanii* sequence databases.

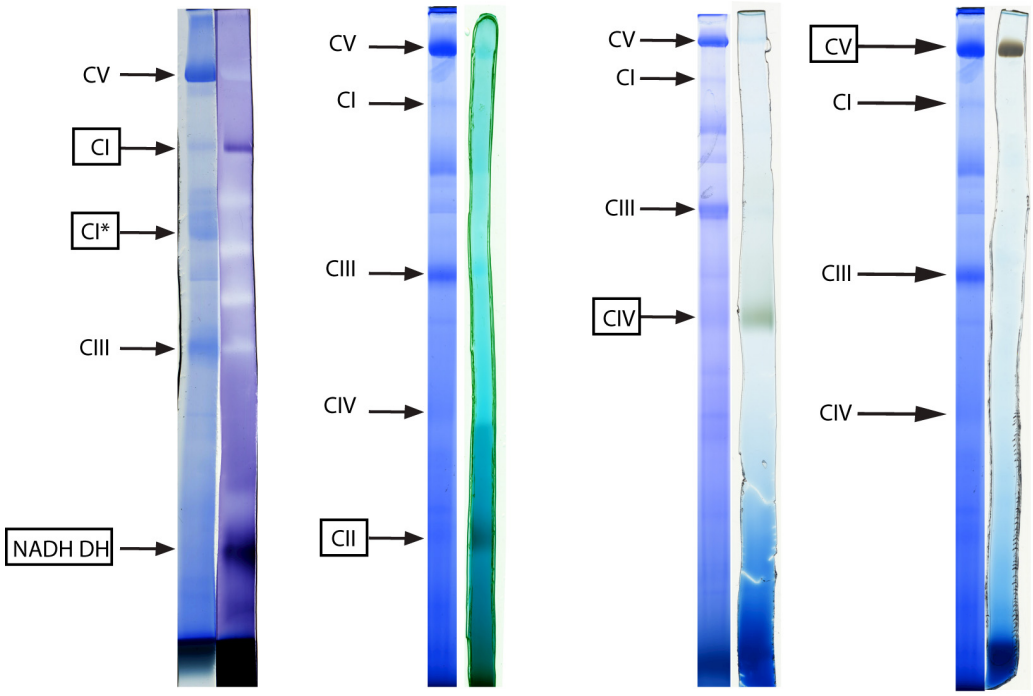


Figure 2.4 BN-PAGE profiles and in-gel activity assays of *A. castellanii* ETC complexes. Gel strips from BN-PAGE experiments stained with either Coomassie Blue R250 (left) or enzyme activity stain (right) for each complex are shown. The relevant ETC complex name in each case is demarcated with a box. No specific CIII activity stain is available, so CIII is indicated only by an arrow in Coomassie Blue R250-stained gels. Profiles are not identical as not all gels consisted of the same acrylamide gradient.

2.4.5.1.2 Detection of CI Subunits via BN-PAGE

MS/MS analyses were performed on both the 940-and 820-kDa CI isoforms isolated via BN-PAGE. The CI subunits detected in each of these bands were similar, although more proteins were identified in the 820-kDa isoform due to its higher abundance. Some differences, including the detection of the 51-kDa subunit in the 940-kDa band, but not the 820-kDa band, were noted; this particular finding is in agreement with in-gel enzyme activity assays, as the 51-kDa subunit forms part of the N module of CI that is responsible for oxidation of NADH [76].

In the 940-kDa complex, nuDNA-encoded subunits that comprise the peripheral arm of CI (corresponding to sectors Iλ and FP of the bovine enzyme) were slightly overrepresented in comparison with nuDNA-encoded constituents of the hydrophobic membrane arm of CI (Iβ and Iγ sectors; 8/11 vs. 9/16 subunits, respectively), whereas the reverse was observed in the 820-kDa complex (5/11 vs. 13/16, respectively). This trend is also apparent with mtDNA-encoded CI subunits; in the native, 940-kDa complex, 0 of 6 mtDNA-encoded subunits of the membrane arm were detected by MS/MS following BN-PAGE, whereas 3 of 3 mtDNA-encoded peripheral arm subunits were identified, all with high ion scores and multiple peptides. On the other hand, Nad1, Nad3, Nad4 and Nad5 were identified in the 820-kDa complex, although by only one convincing peptide in each case. In addition to subunits of the membrane and peripheral domains, I also detected two homologs of γ-type carbonic anhydrases in both CI isoforms, as described elsewhere [182].

Of the 19 putative supernumerary subunits added to CI early in eukaryotic evolution (but see [126]), I identified 17 in one or both of the 940- and 820-kDa complexes via MS/MS after separation by BN-PAGE (only AQDQ and the 13-kDa subunit were not

detected) and 8 in both complexes. Interestingly, among these is SDAP (acyl-carrier protein) – an experimentally validated component of CI in animals, fungi and green algae, but a soluble matrix protein in plants [214]. This subunit was also detected in a CI-enriched BN-PAGE band from the ciliated protozoon *T. thermophila* (my unpublished results). Together, these results suggest that SDAP is truly an ancestral component of mitochondrial CI that may have been 'lost' from CI in the plant lineage. Also, I detected a homolog of the B14.7 subunit, which itself is homologous to the Tim17/Tim22/Tim23 protein import family. An unambiguous orthologous relationship to B14.7 (vs. Tim17/Tim22/Tim23) could not be established based on the sequence alone; however, detection of this protein in both CI isoforms argues strongly that it is B14.7. In addition, more convincing candidates for each of Tim17, Tim22 and Tim23 have been identified (see Table 2.10, below).

Two putative CI subunits identified by MS/MS analysis in both the 940- and 820-kDa complexes, but with no apparent relationship to any other proteins in available sequence databases (as inferred via BLASTp), were annotated as the core eukaryotic CI subunits MWFE and ASHI after HMMER profile searches. MWFE and ASHI are both small proteins associated with the membrane arm of CI, each containing one predicted transmembrane helix. In *D. discoideum*, each of these subunits is more similar to other eukaryotic homologs than is the corresponding *A. castellanii* protein. MWFE and ASHI are conserved across Eucarya (although ASHI was previously thought to be limited to animals and fungi [75]) and are components of CI in animals [107], fungi [103] and plants [108].

Interestingly, BN-PAGE analyses have uncovered homologs of several CI subunits (B14.5b, B15 and NUUM) that had been suggested to be restricted to other

lineages. An *A. castellanii* protein (found only in the 820-kDa complex) was shown to be homologous to a 'plant-specific' CI protein (conserved domain PLN02806) by BLASTp. Further PSI-BLAST searches suggest that proteins of this family are, in turn, homologous to B14.5b (animal-specific). Likewise, B15 (animals + fungi [75]), which was convincingly annotated via BLASTp/PSI-BLAST searches albeit at relatively high E-values, is also found in a wide variety of eukaryotes. Lastly, I detected a homolog of NUUM, previously reported only in the fungus *Pichia pastoris* [215]. Unlike B14.5b and B15, homologs of NUUM were readily identified at relatively low E-values (1 e-10). BLASTp searches using the putative *A. castellanii* NUUM as the query identified homologs in a wide variety of eukaryotes; notably, a homolog of NUUM (but not recognized as such) has been detected as a component of CI in *A. thaliana* [108]. These results suggest that NUUM, B14.5b and B15 were also likely early additions to mitochondrial CI and not recent lineage-specific additions. At least one other putative *A. castellanii*-specific CI subunit was identified by MS/MS (Contig11; Table 2.3, Supplemental File 2.6).

Aside from subunits characterized in several eukaryotic species, BN-PAGE analyses uncovered other potential CI-associated proteins that are functionally annotated in other organisms and that in *A. castellanii* may be lineage-specific CI subunits or proteins that play a role in CI biogenesis (i.e., they may interact transiently with CI). For instance, in both the active and inactive CI isoforms, a rhodanese-related sulfurtransferase was detected. A significant proportion of bovine rhodanese is mitochondrial membrane-associated, where it binds multiprotein complexes and is capable of forming Fe-S clusters [216]. Mitochondrial CI contains 8 Fe-S clusters, forming an electron wire [76]; thus this protein may be involved in CI Fe-S cluster biosynthesis in *A.*

castellanii. Interestingly, an unrelated rhodanese domain-containing protein has been identified as a stable CI subunit in the aerobic yeast, *Yarrowia lipolytica* [103]. In another example, a peptide from an ORF predicted to specify RibB, RibA and RibH (encoding riboflavin synthetic enzymes; the peptide match is to the RibH sequence) was detected by MS/MS in the enzymatically-active CI complex (but not the inactive complex). Riboflavin is a core component of flavoprotein cofactors, including FMN, the NADH-oxidizing cofactor of CI. Although it is unclear whether this is an authentic mitochondrial protein (as it lacks an N-terminal mTP and was not found in the whole proteome MS/MS analysis), the association of active FMN-containing CI with flavin-synthesizing machinery is logical, and an involvement of yeast Rib3 (RibB homolog) in mitochondrial respiration (albeit not flavin synthesis) has been reported [217]. Finally, multiple peptides were confidently identified from a trans-2-enoyl-CoA reductase and an acyl-CoA synthetase. In fungi, trans-2-enoyl-CoA reductase, which plays a role in fatty acid biosynthesis, is essential for mitochondrial respiratory competence [218]. It is tempting to speculate that these proteins function in concert with SDAP in mitochondrial fatty acid metabolism. Ultimately, further experimentation will be required in order to establish whether these proteins are *bona fide* CI subunits in *A. castellanii* and what roles they may play.

Although enzymatically active and inactive CI isoforms were detected, a relatively large number of contaminating non-CI mitochondrial proteins was identified via MS/MS in each. In particular, subunits of other respiratory chain complexes and as well as other abundant mitochondrial proteins were found.

2.4.5.1.3 CI Subunits Detected via Proteome Analysis and/or Bioinformatic Searches

As described above, CI is the largest and most intricate of the respiratory chain complexes, typically comprising 35-45 subunits. Although difficulties in obtaining large quantities of BN-PAGE-purified CI have precluded an exhaustive direct MS/MS analysis of the enzyme complex, homology searches for known CI subunits from model organisms have allowed the identification of most classical CI subunits in the *A. castellanii* mitochondrial proteome dataset and EST sequence database. Importantly, 19/19 proposed eukaryotic core CI subunits and 12/14 prokaryotic core CI subunits (Nad2 and Nad4L were not detected) were identified by MS/MS analysis of mitochondria, suggesting an excellent coverage of CI proteins. Among the proteins identified by MS/MS of mitochondria, but not BN-PAGE isolated CI, are the 24-kDa subunit (NADH-oxidizing domain), AQDQ and Nad6 (mtDNA-encoded).

Bioinformatic analyses have uncovered two putative CI subunits not detected by MS/MS analysis. The predicted 8.8-kDa AGGG subunit was identified by tBLASTn searches of the *A. castellanii* EST dataset. AGGG has been identified in biochemical analyses of purified bovine [107] and *A. thaliana* CI [108]. My BLAST analyses have shown that AGGG can be identified in the nuclear genomes of a wide variety of eukaryotes as well. Lastly, a putative homolog of MLRQ (thought to be animal-specific) was identified via BLAST. Specifically, an MLRQ homolog from *P. pallidum* (conserved domain BLAST analyses indicate similarity to the NADH_ub_rd_NUML protein family at an E-value of 8.9e-4) was used to retrieve a small (~10-kDa) *A. castellanii* protein (*D. discoideum* and *P. polycephalum* homologs are more conserved than the *A. castellanii* counterpart, so that queries using the animal MLRQ sequences do not retrieve the putative *A. castellanii* MLRQ). Bioinformatic searches also identify MLRQ in

Reclinomonas americana, and various opisthokont species. In combination with results presented above and other recent discoveries [182], these results suggest that the ancestral eukaryotic CI may have been even larger than previously believed ($\geq$ 39-40 subunits).

In comparison with amoebozoan relatives (*D. discoideum*, in particular), it appears that, in general, the same subunits of CI are conserved. Apart from the ESSS subunit (and the putative *A. castellanii*-specific subunit), all CI constituents identified in *A. castellanii* can be identified bioinformatically in *D. discoideum*. Notably, however, *A. castellanii* and *D. discoideum* CI protein sequences are generally not more similar to each other than each is to non-amoebozoan CI sequences.

2.4.5.1.4 CI Assembly Proteins

As noted above, CI is under dual genetic control and contains a number of cofactors, including 8 Fe-S clusters and FMN, which must be efficiently integrated for proper function. Although our understanding of CI assembly has suffered due to the absence of CI in *S. cerevisiae*, significant progress has been made recently [219]. MS/MS analysis of *A. castellanii* mitochondria has uncovered a number of characterized CI assembly factors, including homologs of Ind1 [220; 221] (a P-loop NTPase specifically required for assembly of CI Fe-S clusters), C8orf38 [24] (similar to phytoene synthase), C20orf7 [222] and MidA [223] (both containing a methyltransferase domain). Conversely, we failed to detect Cia30/NDUFAF1 [224], NDUFAF3 [225], FOXRED1 [226; 227] and B17.2L [228], although all are nuDNA-encoded in *A. castellanii*. No established CI assembly factors were identified by MS/MS analyses of CI isolated via BN-PAGE. Intriguingly, my MS/MS analyses also identified a number of possible (but not established) CI assembly proteins identified by phylogenetic profiling [24]. Among these proteins are homologs of

3,2 trans-enoyl-CoA isomerase, methylcrotonoyl-Coenzyme A carboxylase β subunit (Mccc2), isovaleryl coenzyme A dehydrogenase and glycerol-3-phosphate acyltransferase, dihydropicolinate synthase and Lyrm5. A homolog of Mccc2, along with Mccc1 (α subunit), was detected by MS/MS in the 820-kDa CI isoform isolated by BN-PAGE, but not the 940-kDa isoform.

Table 2.3 *A. castellanii* CI subunits and CI-associated proteins. Proteins identified via a combination of BN-PAGE-MS/MS analysis of the 940- ("940") and 820-kDa ("820") CI isoforms, as well as MS/MS analyses of whole mitochondria ("MS") are reported. All proteins listed can be found through bioinformatic searches of *A. castellanii* sequence databases. Whether the protein in question is present in other eukaryotes ("Mito."), bacteria ("Bact.") and other amoebozoans ("Amoeb.") is also indicated. The genome in which a given protein is encoded ("Genome") is noted (N; nuclear, M; mitochondrial). Protein sequences for all entries are available in Supplemental File 2.6.

			CI structural proteins					
Protein	**Genome**	**Sector**	**MS**	**940**	**820**	**Mito.**	**Bact.**	**Amoeb.**
Nad1	M	Iγ	+		+	+	+	+
Nad2	M	Iγ/ Iβ				+	+	+
Nad3	M	Iγ/ Iβ	+		+	+	+	+
Nad4	M	Iβ	+		+	+	+	+
Nad4L	M	Iγ/ Iβ				+	+	+
Nad5	M	Iβ	+		+	+	+	+
Nad6	M	Iγ/ Iβ	+			+	+	+
Nad7	M	Iλ	+	+	+	+	+	+
Nad8	N	Iλ	+	+	+	+	+	+
Nad9	M	Iλ	+	+	+	+	+	+
Nad10	N	Iλ	+	+	+	+	+	+
Nad11	M	Iλ	+	+	+	+	+	+
24 kDa	N	FP	+			+	+	+
51 kDa	N	FP	+	+		+	+	+
39 kDa	N	Iγ	+	+	+	+	+[a]	+
B14.7	N	Iλ	+	+	+	+		+[c]
15 kDa	N	Iγ	+		+	+		+
13 kDa	N	Iλ	+			+	+[b]	+
AQDQ	N	Iλ	+			+	+[b]	+
ESSS	N	Iβ	+		+	+		
MWFE[d]	N	Iγ	+	+	+	+		+

Protein	Genome	Sector	MS	940	820	Mito.	Bact.	Amoeb.
PDSW	N	Iβ	+		+	+		+
PGIV	N	Iγ	+	+	+	+		+
SDAP	N	Iγ/Iβ	+	+	+	+	+[a]	+
B22	N	Iβ	+	+	+	+		+
B18	N	Iβ	+		+	+		+
B17.2	N	Iλ	+	+		+	+[b]	+
B16.6	N	Iλ	+		+	+		+
B14	N	Iγ	+	+		+		+
B13	N	Iλ	+	+	+	+		+
B12	N	Iβ	+	+	+	+		+
B8	N	Iλ	+	+		+		+
AcCa1	N	CA	+	+	+	+	+[a]	+
AcCa2	N	CA	+	+	+	+	+[a]	+
ASHI[d]	N	Iβ	+	+	+	+		+
C11	N	?	+	+	+			
B15	N	Iγ/Iβ	+	+	+	+		+
AGGG	N	Iβ				+		+
B14.5b[d]	N	Iβ	+		+	+		+
20.9 kDa	N	Iγ/Iβ	+		+	+		+
NUUM	N	?		+	+	+		+
MLRQ	N	Iγ/Iβ				+		+

Protein	Genome	Sector	MS	940	820	Mito.	Bact.	Amoeb.
Established CI assembly proteins								
Ind1	N	n/a	+			+	+	+
C8orf38	N	n/a	+			+	+	+
C20orf7	N	n/a	+			+	+	+
MidA	N	n/a	+			+	+	+
Ndufaf1	N	n/a				+	+	+
Ndufaf3	N	n/a				+	+	+
Foxred1	N	n/a				+	+	
B17.2L	N	n/a				+		+

[a] These proteins have bacterial homologs but are not known to be components of CI in bacteria.
[b] These proteins have bacterial homologs and are known to be CI components in α-proteobacteria [126].
[c] Other amoebozoan species likely have B14.7 homologs (member of Tim17/Tim22/Tim23 family) but definitive orthologs could not be established on the basis of similarity searches alone.
[d] These subunits were annotated via HMMER 3.0 profile searches.

2.4.5.2 Complex II (Succinate:Ubiquinone Oxidoreductase)

From a structural perspective, CII is the simplest of the ETC complexes, as it is composed of only 4 subunits in bacteria, animals and fungi; however, 8 subunits have been reported in plants [105] and 12 subunits in trypanosomes, including a heterodimeric Fe-S subunit [137; 229]. The classical CII subunits include the hydrophilic subunits SdhA (containing a covalently bound FAD prosthetic group) and SdhB (containing 3 Fe-S

clusters), which catalyze the oxidation of succinate and the transfer of electrons, respectively, plus the membrane-integrated subunits, SdhC and SdhD, which form a hydrophobic binding pocket for ubiquinone [230].

Complex II is unique among the respiratory chain enzyme complexes in several ways. For instance, CII is considered a component of both the ETC and the TCA cycle; it is the only ETC complex that does not actively translocate protons from the mitochondrial matrix to the intermembrane space (excluding CV – ATP synthase – which is not formally a component of the ETC); and, in a large percentage of eukaryotes, all of the genes encoding CII proteins are located in the nucleus.

Consistent with the size of CII in other eukaryotes, I found enzymatically active *A. castellanii* CII to be approximately 130 kDa (Figure 2.4). MS/MS analyses of CII excised from BN-PAGE experiments detected the 4 classical subunits of CII (Table 2.4), all of which are nuDNA-encoded. Moreover, each of these 4 proteins was identified in each of my mitochondrial proteome analyses, indicating that the components of CII are relatively abundant. No *A. castellanii*-specific CII proteins were confidently identified in my MS/MS analysis of CII isolated via BN-PAGE; however, a fairly large number of non-CII mitochondrial proteins were detected in this sample. The reason for this result is likely related to the small physical size of CII: whereas few proteins non-specifically contaminate larger complexes, smaller protein complexes are more prone to contamination in BN-PAGE because many protein monomers and multimers are of a similar molecular size. As a result, it is possible that the detection of novel, hypothetical proteins by MS/MS has been obscured by the presence of other abundant, non-CII mitochondrial proteins. Nonetheless, the small size of the *A. castellanii* complex

suggests that few proteins, if any, were missed, as the combined estimated molecular weights of SdhA-D are in agreement with molecular weight estimates from BN gels of the complex as a whole.

Interestingly, while *sdhA*, *sdhC* and *sdhD* appear to be single-copy genes in *A. castellanii*, two *sdhB* loci encode distinct isoforms of SdhB. Only one of the two SdhB variants was detected in isolated CII or the entire mitochondrial proteome. EST coverage of the undetected isoform is not extensive; hence, it is not clear whether this variant is a constitutively expressed but minor isoform present at all times, whether it is required only under specific conditions, or whether it is a non-functional isoform.

Protein	Genome	MS	BN-MS	Mito.	Bact.	Amoeb.
SdhA	N	+	+	+	+	+
SdhB	N	+	+	+	+	+
SdhC	N	+	+	+	+	+
SdhD	N	+	+	+	+	+

Table 2.4 *A. castellanii* CII subunits. Proteins identified via BN-PAGE-MS/MS ("BN-MS") as well as MS/MS of whole mitochondria ("MS") are reported. Whether the protein in question is present in other eukaryotes ("Mito."), bacteria ("Bact.") and other amoebozoans ("Amoeb.") is also indicated. The genome in which a given protein is encoded ("Genome") is noted (N; nuclear, M; mitochondrial). Protein sequences for all entries are available in Supplemental File 2.7.

2.4.5.3 Complex III (Ubiquinol:Cytochrome *c* Oxidoreductase)

In bacteria, Complex III (CIII) comprises 3 subunits – cytochrome *b* (Cob), the Rieske Fe-S protein (Isp) and cytochrome *c*1 (CytC1) – all of which are catalytic subunits required for the transfer of electrons from ubiquinol to cytochrome *c* and the

translocation of protons across the inner membrane. Mitochondrial CIII is more complex, typically containing a well-conserved set of 6-7 supernumerary subunits that are not directly involved in electron transfer in addition to the 3 bacterial core catalytic subunits [231-233]. Among these supernumerary subunits are the large, so-called "core" proteins (Core1 and Core2; these have homologs in bacteria, but the bacterial homologs are not associated with CIII) that project into the mitochondrial matrix and a collection of smaller, eukaryote-specific proteins (Qcr6-Qcr10). In all CIII-containing eukaryotes studied to date, Cob is mtDNA-encoded, whereas the other subunits are encoded in the nucleus.

2.4.5.3.1 Composition of CIII.

A. castellanii CIII resolved by BN-PAGE was estimated to be ~550 kDa in size based on comparison to high-molecular-weight markers (Figure 2.4). This approximation is in general agreement with the typical size of eukaryotic CIII, and suggests that CIII likely exists as a stable homodimer in *A. castellanii*, as it does in other eukaryotes and prokaryotes.

Initial bioinformatic searches of the *A. castellanii* EST dataset using tBLASTn retrieved all expected proteins except for Qcr10, which is the least conserved of the Qcr proteins in terms of sequence. MS/MS analysis of BN-PAGE-isolated CIII detected all of the subunits identified bioinformatically. Additionally, MS/MS analysis of CIII isolated via BN-PAGE revealed a predicted 10.7-kDa protein that retrieves the human homolog of Qcr10 when used as a BLASTp query. Although slightly larger than Qcr10 homologs from other eukaryotes (8.6 kDa and 6.5 kDa in yeast and bovine mitochondria, respectively), a multiple protein alignment and profile similarity search using authentic

Qcr10 homologs suggest that the 10.7-kDa protein is a *bona fide* Qcr10 homolog. All 10 CIII subunits were detected by MS/MS in my analysis of total *A. castellanii* mitochondrial protein fractions. In sum, these results suggest that CIII in *A. castellanii* is quite similar in composition to the well-described CIII of bovine, yeast and plant mitochondria (Table 2.5).

2.4.5.3.2 Evolution of CIII Core Proteins

In certain land plants, it has been demonstrated that the Core proteins of CIII constitute the mitochondrial processing peptidase activity responsible for the proteolytic cleavage of mitochondrial targeting sequences [131; 234]. In contrast, in *Saccharomyces cerevisiae* [235] and animals [236], an evolutionarily related, heterodimeric mitochondrial processing peptidase (MPP), comprising MPPα (Core2 homolog) and MPPβ (Core1), exists in the mitochondrial matrix. An intermediate scenario has been reported in *Neurospora crassa* [237] and *D. discoideum* [238], as the sole isoform of Core1/ MPPβ exists in both the membrane and the matrix, while MPPα is found only in the matrix (and a separate Core2 is associated with CIII). In animals and *S. cerevisiae*, the catalytic matrix MPPβ subunit possesses the Zn-binding His-x-x-Glu-x-His-x~76-Glu motif characteristic of all active pitrilysin protease family members, whereas the CIII-associated Core1 does not. Conversely, the CIII-associated Core1/MPPβ protein in plant CIII has retained this motif [114].

Similar to plant mitochondria and *D. discoideum*, in *A. castellanii* the Core1 protein associated with CIII is the only isoform of Core1/ MPPβ present (i.e., a distinct MPPβ protein could not be identified in the *A. castellanii* nuclear genome or EST databases). Furthermore, *A. castellanii* Core1/MPPβ has the characteristic pitrilysin motif

described above, suggesting that it is likely to possess mitochondrial processing peptidase activity. There are two separate proteins of the Core2/MPPα family encoded by *A. castellanii*, as is the case in *D. discoideum*; however, only one protein, Core2, was detected in the CIII fraction isolated by BN-PAGE. Core2 appears to be by far the more abundant of the two proteins, as its MS/MS ion scores were among the highest in the entire proteome (strongly detected in each of four fractions), whereas the other homolog was detected by only 1 high-scoring peptide in one of the four separate fractions tested (SWM). It is not clear why there are two distinct Core2/MPPα family proteins in *A. castellanii*, although it is a formal possibility that a fraction of the Core1/MPPβ protein of *A. castellanii* associates with Core2 (in CIII), while another fraction associates with the minor MPPα family protein in the mitochondrial matrix, as reported in *D. discoideum*. CIII-associated Core1/MPPβ and Core2/MPPα were found in abundance in both the membrane-protein-enriched and soluble-protein-enriched fractions by MS/MS analyses, although ion scores were higher in membrane-protein-enriched fractions (for instance, Core1/MPPβ represented the 41st and 6th highest ion scores in soluble and membrane-protein-enriched fractions, respectively).

A maximum likelihood phylogenetic reconstruction of Core1/MPPβ proteins (Supplemental Figure 2.1) suggests that relatively recent gene duplications or losses within established eukaryotic lineages likely account for the duality of the Core1/ MPPβ family, as opposed to an ancient gene duplication early in eukaryotic evolution. Specifically, the human, yeast and *T. thermophila* Core1 proteins are more closely related to MPPβ homologs from the same species, rather than to other Core1 proteins from other species. Thus, it may be that independent gene duplications of the Core subunits resulted

Protein	Genome	MS	BN-MS	Mito.	Bact.	Amoeb.
CIII structural proteins						
Core1	N	+	+	+	+[a]	+
Core2	N	+	+	+	+[a]	+
Cob	M	+	+	+	+	+
Isp	N	+	+	+	+	+
Cytc1	N	+	+	+	+	+
Qrc6	N	+	+	+		+
Qcr7	N	+	+	+		+
Qcr8	N	+	+	+		+
Qcr9	N	+	+	+		+
Qcr10	N	+	+	+		+
CIII assembly proteins						
Bcs1	N	+		+	+[b]	+
Cbp3	N	+		+	+	+

Table 2.5 *A. castellanii* CIII subunits and CIII-associated proteins. Proteins identified via BN-PAGE-MS/MS ("BN-MS") as well as MS/MS of whole mitochondria ("MS") are reported. Whether the protein in question is present in other eukaryotes ("Mito."), bacteria ("Bact.") and other amoebozoans ("Amoeb.") is also indicated. The genome in which a given protein is encoded ("Genome") is noted (N; nuclear, M; mitochondrial). Protein sequences for all entries are available in Supplemental File 2.8.

[a] The Core1 and Core2 proteins have bacterial homologs (M16 family proteases), but are not components of bacterial CIII.
[b] While AAA^+ ATPase homologs are encoded by bacteria, it is doubtful that they are orthologs.

in the establishment of independent separate matrix processing peptidase and the inactivation of the Core subunits in certain lineages (like animals, fungi and ciliates) while the CIII Core1/MPPβ protein retained enzymatic activity in other eukaryotes (plants and possibly *A. castellanii*). Ultimately, it will be important to determine experimentally the suborganellar localization of mitochondrial processing peptidase activity in *A. castellanii* (i.e., whether it is associated with CIII).

2.4.5.3.3 CIII ASSEMBLY PROTEINS

In yeast, several factors aid in the assembly of CIII [239]; among these, Cbp3 and Bcs1 are both conserved in a wide variety of eukaryotes. Bcs1 is a member of the AAA^+ ATPase family and plays a specific role in the insertion of ISP into CIII [240]. The role of Cbp3 is not well understood, although deletion of Cbp3 leads to CIII structural defects and decreased levels of Cob, Qcr7 and Qcr8 [241]. Homologs of both Bcs1 and Cbp3 were identified in *A. castellanii* mitochondria via MS/MS.

2.4.5.4 Complex IV (Cytochrome *c*:O_2 Oxidoreductase)

Complex IV (CIV) is the terminal enzyme complex of the mitochondrial ETC, catalyzing the reduction of O_2 to H_2O. Mitochondria have inherited 3 subunits (Cox1-3) from the protomitochondrial endosymbiont; however, the mitochondrial complex has accrued a number of additional subunits, with a grand total of 13, 12, >10 and 7 subunits reported in human, yeast, *A. thaliana* and *D. discoideum* CIV, respectively [105; 129; 242; 243]. In eukaryotes, the large, multipass membrane-spanning protein Cox1 is universally encoded in the mtDNA where CIV exists (but see [177]), whereas Cox2 and Cox3 are encoded in either mtDNA or nuDNA, depending on the organism [36]. In *A. castellanii*, Cox1-3 are encoded in mtDNA. Atypically, as noted earlier, Cox1 and Cox2 are transcribed from a continuous ORF in *A. castellanii* mtDNA [178] and in the mitochondrial genome of the related amoebozoon *D. discoideum* [244; 245], although western blot investigations indicate that mature Cox1 and Cox2 exist as separate proteins in *A. castellanii* mitochondria [194].

2.4.5.4.1 Analysis of CIV Composition in *A. castellanii*

In-gel enzyme activity assays for CIV after BN-PAGE indicated that the majority of *A. castellanii* CIV exists in a complex of ~360 kDa (Figure 2.4), suggesting that CIV may be predominantly dimeric under the described conditions. The band exhibiting the majority of cytochrome oxidase activity was quite broad in comparison with other bands; notably, activity assays performed after BN-PAGE using a minigel setup demonstrated the presence of two similarly sized, enzymatically active bands (Supplemental Figure 2.2), possibly explaining the breadth of this band in the large-format gel. It is not known how these similarly sized complexes differ in terms of subunit composition, or whether

this observation reflects a physiologically meaningful difference (vs. differential breakdown of the complex during electrophoresis). Similarly, plant mitochondrial CIV exists as two distinct 350- and 280-kDa complexes after BN-PAGE, with the larger complex including an additional subunit homologous to Cox6b [105; 246]. These results are in contrast to previous reports on *A. castellanii* CIV after BN-PAGE [213], which detected enzymatic activity in a smaller, ~160-kDa band. It is unclear why the estimated molecular weight of CIV is different between the two studies, although the fact that mitochondria were harvested from stationary phase cells in the prior report may partially explain this discrepancy.

2.4.5.4.2 CIV SUBUNIT COMPOSITION IN *A. CASTELLANII*

CIV was detected by MS/MS in several different bands isolated from BN-PAGE, although it was not the sole (or dominant) species present in any of these bands. CIV subunits (Table 2.6) were detected in a band immediately below the CIII band (NAD^+ isocitrate dehydrogenase was the dominant protein in this population) after linear sucrose gradient centrifugation and BN-PAGE (not shown) and in the broad ~360-kDa band in which CIV activity was detected. In both CIV-enriched bands, the mtDNA-encoded Cox1/2 and Cox3 proteins along with nuDNA-encoded proteins Cox1-c [177], Cox4 and Cox5 were detected. Additionally, a predicted 8.7-kDa protein, which was not identified in the whole proteome analysis, was detected in both samples. Initially, this protein could not be annotated via BLASTp analysis; however, HMMER searches suggest that this protein may be a homolog of Cox7. In contrast to the above-mentioned subunits, Cox6a and another protein, likely homologous to Cox8, were detected only in the active ~360-kDa band. Interestingly, two Cox8 homologs (69% identical in amino acid

sequence) are encoded in the *A. castellanii* nuclear genome, although only one of these variants was identified via BN-PAGE in my proteome analysis. Two homologs of Cox6b were detected in the analysis of the complete mitochondrial proteome (one is considerably more similar to Cox6b from other eukaryotes); however, neither was detected via BN-PAGE. Note that Cox6b was not detected in an analysis of CIV from *D. discoideum* [242], suggesting that it may dissociate readily from the complex during isolation. In sum, my analyses suggest that *A. castellanii* CIV is likely to consist of $\geq$ 10 subunits.

2.4.5.4.3 CIV Assembly Proteins

Mature CIV contains copper and heme cofactors that play crucial roles in the transfer of electrons from cytochrome *c* to O_2. These metal cofactors are contained within the conserved, hydrophobic, mtDNA-encoded proteins Cox2 (Cu_A; electron entry site) and Cox1 (heme *a* and Cu_B-heme a_3 binuclear centre). The assembly of CIV has been intensively studied in yeast and in excess of 20 factors involved in processes such as translational regulation, heme *a* synthesis, and insertion of copper/heme have been characterized (many of these proteins appear specific to yeast) [247; 248]. MS/MS analyses of *A. castellanii* mitochondria have identified proteins involved in the maturation of copper centres, including Cox11, Cox17, Cox19, Cox23, Cmc1 and Sco1. In addition, MS/MS analyses uncovered Pet100, a CIV chaperone. In contrast, Cox10 and Cox15 – required for the synthesis of heme *a* [248] – as well as the chaperones Shy1 and Pet191, were identified only by bioinformatic analyses.

Table 2.6 *A. castellanii* CIV subunits and CIV-associated proteins. Proteins identified by BN-PAGE-MS/MS ("BN-MS"), MS/MS of whole mitochondria ("MS"), and bioinformatic searches are reported. Whether the protein in question is present in other eukaryotes ("Mito."), in bacteria ("Bact.") and other amoebozoans ("Amoeb.") is also indicated. The genome in which a given protein is encoded ("Genome") is noted (N; nuclear, M; mitochondrial). Protein sequences for all entries are available in Supplemental File 2.9.

Protein	Genome	MS	BN-MS	Mito.	Bact.	Amoeb.
CIV structural proteins						
CoxI/II	M	+	+	+	+	+
CoxIII	M	+	+	+	+	+
CoxI-c	N	+	+	+	+[a]	+[a]
CoxIV	N	+	+	+		+
CoxVa	N	+	+	+		+
CoxVIa	N	+	+	+		+
CoxVIb	N	+		+		+
CoxVII[b]	N		+	+		
CoxVIII	N	+	+	+		+[b]
CIV-associated proteins						
Cox11	N	+		+	+	+
Cox17	N	+		+		+
Cox19	N	+		+		+
Cox23	N	+		+		+
Cmc1	N	+		+		+
Sco1	N	+		+	+	+
Pet100	N	+		+		+
Cox10	N			+	+	+
Cox15	N			+	+	+
Shy1	N			+	+	+
Pet191	N			+		+

[a] CoxI-c is a nuDNA-encoded protein homologous to the C-terminal portion of Cox1. *D. discoideum* also encodes a homolog known to be a component of CIV [242].
[b] The putative CoxVII homolog was annotated based on HMMER 3.0 profile searches.

2.4.5.4.4 Cox1/2 Protein Structure

As noted above, Cox1 and Cox2 are encoded and transcribed as a single unit in *A. castellanii* mitochondria; however, it appears that the proteins exist as separate entities, presumably due to a post-translational cleavage [194]. My analysis provides further support for the conclusion that mature Cox1 and Cox2 do not exist as a fusion protein. Specifically, combined 1D SDS-PAGE and MS/MS experiments have suggested that two individual proteins likely exist, because distinct peptide populations corresponding to predicted mature Cox1 or Cox2 protein sequences were found (Figure 2.5). For instance, in the membrane-enriched fraction, peptides from the predicted mature Cox1 sequence were detected in slices 16-22 (~58 kDa → 100 kDa), whereas peptides belonging to the predicted mature Cox2 were only found in slice 9 (~31 kDa).

MS/MS results also provide insights into other outstanding questions about the structure of mature Cox2. As described elsewhere [194], the *A. castellanii* Cox2 sequence contains several 'insertions' that are predicted to make the mature protein considerably longer than Cox2 from other species (333 vs. 236 amino acid residues in *A. castellanii* and *S. cerevisiae*, respectively); however, *A. castellanii* and yeast Cox2 co-migrate in SDS-PAGE experiments, raising the possibility that the insertion sequences are not a part of the mature Cox2. I have detected peptides from both of these Cox2 'insert' sequences (Figure 2.5), indicating that mature *A. castellanii* Cox2 retains the extra sequences but evidently migrates anomalously faster on SDS-PAGE than would be expected.

MINRLLNNLTSFFTDNRWLFSTNHKDIGTLYLIFGGFSGIIGTIFSMIIRLELAAPGSQILSGNSQLYNV
IITAHAFVMIFFFVMPVMIGGFGNWFVPLMIGAPDMAFPRLNNISFWLLPPSLFLLLCSSLVEFGAGTGW
TVYPPLSSIVAHSGGSVDLAIFSLHLAGISSLLGAINFITTIFNMRVPGLSMHKLPLFVWSVLITAFLLL
FSLPVLAGAITMLLTDRNFNTSFFDPSGGGDPILYQHLFWFFGHPEVYILILPAFGIVSQIIGTFSNKSI
FGYIGMVYAMLSIAVLGFIVWAHHMYTVGLDVDTRAYFTAATMMIAVPTGIKIFSWIATLWGGQIVRKTP
LLFVIGFLILFTLGGLTGIVLSNAGLDIMLHDTYYVVAHFHYVLSMGAVFAFFAGFYYWFWKISGYTYNE
MYGNVHFWLMFIGVNLTFFPMHFVGLAGMPRRIPDYPDNYYYWNILSSFGSIISSVSVIVFFYLIYLAFN
NNNTPKLIK**LVHSIFAPYINTLSKNLLTFASIK**STSDSSFFKFSKFFIFF**MVSLS**VLFIFYDSLLCLNDH
TNSWK**IGFQDPTTPIAYGIIK**LHDHILFFLAVILFVVGYLLLSTYKKFYYGSLNNDLPESKR**ISLFDTLI
NTYKENLSFNVTNR**TYNINHGTTIEIIWTILPAFILLFIAVPSFALLYAMDEIIDPVLTVKVIGHQWYWS
YEYSDYSVVYSNR**MLDYDSIDR**FAAMEMMYKGMGYLKDRSLLSYLYIPMVIPETTIK**FDSYMIHEAELNL
GDLR**LLK**TDMPLFLPK**NTHIR**LLITSSDVLHSWAVPSFGVK**VDAVPGR**LNQTSLYLK**NTGTFYGQCSELC
GVNHAFMPIEVYVVNPVYFYNYVYIYFKNFNLI

Figure 2.5 MS/MS-based evidence for separate Cox1 and Cox2 proteins in *A. castellanii* mitochondria. Identified peptides are inferred from a MASCOT search of mtDNA-encoded proteins from the MPE fraction. **Red** peptides correspond to confidently identified peptides from gel slices 18-22 and appear to be derived from mature Cox1 (~58 kDa - >100 kDa; presumably the protein runs aberrantly due to its highly hydrophobic nature). Note that the final red peptide is actually two different overlapping peptides. **Blue** peptides are from gel slice 9 and seemingly correspond to Cox2 (between 27 and 34.6 kDa). The first 5 residues from proposed protein N-termini are bolded and underlined. These results suggest that the mature C-terminus of Cox1 is located near the '**MVSLS**' believed to be the N-terminus of Cox2. Peptides from the "insert" regions of *A. castellanii* Cox2 (described in [194]) were confidently detected: **ISLFDTLINTYK**, and **ENLSFNVTNR** from one insert and **MLDYDSIDR** from the other. These data suggest that these "inserts" are likely to be retained as a part of the mature Cox2 protein and that the latter may migrate anomalously fast.

2.4.5.5 Complex V (F_OF_1 ATP Synthase)

Complex V (CV) is a multi-subunit enzyme complex that is conserved in bacteria, chloroplasts and the mitochondria of aerobically respiring eukaryotes. CV is responsible for coupling the proton gradient generated by electron transport to the biosynthesis of ATP from ADP and P_i [209]. Structurally, CV is composed of two sectors (F_o and $F_{1)}$ that can be separated from each other under non-denaturing conditions. The F_o sector consists of a stator stalk that prevents futile cycling of the water-soluble, catalytic F_1 α and β head sector subunits that project into the mitochondrial matrix; as well, it contains a membrane-integrated domain that allows protons to re-enter the mitochondrial matrix from the intermembrane space via a channel between Atpa and the oligomeric Atpc ring

[125; 249; 250]. Movement of protons across the inner membrane causes clockwise rotation of the Atpc ring, which in turn rotates the F_1 γ subunit. Rotation of the γ subunit induces conformational changes in the catalytic α and β subunits of the F_1 sector, which project into the mitochondrial matrix [125; 249; 250].

Bacterial (*E. coli*) CV consists of 8 different proteins in total, with the α, β, γ, δ and ε subunits constituting the F_1 sector while a, b and c make up F_o [251]. In general, the subunits of the F_1 head sector are highly conserved across eukaryotes and few additional subunits have been added; conversely, bacterial F_o subunit homologs are conserved in terms of presence/absence, but they are less similar at the sequence level, and the base and stator domains of F_o have gained novel subunits. As in other respiratory complexes, some eukaryote-specific subunits (e.g., Atpd) are conserved across Eucarya while the distribution of others is reported to be more limited, leading to variable numbers of CV subunits in eukaryotes. Mitochondrial CV ranges from ≥9 subunits in jakobid protists [134], 11-13 in plants [132; 246], 16 in *C. reinhardtii* [106], 18 in yeast [125; 130], 16 in humans [128] to ≥ 20 in *T. thermophila* [139] and 22 in *Trypanosoma brucei* [138].

2.4.5.5.1 ATP Synthase Size and Higher Order Structure.

Under the conditions employed in this study, the major, enzymatically-active form (as determined by in-gel ATPase assays) of *A. castellanii* CV is >1 MDa in size (Figure 2.4). Typically in eukaryotes, monomeric CV is ~550-600 kDa; as a result, I interpret the >1-MDa band to represent a dimeric ATP synthase. Dimeric *A. castellanii* ATP synthase appears to be quite stable, even at relatively high detergent concentrations. For instance, dimeric ATP synthase is the only observable enzymatically-active form at

concentrations as high as 2% (w:v) DDM (Supplemental Figure 2.2); however, at a concentration of 5% DDM, much of the complex breaks down into a smaller, ~820-kDa isoform with weak ATPase activity.

The high stability of the *A. castellanii* CV dimer is in stark contrast to dimeric CV in yeast [130] and bovine [252] mitochondria, which must be prepared under gentle conditions, employing lower concentrations of mild delipidating detergents such as digitonin. The monomers that comprise these dimeric complexes in *S. cerevisiae* are held together by dimer-specific subunits, including Atpe, Atpg, Atpi/j and Atpk [130; 253]. Conversely, the stable dimeric nature of *A. castellanii* CV is reminiscent of mitochondrial ATP synthase from the green algae *C. reinhardtii* and *Polytomella* sp. [106]. In *Polytomella* sp., an ~1600-kDa dimeric CV is the predominant enzymatically-active form of the complex and dissociates only upon heat treatment or high detergent concentrations (>10% DDM) [254]. Notably, CV from these chlorophyte algae contains 8 novel subunits that play a role in dimerization of the complex [106; 255].

1D BN-PAGE (Figure 2.4) and 2D BN/SDS-PAGE [182] profiles suggest that CV is extremely abundant in *A. castellanii* under the growth conditions employed here: CV is by far the most intensely staining band in BN-PAGE profiles of *A. castellanii* mitochondrial protein complexes and the constituents of CV are also obviously more abundant than subunits of other respiratory complexes in silver-stained 2D gels. Whole proteome analysis confirmed these observations: although not a direct measure of the abundance of a protein, ion scores for the F_1 α and β subunits were much higher than for any other protein in the WM fraction. It is not yet clear why CV is so abundant in *A.*

castellanii, whether the complex is equally as abundant in all growth conditions, or what implications this high concentration has for the physiology of the organelle.

2.4.5.5.2 Analysis of ATP Synthase Composition in *A. castellanii*

The subunit composition of CV (Table 2.7) was investigated via an LC-MS/MS analysis of in-gel digested dimeric CV obtained via 1D BN-PAGE (Figure 2.4) and various other CV isoforms detected by BN-PAGE after separation in a linear 10-40% sucrose gradient (Supplemental Table 2.1).

2.4.5.5.2.1 F_1 Subunits

All of the expected subunits that constitute the catalytic F_1 head (α, β, γ, δ and ε subunits) in yeast and bovine CV were detected in dimeric *A. castellanii* CV (Table 2.7). A smaller, low-abundance subcomplex, isolated via BN-PAGE after linear sucrose gradient centrifugation, also contained the α, β, γ, δ and ε subunits. This subcomplex likely corresponds to the dissociated F_1 head sector, suggesting that the F_1 sector is similar to those in other eukaryotes and that most, if not all, of the other identified subunits are associated with the F_o base or stator stalk. However, no isolated F_o sector was detected on BN gels, so this inference awaits experimental confirmation.

2.4.5.5.2.2 F_o Subunits

MS/MS analysis revealed several of the expected F_o subunits, both mtDNA-encoded and nuDNA-encoded, in dimeric *A. castellanii* CV (Table 2.7). Subunits comprising the membrane-embedded section of CV (e.g., Atpc) and stator stalk (e.g., the oligomycin sensitivity-conferring protein, OSCP) were detected. Moreover, MS/MS analysis provided strong support for previous suggestions that certain mtDNA-encoded ORFs are constituents of CV. In particular, OrfB (Orf142) and Orf25 (Orf124) were both

confidently identified as components of CV. OrfB is suggested to be homologous to the A6L (Atp8) ATP synthase subunit [36; 132]; BLASTp searches of the NCBI nr database using *A. castellanii* OrfB as query convincingly identify A6L homologs from plant and protist mtDNA sequences with modest statistical support. Intriguingly, BLAST searches using protist Atp8/OrfB as query also retrieve α-proteobacterial ATP synthase B′ subunit – a component of the stator stalk not present in *E. coli* CV (to my knowledge, this fact has not been noted previously). On the other hand, Orf25 – thought to be related to subunit b or Atp4 of the stator stalk [134] – bears very little similarity to the Atpb subunit from other eukaryotes or prokaryotes. BLAST analyses do not strongly suggest a relationship between Orf25 and Atpb; however, the fact that Orf25 (i) is mtDNA-encoded (as Orf25/Atpb homologs are in certain other eukaryotes [36]); (ii) has two predicted N-terminal transmembrane helices (according to TMHMM [256]), much like Atpb encoded by the minimally-derived mtDNA of the protist *R. americana* [37]; and (iii) is a confirmed component of CV, together provide strong support for the assertion that Orf25 is homologous to Atpb. On the other hand, the mtDNA-encoded subunit Atpa was not found in any BN-PAGE-isolated CV isoforms, but was detected by one acceptable peptide in the MPE mitochondrial proteome fraction. It is unclear why peptides from this particular protein are relatively refractive to MS/MS analysis.

2.4.5.5.2.3 'Hypothetical' CV Proteins

A predicted 56-kDa CV subunit with no obvious homologs in the NCBI nr database (including in *D. discoideum*) was deduced to be the Atpd subunit - a component of the stator stalk - when the *A. castellanii* mitochondrial protein set was searched with an Atpd profile using HMMER. Most notably, a short C-terminal region of the putative *A.*

castellanii Atpd is similar to homologs from other eukaryotes (Figure 2.6A). A multiple protein alignment demonstrates that the putative *A. castellanii* Atpd protein is quite divergent and significantly larger than other Atpd homologs. Specifically, all other identified Atpd homologs, including Atpd from *D. discoideum*, are predicted to be ~20 kDa in size. These observations indicate that Atpd underwent a considerable size increase during the evolutionary separation of members of Amoebozoa. Although the highly divergent nature of *A. castellanii* Atpd makes analysis of multiple alignments problematic, it appears that the increased size of the protein is due to multiple internal 'insertion' sequences, for which peptide data were recovered. Interestingly, Atpd is not the only *A. castellanii* CV subunit that is considerably larger than its homologs; for instance, the OSCP (Atp5) subunit and the otherwise highly-conserved F_1 β subunit (Atp2) have C-terminal extensions of ~165 and ~140 amino acids, respectively. In each case, the C-terminal extension is predicted to contain additional helical components. Unlike the case of Atpd, the OSCP and β subunits of other amoebozoans also appear to have C-terminal extensions.

Other apparently *A. castellanii*-specific, or 'hypothetical', CV subunits were identified by MS/MS analysis; in some of these cases, putative homology inferences have been made after detailed bioinformatic searches (3 proteins remain without annotation). A number of these subunits appear to be related to proteins characterized as CV dimer-specific proteins in yeast. For instance, a predicted 10.6-kDa protein was identified as a homolog of yeast Atpi/j by profile HMM similarity searches. Recently, Atpi/j was characterized as important in the stepwise dimerization and/or oligomerization of CV in yeast [253]. Interestingly, HMMER searches also uncovered a previously unidentified

homology; specifically Atpi/j is related to the 6-kDa plant ATP synthase subunit [257], which has multiple protist homologs as well. This conclusion is supported by a multiple protein alignment, which shows conserved regions between the plant/protist and yeast proteins (Figure 2.6B). Atpi/j was identified as a protein specific to yeast ATP synthase, as no homologs could be detected in mammalian or bacterial CV. However, these results demonstrate that it is likely a very ancient component of CV in eukaryotes. In HMMER searches, another *A. castellanii* CV subunit that could not be annotated by conventional BLASTp retrieves subunit Atpg, also a protein identified in yeast CV as a dimer-specific subunit [130]. The *A. castellanii* protein is similar in size to the yeast and human homologs of Atpg. Atpg homologs were detected in a wide variety of eukaryotic groups, including *D. discoideum* (but *D. discoideum* homologs are more highly conserved than their *A. castellanii* counterparts). Although BLAST similarity scores are low, multiple alignments demonstrate similarity in the C-terminal region of the protein. Taken together, these results suggest that *A. castellanii* likely encodes homologs of certain proteins required for dimerization of yeast ATP synthase and that these proteins are likely widespread across eukaryotes. In contrast to yeast CV, these subunits were identified by MS/MS analysis of a CV isoform likely representing a roughly monomeric CV) from BN-PAGE experiments; however, this particular complex was isolated as a breakdown product of dimeric CV that was partially purified by linear sucrose gradient centrifugation, so it may not truly represent monomeric CV. Dimerization of CV has been described in animals [252], fungi [130], plants [246], green algae [106; 118; 254] and protists [138; 139], suggesting that the dimerization of ATP synthase may be functionally important. In yeast, deletion of genes encoding dimer-specific subunits e and g dramatically alters the

morphology of the mitochondrial membranes, resulting in a lack of classical cristae and formation of onion-like membranes [258].

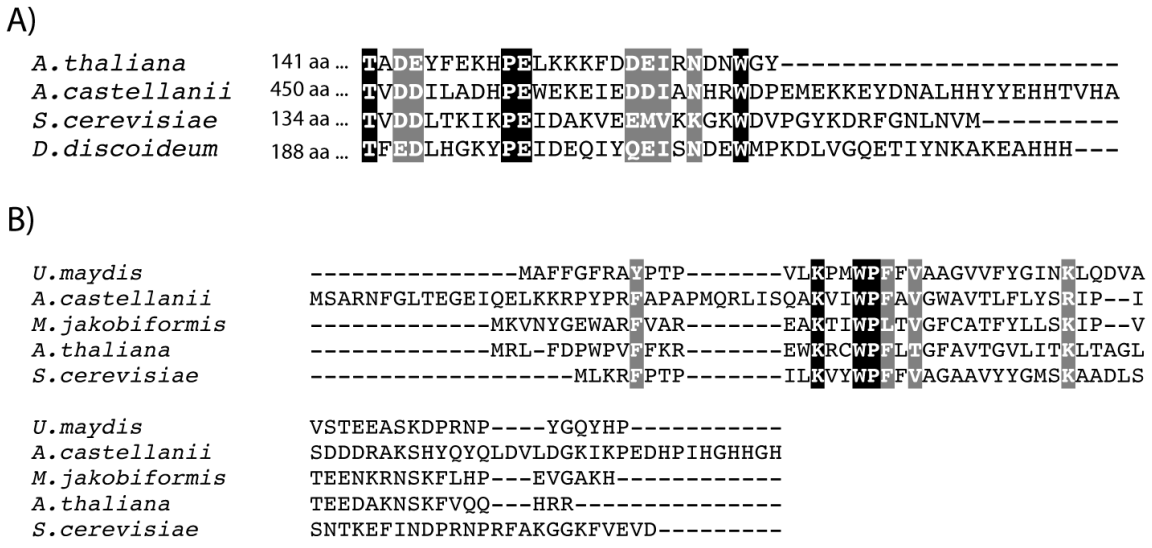


Figure 2.6 *A. castellanii* CV proteins annotated via HMMER 3.0 searches. A) Alignment of the C-terminal region of an expanded and divergent *A. castellanii* Atpd homolog with the corresponding region of Atpd from other eukaryotes. Numbers refer to the N-terminal amino acids omitted from the alignment. B) An *A. castellanii* CV protein is homologous to yeast Atpi/j (Atp18) and a 6-kDa component of plant CV. A multiple protein alignment of putative Atpi/j homologs from several species provides evidence that Atpi/j was an early addition to mitochondrial CV. Black and gray shading represent 100% and $\geq$ 75% similar sites, respectively.

Unexpectedly, malate dehydrogenase (MDH) - the terminal enzyme of the TCA cycle that catalyzes the reversible NAD^+-dependent oxidation of malate to oxaloacetate - was found to be associated with CV in *A. castellanii*. Initially, I suspected that detection of MDH simply reflected contamination of dimeric CV by an abundant mitochondrial protein, as has been the case for other *A. castellanii* respiratory complexes isolated by BN-PAGE. However, MDH was found to be associated with a number of distinct isoforms of CV, including the most abundant, dimeric one (>1 MDa), a larger (possibly

trimeric or tetrameric) form, an ~820-kDa variant (the main breakdown product when CV is treated with high concentrations of detergent), and an ~520-kDa complex (Supplemental Table 2.1). Moreover, due to the high abundance of CV in *A. castellanii*, few other contaminating proteins were found in MS/MS analyses of isolated CV, and only one suspected contaminant was found in more than one CV isoform (AcCa2, a CI subunit, was identified in the CV tetramer and 820-kDa isoforms). Additionally, and in contrast to other likely contaminants of CV, MDH ion scores were consistently among the highest for CV, usually only eclipsed by the α and β subunits of CV. However, MDH was not detected in a different complex that contained only the α, β, γ, δ and ε subunits. This subcomplex likely corresponds to the soluble F_1 sector, suggesting that MDH (along with other novel proteins) may be associated with the F_o sector. To the best of my knowledge, an association of MDH with CV has not been reported in other organisms; further experimentation (e.g., isolation of CV via other biochemical techniques) is required to establish whether MDH is a *bona fide* component of CV or an unusually persistent contaminant.

My results suggest that CV of *A. castellanii* is predominantly dimeric and comprises ~18 subunits, adding up to a molecular weight in excess of 1 MDa (calculations based on identified subunits with putative mTPs removed suggest that dimers likely range from ~1360 to ~1450 kDa in size, depending whether or not MDH is included). The stable interaction of CV monomers, even at relatively high detergent concentrations, is reminiscent of the well-studied CV of *C. reinhardtii* and *Polytomella* sp. Also, like green algal CV [106] (but to a less dramatic extent), subunits of the stator stalk, like the expanded Atpd and OSCP and divergent putative Atpb (Orf25), appear to

be especially unusual. As the stator stalk serves a primarily structural role – it does not contain any subunits directly involved in catalysis – the conservation of protein sequence may not be important for CV function. Further investigation will be required to understand the mechanisms of CV dimerization in *A. castellanii:* it is unclear whether this is achieved by interaction between novel and divergent stator stalk elements, as in green algae, or via putative dimer-specific subunits present in other eukaryotes.

2.4.5.5.3 CV Assembly Proteins

MS/MS analyses of *A. castellanii* mitochondria have identified factors putatively involved in the expression, regulation and assembly of CV.

A homolog of yeast NCA2, which plays a role in the expression of mtDNA-encoded F_o subunits 6 and 8 (Atpa and Atpa6l/OrfB) [259], was detected in *A. castellanii* by MS/MS. No other homologs of yeast CV expression-related proteins could be identified via bioinformatic searches.

A homolog of yeast INH1 (IF_1 in mammals), an endogenous F_oF_1 ATP synthase inhibitor, was detected by MS/MS, although the primary sequence is fairly divergent in *A. castellanii*. During mitochondrial stress conditions (represented by matrix acidification and depletion of the electrical membrane potential, $\Delta\psi_m$), CV can run 'in reverse', hydrolyzing ATP and pumping protons from the matrix in order to restore $\Delta\psi_m$ [260]. The binding of INH1 to the F_1 sector during stress conditions limits ATP hydrolysis. Although we detected a homolog of INH1 in *A. castellanii* mitochondria by MS/MS, it was not associated with any isoforms of CV examined.

Assembly proteins of both the F_o and F_1 sectors were identified by MS/MS. Homologs of Atp10 [261] and Atp23 [262] were identified by MS/MS analysis; each serves as

a chaperone in the assembly of the membrane-imbedded F_o sector (and Atp23 has an additional role in the proteolytic processing of Atp6). In addition, MS/MS detected homologs of yeast Atp11 and Atp12, which bind to and prevent the aggregation of the $F_1\beta$ and α subunits, respectively [263]. No assembly factors were identified by analysis of any CV isoforms purified by BN-PAGE and no other assembly factors could be uncovered using bioinformatic interrogations.

Table 2.7 *A. castellanii* CV subunits and CV-associated proteins. Proteins identified via BN-PAGE-MS/MS ("BN-MS") as well as MS/MS of whole mitochondria ("MS") are reported. Whether the protein in question has been reported in other eukaryotes ("Mito."), bacteria ("Bact.") and other amoebozoans ("Amoeb.") is also indicated. The genome in which a given protein is encoded ("Genome") is noted (N; nuclear, M; mitochondrial). Protein sequences for all entries are available in Supplemental File 2.10.

Protein	Genome	MS	BN-MS	Mito.	Bact.	Amoeb.
CV structural proteins						
F_1 α (Atp1)	M	+	+	+	+	+
F_1 β (Atp2)	N	+	+	+	+	+
F_1 γ (Atp3)	N	+	+	+	+	+
F_1 ε (Atp15)	N	+	+	+		+
F_1 δ (Atp16)	N	+	+	+	+	+
F_o a (Atp6)	M	+		+	+	+
F_o b (Orf25; Atp4)	M	+	+	+	+	+
F_o c (Atp9)	M	+	+	+	+	+
F_o OSCP (Atp5)	N	+	+	+	+	+
F_o a6l (OrfB; Atp8)	M	+	+	+	+[a]	+
Atpd[b] (Atp7)	N	+	+	+		+
Atpg[b] (Atp20)	N	+	+	+		+
Atpi/j[b] (Atp18)	N	+	+	+		+
Atpf (Atp17)	N	+	+	+		+
15510[c]	N	+	+			
31[c]	N	+	+			
24711[c]	N	+	+	+		
MDH[d]	N	+	+	+	+	+
CV-associated proteins						
INH1	N	+		+		+
Atp10	N	+		+		+
Atp11	N	+		+		+
Atp12	N	+		+	+	+
Atp23	N	+		+		+
NCA2	N	+		+		+

[a] BLAST searches using protist Atp8/A6L as a query retrieve α-proteobacterial ATP synthase B′ – a stator stalk subunit not present in *E. coli* CV.
[b] These subunits were annotated via HMMER 3.0 profile searches.
[c] These proteins do not have identifiable homologs in other organisms. Names/numbers reflect the EST contig numbers to facilitate retrieval from the proteome database.
[d] Malate dehydrogenase was found to be associated with CV in *A. castellanii*. MDH homologs are present in other eukaryotes and bacteria, but have not been reported to be associated with CV in these cases.

2.4.5.6 Branched Respiratory Chain

In contrast to mammalian mitochondria, electron transport chains of fungi, plants and protists are often branched [123; 124; 264-266], containing external and internal rotenone-insensitive NADH:ubiquinone oxidoreductases (facing the cytoplasm and matrix, respectively) and a cyanide-insensitive alternative oxidase (AOX), which accepts electrons from ubiquinol and reduces molecular oxygen (normally catalyzed by CIV). These steps are not energy conserving (i.e., they are not associated with proton-pumping) and therefore result in decreased ATP synthesis in comparison to the classical ETC.

In *A. castellanii* mitochondria, I identified 3 distinct proteins that are homologous to rotenone-insensitive NADH dehydrogenase (Supplemental File 2.11). The inferred molecular weights of the preproteins are ~53, 66 and 70 kDa, and all are predicted to have N-terminal mTPs. BLAST analyses could not provide significant insight into the precise function of any given homolog (i.e., whether it is an external or internal NADH dehydrogenase).

An AOX protein was identified in each of the four conditions employed in this study (Supplemental File 2.11). *A. castellanii* AOX has been described previously [266; 267]; two isoforms – A and B – are encoded by genes that differ only by an insertion/deletion of 12 nucleotides in the portion of exon 1 that encodes the mTP [268]. MS/MS did not identify any peptides corresponding to the mTP portion of AOX, therefore I was unable

to determine whether only one or both isoforms were detected in the analysis. Additionally, MS/MS did not detect an association of AOX with any of the ETC complexes, although previous investigations suggest that it physically interacts with CIII [213].

2.4.6 Metabolism

2.4.6.1 Mitochondrial Glyoxylate Cycle

The glyoxylate cycle is an anabolic variant of the TCA cycle that permits the synthesis of carbohydrates from acetyl-CoA derived via β-oxidation of fatty acids. In eukaryotes, the cycle typically takes place in glyoxysomes [269; 270] - specialized peroxisomes - although steps also occur in the cytosol in some species [271]. Many of the reactions comprising the glyoxylate cycle are the same as those of the TCA cycle, except that glyoxysome-specific isoforms of TCA-cycle enzymes catalyze the glyoxysomal reactions; however, the decarboxylating enzymes isocitrate dehydrogenase and 2-oxoglutarate dehydrogenase (along with succinyl-CoA synthetase) are absent. In particular, the glyoxylate cycle deviates from the TCA cycle where the glyoxylate cycle-specific enzyme isocitrate lyase (ICL) cleaves isocitrate into succinate and glyoxylate. Subsequently, the other glyoxylate cycle-specific enzyme, malate synthase (MS), condenses glyoxylate with acetyl-CoA, yielding malate. Succinate is transported to mitochondria for oxidation and oxaloacetate (derived from malate via malate dehydrogenase) is converted to phosphoenolpyruvate by phosphoenolpyruvate carboxykinase (PEPCK) in the rate-limiting step of gluconeogenesis. Interestingly, MS and ICL are required for *Candida albicans* pathogenesis after ingestion by cultured macrophages [272].

In *A. castellanii* mitochondria, MS/MS analysis detected a predicted ~128-kDa bifunctional MS-ICL fusion protein. Specifically, multiple high-confidence peptides matching an ORF with an N-terminal MS domain and a C-terminal ICL domain (Figure 2.7A) were detected in the SWM and SPE analyses. The MS-ICL fusion protein possesses a strong mTP (probability of 0.947 with TargetP [90]), suggesting that it is a mitochondrial (and not peroxisomal) protein.

I infer that the fusion protein represents mature MS and ICL in *A. castellanii* (as opposed to its being cleaved post-translationally) on the basis of several observations, aside from the predicted protein sequence. MS and ICL are both usually ~60 kDa in size; however, peptides were identified from excised SDS-PAGE bands that contained proteins larger than 100 kDa, in agreement with the proposed molecular mass of the MS-ICL fusion protein. Peptides corresponding to the classical MS protein sequence and the ICL sequence were detected from the same >100-kDa band; in addition, MS/MS revealed several peptides from the region that spans the inferred junction between MS and ICL sequences. These observations robustly support the conclusion that the MS and ICL are fused in the mature protein. MS/MS analysis also detected one peptide from another, monomeric, MS homolog. This protein does not have a classical mTP, or any other information that indicates its subcellular localization; conversely, no other homologs of ICL could be detected bioinformatically in *A. castellanii*.

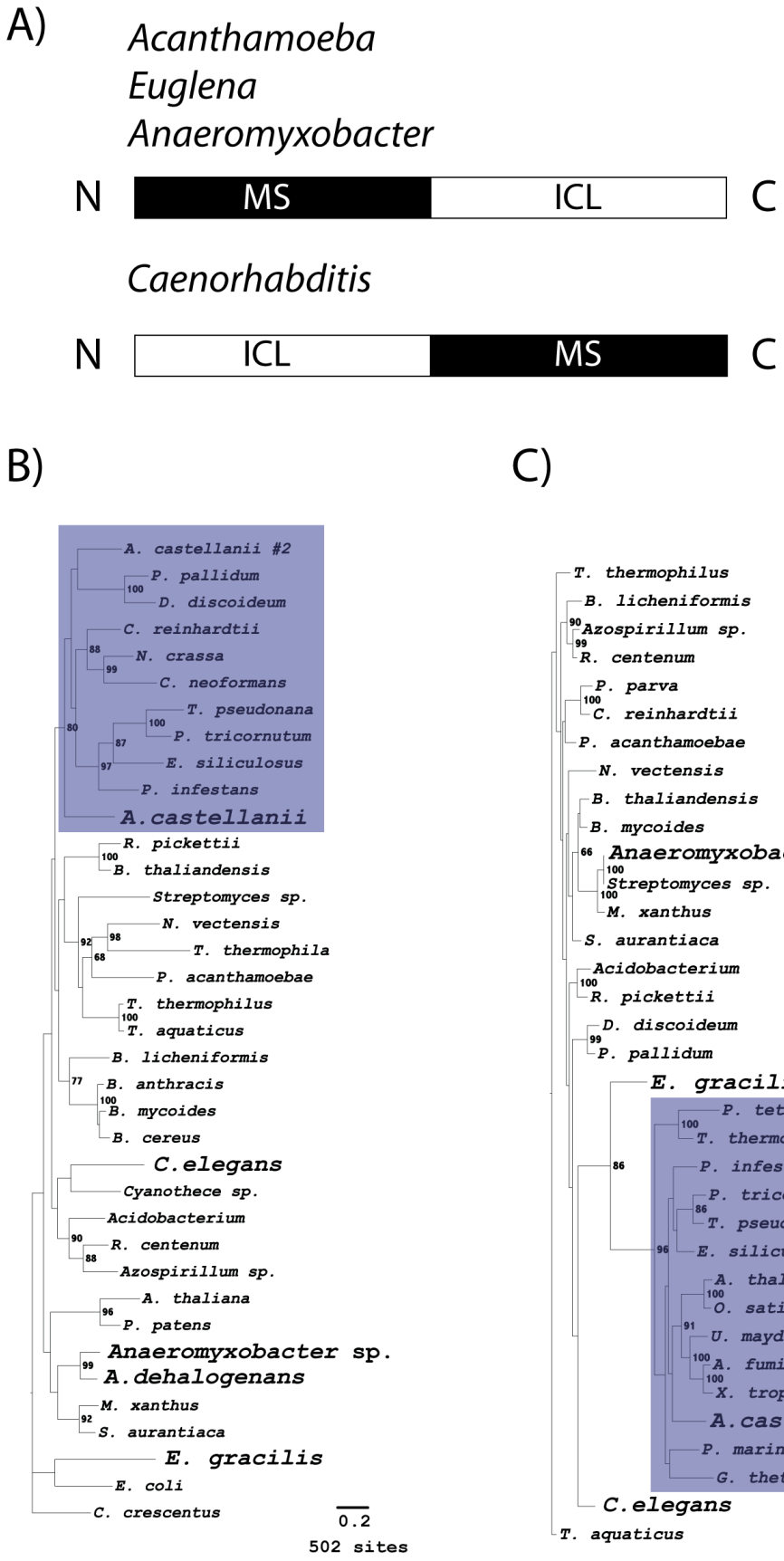
My identification of an MS-ICL fusion protein is not without precedent. An MS-ICL fusion protein has been described in the flagellated protozoon *Euglena gracilis* [273; 274], the only other established case of mitochondrial glyoxylate cycle enzymes, and in δ-proteobacteria of the genus *Anaeromyxobacter*. Moreover, an ICL-MS fusion protein

(opposite orientation from the *A. castellanii* and *E. gracilis* proteins) exists in the nematode *Caenorhabditis elegans* [275; 276]. Phylogenetic analyses of *E. gracilis* MS-ICL and *C. elegans* ICL-MS indicate that these fused genes/proteins are the result of independent lateral transfer events from bacteria [276; 277]. These conclusions also make sense in light of the arrangement of MS and ICL in the bacterial ACE operon; in ACE operons from some bacteria, MS is 5′ to ICL, while in other bacteria ICL is 5′ to MS, suggesting that the *E. gracilis* MS-ICL fusion protein may be derived from a bacterial ACE operon in the former arrangement [276].

In order to better understand the origin of the *A. castellanii* MS-ICL, I performed separate maximum likelihood (ML) phylogenetic reconstructions for the regions that correspond to MS and ICL. Surprisingly, both MS and ICL from *A. castellanii* cluster with eukaryotic homologs, rather than with bacterial ones, indicating the fusion of two genes encoding eukaryotic-type proteins (Figure 2.7B, 2.7C). Although bootstrap values are low, the results demonstrate that MS and ICL have probably fused at least four times independently, raising the question of how reliable supposedly rare gene fusion and fission events are as evolutionary markers. MS and ICL from other amoebozoan species for which genome sequences are available have evolutionary histories distinct from the *A. castellanii* sequences, precluding insight into the origin of the MS-ICL fusion. Like *A. castellanii*, *D. discoideum* and *P. pallidum* have a typical eukaryotic MS; however, ICL in these lineages appears to be of a separate, bacterial origin. Additionally, MS and ICL are encoded on different chromosomes in each of the latter two organisms, thereby providing no clues as to how the genes may have become fused in *A. castellanii*.

Consistent with the presence of the anabolic glyoxylate cycle, *A. castellanii* mitochondria also contain the classical mitochondrial proteins pyruvate carboxylase, aspartate aminotransferase and malate dehydrogenase, which generate oxaloacetate, the first substrate of gluconeogenesis. Additionally, GTP-specific and ATP-specific isoforms of PEPCK, which catalyze the rate-limiting first step in gluconeogenesis to yield phosphoenolpyruvate, were detected by MS/MS; both have strong N-terminal mTPs. A putative cytosolic, 'double' PEPCK (the predicted protein sequence consists of two classical GTP-specific PEPCK proteins back-to-back), was not detected by MS/MS analysis. Together, my results suggest that *A. castellanii* mitochondria play a role in the net synthesis of malate via an MS-ICL fusion protein along with the first steps of gluconeogenesis.

Figure 2.7 An MS and ICL fusion protein is present in *A. castellanii* mitochondria**.** A) MS and ICL fusion proteins are present in *Acanthamoeba*, *Euglena*, *Anaeromyxobacter* and *Caenorhabditis*. In *Acanthamoeba*, *Euglena* and *Anaeromyxobacter*, the MS portion is N-terminal and whereas the ICL portion is C-terminal; this arrangement is reversed in *Caenorhabditis*. B) Maximum likelihood (ML) phylogenetic tree of MS. C) ML phylogenetic tree of ICL. In each phylogenetic reconstruction, the names of lineages with MS-ICL (or ICL-MS) fusion proteins are bolded and slightly enlarged. ML trees do not support a specific relationship between the various MS-ICL fusion proteins. Bootstrap values (from 100 replicates) are presented; values less than 65 are not shown.


A)
Acanthamoeba
Euglena
Anaeromyxobacter
N
MS
ICL
C
Caenorhabditis
N
ICL
MS
C
B)
A. castellanii #2
P. pallidum
D. discoideum
C. reinhardtii
N. crassa
C. neoformans
T. pseudonana
P. tricornutum
E. siliculosus
P. infestans
A.castellanii
R. pickettii
B. thaliandensis
Streptomyces sp.
N. vectensis
T. thermophila
P. acanthamoebae
T. thermophilus
T. aquaticus
B. licheniformis
B. anthracis
B. mycoides
B. cereus
C.elegans
Cyanothece sp.
Acidobacterium
R. centenum
Azospirillum sp.
A. thaliana
P. patens
Anaeromyxobacter sp.
A.dehalogenans
M. xanthus
S. aurantiaca
E. gracilis
E. coli
C. crescentus
0.2
502 sites
C)
T. thermophilus
B. licheniformis
Azospirillum sp.
R. centenum
P. parva
C. reinhardtii
P. acanthamoebae
N. vectensis
B. thaliandensis
B. mycoides
Anaeromyxobacter sp.
Streptomyces sp.
M. xanthus
S. aurantiaca
Acidobacterium
R. pickettii
D. discoideum
P. pallidum
E. gracilis
P. tetraurelia
T. thermophila
P. infestans
P. tricornutum
T. pseudonana
E. siliculosus
A. thaliana
O. sativa
U. maydis
A. fumigatus
X. tropicalis
A.castellanii
P. marinus
G. theta
C.elegans
T. aquaticus
0.2
452 sites

2.4.6.2 LYSINE BIOSYNTHESIS

In eukaryotes, two distinct pathways, diaminopimelate and α-aminoadipate, carry out the biosynthesis of lysine [278]. Whereas the diaminopimelate pathway is present in plants and green algae, the α-aminoadipate pathway has only been characterized comprehensively in fungi. However, recent bioinformatic analyses have described homologs of a particular α-aminoadipate pathway enzyme, α-aminoadipate reductase (AAR/Lys2), in other opisthokonts, *A. castellanii* and the excavates *N. gruberi* and *E. gracilis* [279]. Thus, the α-aminoadipate pathway appears to be limited to the unikonts (opisthokonts + amoebozoans) and excavates.

In fungi, the first half of the α-aminoadipate pathway takes place in mitochondria and the second half (starting with AAR/Lys2) occurs in the cytosol [278]. Among the mitochondrial enzymes are homoisocitrate synthase (Lys20), homoaconitase (Lys4), a copper chaperone for superoxide dismutase (Lys7), homoisocitrate dehydrogenase (Lys12) and α-ketoadipate glutamate aminotransferase (Aro8; may be in both mitochondria and cytosol). I have identified homologs of Lys20, Lys4, Lys12 and Aro8 (but not Lys7) in *A. castellanii* mitochondria and all of these have predicted N-terminal mTPs (Supplemental File 2.12). Moreover, I have identified homologs of saccharopine reductase (Lys9) and saccharopine dehydrogenase (Lys1); however, these enzymes may not be part of the biosynthetic pathway as other isoforms not predicted to be mitochondrial (and somewhat more similar to their yeast counterparts) are encoded in the *A. castellanii* nuclear genome. Moreover, Lys9 is not specific to the α-aminoadipate pathway, as homologs are present in animals (as α-aminoadipic semialdehyde synthase, which catalyzes the first two steps of mitochondrial lysine degradation). As reported

elsewhere, other amoebozoan species (*D. discoideum* and *P. pallidum*) apparently do not employ the α-aminoadipate pathway for lysine biosynthesis; a homolog of LysA (diaminopimelate pathway) has been reported in *D. discoideum* [279]. Notably, the *A. castellanii* genome encodes some components of the diaminopimelate pathway (DapE x 2, DapF and LysA), suggesting that lysine biosynthesis may be quite complex in this protist.

2.4.6.3 Nucleoside/Nucleotide Metabolism

MS/MS analysis revealed a relatively large number of enzymes involved in anabolic and catabolic metabolism of pyrimidine and purine nucleosides/nucleotides. Of particular interest is a Type II (adenosylcobalamin-dependent) ribonucleotide reductase (RNR) bearing an N-terminal mTP. Typically, eukaryotes possess only Type I RNRs, whereas Type II RNRs are found in aerobic and anaerobic prokaryotes and Type III RNRs in anaerobic prokaryotes. A Type II RNR homolog has been reported in *E. gracilis* [280], and BLAST analyses uncover evident homologs in certain other protists, including slime molds, *N. gruberi* and some stramenopiles. Intriguingly, it has been postulated that Type II RNR, which in contrast to Type I RNR is not sensitive to O_2, facilitates life in anaerobic conditions [280]. *A. castellanii* also encodes a Type I RNR, but it is not predicted to be mitochondrial.

2.4.7 Replication, Transcription

2.4.7.1 DNA Metabolism

In animals and fungi, the enzyme responsible for replication of mtDNA is DNA polymerase γ (Polγ) [281]. Importantly, however, Polγ homologs are not found in other eukaryotic groups, and little is known about the replicative helicase across the breadth of

Eucarya. In *A. castellanii* mitochondria, my MS/MS analysis identified a DNA polymerase predicted to possess an N-terminal mTP and related to α-proteobacterial PolI. Notably, the *A. castellanii* DNA polymerase is closely related to a *D. discoideum* protein annotated as the mitochondrial DNA polymerase (although the basis of this assertion is not clear), to PolI homologs dual-targeted to mitochondria and chloroplasts in *A. thaliana* [282] and *Cyanidioschyzon merolae* [283], and to the mitochondrial DNA polymerase of *T. thermophila* [284]. The predicted protein sequence of *A. castellanii* PolI includes domains (from N-terminus to C-terminus) corresponding to 5′-3′ and 3′-5′ exonuclease activities along with polymerase activity. Similarly, PolI homologs from other amoebozoans (*D. discoideum* and *P. pallidum*) and *N. gruberi* (predicted mTP) share all of these domains, whereas other eukaryotic PolI homologs lack the domain corresponding to 5′-3′ exonuclease activity. Intriguingly, in certain eukaryotes lacking the latter domain, including *A. thaliana*, *Oryza sativa*, *Phytophthora infestans* and *Ectocarpus siliculosus*, separate proteins with a 5′-3′ exonuclease domain exist, and these contain predicted N-terminal mTPs.

Several other proteins involved in DNA binding were detected by MS/MS analysis, including the mitochondrial replicative DNA helicase Twinkle [71], mitochondrial single-strand DNA-binding protein, a DNA topoisomerase III homolog and DNA-binding protein Hu.

2.4.7.2 RNA METABOLISM

A number of proteins involved in mitochondrial transcription, RNA binding and RNA degradation were identified by MS/MS analysis.

Transcription in mitochondria, with the possible exception of jakobid flagellates [37], is achieved by a single-subunit DNA-directed RNA polymerase likely derived from a T3/T7-like bacteriophage [72]. I detected multiple peptides from the *A. castellanii* mitochondrial RNA polymerase (predicted to have a strong N-terminal mTP), as our group did previously in the ciliate *T. thermophila* [23], providing further support for its role in mitochondrial transcription across Eucarya. Conversely, MS/MS analysis failed to detect any mitochondrial transcription factors by MS/MS, although *A. castellanii* encodes a homolog of mitochondrial transcription factor B, as reported elsewhere [285].

2.4.7.3 Pentatricopeptide Repeat (PPR) Domain Proteins

PPR domain proteins comprise a family characterized by a somewhat degenerate ~35 amino acid motif repeated up to 30 times [286]. PPR domain-containing proteins either contain only PPR motifs (P class) or other C-terminal motifs designated E and DYW (PLS class). Structurally, the PPR motifs are believed to form α-helices, although no direct crystallographic data are available to confirm this inference. In plants, all PPR proteins characterized to date play a role in RNA metabolism in mitochondria and plastids [112]. In particular, it is thought that each PPR domain protein binds to a specific RNA target site [287] and contributes to one (or more) of a range of functions, including RNA editing, intron splicing, transcript stabilization and translational control [112].

From a phylogenetic perspective, the abundance of predicted PPR domain-containing proteins varies considerably among eukaryotic groups. Whereas a small number (~5-12) of PPR domain proteins is encoded by animals, fungi and green algae, the PPR family apparently expanded enormously in the land plant radiation, with approximately 450 PPR proteins encoded by *A. thaliana* [112]. In *A. castellanii*

mitochondria, I identified 13 PPR domain proteins via MS/MS and an additional 14 PPR proteins bioinformatically (Supplemental File 2.13), suggesting a relatively large complement of PPR domain proteins in this protist. Note, however, that we do not have full sequence data for several of these proteins, as they appear to be poorly represented in the EST database, so the actual number of *A. castellanii* PPR proteins could potentially be more than 27 (i.e., some may not have been detected because they were not in the database) or less (two or more PPR domain proteins may be mistakenly predicted from non-contiguous ESTs that actually represent a single gene). Intriguingly, an emerging picture is that the number of PPR domain proteins encoded in the nuDNA of several protists appears to be intermediate between that found in opisthokonts (animals and fungi) + green algae on the one hand and in land plants on the other. Specifically, 28, 41, >80 and ~27 PPR-domain proteins are encoded by *T. brucei,* the heterolobosean *N. gruberi*, the brown alga *E. siliculosus* and the amoebozoan *A. castellanii*, respectively [288-290].

For the most part, the *A. castellanii* PPR domain proteins appear to belong to the P class, although one PLS class protein containing a C-terminal DYW motif was identified via bioinformatics. This observation is in contrast to a recent description of PPR domain proteins from *N. gruberi*, which reported 11 DYW-type (PLS class) and 30 P class PPR domain proteins [288]. Interestingly, the DYW domain component of PLS class proteins has been suggested to be the catalytic engine of C deamination in plant organellar C-to-U RNA editing, based on distant structural similarity to cytidine deaminases and a correlation between the number of DYW-type PPR domain proteins and number of C-to-U editing sites in plant organelles [291].

Some, but not all, of the *A. castellanii* PPR domain proteins possess strong predicted N-terminal mTPs, although the virtually exclusive roles of PPR proteins in organellar RNA metabolism in plants, animals and fungi argue strongly that most or all of the identified *A. castellanii* PPR domain proteins are all likely to be mitochondrial. Furthermore, a recent study of trypanosome PPR proteins has demonstrated that they are generally mitochondrial proteins that stabilize RNA and are essential for respiration [289], providing further support for the organellar localization of PPR domain proteins in eukaryotes. Ultimately, further investigation will be required to determine the exact number of PPR proteins targeted to *A. castellanii* mitochondria and to uncover their precise molecular functions.

2.4.8 Mitochondrial Translation and Ribosomes

The maintenance of a mitochondrial translation system is required in order to synthesize polypeptides encoded in mtDNA. Although derived from the α-proteobacterial translation system [1], mitochondrial rRNA and ribosomal protein composition differ significantly from their evolutionary antecedents [292]. The RNA components of mitochondrial ribosomes can vary drastically from their prokaryotic counterparts, as they can be fragmented and rearranged [143; 151], severely truncated [293], and/or lacking a mtDNA-encoded 5S rRNA [294]. Systematic characterizations of mitochondrial ribosomal proteins have been performed in relatively few species, notably humans, yeast and trypanosomes. Human and yeast mitoribosomes lack homologs of a number of bacterial proteins; however, a larger number of novel eukaryotic or lineage-specific proteins has accrued in these species [77]. In sum, *E. coli* ribosomes contain 55 proteins, whereas (at least) 83, 85 and 133 proteins are found in mitoribosomes from

yeast [295], human [296] and *Trypanosoma brucei* [297], respectively, typically resulting in an increased protein:RNA mass ratio in comparison with bacterial ribosomes [298]. It has been suggested that expansion of the protein complement of mitoribosomes has compensated for rRNA truncation or alteration. Crystallographic studies of *Leishmania* mitoribosomes, in which novel protein subunits are positioned in regions normally occupied by RNA in a conventional ribosome, support this suggestion [299]. Here, I report a combined proteomic and bioinformatic investigation of the mitochondrial translation apparatus in *A. castellanii*.

2.4.8.1 Translation-Associated Factors

One mitochondrial translation initiation factor – IF3 – was identified by MS/MS analysis, whereas putative homologs of IF2 were retrieved in my bioinformatics survey (Table 2.8). Conversely, a relatively large number of translation elongation factors were identified via MS/MS in *A. castellanii*: EF-Tu and EF-Ts, EF-G and EF-P were all confidently detected. To my knowledge, the existence of EF-P, which stimulates peptidyltransferase activity in bacteria [300], has not been reported in other eukaryotes, although homologs can be identified by BLAST in other amoebozoons, land plants and *N. gruberi*.

I have also detected a homolog of bacterial TypA, a widely distributed but poorly understood GTPase homologous to EF-G that may serve a function in bacterial stress responses [301]. TypA was identified via MS/MS in our earlier analysis of *T. thermophila* mitochondria [23] and this protein appears to be widespread in eukaryotes. Bioinformatics searches also identified homologs of several other bacterial translation-associated

GTPases, including Guf1 (a homolog of bacterial LepA [302]), EngA, EngB and EngD/YchF [303].

I did not identify any translation termination factors via MS/MS, but I did retrieve a homolog of RF-1 (but not RF-2) by BLAST, consistent with the need of the *A. castellanii* mitochondrial translation system to recognize only UAA and UAG termination codons (UGA specifies Trp in this system) [178] (Table 2.8). A single peptidyl-tRNA hydrolase, responsible for cleaving the peptidyl-tRNA ester bond, was identified by MS/MS. In regards to ribosome disassembly, a ribosome recycling factor was identified by MS/MS, whereas BLAST searches uncovered a homolog of human mEFG2 (closely related to EF-G), a factor involved in ribosome recycling in human mitochondria [304].

Ten distinct mitochondrial aminoacyl tRNA synthetases were identified by MS/MS analysis and a further 2 were identified bioinformatically (Table 2.8). Peptides were also identified from homologs of the A and B subunits of the heterotrimeric bacterial glutamyl-tRNA^{Gln} amidotransferase system (the C subunit was found bioinformatically), which together generate glutaminyl-tRNA^{Gln} via an indirect pathway that involves transamidation of glutamyl-tRNA^{Gln} bound to a non-discriminating glutamyl-tRNA synthetase [305]. These results suggest that *A. castellanii* mitochondrial glutaminyl-tRNA^{Gln} is generated by the indirect pathway, as is the case in human, yeast and *A. thaliana* mitochondria [306-308].

Table 2.8 *A. castellanii* mitochondrial translation-associated factors. Proteins identified by MS/MS of mitochondria ("MS") and by bioinformatic searches ("Bioinf.") are reported. "Identifier" refers to a unique identifier (from clustered EST or PGP datasets) that can be used to retrieve the relevant entry from the associated FASTA sequence file (Supplemental File 2.14).

Protein	Identifier	MS	Bioinf.
Initiation			
IF2	g476.t1		+
IF3	19975	+	
Elongation			
EF-G	26230	+	
EF-Tu	9759	+	
EF-Ts	3833	+	
EF-P	26691	+	
Translational GTPases			
TypA/BipA	1617	+	
GUF1	g3637.t1		+
YchF	g6187.t1		+
EngA	g13051.t1/g13050.t1		+
EngB	g15942.t1		+
Release/recycling			
Rf1	g10217.t1		+
Peptidyl-tRNA hydrolase	8428	+	
Ribosome recycling factor	11098	+	
EF-G2	g6016.t1		+
Aminoacyl tRNA synthases			
Alanine	25588	+	
Asparagine	8636	+	
Aspartate	12808	+	
Glutamate	20313	+	
Isoleucine	17388	+	
Leucine	9820	+	

Protein	Identifier	MS	Bioinf.
Phenylalanine	21167	+	
Serine	g3640.t1	+	
Threonine	6211	+	
Valine	19665	+	
Proline	g100.t1		+
Tryptophan	g3676.t1		+
Others			
Glutamyl-tRNA (Gln) amidotransferase, A subunit	7721	+	
Glutamyl-tRNA (Gln) amidotransferase, B subunit	16382	+	
Glutamyl-tRNA (Gln) amidotransferase, C subunit	g5738.t1		+
Translation inhibitor protein	g7864.t1		+
Ribosome binding factor A	g6327.t1		+

2.4.8.2 MITOCHONDRIAL RIBOSOMAL PROTEINS

Although *A. castellanii* mitochondrial LSU and SSU [178; 309] and 5S [183] rRNAs have been characterized (with the LSU and SSU rRNAs being more conservative than their human and yeast counterparts), relatively little is known about the mitochondrial ribosomal protein complement of *A. castellanii*, aside from the 10 small and 6 large subunit ribosomal proteins identified through direct sequencing of the mtDNA [178].

Our analysis has uncovered a total of 22 putative SSU mitoribosomal proteins in *A. castellanii* (Table 2.9). Of these, 11 (including mRps16, newly identified here; see below) are encoded in mtDNA; all but 3 – mRps12 (mtDNA-encoded) and mRps33 and mRsm22 (nuDNA-encoded) – were detected via MS/MS analysis. Conversely, I

identified 39 likely LSU mitoribosomal proteins in *A. castellanii*, 7 of which are encoded by mtDNA (including newly identified mRpl19) and 26 of which were identified by MS/MS (Table 2.9). As I have identified a total of only 61 putative mitoribosomal proteins in *A. castellanii* by a combined MS/MS and bioinformatic approach, it is likely that further mitoribosomal protein subunits too divergent to identify by sequence similarity remain undetected and that other novel proteins may have been added to the *A. castellanii* mitochondrial ribosome.

The analysis of *A. castellanii* mitoribosomal proteins closely reflects the reported mitoribosomal protein complement of *D. discoideum* [77]; however, my results differ in several ways. For instance, I failed to detect mRps6, mRps35, mRpl10 and mRpl56 reported from *D. discoideum* (although *D. discoideum* mRps6 and mRps35 were only identified in the genome sequence and not in the EST or protein databases and so may not be authentic proteins), whereas I identified several *A. castellanii* mitoribosomal proteins, including mRps11, mRpl5, mRpl19, mRpl34, mRpl37 and mRpl40, not reported in *D. discoideum* (note, however, that BLAST searches using *A. castellanii* mRpl34 and mRpl40 retrieve probable *D. discoideum* homologs while mRpl5 and mRps11 are annotated in the GenBank entry for the *D. discoideum* mitochondrial genome sequence).

Of particular note, MS/MS analysis uncovered a homolog of bacterial Rpl25 – a protein that binds to Loop E of 5S rRNA [310]. As described previously, *A. castellanii* mitochondria encode a divergent 5S rRNA [183]; moreover, other amoebozoan species encode presumptive and highly unusual 5S rRNA species [311], and homologs of Rpl25 can be detected in *D. discoideum* [77] and *P. pallidum* nuDNA sequences. Thus, it is possible that the occurrence of an mRpl25 homolog in a given nuclear genome is a marker for a

mitochondrial 5S rRNA. Few other organisms are known to harbor a mitochondrial 5S rRNA, although mitochondrial genomes from land plants [312], jakobid flagellates [37], some green algae [313] and red algae [314] encode 5S rRNA. Correspondingly, mRpl25 homologs were detected bioinformatically in plants, jakobids and green algae by BLAST analysis. Interestingly, used as a PSI-BLAST query, the *A. castellanii* mRpl25 protein also retrieved a number of putative mRpl25 homologs (bearing N-terminal mTPs) in species not known to contain a mitochondrial 5S rRNA, including *P. infestans*, *B. hominis* and *Plasmodium falciparum*, suggesting that mitochondria of these species may encode (or import) mitochondrial 5S rRNA homologs.

A comparison of MS/MS-identified peptides to a 6-frame translation of *A. castellanii* mtDNA demonstrated the existence of two directly adjacent, previously unknown ORFs of 92 and 99 amino acid residues. Unlike all other genes in *A. castellanii* mtDNA, these two ORFs – located between *atp9* and *trnL2* – are encoded on the negative strand, but do not overlap with any other mtDNA-encoded genes. A BLAST query using the 99-amino acid sequence retrieves bacterial and mitochondrial Rpl19 homologs, and multiple protein alignments strongly suggest that this is an authentic mtDNA-encoded mRpl19 subunit (Figure 2.8A). Conversely, similarity searches using the 92-amino acid ORF do not identify any putative homologs; however, other lines of evidence lead us to believe that this ORF corresponds to mRps16. Firstly, mtDNA-encoded mRpl19 (102 amino acids) and mRps16 (84 amino acids) from the related lobose amoeba *H. vermiformis* are adjacent to each other, albeit on the positive strand (Figure 2.8B). Importantly, mRps16 from *H. vermiformis* is much more highly conserved than is the putative mRps16 homolog from *A. castellanii*. Apart from *H. vermiformis*, an Rps16

homolog does not appear to be present in the mtDNA of any other sequenced mitochondrial genomes, including *R. americana*, so no further mtDNA gene order comparisons could be carried out. However, in several proteobacterial species, Rps16 and Rpl19 are encoded in close proximity to each other and often far from other ribosomal proteins (although, in other α-proteobacteria such as *Rickettsia*, the genes are not closely linked; Figure 2.8C). Typically in these species, Rps16 and Rpl19 are separated by 1-3 genes, usually including *rimM* (a 16S rRNA processing factor) and *trmD* (tRNA guanine-N^1-methyltransferase). Lastly, conserved domain BLAST searches using *A. castellanii* Orf92 align to Rps16, albeit at very high E-values (9.8). Taken together, I suggest that these lines of evidence provide solid support for the assertion that *A. castellanii* Orf92 is an Rps16 homolog.

Figure 2.8 Two novel mtDNA-encoded ORFs likely encode mitoribosomal proteins mRpl19 (Orf99) and mRps16 (Orf92). A) Multiple protein alignments suggest that Orf99 is homologous to mRpl19. The sole confidently identified peptide from mRpl19 is marked in red. Black and gray shading represent 100% and 80% identical sites, respectively. B) Orientation of genes encoding Rps16 and Rpl19 in the mtDNA of the amoebozoon, *H. vermiformis*, in relation to *orf92* and *rpl19* genes in *A. castellanii* mtDNA. The reading frame (+2, +3, -3) in which each gene is encoded and predicted stop codons are labeled. C) Rps16 and Rpl19 are encoded in proximity to each other in a wide variety of proteobacterial species as well as in *H. vermiformis* mtDNA.

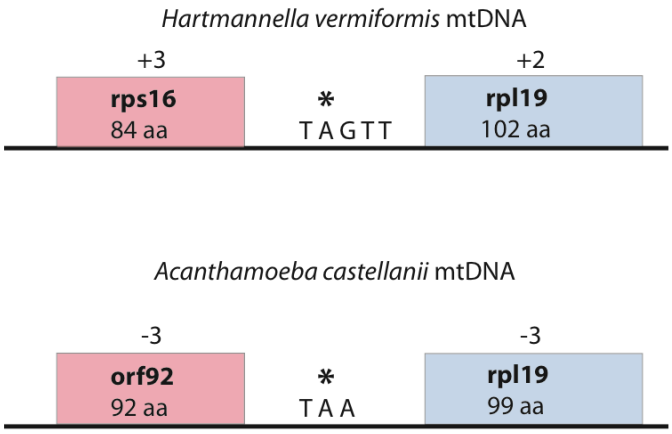
A)

```
A.castellanii      --------MAINQINKKRKNNLNNNVIKTGMVLYIKYLLLKEDEIKISSFIGLCVIKKKK
H.vermiformis      -------------------MLIDYPKTSPGDILDVKLFING----FAKYFKGLCISFKGR
M.jakobiformis     -MNIIKILNNELIKQSQKKSNLKTSSLKIGNTVSIRIILSK-KLKRKQVFMGVCIKKKNK
R.americana        MLNLIQTLR----MEATQKSRHKKTQIQSGDILSITTTQYK-NKKRKQNLKGICIGIKKR

A.castellanii      -----SNTIVLKNNIKKEDIKLVINIHSPLIINIKIVKRYKK--KFRLSKLYYK------
H.vermiformis      SMLNRNSNLTVRNYLNKNGVEFCFFLHLNVVLSFNLYDYQRKRFAYRASKLFYLRNKLNR
M.jakobiformis     GF---RSTFTLKNIIDNVTVIKTFPLYSSMITEILVK----------------------
R.americana        IG---YTTIQLRNFIGGVSQEQSFILESPIINNIEIIGKIKG--NTKAKKYYLRTKSPSE

A.castellanii      -----
H.vermiformis      ESLVF
M.jakobiformis     -----
R.americana        NKV--
```

B)



C)

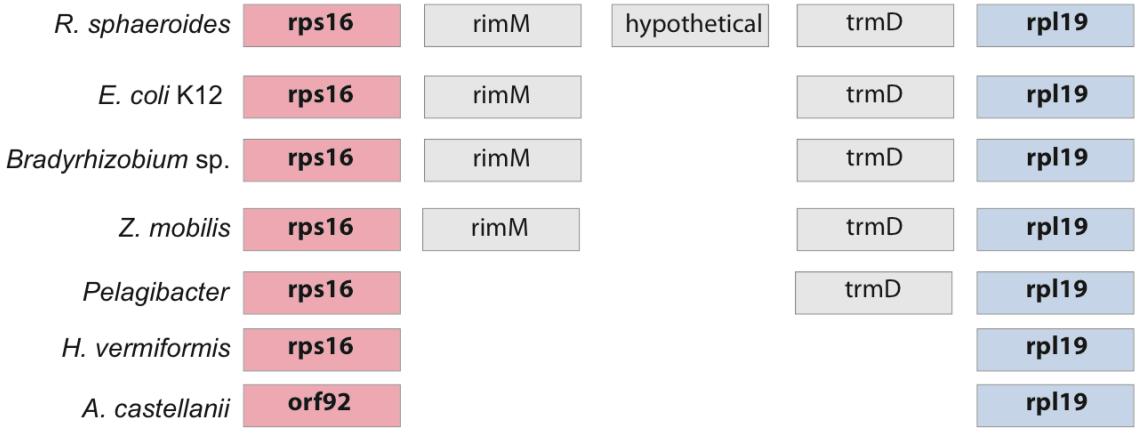


Table 2.9 *A. castellanii* mitochondrial ribosomal proteins. Proteins identified by MS/MS of mitochondria ("MS") and bioinformatic searches ("Bioinf.") are reported. Protein sequences for all entries are available in Supplemental File 2.15.

Protein	Genome	MS	Bioinf.
S2	M	+	
S3	M	+	
S4	M	+	
S5	N	+	
S7	M	+	
S8	M	+	
S9	N	+	
S10	N	+	
S11	M	+	
S12	M		+
S13	M	+	
S14	M	+	
S15	N	+	
Orf92/S16	M[a]	+	
S17	N	+	
S18	N	+	
S19	M	+	
S29	N	+	
S33	N		+
S34	N	+	
S36	N	+	
Rsm22	N		+
L2	M	+	
L3	N	+	
L4	N	+	

Protein	Genome	MS	Bioinf.
L5	M	+	
L6	M		+
L7/L12	N	+	
L9	N	+	
L11	M	+	
L13	N	+	
L14	M	+	
L15	N	+	
L16	M	+	
L17	N		+
Orf99/L19	M[b]	+	
L20	N	+	
L21	N		+
L22	N	+	
L23	N	+	
L24	N	+	
L25	N	+	
L27	N	+	
L28	N	+	
L30	N		+
L32	N		+
L33	N		+
L34	N		+
L36	N		+
L37	N		+
L38	N		+

Protein	Genome	MS	Bioinf.
L40	N	+	
L41	N		+
L43	N	+	
L45	N	+	
L46	N	+	
L47	N	+	
L49	N	+	
L52	N	+	
L53	N		+
L54	N		+

[a] Putative novel mRpl19 encoded on 'opposite' strand of mtDNA. Annotated via BLASTp.

[b] Putative novel mRps16 encoded on 'opposite' strand of mtDNA. Annotated via comparisons to gene order in *H. vermiformis* mtDNA and proteobacteria.

2.4.9 Mitochondrial Protein Import

The initial transfer of genetic material from the protomitochondrion to the host nuclear genome – along with the establishment of a rudimentary protein import system – was a seminal event in eukaryotic evolution, as it inextricably tied the fitness of the protomitochondrial endosymbiont to that of the host cell. Since the inception of mitochondrial protein import, a series of multisubunit, modular machines that efficiently recognize and sort proteins bearing N-terminal or internal targeting sequences have evolved [43; 315]; these complexes are now responsible for importing the vast majority (~95-99%) of the ~1000 mitochondrial proteins from the cytosol and directing them into the

appropriate suborganellar compartment (i.e., outer membrane, intermembrane space, inner membrane or matrix).

The mitochondrial protein import machinery is an evolutionary mosaic comprising proteins 'invented' by eukaryotes and others that were inherited from the protomitochondrion and co-opted for this novel function [60]. Although nearly all of the initial functional characterizations were carried out with the fungi *S. cerevisiae* and *N. crassa*, an increasing number of experimental and bioinformatic analyses of animal, plant and protistan systems has led to an emerging picture of an evolutionarily and functionally conserved core (i.e., typically proteins involved directly in import and processing) while accessory proteins (typically receptors) bear a mostly lineage-specific distribution [60] (but see [316]). Here I present the results of my combined MS/MS and bioinformatic survey of the *A. castellanii* protein import system.

2.4.9.1 MITOCHONDRIAL TARGETING SEQUENCES

Proteins are marked for import into mitochondria via *cis*-acting elements that are most often present as either cleavable N-terminal or non-cleavable, internal signals [43; 315; 317]. Although internal signals remain poorly understood, cleavable N-terminal mTPs are thought to have several features in common, including an enrichment of basic and certain hydrophobic residues, a paucity of acidic amino acids [57; 318] and the ability to form amphipathic α-helices [54; 55; 319]. In yeast, estimates of the fraction of mitochondrial proteins possessing cleavable N-terminal mTPs range from 43% [19] to ~70% [65].

In general, subcellular prediction algorithms such as TargetP [90] quite accurately (and with relatively high confidence) predict *A. castellanii* mitochondrial proteins

harboring canonical N-terminal mTPs. For instance, all of the 30 TCA cycle-associated proteins identified are predicted to contain mTPs, with a confidence of $\geq$ 80%. Similarly, as observed in other eukaryotes, an analysis of the amino acid composition of the first 15 and 30 residues of 299 high-confidence *A. castellanii* mitochondrial proteins bearing mTPs reveals that the latter are enriched in Arg, Ser and Ala relative to the full protein sequences, but heavily depleted of Asp and Glu. These trends are readily visualized in a sequence logo depiction of the first 30 residues of mTPs, which demonstrates the abundance of Arg, Ser and Ala (Figure 2.9).

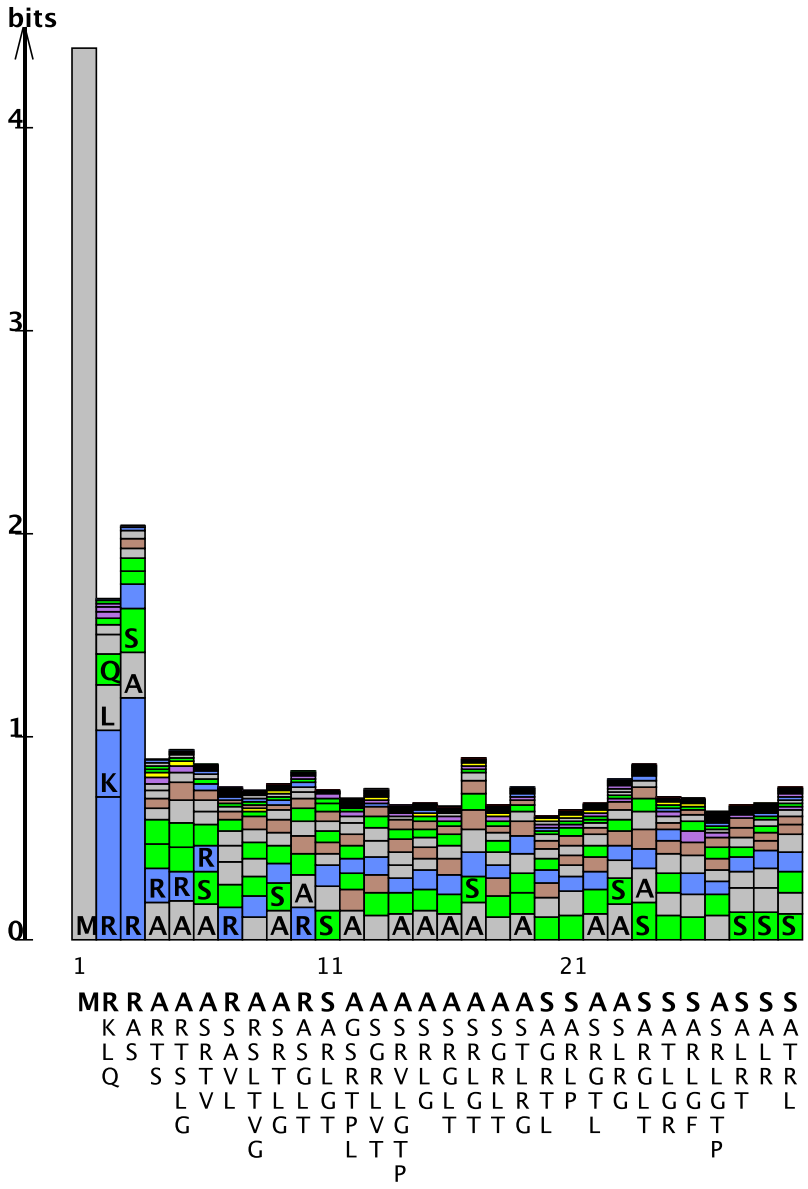


Figure 2.9 Composition of 299 canonical N-terminal mTPs from *A. castellanii*. Sequence logo depicts the relative positional frequencies of amino acids in the first 30 positions of mTPs identified with high confidence.

As described above, not all mitochondrial proteins have N-terminal mTPs. Of the 627 *A. castellanii* mitochondrial proteins that likely have full N-terminal sequence data (including hypothetical proteins), TargetP predicts with ≥ 75% confidence that 371 (~59%) have an N-terminal mTP. Proteins not predicted to be mitochondrial are often quite small and/or members of the mitochondrial carrier, protein import system or respiratory chain systems. For instance, TargetP predicts (75% confidence) that only 45% of putative nuDNA-encoded Complex I subunits, and 38% of protein import subunits, are mitochondrial. Furthermore, as expected [43; 315] many intermembrane space (e.g., cytochrome *c*) and outer membrane proteins (e.g., Tom40, Sam50) lack N-terminal mTPs.

2.4.9.2 TRANSLOCASE OF THE OUTER MITOCHONDRIAL MEMBRANE (TOM)

With the exception of a small number of tail-anchored outer membrane proteins, all mitochondrial proteins are initially imported via the TOM complex. In yeast, the TOM complex is composed of 7 distinct polypeptides that interact with cytosolic chaperones, recognize mitochondrial proteins bearing either N-terminal or internal targeting signals, and transport them across the outer membrane via the 'general insertion pore' formed by the β-barrel protein, Tom40. Previous phylogenetic analyses have suggested that 3 of these 7 proteins – Tom40, Tom22 and Tom7 – are conserved across eukaryotes; these conclusions are supported by my investigation (Table 2.10). In particular, I detected Tom40 and a putative Tom22 homolog by MS/MS analysis, while the small (~7-kDa) protein Tom7 was found by tBLASTn searches of the *A. castellanii* EST dataset. While Tom40 and Tom7 displayed a relatively high degree of sequence similarity to homologs from other eukaryotes, a profile HMMER search was required in

order to identify a putative *A. castellanii* Tom22 homolog. Identification of the latter protein is supported by structural considerations, as the protein has one predicted transmembrane helix that aligns with other Tom22 homologs. Tom22 homologs have previously been identified in a wide variety of eukaryotes [320]. These results suggest that the *A. castellanii* TOM complex contains Tom40, Tom22 and Tom7; however, it likely contains further subunits, as reports of an ~500 kDa TOM complex from *A. castellanii* (similar in size to yeast TOM) have been published [321].

2.4.9.3 Sorting and Assembly Machinery (SAM) Complex

Discovered relatively recently [322; 323], the SAM (Sorting and Assembly Machinery) complex is required for the insertion of β-barrel proteins, including Tom40, into the OMM (Table 2.10). The central component of the complex is Sam50, a member of the Omp85 protein family that also includes BamA, the core component of the BAM complex that inserts β-barrel proteins into the outer membrane of Gram-negative bacteria. Conversely, the other components of the yeast SAM complex – Sam35, Sam37 (Sam35 and Sam37 are members of the metaxin family), Mdm10 and Mim1 – do not have apparent bacterial homologs. In a bioinformatic analysis, it was suggested that Sam50 is the only component of the SAM complex with a broad distribution across eukaryotes [60]. On the contrary, metaxin family proteins, which function in the docking and release of proteins into the outer membrane in yeast [324], have been detected as components of the TOM complex of animals, fungi and land plants [325; 326]. Via MS/MS analysis, I detected Sam50, along with two distinct proteins exhibiting high similarity to animal metaxins (but with lower similarity to fungal Sam35/37), suggesting that metaxins are likely to be widespread across eukaryotes as well. More surprisingly, I also detected a homolog of

Mdm10 via MS/MS, although Mim1 could not be identified. Mdm10 was previously thought to be specific to fungi; however, its presence in members of Amoebozoa suggests that the common ancestor of unikonts (Opisthokonta + Amoebozoa) encoded an Mdm10 homolog.

Notably, Mdm10 appears to have at least two roles in the biogenesis of the OMM, as it was first recognized as a regulator of mitochondrial morphology and inheritance [327] and was only later shown to play a role in TOM complex assembly [328]. Together with Mdm12 and Mmm1 [329], Mdm10 also forms a distinct complex that functions downstream of the core SAM complex in β-barrel assembly (Mmm1 was detected here by MS/MS, Mdm12 was identified via BLASTp). These same proteins, along with Mdm34 (identified here by MS/MS) were detected in a synthetic biology screen [200] as a tether that connects the OMM to the ER in yeast (in fact, Mmm1 is formally an ER protein in yeast, and it has a strongly predicted ER signal peptide in *A. castellanii*). This complex, termed ERMES (ER-mitochondria encounter structure) is thought to play roles in shaping mitochondrial morphology, along with maintaining calcium and lipid homeostasis [200; 330]. Together, these results point to important links between the assembly of the OMM and functional interactions between mitochondria and the ER. To date, no homologs of Mdm34, Mmm1 or Mdm12 have been reported outside of fungi [61]. However, genes encoding clear homologs of these proteins can be detected in the slime mold *D. discoideum*, and a potential homolog of Mdm34 is also encoded by the heterolobosean *N. gruberi*. As is the case with Mdm10, these results underscore the possible antiquity of this system, as it was certainly present in the ancestor of Opisthokonta and Amoebozoa, and possibly prior to the divergence of these eukaryotic supergroups from one another.

2.4.9.4 TINY TIMS

Tiny Tims are small protein chaperones of the intermembrane space that shuttle hydrophobic mitochondrial preproteins from the TOM complex to either the SAM complex or TIM22 (see below). The Tiny Tim system comprises two distinct but evolutionarily related hexameric complexes: Tim9/Tim10 and Tim8/Tim13. In *A. castellanii*, I can identify bioinformatically three isoforms of Tim9 and two of Tim10 (Table 2.10). However, I detected only one Tim9 isoform by MS/MS and neither of the Tim10 homologs. In contrast, I detected both of the *A. castellanii* Tim8 isoforms and the sole Tim13 homolog via MS/MS searches. Why there is a high degree of functional redundancy in the *A. castellanii* Tiny Tim protein import system is not known.

2.4.9.5 DISULFIDE RELAY

The disulfide relay system, composed of Mia40, Erv1 and Hot13 in yeast, utilizes redox chemistry to import small, cysteine-rich proteins – including Tiny Tims – into the mitochondrial intermembrane space [315; 331]. In particular, Mia40 drives protein import using its redox-active Cys-Pro-Cys motif to form a mixed disulfide bond with a Cys residue on the substrate protein. Mia40 transfers the disulfide linkage to the substrate; the flavin-linked sulfhydryl oxidase, Erv1, with assistance from Hot13, then oxidizes the reduced sulfhydryl on Mia40, reforming the disulfide bridge, and subsequently transfers electrons to cytochrome *c* of the respiratory chain [315]. Surprisingly, in *A. castellanii*, I could only confidently identify Erv1 (detected by MS/MS; a potential, but not definitive, homolog of Hot13 was identified by BLASTp query); no Mia40 homolog could be identified with conventional BLASTp or HMMER searches (Table 2.10). Although Mia40 homologs cannot be identified in the nuclear genomes of certain anaerobic,

parasitic protists that lack the typical Mia40 substrates and other protists that do encode classical Mia40 substrates [332], my inability to identify a Mia40 homolog in *A. castellanii* is more surprising, as homologs are easily detected in a wide variety of amoebozoans, including *D. discoideum*, *P. pallidum*, *P. polycephalum* and *Hyperamoeba* sp.. It may be that the sequence encoding Mia40 is simply absent from the genome and EST databases or that the sequence has diverged drastically in *A. castellanii*; however, it is also a distinct possibility that Mia40 was lost at some point within the Amoebozoa supergroup.

2.4.9.6 Translocase of the Inner Mitochondrial Membrane 22 (TIM22)

The TIM22 complex inserts nuDNA-encoded polytopic proteins containing internal mitochondrial targeting information, including the ADP/ATP translocase, into the inner mitochondrial membrane (IMM). In *A. castellanii*, I identified two homologs of Tim22, the channel-forming subunit of the complex (Table 2.10). I detected only one of these Tim22 homologs via MS/MS; the other was identified by BLASTp query. The MS/MS-identified Tim22 isoform is less similar to Tim22 proteins from other organisms than the Tim22 identified bioinformatically. Interestingly, the Tim22 homolog identified by MS/MS was also found to be a constituent of respiratory CI, suggesting that it is the B14.7 subunit (which is related to Tim17/Tim22/Tim23; Table 2.3) and not a Tim22 ortholog.

2.4.9.7 Translocase of the Inner Mitochondrial Membrane 23 (TIM23)

TIM23 is responsible for the import of mitochondrial proteins bearing classical mTPs into the matrix or the inner membrane. The complex comprises a core preprotein translocase capable of inserting proteins into the inner membrane and a preprotein translocase-associated motor that drives protein import into the mitochondrial matrix. In

general, the proteins that make up the preprotein translocase are eukaryotic inventions whereas most components of the PAM module were likely encoded by the protomitochondrial genome (mtHsp70, Pam18 and Tim44 all have homologs in α-Proteobacteria). In *A. castellanii*, 8 of 9 TIM23 subunits characterized in yeast were identified, 7 of which were detected by MS/MS (Table 2.10).

2.4.9.8 OXA TRANSLOCASE

Inner membrane proteins encoded in the mitochondrial genome and certain nuDNA-encoded inner membrane proteins with N-termini facing the intermembrane space are inserted into the membrane by the OXA complex [333]. Oxa1, the core component of the complex, is homologous to the protein YidC, which mediates membrane protein insertion in bacteria [334]. A homolog of Oxa1 was identified by MS/MS in *A. castellanii* (Table 2.10). The other proteins that constitute the OXA complex are involved in docking of mitochondrial ribosomes at the inner membrane for co-translational insertion of membrane proteins [43]. At least 4 proteins were identified that have sequence similarity to Mdm38 and Ylh47 (these proteins are very similar to each other and are homologous to human Letm1). Only one of these proteins was detected by MS/MS; however, all have a predicted mTP. No homologs of Mba1 could be identified.

2.4.9.9 MATRIX PROCESSING PEPTIDASE (MPP)

Mitochondrial preproteins bearing classical N-terminal mTPs are proteolytically processed by the matrix processing peptidase (MPP), comprising the regulatory α and catalytic β subunits. As described in Section 2.4.5.3 (above), I detected two homologs of MPPα and one of MPPβ by MS/MS (Table 2.5, Table 2.10). In the case of MPPβ and one of the putative MPPα subunits, many high-confidence MS/MS peptides were

identified; only one peptide was confidently identified in the other MPPα homolog. The two proteins with abundant MS/MS peptides were identified by BN-PAGE as components of respiratory Complex III; thus, it is formally possible that *A. castellanii*, like land plants [131], uses a CIII-integrated MPP instead of the soluble, matrix-associated MPP observed in fungi and animals, although further experimentation is required to test this inference.

2.4.9.10 INNER MEMBRANE PEPTIDASE (IMP)

Certain nuDNA-encoded proteins (often targeted to the intermembrane space [335], but also to the inner membrane and matrix [336]) possess a second cleavable peptide that subsequent to release of the more N-terminal mTP is proteolytically removed by the inner membrane peptidase (IMP), comprising the catalytic subunits Imp1, Imp2 and a regulatory subunit, Som1. In addition, certain mtDNA-encoded proteins are processed by IMP [337]. Imp1 and Imp2 are related to bacterial leader peptidases, which cleave the signal peptide during protein export. I identified Imp1 and Imp2, but not Som1, in *A. castellanii* by BLAST search. None of these proteins was identified by MS/MS (Table 2.10). Interestingly, by MS/MS search I also detected an mTP-bearing homolog of the bacterial signal peptide peptidase SppA – a protein with no homologs in animals or fungi (plastid homologs appear to be present in plants/algae). In *E. coli*, SppA further digests the cleaved signal peptide. It is tempting to speculate that the *A. castellanii* SppA homolog participates with Imp1 and Imp2 in degradation of sorting signals.

Table 2.10 The *A. castellanii* mitochondrial protein import apparatus. Proteins identified via MS/MS of whole mitochondria ("MS") and bioinformatic searches ("Bioinf.") are reported. "Identifier" refers to a unique identifier (from clustered EST or PGP datasets) that can be used to retrieve the relevant entry from the associated FASTA sequence file (Supplemental File 2.16).

Name	Subcomplex	Identifier	MS	Bioinf.
TOM complex				
Tom40	Core translocase	13048	+	
Tom22[a]	Core translocase	16787	+	
Tom7	Core translocase	18214		+
Tom6	Small subunit	-		
Tom5	Small subunit	-		
Tom70	receptor	-		
Tom20	receptor	-		
SAM complex				
Sam50	Core translocase	10559	+	
Sam35	Metaxin	19964	+	
Sam37	Metaxin	18718	+	
Mdm10	Mdm10	673	+	
Mdm12	Mdm12/Mmm1	7451		+
Mmm1[b]	Mdm12/Mmm1	7284	+	
Mim1	Mim1	-		
Tiny Tims				
Tim9[c]	Core complex	18985	+	
Tim10[d]	Core complex	2235/8351		+
Tim8[e]	Core complex	15686/ g6199.t1	+	
Tim13	Core complex	20011	+	
Disulfide relay				
Erv1	Disulfide relay	2125	+	

Name	Subcomplex	Identifier	MS	Bioinf.
Mia40[f]	Disulfide relay	-		
Hot13	Disulfide relay	3405		?
TIM22 complex				
Tim22[g]	Core translocase	22734		+
Tim12	Peripheral Tim	-		
Tim54	Accessory	-		
Tim18	Accessory	-		
Tim23 complex				
Presequence translocase				
Tim23	Core translocase	27058	+	
Tim17	Core translocase	25569	+	
Tim21	Core translocase	1844	+	
Tim50	Tim50	18694	+	
Presequence translocase-associated motor				
Pam18	PAM complex	13309		+
Pam16	PAM complex	10914	+	
Pam17	PAM complex	-		
Tim44	PAM complex	12315	+	
mtHsp70[h]	PAM complex	3529/14215	+	
Oxa complex				
Oxa1	Core chaperone	3549	+	
Mba1	Ribosome receptor	-		
Mdm38	Ribosome receptor	14024/17709	+	

Name	Subcomplex	Identifier	MS	Bioinf.
Ylh47	Ribosome receptor	-		
Imp complex				
Imp1	Core peptidase	19247		+
Imp2	Core peptidase	6343		+
MPP complex				
Mas1[i]	Core peptidase	18201	+	
Mas2	Core peptidase	18841/27020	+	

[a] Putative Tom22 was identified via HMMER 3.0 profile searches.

[b] Mmm1 is considered to be a component of the mitochondrial SAM complex [93]; however, it is actually an ER protein that links to the OMM. In *A. castellanii*, unlike yeast, Mmm1 has a strongly predicted ER signal peptide.

[c] A Tim9 homolog (18985) was identified by one peptide. Two other Tim9 homologs (4832 and 26620) were not identified by MS/MS.

[d] Two Tim10 homologs (2235 and 8351) can be identified by BLAST. Neither was confidently identified by MS/MS. Additionally, it is possible that one of these proteins is actually homologous to Tim12 of the Tim22 complex, as these proteins are closely related.

[e] Two Tim8 homologs were identified by MS/MS: 15686/20235 was identified in ESTs and another homolog, g6199.t1, was confidently identified in the PGPs.

[f] No homolog of Mia40 could be detected in the *A. castellanii* EST, genome, or PGP data sets, even via HMMER profile searches; however, homologs were readily identified in other amoebozoan species by conventional BLASTp. It is unclear whether Mia40 is simply not represented in either the EST or genome datasets, or whether it is extremely divergent in *A. castellanii*, or whether has been lost altogether in this protist.

[g] Two homologs of Tim22 are present in *A. castellanii*. Only one (22734) was confidently detected by MS/MS; the other, 19250, was found in a PGP search, but with a low ion score of 32. However, 19250 is a better candidate for a true Tim22 ortholog, as BLAST scores are better and 22734 was detected in the 940-and 820-kDa isoforms of CI via BN-PAGE, suggesting it is the B14.7 subunit of CI instead.

[h] Two mtHsp70 homologs (3529 and 14215, 49% identical) were detected by MS/MS. It is unclear if/how their roles differ.

[i] There are two homologs of Mas2 and only one of Mas1. Mas1 (18201) and one Mas2 homolog (18841) are actually subunits of respiratory CIII. The role of the other Mas2 homolog is not clear (it is not a CIII protein).

2.4.10 Mitochondrial Protein Degradation

The mitochondrial proteome is a metabolically dynamic entity that is under the genetic control of two distinct genomes. Mitochondria must maintain appropriate levels of properly functioning nuDNA- and mtDNA-encoded proteins, and selective degradation of misfolded and excess polypeptides is an important means by which mitochondria achieve quality control [338]. The mitochondrial i-AAA^+ ATPase (Yme1) and m-AAA^+ ATPase (Yta10, Yta12) are widely distributed inner membrane proteases whose active sites face the intermembrane space and matrix, respectively [339]. These proteases, which form multimeric complexes, utilize energy derived from ATP hydrolysis to induce conformational changes in substrate proteins and subsequently degrade or proteolytically process them [338]. In *A. castellanii*, I detected a homolog of the yeast i-AAA^+ ATPase Yme1 and the highly similar m-AAA^+ ATPase subunits Yta10 and Afg3 via MS/MS. Moreover, I detected homologs of both yeast prohibitin proteins, Phb1 and Phb2, which regulate proteolytic degradation of IMM proteins by m-AAA^+ ATPase in yeast. Misfolded mitochondrial matrix proteins are degraded by the highly conserved, heptameric Pim1 (Lon protease) in yeast. I identified two distinct Pim1 homologs in *A. castellanii* via MS/MS.

Interestingly, MS/MS analyses also uncovered homologs of the bacterial proteins HslU and HslV, which together form a large multimeric complex. HslU is an AAA^+

ATPase whereas HslV is a catalytic peptidase that is homologous to the β subunit of the 26S eukaryotic proteasome [340], the engine that degrades proteins marked by the ubiquitin pathway. In bacteria, the HslUV complex comprises an active peptidase, though its function may be limited, or largely redundant, as deletion of HslUV affects growth of *E. coli* only at high temperatures [341]. HslUV proteins are encoded by a wide variety of eukaryotes [342; 343], are most closely related to α-proteobacterial homologs [343] and are presumably mitochondrial in these systems. In *T. brucei* mitochondria, the HslUV complex is associated with the kinetoplast (mitochondrial) DNA, where it plays a role in the replication and segregation of this complex organellar genome [344]. However, neither component of the HslUV complex is encoded by animals, fungi or plants, so in a phylogenetically broad sense an understanding of the precise molecular function in lacking.

In addition to general degradative proteases of the matrix and IMM, I also identified homologs of Mop112, Prd1 and DegP (2 homologs 40% identical in amino acid sequence), all peptidases of the intermembrane space [345], as well as 3 distinct aminopeptidases (including a homolog of yeast ICP55) [65; 346].

In sum, these results demonstrate the presence of a significant number of proteases involved in the processing of mitochondrial proteins from a variety of suborganellar compartments (Supplemental File 2.17).

2.4.11 Iron-Sulfur (Fe-S) Cluster Biosynthesis

The synthesis (and export) of Fe-S clusters is a critical function of mitochondria as Fe-S clusters are central components of a variety of mitochondrial and non-

mitochondrial proteins/protein complexes [347; 348]. In mitochondria, Fe-S clusters are components of respiratory Complexes I-III, aconitase and ferredoxin, among others. Moreover, in *A. castellanii* mitochondria, this pathway is critical for the function of PFO and possibly [FeFe]-hydrogenase [203]. The biosynthesis of Fe-S clusters appears to be a ubiquitous function of mitochondria and MROs, occurring in hydrogenosomes [349] and highly reduced mitosomes [13; 350] as well as mitochondria.

I have identified a large proportion of *A. castellanii* mitochondrial Fe-S biosynthetic proteins by direct MS/MS analysis and the remainder via bioinformatic searches (Table 2.11). MS/MS analyses failed to detect only Isu1/Isu2, Arh1 (ferredoxin reductase) and one of two potential homologs of both Isd11 and Isa1, although all of these proteins are predicted to be mitochondrial.

I have also identified components of the Fe-S cluster export system. Specifically, homologs of Atm1 (an ABC transporter) and the sulfhydryl oxidase Erv1 (also involved in mitochondrial protein import) were identified by MS/MS analysis. Bioinformatic interrogations also uncovered a homolog of mitoferrin – a mitochondrial carrier protein that participates in the import of Fe^{2+} into mitochondria. A mitoferrin homolog was previously identified bioinformatically in *D. discoideum* [351].

Table 2.11 The Fe-S cluster biosynthesis machinery of *A. castellanii*. Proteins identified by MS/MS analysis ("MS") and bioinformatic interrogations ("Bioinf.") are reported. "Identifier" refers to a unique identifier (from clustered EST or PGP datasets) that can be used to retrieve the relevant entry from the associated FASTA sequence file (Supplemental File 2.18).

Protein	MS	Bioinf.	Identifier
ISC-assembly machinery			
Nfs1 (cysteine desulfurase)	+		23289
Isd11	+		21283
		+[a]	7577
Isu1/Isu2		+[b]	20119
Isa1 (Isca1)		+[b]	2426
Isa1 (Isca2)	+		20123
Iba57	+		13880
Yah1 (ferredoxin)	+		18230
Arh1 (ferredoxin reductase)		+	15564
Yfh1 (frataxin)	+		119
Nfu1	+		10864
Grx5 (glutaredoxin)	+[c]		25663
	+		10896
		+	25994
Ssq1 (mtHsp70)	+[d]		3529
	+[d]		14215
Jac1 (co-chaperone HscB)	+		22279
Mge1 (co-chaperone Grpe)	+		881
ISC-export machinery			
Atm1	+		14471
Erv1	+		2125
Fe^{2+} import machinery			
Mitoferrin		+	13335

[a] Two homologs of the Nfs1-interacting protein Isd11 are present in *A. castellanii*. One of these (21283) was detected by MS/MS; however, the other (20119), which is more similar to yeast Isd11, was not. Both belong to the Complex I_LYR superfamily.

[b] Two homologs of Isa1 are present in *A. castellanii*, but only one was detected by MS/MS. One is similar to human Isca1 (not detected via MS/MS), while the other is more similar to Isca2. Both homologs are more similar to yeast Isa1 than to Isa2 (though Isca2 is the highest BLAST hit in the *A. castellanii* EST dataset when yeast Isa2 is used as query). Both of these proteins have predicted mTPs.

[c] A larger number (~7) of glutaredoxin-like proteins are encoded by *A. castellanii*, but only ones with predicted mTPs are reported. The first (25563) is the most similar to yeast Grx5.

[d] Ssq1 is an mtHsp70 isoform that is the result of a fungi-specific gene duplication. Two mtHsp70 homologs (3529 and 14215, 49% sequence identity) were detected by MS/MS. It is unclear which of these isoforms participates in Fe-S cluster biosynthesis.

2.4.12 Evolutionary Analyses

As described elsewhere [23; 25; 63], the mitochondrial proteome has a dual evolutionary origin, consisting of proteins derived from prokaryotes (including about 15% that can be confidently assigned as protomitochondrial contributions based on similarity to α-proteobacterial homologs) and other proteins that are either broadly distributed or lineage-specific eukaryotic inventions. Indeed, an analysis of the yeast mitochondrial proteome (using as a criterion, E-values < 1 e-10) suggested that only 50% of yeast mitochondrial proteins have prokaryotic homologs, whereas the remainder is eukaryote-specific, with approximately 18% restricted to yeast and other fungi [63]. In humans, ~75% of mitochondrial proteins are reported to have prokaryotic homologs (at an E-value < 1e-3), and ~57% have bacterial best-bidirectional orthologs; in excess of 10% of human mitochondrial proteins are restricted to animals [24].

I have performed BLASTp searches in order to understand the phylogenetic distribution of *A. castellanii* mitochondrial proteins. Specifically, I have catalogued the

number of proteins with homologs in bacteria and a number of experimental mitochondrial proteomes, including yeast [22], human [24], *A. thaliana* [21], *C. reinhardtii* [25] and *T. thermophila* [23], using as cutoff an E-value < 1e-10 (although exceptions were made for certain small, divergent proteins). Due to limitations associated with the datasets (i.e., the lack of a complete nuclear genome sequence and high-quality gene predictions), my analysis reports homologs and not necessarily orthologs, as best-bidirectional orthology analyses are not yet feasible.

I found that ~60% of *A. castellanii* mitochondrial proteins have bacterial homologs, whereas ~40% are limited to eukaryotes. These numbers are similar to those from yeast and human mitochondria. In excess of 20% of all proteins identified are those lacking annotation (i.e., 'hypothetical' proteins). Typically, these are *A. castellanii*-specific, although a handful is conserved in other amoebozoan species or other eukaryotic groups. Other eukaryote-specific proteins are novel subunits of ancient (i.e., protomitochondrion-derived) protein complexes, such as the respiratory chain and mitoribosome, and proteins that perform functions required of integrated organelles but not free-living bacteria (e.g., an apparatus for the import of nuDNA-encoded proteins and a large family of metabolite-carrier-family proteins). Conversely, proteins that have eubacterial homologs include ancestral components of mitochondrial complexes (such as respiratory chain complexes and mitoribosomes/translation-associated factors), the majority of metabolic enzymes (including TCA cycle enzymes), chaperones and proteases and Fe-S cluster biosynthetic enzymes.

The overlap between my *A. castellanii* mitochondrial protein dataset and other experimental mitochondrial proteomes was also investigated. It should be noted that not

all experimental proteomes are equal in terms of their coverage; for instance, the *C. reinhardtii* proteome contains 349 high-confidence proteins whereas this number is 851 and 1098, respectively, for the yeast and human proteomes. As a result, the overlap does not, strictly speaking, have direct implications with regard to the evolutionary relatedness of the various proteomes; however, the results do permit some insight into the composition of the mitochondrial proteome in a phylogenetically broad sense.

My *A. castellanii* mitochondrial protein dataset had the most overlap with the human dataset (homologs of ~50% of identified *A. castellanii* mitochondrial proteins had homologs at the specified E-value cutoff in this dataset), whereas the least overlap was with *C. reinhardtii*. As emphasized above, this number reflects the completeness of the experimental proteomes more than it does the actual degree of similarity between mitochondrial proteomes.

In excess of 100 proteins were found to have homologs in all 6 experimental mitochondrial proteomes. This cohort includes abundant metabolic enzymes (like pyruvate dehydrogenase, some TCA cycle enzymes and aspartate aminotransferase), core respiratory chain components (SdhA, cytochrome *c1*, ATP synthase β subunit), protein import subunits (Tom40), translation-associated proteins (EF-Tu), metabolite carriers (ADP/ATP translocase) and chaperones (Hsp70, Hsp60 and Hsp10), among others.

Conversely, many proteins were found in *A. castellanii* and 4 or fewer reference proteomes. Proteins not found in any other mitochondrial proteomes are largely lineage-specific, or highly divergent constituents of the *A. castellanii* mitochondrial proteome. However, certain proteins identified in *A. castellanii* mitochondria and not in other experimental mitochondrial proteomes are particular to *A. castellanii* due to novel

biochemical activities associated with *A. castellanii* mitochondria. For instance, none of the other organisms is known to encode mitochondrion-targeted homologs of PFO, HydF (although *C. reinhardtii* encodes a PFO homolog) or a Type II ribonucleotide reductase. These data underscore the evolutionarily dynamic nature of the mitochondrial protein complement. Failure to detect certain other proteins in all reference proteomes might also stem from incomplete proteome coverage: I detected both HslUV protease subunits in *A. castellanii*, but they were not detected via MS/MS in *T. thermophila* or *C. reinhardtii*, although they are nuDNA-encoded in both of these organisms. In other instances, e.g., certain eukaryote-specific subunits of CI, the proteins may simply be too divergent to recognize in a particular lineage (especially *T. thermophila*) via bioinformatics, or absent altogether (yeast, e.g., has lost CI in its entirety). Consequently, one would expect the 'absence' of a certain number of subunits, even if the proteome coverage were complete.

2.5 Conclusions

I have undertaken a comprehensive MS/MS analysis of mitochondria from the amoebozoon, *A. castellanii* – the first such investigation of a free-living member of the Amoebozoa supergroup. The study employed a combined analysis of whole mitochondria along with membrane-protein-enriched and soluble-protein-enriched fractions in order to maximize the coverage of all classes of mitochondrial proteins. In addition, selected bioinformatic analyses were used to augment our comprehension of certain mitochondrial pathways and complexes, although these proteins were not included in the mitochondrial dataset. Nearly all nuDNA-encoded mitochondrial proteins were identified via searches of a 6-fold translation of a high-coverage clustered EST

dataset, as the *A. castellanii* nuclear genome sequence is not yet complete and predicted gene models have not been released. Proteins encoded by mtDNA were identified by searches of predicted *A. castellanii* mtDNA-encoded proteins as well as a 6-fold translation of the complete *A. castellanii* mtDNA sequence. In this way, a relatively large number of both nuDNA-encoded (716) and mtDNA-encoded (32) mitochondrial proteins has been identified (113 putative mitochondrial proteins were identified by bioinformatics interrogations in the course of characterizing mitochondrial protein pathways and complexes – thus, the integrated total is 748 + 113 = 861 proteins). These proteins represent hydrophobic and hydrophilic components of all four submitochondrial compartments and span a wide spectrum of functional categories (Figure 2.3).

As was observed in analyses of the *T. thermophila* mitochondrial proteome [23], interrogations of 6-fold-translated mtDNA sequences using MS/MS spectra have been valuable in the further annotation of mtDNA-encoded proteins. In *A. castellanii* mitochondria, I identified two novel ORFs encoded on the strand opposite to that specifying all other 41 mtDNA-encoded proteins. I infer that these newly identified genes likely code for components of the mitoribosome in *A. castellanii* (Figures 2.7A-C).

A combined BN-PAGE-MS/MS and bioinformatic interrogation of the mitochondrial respiratory chain complexes has provided significant insight into the structure and composition of the *A. castellanii* respiratory chain. From a structural perspective, the ETC complexes are fairly conservative, aside from CV, which in *A. castellanii* is enzymatically active as a stable dimer reminiscent of CV from green algae [106]. In many ways, the composition of these multiprotein complexes mirrors this structural conservatism, as most conserved subunits of CI-CIV were readily identified

(although sensitive bioinformatic techniques were required to annotate several of these proteins, such as CI subunits MWFE, ASHI and B14.5b), whereas many CV subunits were more difficult to identify. Nonetheless, my current analyses of subunits from the more conservative ETC complexes, along with other recent reports on *A. castellanii* ETC proteins [177; 182], have shed light on outstanding problems (e.g., the MS-based confirmation of 'insert' sequences in mature Cox2) and have posed further intriguing questions about the evolution of the mitochondrial respiratory chain.

The phylogenetic distribution of *A. castellanii* mitochondrial proteins is similar to that found in certain other eukaryotes [25; 63] in that many proteins (>20%) are 'hypothetical' (i.e., lacking homologs of established function in other organisms) and over 40% are limited to the eukaryotic domain, as inferred by BLAST analysis. Interestingly, although *A. castellanii* belongs to the Amoebozoa supergroup, a sister to the Opisthokonta supergroup that includes animals + fungi, the *A. castellanii* mitochondrial proteome does not specifically resemble opisthokont mitochondrial proteomes. In some ways, it is similar to fungal mitochondria (e.g., the presence of ERMES complex subunits, lysine biosynthesis via the α-aminoadipate pathway), but in others it more closely resembles mitochondria from plants/green algae and other protists (e.g., a modest number of PPR-domain proteins, PolI-like DNA polymerase, γ-type carbonic anhydrases as components of CI) or distantly related organisms (MS-ICL fusion), likely via convergent mechanisms. Moreover, I detected several *A. castellanii* mitochondrial proteins, including PFO, ASCT and HydF (but not [FeFe]-hydrogenase or HydE or HydG) that participate in metabolic pathways characteristic of protists adapted to anaerobic or microaerophilic environments. These findings, in combination with my

description of a relatively 'normal' ETC, highlight the metabolic versatility of the organelle and pose important questions about the regulation of mitochondrial energy generation in *A. castellanii*.

In sum, my results support previous analyses that portray the mitochondrial proteome as a diverse metabolic and evolutionary entity built around a limited core set of highly conserved components, mostly inherited from the α-proteobacterial ancestor of mitochondria. Further investigation of mitochondria from protists occupying diverse ecological niches and evolutionary positions is certain to provide novel and more complete insight into the function and evolution of mitochondria.

2.6 Acknowlegements

I would like to thank David Spencer for assembly of the CAP3 EST dataset and helpful discussions. Also, thanks to Rob Eveleigh for generating the PGP dataset from genomic DNA sequence and Martin Kolisko for aiding in the assembly of the MIRA EST database.

Chapter 3 Evidence For An Early Evolutionary Emergence Of γ-Type Carbonic Anhydrases As Components Of Mitochondrial Respiratory Complex I

This chapter contains work published in **Gawryluk, R. M. R.** & Gray, M. W. (2010). Evidence for an early evolutionary emergence of γ-type carbonic anhydrases as components of mitochondrial respiratory complex I. *BMC Evol Biol* **348**, 857-870.

3.1 Abstract

Background. The complexity of mitochondrial complex I (CI; NADH:ubiquinone oxidoreductase) has increased considerably relative to the homologous complex in bacteria. Comparative analyses of CI composition in animals, fungi and land plants/green algae suggest that novel components of mitochondrial CI include a set of 18 proteins common to all eukaryotes and a variable number of lineage-specific subunits. In plants and green algae, several purportedly plant-specific proteins homologous to γ-type carbonic anhydrases (γCA) have been identified as components of CI. However, relatively little is known about CI composition in the unicellular protists, the characterizations of which are essential to our understanding of CI evolution.

Results: I have performed a tandem mass spectrometric characterization of CI from the amoeboid protozoon *Acanthamoeba castellanii*. Among the proteins identified were two γCA homologs, AcCa1 and AcCa2, demonstrating that γCA proteins are not specific to plants/green algae. In fact, by bioinformatic searches I detected γCA homologs in diverse protist lineages, and several of these homologs are predicted to possess N-terminal mitochondrial targeting peptides.

Conclusions: The detection of γCAs in CI of *Acanthamoeba*, thought to be a closer relative of animals and fungi than plants, suggests that γCA proteins may have been an ancestral feature of mitochondrial CI, rather than a novel, plant-specific addition. This assertion is supported by the presence of genes encoding γCAs in the nuclear genomes of a wide variety of eukaryotes. Together, these findings demonstrate the importance of a phylogenetically broad characterization of CI for elucidating CI evolution in eukaryotes.

3.2 Background

Mitochondrial complex I (CI; NADH:ubiquinone oxidoreductase) is a multi-subunit and intricate proton pump that is responsible for the first step in the canonical respiratory chain – the oxidation of NADH and subsequent reduction of ubiquinone. Investigations of CI structure in animals and fungi portray CI as an L-shaped complex consisting of a hydrophobic domain integrated into the inner mitochondrial membrane and a hydrophilic mitochondrial matrix domain that oxidizes NADH and transfers electrons along 8-9 Fe-S clusters [76]. Mitochondrial CI comprises 35-45 proteins (~900-1000 kDa), which are encoded in both the mitochondrial and nuclear genomes. Conversely, the homologous ~550-kDa bacterial CI consists of only 14 subunits [352], suggesting a massive expansion of CI in eukaryotes.

The subunit composition of mitochondrial CI has been intensively studied in animals [107], fungi [103; 353] and land plants/green algae [104; 354]. Comparative proteomic and genomic analyses of CI composition in these groups have provided insight into the evolution of mitochondrial CI: specifically, in all three groups, the 14-subunit 'bacterial core' (corresponding to the subunits inherited from the α-proteobacterial ancestor of mitochondria; [1]) has been retained, while an additional 18-subunit 'eukaryotic core'

(made up of proteins shared by these eukaryotic lineages but not by bacterial CI) along with a variable number of lineage-specific proteins has been added [75]. Few of the lineage-specific proteins are similar to other proteins of known function in available databases; however, in plants [354; 355] and green algae [104], CI contains multiple proteins with high similarity to γ-type carbonic anhydrases (γCAs), first described as a homotrimeric complex (Cam) in the methanogenic archaeon, *Methanosarcina thermophila* [356].

Carbonic anhydrases are ubiquitous metalloenzymes that catalyze the reversible hydration of CO_2 to HCO_3^-. Five evolutionarily unrelated CA classes (α, β, γ, δ, and ζ) are currently known, suggesting that this important enzymatic mechanism has been invented independently multiple times. The γCA class is among the most ancient, with homologs widespread in Archaea and Bacteria [357]; however, in eukaryotes, γCA homologs have been described from a phylogenetically narrow collection of species, comprising almost exclusively green algae and land plants, in association with CI. In this group, single-particle electron microscopy studies have demonstrated the presence of an additional CI domain [115; 117]. This novel structure, likely representing the γCA proteins, is associated with the inner membrane portion of CI and projects into the mitochondrial matrix. The function of γCA proteins in CI remains enigmatic: thus far, carbonic anhydrase activity has not been detected biochemically. However, experiments in *Arabidopsis* have demonstrated that expression of plant γCAs is affected by CO_2 concentration (see [358]) and that plant γCA proteins bind inorganic carbon [359]. These functional observations, coupled to the apparent restriction of γCAs to mitochondria of plant/green algal lineages, have prompted the proposal that CI γCA proteins may have

plant-specific functions; for instance, Braun & Zabaleta [127] have suggested that γCA proteins play an important role in chloroplast function via the generation of HCO_3^- in the recycling of inorganic carbon for CO_2 fixation.

Although γCA proteins are not components of CI in animals and fungi (see [75]), little is known about the composition of CI in the predominantly unicellular protists. Because protists comprise the bulk of eukaryotic diversity, it is important to characterize CI from diverse lineages in order to understand fully the evolution of this protein complex. In particular, we cannot conclude that supposedly lineage-specific CI subunits are truly lineage-specific without knowledge of CI composition in the majority of eukaryotes. To this end, I have isolated an enzymatically active, ~940-kDa native CI and an inactive, ~820-kDa CI subcomplex (CI*) from the amoeboid protist *Acanthamoeba castellanii* via blue native polyacrylamide gel electrophoresis (BN-PAGE) and have carried out a characterization of subunit composition via tandem mass spectrometry (MS/MS). Here I describe the detection of two γCA homologs, AcCa1 and AcCa2, in both forms of *Acanthamoeba* CI. *Acanthamoeba* is a member of the eukaryotic supergroup Amoebozoa and is not a close relative of plants [111]; rather, it is currently hypothesized that Amoebozoa comprises the evolutionary sister group to the opisthokonts (animals, fungi and their specific unicellular relatives). The unexpected finding of γCA homologs in mitochondrial CI of *Acanthamoeba* demonstrates that these proteins are not limited to CI of plants/green algae, as was previously suggested [355]. In fact, I provide evidence here that γCA homologs having predicted mitochondrial targeting peptides (mTPs) are widespread throughout the domain Eucarya.

3.3 Results and Discussion

3.3.1 γ Carbonic Anhydrase Proteins in Acanthamoeba CI

An in-gel assay for NADH dehydrogenase activity was performed on *Acanthamoeba* mitochondrial protein complexes separated by BN-PAGE, yielding an insoluble formazan precipitate at ~940 kDa (Figure 3.1A). This molecular weight estimate is in agreement with the findings of Navet *et al*. [213] and corresponds to the general size of CI across eukaryotes. Together, these observations suggest that the identified NADH dehydrogenase activity corresponds to the native, enzymatically active CI of *Acanthamoeba*. Furthermore, analysis of *Acanthamoeba* mitochondrial protein complexes by two-dimensional BN/SDS-PAGE revealed another, more abundant complex (CI*) of ~820 kDa that has a very similar protein profile to intact CI (Figure 3.1B). Based on the two-dimensional BN/SDS-PAGE profile and lack of enzyme activity following one-dimensional BN-PAGE, I infer that this complex is likely a partially dissociated CI missing a portion of the NADH-oxidizing mitochondrial matrix domain. The dissociation of CI into subcomplexes during BN-PAGE is not uncommon, and the observed breakdown into a slightly smaller, enzymatically inactive subcomplex is reminiscent of the situation in the green alga *Polytomella* [115].

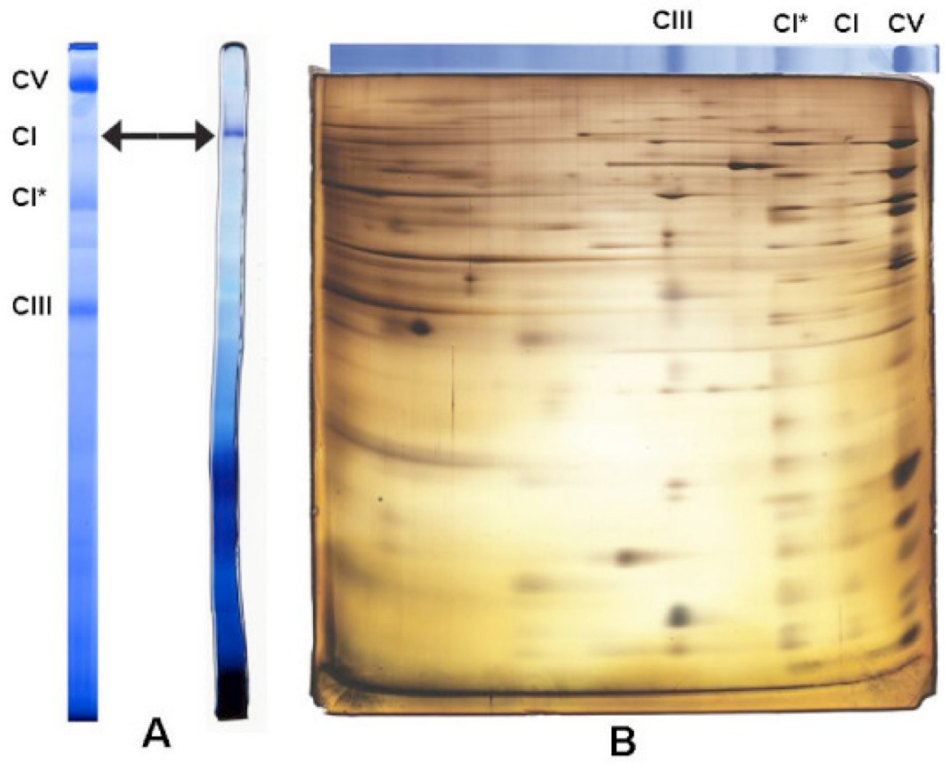


Figure 3.1 BN-PAGE, CI enzyme activity and two-dimensional BN/SDS-PAGE profiles. Panel A shows the characteristic Coomassie-stained BN-PAGE profile of *Acanthamoeba* mitochondrial complexes (left) and in-gel assay of these complexes (*Acanthamoeba* BN-PAGE lane from the same gel), identifying an NADH dehydrogenase activity at ~940 kDa (right). This band is considered to represent intact CI from *Acanthamoeba*. Panel B shows the protein profile observed after two-dimensional BN/SDS-PAGE of *Acanthamoeba* mitochondrial complexes resolved in panel A. A list of proteins identified in CI and CI* isolated from BN-PAGE bands is presented in Supplemental Table 3.1. CI, intact complex I (940 kDa); CI* = CI subcomplex (820 kDa); CIII, complex III; CV, complex V.

Tandem mass spectrometric analysis of the 940-kDa and 820-kDa bands demonstrated that both were enriched in CI subunits known from other organisms, supporting my hypothesis that both represent CI (see Supplemental Table 3.1**)**. A number of contaminating, high-abundance mitochondrial proteins from other complexes were also present; this is not unexpected, as the relative abundance of CI is low in comparison with other complexes, and CI* appears to co-migrate with other abundant

complexes (Figure 3.1B). Notably, two γCA homologs were detected in both the 940-kDa and 820-kDa samples. These proteins, here named AcCa1 and AcCa2, are 36% identical to each other and are the only γCA homologs represented in the *Acanthamoeba* TBestDB EST library [185] and the ongoing *Acanthamoeba* nuclear genome project (Baylor College of Medicine; http://www.hgsc.bcm.tmc.edu/microbial-detail.xsp?project_id=163). Between the two samples analyzed by MS/MS, 31% and 47% of the AcCa1 and AcCa2 protein sequences were covered, respectively (see Supplemental File 3.1). These findings constitute the first report of γCA homologs as CI components outside of the plant/green algal lineage (see Figure 3.2 for multiple alignment). Based on these results, I hypothesize that AcCa1 and AcCa2 are matrix-facing components of the inner membrane arm of CI in *Acanthamoeba*, as are their homologs in plants/green algae [115; 117].

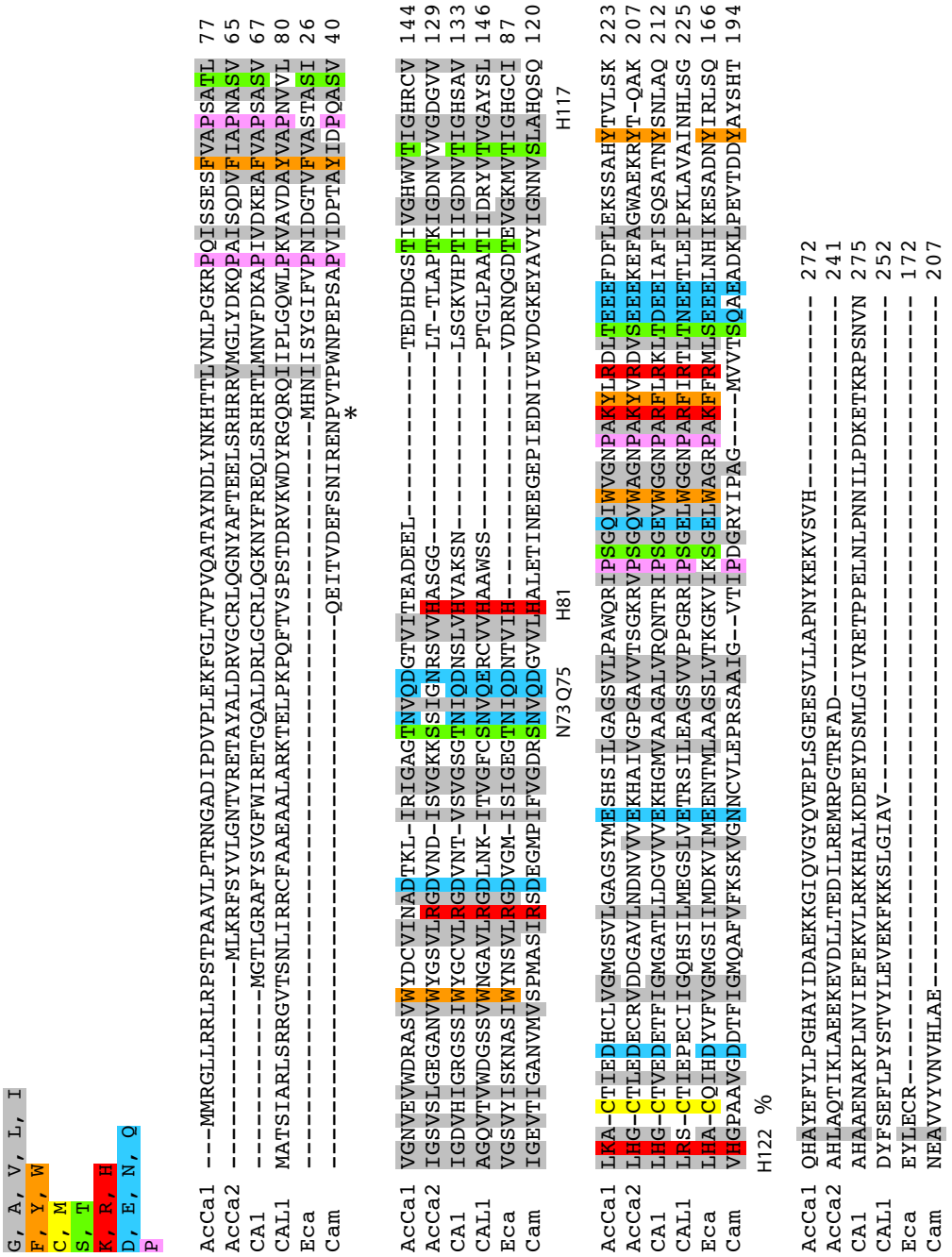


Figure 3.2 Alignment of AcCa1 and AcCa2 sequences with homologs from *Arabidopsis* and prokaryotes**.** *Acanthamoeba* AcCa1 and AcCa2 are aligned with *Arabidopsis* CA1 and CAL1 along with an α-proteobacterial (*Ehrlichia canis*) γCA homolog and archaeal Cam (*M. thermophila*). Critical amino acids described in text are referred to according to the Cam nomenclature. H42 and C137 of CA1 are denoted by * and %, respectively. Shading of columns according to the scheme presented above the alignment reflects amino acid similarity of ≥ 80%.

Because not all expected CI subunits were detected in the MS/MS analysis and because the main contaminants of my CI proteins were ATP synthase proteins, I considered the possibility that the γCA proteins might actually be ATP synthase (CV) subunits. The enzymatically active CI complex is present in very low abundance in *A. castellanii*, so that detection of all CI subunits is proving to be a challenge. Conversely, CV is extraordinarily abundant, as is evident in Figures 3.1A and 3.1B; as a result, MS/MS data have revealed all CV subunits in *A. castellanii*, whereas no γCA proteins were detected in this complex. These considerations combined with the fact that γCA proteins have been characterized as *bona fide* CI components in other organisms effectively eliminates the possibility that they might be CV subunits instead in *Acanthamoeba*.

A small number of γCA homologs have been described in eukaryotes outside of the plant supergroup. A cDNA encoding a γCA homolog from the haptophyte alga *Emiliania huxleyi* was identified previously [360]; however the authors did not investigate the subcellular localization of the protein. Our group has previously identified three different γCA homologs in a proteomic investigation of highly purified mitochondria from the ciliated protozoon, *Tetrahymena thermophila* [23]. Interestingly, two of these three γCA homologs were detected in mass spectrometric analyses of a 1 MDa aggregate from *Tetrahymena* mitochondrial membrane proteins separated by BN-PAGE (Gawryluk *et al*., unpublished results). This aggregate is highly enriched in known protein components of complexes I, III and V, suggesting that the identified γCA homologs may be components of CI in *Tetrahymena* as well.

From an evolutionary perspective, my findings are somewhat surprising: the Amoebozoa supergroup (to which *Acanthamoeba* belongs) is thought to be sister to the opisthokonts (animals + fungi), and is not closely related to the Plantae supergroup [111]. Since no authentic γCA homologs were detected in opisthokont nuclear genome or EST sequence databases (and no γCA homologs are part of opisthokont CI), these results suggest that γCA proteins may have been specifically lost from opisthokont CI, rather than recently added to the plant/green algal lineage. Moreover, the finding that γCA homologs may be components of CI in other eukaryotic groups (ciliates) raises the possibility that γCA homologs were an ancestral feature of mitochondrial CI (i.e., a component of the eukaryotic core of CI lost in animals and fungi). I hypothesized that if this were the case, γCA homologs should be detected in other protist lineages.

3.3.2 γ Carbonic Anhydrase Proteins are Found Throughout Eucarya

In order to better understand the distribution and evolution of γCA homologs across the domain Eucarya, I employed a systematic bioinformatics approach to search nuclear genomic and EST sequence databases from diverse eukaryotes. In contrast to prior reports that γCA proteins are largely limited to plants/green algae, I detected (often multiple) homologs in the majority of eukaryotic groups (Table 3.1; see also Supplemental File 3.2, Supplemental Figure 3.1 and Supplemental Table 3.2). γCA homologs are found in other major groups within the Plantae supergroup, including one in the red alga *Cyanidioschyzon merolae* and at least two in the glaucophyte alga, *Cyanophora paradoxa*. Aside from the opisthokonts, homologs could be detected in most eukaryotic supergroups, including Amoebozoa (*Acanthamoeba*, *Dictyostelium*, *Hartmannella, Hyperamoeba*, *Polysphondylium*), Chromalveolata (*Tetrahymena*,

Phytophthora, *Phaeodactylum*, *Blastocystis*, *Emiliania*, *Guillardia*), Excavata (*Reclinomonas*, *Seculamonas*, *Naegleria*, *Euglena*, *Trypanosoma*, *Malawimonas*) and possibly Rhizaria (an EST homolog was identified in *Bigelowiella natans*; however, several stop codons interrupt the reading frame, suggesting that the EST in question may represent an incompletely spliced transcript).

Of the retrieved γCA protein sequences considered to be complete at their N-termini (where mTPs would be located; [43]), many are predicted to be mitochondrial by the programs TargetP [90] and MitoProt II [88] above a confidence level of 70%, suggesting that γCA proteins may be targeted to mitochondria in a wide variety of eukaryotes (see Supplemental Table 3.3). Notably, several of the known mitochondrial γCA proteins are not predicted to have mTPs with a given program (e.g., CA3 and CAL2 from *Arabidopsis*, AcCa2 from *Acanthamoeba* and all three γCAs from *Tetrahymena* with TargetP), so the lack of an apparent mTP does not preclude a mitochondrial localization. Genes encoding γCA proteins are absent from the nuclear genome sequences of several eukaryotes that do not have conventional mitochondria (e.g., *Trichomonas*, *Giardia*) and are also absent from at least one eukaryote (*Plasmodium*) that specifically lacks CI, further suggesting that γCA proteins may be targeted to mitochondria – and likely to CI – in most eukaryotes. Surprisingly, a γCA homolog was detected in the nuclear genome of *Entamoeba histolytica* [361], an amoebozoon that does not possess conventional mitochondria (i.e., it lacks an electron transport chain). This homolog, however, does not appear to be closely related to other eukaryotic γCAs, and may represent a lateral gene transfer from bacteria (data not shown).

Table 3.1 Distribution of γCAs throughout eukaryotes. Organisms are arranged according to supergroup membership [111]. 'Number of γCA homologs' refers to the number of distinct γCA proteins encoded by an organism as determined by searches of EST or nuclear genomic databases. Inferred protein sequences are available in Supplemental Files 3.2 and Supplemental Figure 3.1.

Supergroup	Taxon	γCA homolog?	# γCA homologs	
Plantae	*Arabidopsis*	Yes	5	All are CI subunits
	Chlamydomonas	Yes	3	All are CI subunits
	Cyanidioschyzon	Yes	1	
	Cyanophora	Yes	≥ 2	
Amoebozoa	*Acanthamoeba*	Yes	2	Both are CI subunits
	Dictyostelium	Yes	2	
	Hartmannella	Yes	≥ 1	
	Entamoeba	Yes	1	No ETC
	Polysphondylium	Yes	2	
Opisthokonta	*Saccharomyces*	No	0	
	Neurospora	No	0	
	Ustilago	No	0	
	Yarrowia	No	0	
	Bos	No	0	
	Drosophila	No	0	
	Monosiga	No	0	
Chromalveolata	*Tetrahymena*	Yes	3	All are mitochondrial
	Plasmodium	No	0	Lacks CI
	Phytophthora	Yes	2	
	Phaeodactylum	Yes	≥ 2	
	Blastocystis	Yes	≥ 1	
	Guillardia	Yes	≥ 3	
	Emiliania	Yes	≥ 1	

Supergroup	Taxon	γCA homolog?	# γCA homologs	
Rhizaria	*Bigelowiella*	Yes	≥ 1	
Excavata	*Reclinomonas*	Yes	≥ 1	
	Naegleria	Yes	2	
	Euglena	Yes	≥ 1	
	Trypanosoma	Yes	2	
	Malawimonas	Yes	≥ 2	
	Trichomonas	No	0	Lacks ETC
	Giardia	No	0	Lacks ETC

To assess evolutionary relationships, I performed a phylogenetic analysis of eukaryotic and bacterial γCA proteins, using the archaeal Cam sequence as outgroup. In the resulting phylogenetic trees (not shown), virtually all eukaryotic γCA proteins formed a large clade branching as a sister group to α-Proteobacteria (the *B. natans* sequence branched within the latter clade whereas the *E. histolytica* sequence mentioned above, as well as a *G. theta* one, did not affiliate with the other eukaryotic γCA sequences, but neither did they branch robustly with any particular bacterial clade). However, these trees were poorly resolved, with very little statistical (bootstrap) support for most partitions, so the phylogenetic conclusions that can be drawn from them are limited. Although the available data (Table 3.1) suggest an early γCA gene duplication during eukaryotic evolution, phylogenetic analysis was not able to delineate the point at which this may have occurred. It can be inferred, however, that the duplication predated the divergence of the amoebozoan taxa investigated here; moreover, the amoebozoan

sequences orthologous with either AcCa1 or AcCa2 were identified (Supplemental Table 3.2).

3.3.3 Analysis of Primary Protein Structure of γCA Homologs

As earlier noted, plant and green algal γCA proteins constitute a CI domain that is associated with the matrix face of the inner membrane-integrated arm [115; 117]. *Arabidopsis* encodes five γCA homologs (CA1, CA2, CA3, CAL1 and CAL2), all of which interact with CI [362], while *Chlamydomonas* CI contains three different γCA homologs [104]. Although *Arabidopsis* γCA1-3 (but not CAL1, CAL2) has retained nearly all residues known to be structurally or functionally important in the archetypal γCA (Cam) from the archaeon *M. thermophila* [356], including the Zn^{2+}-binding residues H81, H117 and H122 [355], the function of γCA homologs in plant mitochondria is still largely unknown. Plant γCA proteins lack two Asp residues (D61, D84) critical to the proton transfer mechanism of Cam and all previous attempts to demonstrate carbonic anhydrase activity in *Arabidopsis* mitochondrial extracts and in sucrose gradient-enriched plant CI [358] and recombinant γCA2 homotrimers [359] have been unsuccessful. Like the plant homologs, *Acanthamoeba* γCAs lack D61 and D84; moreover, *Acanthamoeba* γCAs conserve fewer of the other residues critical to Cam function than do plant γCAs. In terms of Zn^{2+}-binding His residues, AcCa1 conserves only H117 whereas AcCa2 retains H81 and H122 (Figure 3.2). However, it should be noted that the three His residues required to bind Zn^{2+} in Cam are not localized within the same γCA monomer; rather, Zn^{2+} is coordinated by H81 and H122 of one monomer and H117 of a second monomer within the trimeric Cam structure [363]. It is noteworthy that the three His residues are distributed between the two γCA homologs found in *Acanthamoeba*, so it may be that the

latter are able to bind Zn^{2+} in concert. However, in AcCa2, S and G replace N73 and Q75, respectively; in Cam, N73 and Q75 are proposed to be critical components of the catalytic mechanism [363]. Taken together, these findings suggest that the γCA homologs of *Acanthamoeba* CI may not possess a carbonic anhydrase activity. On the other hand, carbonic anhydrase activity has been reported for a comparably divergent, recombinant γCA from the haptophyte *Emiliania huxleyi* [360].

As is the case with the *Acanthamoeba* proteins, the vast majority of γCA proteins from eukaryotes outside of the plant supergroup do not individually conserve all three metal-binding H residues, although all three are often found among different γCAs where multiple isoforms are present. Thus, it is still unclear whether carbonic anhydrase activity or Zn^{2+} binding is likely to be a general feature of γCAs across eukaryotes. Interestingly, several other positions in my phylogenetically broad multiple alignment are fairly well conserved across eukaryotes (and in some cases, bacteria), but not in Cam; it is possible that the conservation of these residues points to novel, potentially important functional sites. For instance, a His residue (corresponding to H42 of *Arabidopsis* γCA1) is present in 27 of the 40 eukaryotic homologs with complete N-terminal sequence that I identified in this study, but not in bacterial/archaeal γCA homologs, while a Cys residue (*Arabidopsis* γCA1 C137) is highly conserved in eukaryotes and bacteria (41/48), but not in Cam (see Supplemental Figure 3.1). Additionally, many of the eukaryotic γCA homologs (including the *Acanthamoeba* ones) have C-terminal extensions relative to their prokaryotic homologs. In land plants, these extensions have been implicated in integration into the inner mitochondrial membrane [115], further supporting the idea that γCA proteins may be components of CI in many diverse eukaryotes.

Interestingly, inspection of the N-terminal portion of γCAs from diverse eukaryotes demonstrates that there may be, in general, at least two distinct classes of γCA proteins. In particular, the N-terminal ~60 amino acids of AcCa2, along with various protist γCAs, are highly similar to the corresponding region of *Arabidopsis* CA1-3, whereas the N-termini of AcCa1 and *Arabidopsis* CAL1-2 are not highly conserved (Figure 3.3). This observation suggests that there may be distinct roles or sub-localizations of the two classes of γCA within CI; specifically, the high degree of conservation of the N-terminal region of the AcCa2/CA1-3 class, but not AcCa1/CAL1-2 class, implies that the functional constraints at protein N-termini are stronger in the former class.

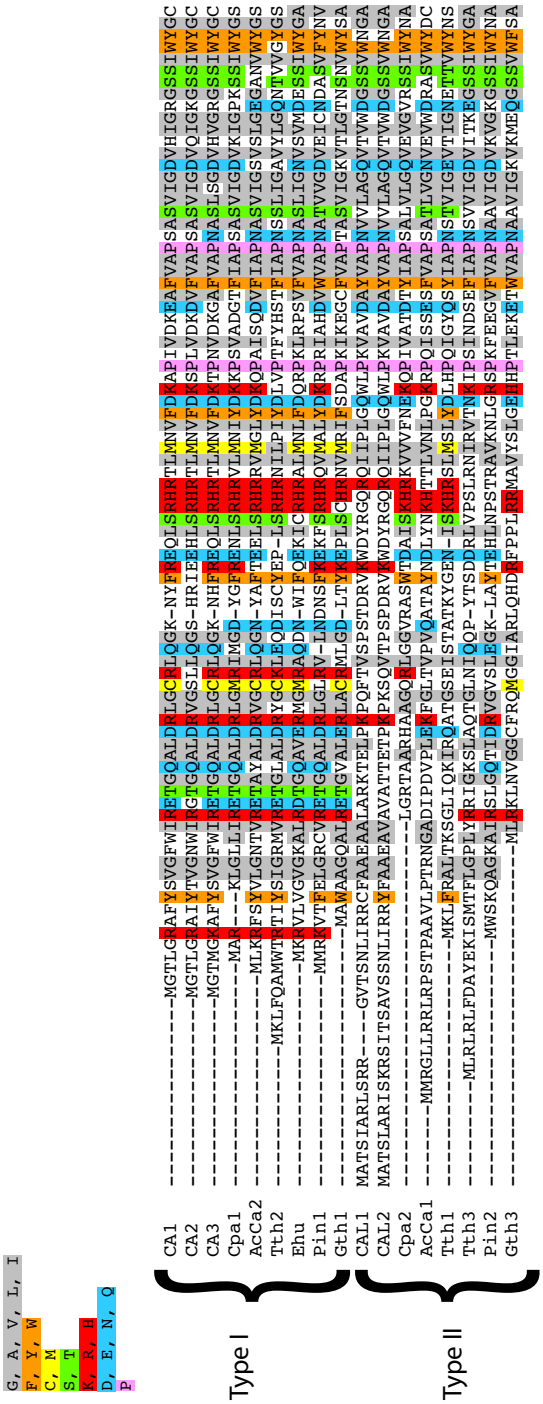


Figure 3.3 Distinct subtypes of eukaryotic γCAs**.** A truncated alignment of γCA proteins demonstrates that the N-terminal regions of certain isoforms (termed Class I) are highly conserved throughout Eucarya (see text). Species abbreviations: CA1-3, *Arabidopsis;* Cpa, *Cyanophora paradoxa;* AcCa, *Acanthamoeba castellanii;* Tth, *Tetrahymena thermophila;* Ehu, *Emiliania huxleyi;* Pin, *Phytophthora infestans;* Gth, *Guillardia theta*. Shading of columns according to the scheme presented above the alignment reflects amino acid similarity of ≥ 40%.

3.3.4 Possible Function(s) of γCA Proteins in CI

Although *in silico* reconstructions of *Arabidopsis* γCA proteins and comparisons to Cam suggest the possibility of carbonic anhydrase activity, the biochemical function of γCA proteins homologs in plant/green algal CI is controversial as there are currently no experimental data confirming that plant/green algal γCA proteins are carbonic anhydrases. Recently, however, it has been shown that *Arabidopsis* γCA2 trimers are capable of binding inorganic carbon [359], and microarray studies indicate that expression of *Arabidopsis* γCA1 and γCA2 is down-regulated under high CO_2 concentrations (see [358]). These results suggest that γCA homologs might play an important role in the metabolism or transport of one-carbon compounds in land plants. Although these observations provide evidence that γCAs function in relation to inorganic carbon metabolism in plants, it is still unclear whether these observations can be generalized across eukaryotes. As discussed above, most eukaryotes do not conserve all three His residues required for Zn^{2+} coordination (and therefore binding of CO_2 and HCO_3^-). Consequently, it may be that the function of γCA proteins is somewhat different in plants and other eukaryotes, and that CI γCA proteins in non-plant species serve a predominantly structural role in the complex. Ultimately, in order to distinguish between these possibilities, it will be important to determine whether or not non-plant γCA proteins are capable of binding Zn^{2+}.

Because CI γCA proteins were believed to be plant-specific, it has been proposed that they are involved in plant-specific processes, namely HCO_3^- formation (and possibly HCO_3^- transport across the inner mitochondrial membrane), ultimately for CO_2 fixation in chloroplasts [127]. Although I cannot rule out the possibility that CI γCAs play an active

role in chloroplast metabolism in plants, the presence of γCA homologs in CI of ancestrally aplastidic eukaryotes (e.g., *Acanthamoeba*) demonstrates that this could not have been the ancestral role of γCAs in CI, and that these proteins must perform some role that is not related to chloroplast function. Other data suggest that γCA proteins in plants may play a role in expression and assembly of mitochondrial CI. For instance, it was demonstrated that total levels of CI, along with abundance of CI+CIII supercomplexes, are dramatically decreased (>80%) in *Arabidopsis* γCA2 and γCA3 knockout lines grown in suspension culture [358]. A reduction in total CI protein was detected (i.e., not just aberrant CI assembly), suggesting that γCA proteins play an important role in the expression/stability of CI subunits. Recently, Tripp *et al*. [364] demonstrated that the active site of Cam contains Fe^{2+} instead of Zn^{2+} when the enzyme is reconstituted under anaerobic conditions. As noted by Parisi *et al*. [355], CI (along with the rest of the respiratory chain) contains multiple iron centers, so the Fe-binding ability of Cam may point to a role for γCA proteins in mitochondrial respiratory control. Notably, *Arabidopsis* γCA2, a physical component of CI, has been annotated as a transcription factor involved in the anther-specific expression of nucleus-encoded CI proteins (see GenBank accession no. AAK28403). The prospect that γCA proteins may function as CI-specific transcription factors is quite intriguing, as it would suggest a novel mechanism of communication between the mitochondrial electron transport chain and the nucleus, and would indicate a function for CI γCA proteins that could be applicable across all aerobic CI-containing eukaryotes, and not just photosynthetic ones.

3.4 Conclusions

The number of subunits comprising CI in eukaryotes has expanded markedly relative to the homologous bacterial complex. Previous comparative genomic analyses have suggested that 18 proteins were added to CI very early in eukaryotic evolution, while a smaller proportion is specific to particular lineages within the eukaryotic domain [75]. Although several of the novel, 'lineage-specific' proteins are likely to be encoded by a phylogenetically restricted group of related organisms, others are currently deemed to be 'lineage-specific' only because our knowledge of CI composition across the eukaryotes is incomplete. I have detected two γCA homologs – previously thought to be plant/green algal CI proteins – as components of CI in an amoeboid protozoon, *Acanthamoeba castellanii*, an organism not specifically related to the plant/green algal lineage. The most parsimonious interpretation of my observations is that γCA proteins were part of the eukaryotic core of CI that was added early in eukaryotic evolution, and that they were subsequently lost in certain groups, most notably the opisthokonts. Moreover, γCA proteins are known to be mitochondrial in ciliates (and may be CI proteins as well) and bioinformatics searches have revealed a large repertoire of γCA homologs in other major eukaryotic groups. These observations underscore the importance of characterizing the composition of mitochondrial protein complexes from a wide variety of organisms in order to understand fully and accurately their function and evolution.

3.5 Methods

3.5.1 Cell Growth and Isolation of Mitochondria

Mitochondria were prepared from two 500-ml cultures of *Acanthamoeba castellanii* (strain Neff) cells grown to an OD_{580} of ~1.0, essentially as reported by Lohan and Gray [179]. The present method differed in that cells were not washed with phosphate buffered saline prior to lysis, the crude mitochondrial pellet was washed only once with mitochondrial wash buffer (10 mM Tris·HCl, 10 mM Na_2EDTA, 0.35 M sucrose, 1 mM dithiothreitol and 0.1% BSA) and mitochondria were purified in a SW27Ti rotor centrifuged at 22,000 rpm. The protein concentration was assayed with the BioRad D_C Protein Assay kit and whole mitochondria were diluted to a final protein concentration of 25 mg/ml.

3.5.2 Blue Native Polyacrylamide Gel Electrophoresis (BN-PAGE)

BN-PAGE was carried out according to Schägger and von Jagow [140]. A 4-µl aliquot of purified mitochondrial fraction (corresponding to 100 µg protein) was solubilized with 40 µl of a solution containing 0.5% *n*-dodecyl-β-D-maltoside in 750 mM 6-aminocaproic acid/50 mM bis-tris [bis-(2-hydroxyethyl)-amino-tris(hydroxymethyl)-methane] for 15 min on ice. The solution was centrifuged at 18,000 × *g* for 30 min at 4°C in order to sediment insoluble material. The supernatant was supplemented with 1.5 µl 5% Coomassie Brilliant Blue G250 in 750 mM 6-aminocaproic acid/50 mM bis-tris and 5.0 µl 50% (v:v) glycerol. Samples were loaded onto a 1.5-mm thick, 4-12% polyacrylamide linear gradient gel without a stacking gel and electrophoresed for 2 hr at 150 V, 2 hr at 350 V and ~5 hr at 500 V at 4°C. Gels were either a) stained with a solution of 0.025% Coomassie Brilliant Blue R250, 40% ethanol, 10% acetic acid or b)

gel lanes were excised for in-gel enzyme activity assays or two-dimensional BN/SDS-PAGE. Molecular weights were estimated using the NativeMark unstained protein standard (Invitrogen).

3.5.3 In-gel Enzyme Activity Assay

In order to identify enzymatically active CI, in-gel activity assays were performed. A lane from a BN-PAGE experiment was excised and incubated overnight in 5 ml 50 mM MOPS·NaOH, pH 7.4 buffer containing 0.1 mg/ml reduced β-NADH and 1 mg/ml nitroblue tetrazolium. Active CI was identified by the formation of a purple-blue formazan precipitate at approximately 940 kDa.

3.5.4 Two-dimensional BN/SDS-PAGE

After one-dimensional BN-PAGE, a gel strip was excised and treated for 45 min with 10 ml of a 0.125 M Tris·HCl buffer, pH 6.8 (1× SDS-PAGE stacking buffer), containing 2% sodium dodecyl sulfate (SDS) and 1% β-mercaptoethanol. The strip was incubated for 10 min in 10 ml of the same solution, minus β-mercaptoethanol (a potent inhibitor of polymerization). The BN gel strip was encased in 4% SDS-PAGE stacking gel poured on top of a 15% resolving gel. Electrophoresis was carried out at room temperature at a constant current of 30 mA for approximately 6 hr. The gel was stained with the BioRad Silver Stain Plus kit according to manufacturer's protocols.

3.5.4 Reversed-phase HPLC and MS/MS

The enzymatically active CI and inactive CI complexes were excised and each was placed in a separate microcentrifuge tube. The proteins were reduced with 10 mM dithiothreitol in 100 mM ammonium bicarbonate (AB) for 30 min at 56°C and alkylated with 100 mM iodoacetamide in 100 mM AB for 30 min in the dark at room temperature.

Proteins were incubated with trypsin overnight (~16 hr) at 37°C with 12.5 ng/µl modified porcine trypsin (Promega) in 100 mM AB. Peptides were extracted once with 100 mM AB and twice with a 50:50 (v:v) acetonitrile:H_2O solution containing 0.2% formic acid. All samples were analyzed by LC-MS/MS using an Agilent 1100 HPLC system equipped with a 15 cm × 100 mm Onyx Monolythic C_{18} column (Phenomenex, Torrance, California). The separation was carried out using the following gradient: 2% B for 3 min increasing to 25% B over 45 min and 95% B over 10 min (A: 0.1% formic acid in water; B: 0.1% formic acid in acetonitrile) at 2 µl/min. The HPLC was interfaced to an AB/MDS-SCIEX QTrap 4000 mass spectrometer via a nanoflow source. Data were acquired in the information-dependent acquisition mode, i.e., the m/z values of the tryptic peptides were measured using an MS scan, followed by enhanced resolution scans of the three most intense peaks and finally three tandem MS scans. The tandem MS spectra were submitted to the database search program MASCOT [365] in order to identify the proteins. Data files were searched against *A. castellanii* EST clusters from TBestDB [185], a collection of *Acanthamoeba* ESTs generated by 454 pyrosequencing (kindly made available by BJ Loftus, University College Dublin), the current nuclear genome assembly (*A. castellanii* genome project), and the *A. castellanii* mitochondrial genome sequence [178]. The current *A. castellanii* genome assembly and Loftus EST data set are both publicly available from the Human Genome Sequencing Center, Baylor College of Medicine (http://www.hgsc.bcm.tmc.edu/projects/microbial/microbial-pubPreview.xsp?project_id=163).

3.5.5 Homology Searches, Alignments, mTP Prediction

The amino acid sequence of *Arabidopsis thaliana* γCA1 (AT1G5980) was used to query the nr protein database at NCBI with BLASTp, as well as EST and nuclear genome databases from a wide variety of eukaryotes with tBLASTn [189]. Nucleic acid sequences were translated using the transeq program of the EMBOSS package and protein sequences were inferred manually. Protein homologs were aligned with Muscle v3.6 [190] using default parameters and edited with the BioEdit Sequence Alignment Editor.

TargetP [90] and MitoProt II [88] were used to assess the probability of mitochondrial localization for eukaryotic γCA homologs. When TargetP was used, the 'Plant' organism group was selected for organisms with primary or secondary plastids in order to include the possibility of plastid-targeted γCA proteins; for species lacking a plastid, the 'Animal' organism group was selected. MitoProt II does not have an option for assessing plastid localization.

3.5.6 Phylogenetic Analysis

The AcCa1 protein sequence was used to query the nr database at NCBI via BLASTp. The best scoring homologs (i.e., ones having the lowest E- values) from a variety of major bacterial groups were retrieved and aligned with eukaryotic γ-type CA homologs using Muscle v.3.6 [190]. The alignment was edited manually with eBioX and was then used to reconstruct the maximum likelihood phylogenetic tree using RAxML-HPC [193] under the WAG + Γ model (using the PROTGAMMAWAGF option) with 25 categories of substitution rate variation. One hundred bootstrap replicates were performed as a measure of statistical support for inferred nodes.

3.6 List of Abbreviations

CI, complex I of the respiratory chain; γCA, γ-type carbonic anhydrase; Cam, archetypal γCA from *Methanosarcina thermophila*; AcCa1 and AcCa2, names given to γ-type carbonic anhydrases from *Acanthamoeba*; NADH, nicotinamide adenine dinucleotide (reduced); BN-PAGE, blue native polyacrylamide gel electrophoresis; mTP, mitochondrial targeting peptide; EST, expressed sequence tag; ETC, electron transport chain

3.7 Author's Contributions

RMRG performed the mitochondrial isolations, BN-PAGE, two-dimensional BN/SDS-PAGE and sample preparation for MS/MS. RMRG performed bioinformatic analyses. RMRG and MWG prepared the manuscript. Both authors read and approved the final manuscript.

3.8 Acknowledgements

RMRG was supported by a Canada Graduate Scholarship (Doctoral) from the Natural Sciences and Engineering Research Council (NSERC) and a Predoctoral Scholarship from the Killam Trusts. This work was supported by operating grant MOP-4124 to MWG from the Canadian Institutes of Health Research. The authors thank DM Pinto and KA Chisholm (National Research Council Canada, Institute for Marine Biosciences, Halifax, NS) for assistance with mass spectrometry.

Chapter 4 A Split And Rearranged Nuclear Gene Encoding The Iron-Sulfur Subunit Of Mitochondrial Succinate Dehydrogenase In Euglenozoa

This chapter includes work published in **Gawryluk, R.M.R.** & Gray, M.W. (2009). A split and rearranged nuclear gene encoding the iron-sulfur subunit of mitochondrial succinate dehydrogenase in Euglenozoa. *BMC Res Notes* **2**, 16.

4.1 Abstract

Background: Analyses based on phylogenetic and ultrastructural data have suggested that euglenids (such as *Euglena gracilis*), trypanosomatids and diplonemids are members of a monophyletic lineage termed Euglenozoa. However, many uncertainties are associated with phylogenetic reconstructions for ancient and rapidly evolving groups; thus, rare genomic characters become increasingly important in reinforcing inferred phylogenetic relationships.

Findings: I discovered that the iron-sulfur subunit (SdhB) of mitochondrial succinate dehydrogenase is encoded by a split and rearranged nuclear gene in *Euglena gracilis* and trypanosomatids, an example of a rare genomic character. The two subgenic modules are transcribed independently and the resulting mRNAs appear to be independently translated, with the two protein products imported into mitochondria, based on the presence of predicted mitochondrial targeting peptides. Although the inferred protein sequences are in general very divergent from those of other organisms, all of the required iron-sulfur cluster-coordinating residues are present. Moreover, the discontinuity in the euglenozoan SdhB sequence occurs between the two domains of a typical, covalently

continuous SdhB, consistent with the inference that the euglenozoan 'half' proteins are functional.

Conclusions: The discovery of this unique molecular marker provides evidence for the monophyly of Euglenozoa that is independent of evolutionary models. My results pose questions about the origin and timing of this novel gene arrangement and the structure and function of euglenozoan SdhB.

4.2 Background

Succinate dehydrogenase (SDH, Complex II) is a membrane-anchored protein complex of the mitochondrial and bacterial electron transport chain that catalyzes the oxidation of succinate to fumarate (although it is capable of the reverse reaction under favorable conditions) and the reduction of ubiquinone. High-resolution crystal structures of Complex II from bacterial (*E. coli*; [366]), avian (chicken; [367]) and mammalian (pig; [368]) sources demonstrate that it is a heterotetramer consisting of the succinate-oxidizing, matrix-associated, flavoprotein subunit (SdhA), an electron transfer iron-sulfur subunit (SdhB) and two hydrophobic membrane anchors (SdhC and SdhD) that provide the binding site for ubiquinone and are required for integration of the complex into the inner mitochondrial membrane. SdhA-D are nucleus-encoded in a wide variety of eukaryotes, including mammals, whereas SdhB-D are specified by the gene-rich mitochondrial genomes of certain protists such as red algae and jakobid flagellates [369]. SdhA invariably appears to be nucleus-encoded.

Euglena gracilis is a free-living, flagellated eukaryotic microbe that contains a plastid likely acquired through the engulfment of a green alga [370]. A monophyletic

'Euglenozoa' clade comprising *Euglena* (and related euglenids) along with two aplastidic lineages, the kinetoplastids (encompassing trypanosomatids and bodonids) and the predominantly free-living diplonemids, has been postulated principally on the basis of shared ultrastructural features, including disc-shaped mitochondrial cristae and flagellar paraxonemal rods [371]. Phylogenetic reconstructions based on small subunit ribosomal RNA (SSU rRNA; [372]) and protein [373] sequences established that these physiologically and ecologically disparate taxa likely comprise a (potentially early-branching) monophyletic group. However, the well-documented effects of rapid rates of sequence change, along with the acquisition of a secondary endosymbiont [370] (evidenced by the presence in *Euglena* of a plastid with three surrounding membranes) and possible ephemeral, cryptic endosymbioses [68], have complicated reconstructions of *Euglena*'s evolutionary history. In particular, the transfer of endosymbiont-derived genes to the nucleus has, in effect, yielded a mosaic nuclear genome displaying characteristics of all constituent sources [374]. Moreover, the internal branching patterns within Euglenozoa are still not completely resolved [375], although phylogenies based on conserved protein genes seem to be consistent in placing euglenids at the base of Euglenozoa, with diplonemids and kinetoplastids forming a later diverging sister group [68].

Here I report that in *E. gracilis*, the nucleus-encoded *sdhB* gene is split into two independently transcribed (and presumably independently translated) subgenic modules whose products correspond to the N-terminal and C-terminal halves (referred to here as SdhB-n and SdhB-c, respectively) of a typical SdhB protein. Moreover, in various trypanosome species, I have identified separate genes encoding predicted proteins corresponding to SdhB-n and SdhB-c. The splitting of *sdhB* in *Euglena* and

trypanosomatids is an example of a unique molecular character that specifically unites these two phylogenetic groups and raises interesting questions about the evolution and function of euglenozoan SdhB.

4.3 Results and Discussion

Relatively few genomic data are available for *Euglena.* Neither nuclear nor mitochondrial genome sequencing projects are currently being undertaken, and only three mitochondrion-encoded protein-coding genes (*cox1*, *cox2* and *cox3*) have been identified thus far [[376; 377], GenBank:AF156178]. Nevertheless, the construction and sequencing of EST libraries generated from mature mRNAs is being exploited to better understand the biochemistry and evolution of this organism. The conserved 24-nucleotide 5′ spliced leader (SL) sequence characteristic of *Euglena* nucleus-encoded mRNAs [378] confers a specific advantage in that its presence in an EST confirms that the translated sequence encompasses the complete N-terminus of the corresponding protein. This information is important in predicting the subcellular localization of a given protein product, as the signals required for targeting proteins to various subcellular compartments, including mitochondria, are frequently located at protein N-termini.

Analysis of *Euglena gracilis* EST data demonstrates that SdhB is expressed as separate N- and C-terminal units. The EST clusters from *Euglena* are considered to be complete, as those representing both *sdhB-n* and *sdhB-c* each contains at least 10 of the 3′-most nucleotides (‘TTTTTTTTCG’) of the conserved SL sequence at the 5′ end (Figures 4.1A,B), an ATG initiation codon a few nucleotides further downstream, and a stop codon near the 3′ end of the EST sequence. Moreover, the presence of an SL in *Euglena* ESTs demonstrates that *sdhB-n* and *sdhB-c* are nucleus-encoded in this protist,

as mitochondrial transcripts are not known to contain spliced leaders. In total, I identified 8 and 4 ESTs corresponding, respectively, to *Euglena sdhB-n* and *sdhB-c*. Nearly complete *sdhB-n and sdhB-c* ESTs were also found for the related species, *Euglena* (*Astasia*) *longa.* SdhB-n and SdhB-c protein sequences from *E. gracilis* and *E. longa* are 93% and 91% identical, respectively. Although the *E. longa* ESTs lack the SL and the sequence corresponding to the extreme N-termini of the two proteins, these ESTs provide further evidence that the *sdhB* modules are transcribed separately in the nucleus of euglenids. Similarly, SdhB sequences inferred from the genome sequences of several trypanosome species indicate that SdhB is also expressed as two separate pieces in these organisms (see Figure 4.2 for partial protein alignments and Supplemental Figures 4.1 and 4.2 for more extensive alignments). In fact, the two SdhB pieces are encoded on separate chromosomes in the nuclear genomes of trypanosomatids; in *T. brucei*, *sdhB-n* is on chromosome 8 while *sdhB-c* is on chromosome 9, whereas in *L. major*, *sdhB-n* is on chromosome 23 and *sdhB-c* is on chromosome 15. The fragmented nature of *sdhB* in trypanosomatids was evidently not previously noted, as the relevant coding regions in both *T. brucei* and *L. major* are annotated simply as 'succinate dehydrogenase subunits' or 'hypothetical proteins'. BLAST searches did not retrieve any *sdhB* transcripts from the limited diplonemid EST libraries available in TBestDB (http://amoebidia.bcm.umontreal.ca/pepdb/searches/login.php?bye=true), including those of *D. papillatum*, *D. ambulator* and *Rhyncopus*; moreover, no mitochondrion-encoded *sdhB* gene was identified during sequencing of the mitochondrial genome of *Diplonema papillatum* [379]. Nevertheless, parsimony considerations argue that SdhB is nucleus-encoded and bipartite in diplonemids as well, given phylogenetic evidence indicating that

diplonemids and trypanosomatids [373] or diplonemids and euglenids [372] are sister groups. Exhaustive searches of available genomic and EST data did not turn up evidence of this split SdhB gene arrangement anywhere outside Euglenozoa.

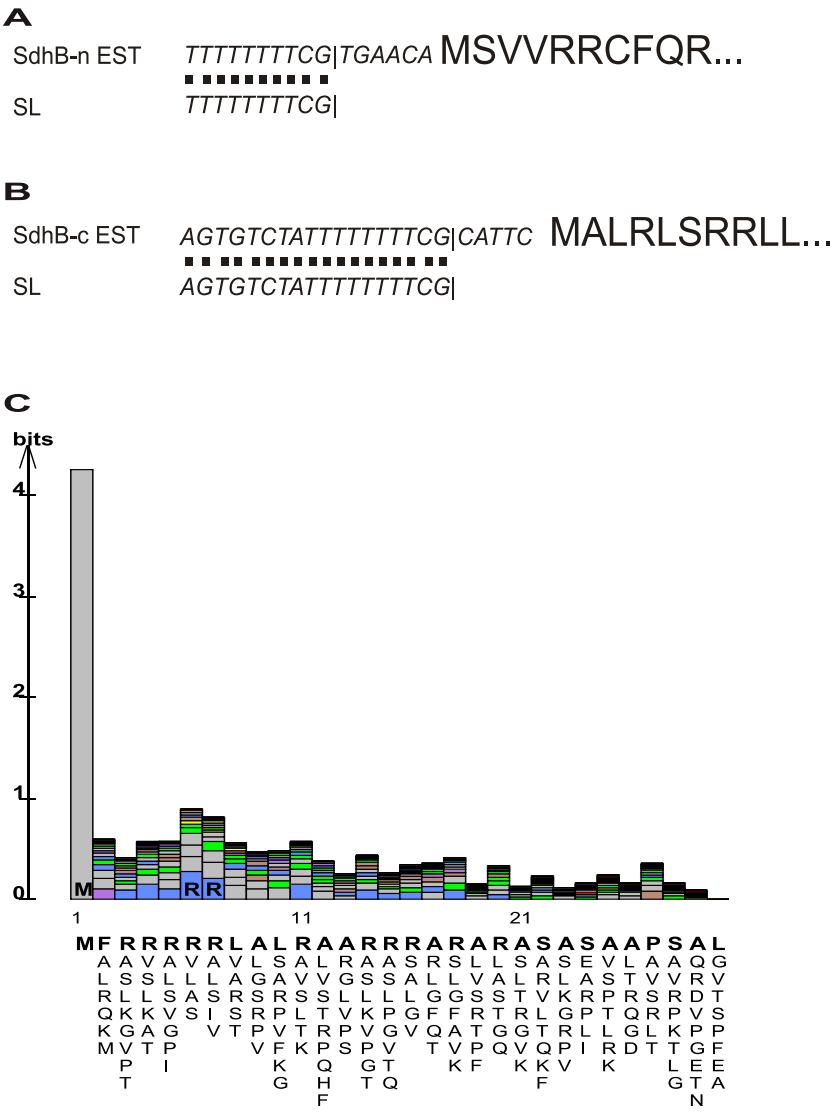
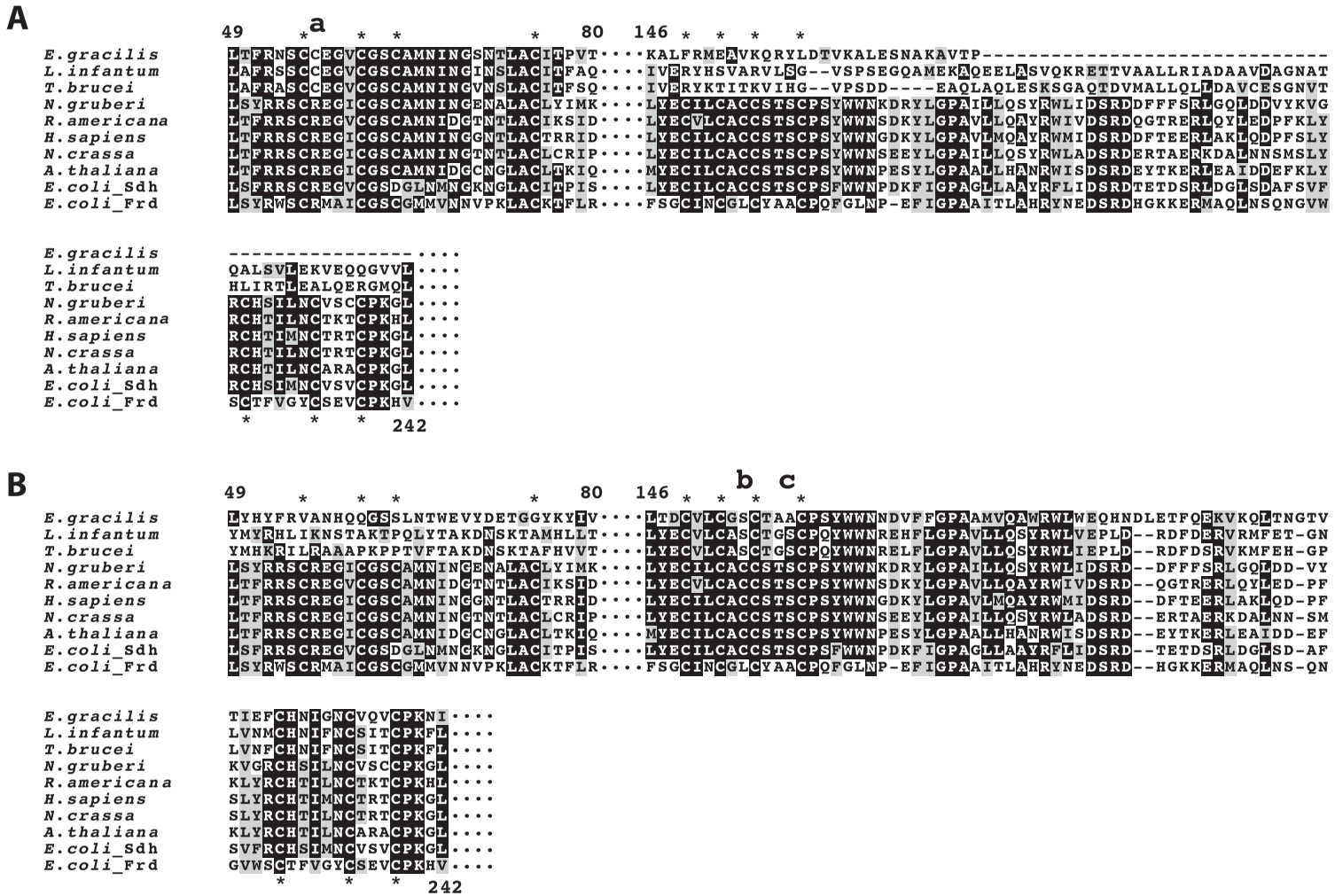


Figure 4.1 *E. gracilis* consensus EST clusters contain sequences corresponding to spliced leaders. A. SdhB-n. B. SdhB-c. Letters in italics and smaller font represent nucleic acid sequences, while larger bolded letters represent the N-terminus of the inferred protein sequences. The | character denotes the 3′ end of the spliced leader. 'EST' refers to the sequence of the expressed sequence tag and 'SL' refers to the sequence of the spliced leader. C. Consensus mTP profile generated from de-gapped alignment of the N-terminal 30 residues from 107 *E. gracilis* predicted mitochondrion-targeted proteins.

Both of the deduced *Euglena* SdhB pieces are predicted to contain mitochondrial targeting peptides (mTPs). TargetP [90] predicts a mitochondrial localization for SdhB-n and SdhB-c with 91.5% and 96.1% confidence, respectively, whereas the confidence levels with MitoProtII [88] are 99.8% and 95.7%, respectively. Moreover, the N-terminal sequences of SdhB-n and SdhB-c, which are rich in Arg, Ser and hydrophobic residues, closely resemble a consensus *E. gracilis* mTP profile generated from an alignment of 107 predicted mitochondrion-targeted proteins (Figure 4.1C). That both SdhB-n and SdhB-c contain predicted mTPs is strong evidence that *Euglena* imports both of these separate proteins independently into mitochondria, where they presumably form a heterodimer that effects the role of the classical, covalently continuous SdhB. In trypanosomatids, only SdhB-n is confidently predicted to possess a mTP. The significance of this observation is unclear, although it is possible that SdhB-c is only imported into mitochondria under certain developmental or physiological conditions, or that the protein is imported in a fashion that does not require a cleavable mTP [380].

Figure 4.2 Alignment of selected regions of *Euglena* and trypanosome SdhB-n and SdhB-c with SdhB and FrdB homologs from other eukaryotes and bacteria. A. SdhB-n. B. SdhB-c. Euglenozoan SdhB-n and SdhB-c proteins were aligned with the same set of SdhB and FrdB proteins from other species. Alignments were edited to emphasize regions encompassing conserved Cys residues (denoted by *) responsible for co-ordinating Fe-S clusters. Numbering reflects corresponding amino acid residues in *E. coli* SdhB. Shading of columns indicates amino acid identity of 60% or higher. The letters **a**, **b** and **c** highlight particular residues in the alignment that are discussed in the text. Full organism names and database accession numbers are listed in 'Supplemental Table 4.1: Accession numbers and allied information'.


A
49
a
80
146
E.gracilis
L.infantum
T.brucei
N.gruberi
R.americana
H.sapiens
N.crassa
A.thaliana
E.coli_Sdh
E.coli_Frd
242
B
b
c

The discovery of split genes encoding proteins that function within mitochondria is not without precedent. For instance, cytochrome oxidase subunit 2 (Cox2) in the green algae *Chlamydomonas reinhardtii* and *Polytomella* sp. [135] and in several apicomplexan parasites [146] and dinoflagellates [147] is a nucleus-encoded heterodimer specified by two separate subgenic modules. In *Chlamydomonas*, the N-terminal portion of Cox2 has been shown to contain a cleavable N-terminal mTP, whereas the C-terminal unit does not [135]; this situation parallels that reported here for trypanosome SdhB-c, which does not appear to contain a canonical cleavable mTP. In the case of chlamydomonad algae, it has been proposed that a 20-amino acid C-terminal extension in Cox2a (the N-terminal unit) and a 42-amino acid N-terminal extension in Cox2b might facilitate the functional interaction of these two subunits [135]. In the absence of biochemical evidence confirming the length of the mitochondrial targeting peptide, it is not possible to determine unequivocally whether or not trypanosome SdhB-c has an N-terminal extension. On the other hand, SdhB-n from *Euglena* does possess a C-terminal extension of ~35 amino acids, whereas the corresponding trypanosome SdhB-n C-terminal extension is ~105 residues long. Sequence alignments do not indicate any significant similarity between the *Euglena* and trypanosome extensions. As proposed for Cox2 in chlamydomonads, these extensions might allow the dimerization of SdhB-n and SdhB-c in euglenozoans, although bioinformatic analysis does not suggest the presence of obvious protein-protein interaction domains.

From a protein domain perspective, the split in the SdhB sequence in *Euglena* and trypanosomatids occurs in a region that might be particularly tolerant of such disruption (Figure 4.3). SdhB contains three iron-sulfur (Fe-S) centers, arranged in a linear chain,

that function to transport electrons from SdhA to the membrane-integrated subunits [366]. SdhB from *E. coli* is organized into two domains: an N-terminal domain containing a [2Fe-2S] cluster that forms a fold similar to plant-type ferredoxins and a C-terminal domain that houses the [4Fe-4S] and [3Fe-4S] clusters with a fold similar to bacterial ferredoxins [381]. SdhB-n from *Euglena* contains a predicted Fer2 domain whereas SdhB-c is predicted to contain two Fer4 domains, indicating that the break between *Euglena* SdhB-n and SdhB-c occurs in a region corresponding to the junction between the two *E. coli* domains. Moreover, protein alignments demonstrate that all of the Cys residues required for co-ordination of the three Fe-S clusters in *E. coli* SdhB are accounted for when both SdhB-n and SdhB-c from Euglenozoa are considered. These observations lend further support to the notion that these separate protein halves are functional, as rearrangement occurring within protein domains and/or loss of Fe-S cluster ligands would likely not be tolerated.

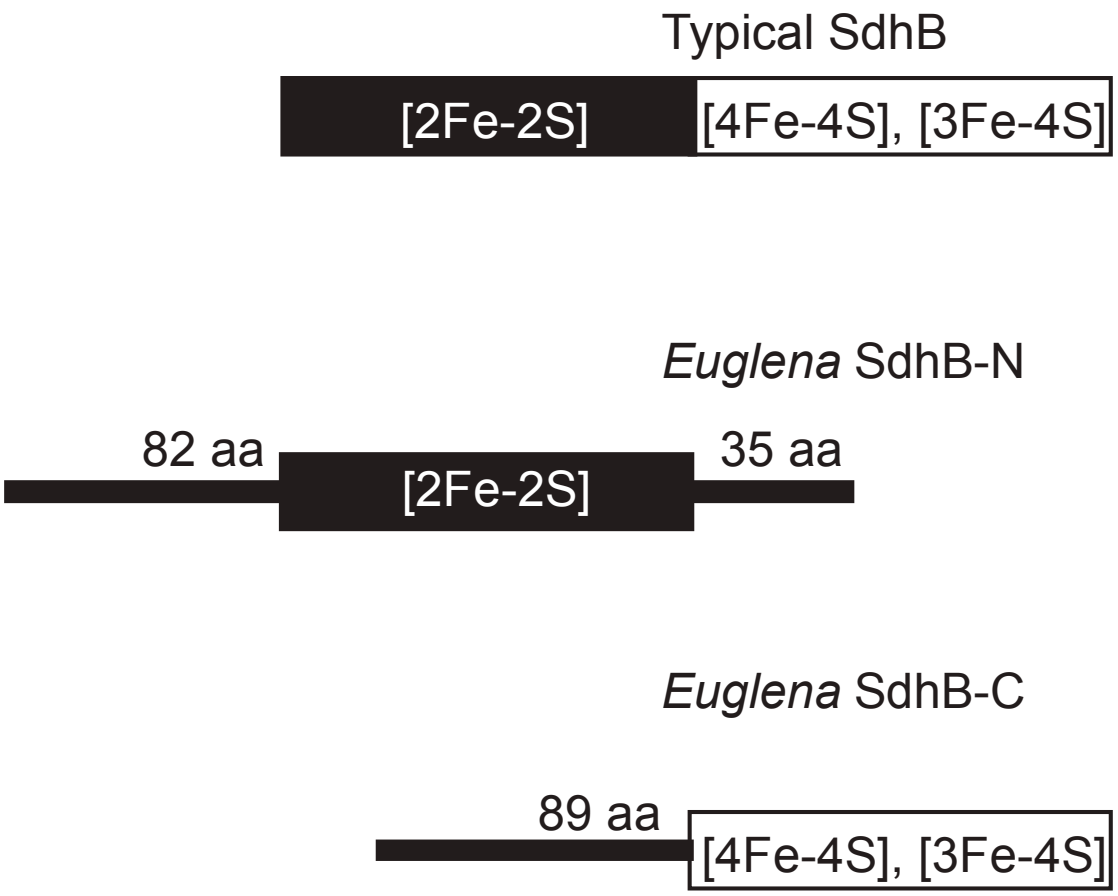


Figure 4.3 Organization of protein domains in typical SdhB proteins versus SdhB-n and SdhB-c from *Euglena* and trypanosomatids. SdhB-n contains the [2Fe-2S]-cluster-binding domain characteristic of the N-terminus of typical SdhB. SdhB-c contains [3Fe-4S]- and [4Fe-4S]-cluster-binding domains that are present in the C-terminal half of SdhB.

Notably, the amino acid sequences of SdhB-n and SdhB-c from Euglenozoa are exceptionally divergent in comparison with SdhB characterized to date in any other organism. In fact, many of the otherwise universally (or nearly universally) conserved residues have been substituted with different ones in Euglenozoa. For instance, a universally conserved Arg (R56 in *E. coli*) is Cys in SdhB-n of both *Euglena* and

trypanosomatids (Fig. 4.2A, **a**). Conversely, the conserved Cys corresponding to C154 in *E. coli* is Ser in *Euglena* and trypanosome SdhB-c (Fig. 4.2B, **b**), as well as in SdhB from the unrelated malaria parasite, *Plasmodium falciparum*. The nearby Ser-Thr-Ser motif present in all other SdhB sequences examined here (corresponding to *E. coli* residues 156-158; Fig. 4.2B, **c**) is Thr-Ala-Ala in *Euglena*. Although *E. coli* C154 is not directly responsible for coordinating Fe-S clusters in SdhB, the crystal structure suggests that it contributes a hydrogen bond to the thiol group, important in stabilizing the [4Fe-4S] cluster ligand C152 [366]. It is thought that this H-bond maintains a higher midpoint potential in the [4Fe-4S] cluster. Interestingly, Cheng *et al*. [382] found a direct relationship between the midpoint potential of the [4Fe-4S] cluster and the turnover rates of succinate dehydrogenase, whereas Hudson *et al.* [383] found the inverse for the Fe-S subunit of *E. coli* fumarate reductase (Frd; an homologous enzyme that catalyzes the reduction of fumarate to succinate). Thus, the presence of C154 may favor the *in vivo* oxidation of succinate to fumarate, as opposed to the reverse reaction [382]. *E. coli* FrdB, which has a lower [4Fe-4S] cluster midpoint potential than does *E. coli* SdhB, has a Leu residue instead of the *E. coli* C154 equivalent (Fig. 4.2B, **b**) and a Tyr-Ala-Ala motif (Thr-Ala-Ala in *Euglena*) instead of Ser-Thr-Ser (Fig. 4.2B, **c**). Thus, there exist some interesting parallels between the euglenozoan SdhB and *E. coli* FrdB sequences, although phylogenetic analyses (see Supplemental Figure 4.3: SdhB phylogenetic tree) clearly demonstrate that SdhB-n and SdhB-c are SdhB (and not FrdB) homologs. Moreover, it is quite possible that the Ser in euglenozoan SdhB-c contributes a stabilizing hydrogen bond to the [4Fe-4S] cluster (equivalent to the function of C154 of *E. coli*) whereas Leu in FrdB could not. Taken together, the euglenozoan SdhB structure and sequence are intriguing, and emphasize the

need for biochemical investigations to fully understand the function and structure of these split proteins.

4.4 Methods

Expressed sequence tags (ESTs) from *E. gracilis* strain Z were prepared as described in [384]. ESTs encoding *Euglena* SdhB were identified by a tBLASTn [189] search of the taxonomically broad EST database (TBestDB; [185]) and GenBank, using SdhB from *Reclinomonas americana* (gi:11466549) as query. Consensus EST sequences specifying SdhB-n and SdhB-c were translated and the inferred protein sequences were subsequently used to query the non-redundant protein sequence database at NCBI (using BLASTp) along with the non-human, non-mouse EST database (est_others) and TBestDB (using tBLASTn). The programs TargetP [90] and MitoProt II [88] were used to assess the probability of mitochondrial localization for *Euglena* and trypanosome SdhB-n and SdhB-c. When using TargetP for *Euglena* proteins, I selected the 'Plant' organism group in order to include the possibility of plastid-targeting, whereas I selected the 'Animal' organism group for trypanosomatids, as the latter do not contain plastids. MitoProt II contains no option for assessing plastid localization. The consensus *E. gracilis* mTP profile was generated using LogoBar-0.9.12 [192] from a de-gapped alignment of the 30-most N-terminal residues from 107 predicted *E. gracilis* mitochondrion-targeted proteins.

Conserved domains were identified by searching the Pfam and SMART databases at the SMART server [385], using *E. gracilis* SdhB-n and SdhB-c as queries. Protein alignments were constructed using Muscle v3.6 [190] with default parameters and edited with the BioEdit Sequence Alignment Editor. The editing function was used to remove gaps from the non-homologous euglenozoan protein extensions. However, regions

corresponding to likely mTPs were left unedited. In the alignment, shading of a given column reflects a minimum of 60% identity.

4.5 List of Abbreviations

FAD, flavin adenine dinucleotide (oxidized form); $FADH_2$, flavin adenine dinucleotide (reduced form); Fe-S, iron-sulfur; mTP, mitochondrial targeting peptide; SDH, succinate dehydrogenase (succinate-ubiquinone oxidoreductase); SL, spliced leader

4.6 Competing Interests

The authors declare that they have no competing interests related to the contents of this manuscript.

4.7 Authors' Contributions

RMRG discovered the EST and gene sequences corresponding to bipartite SdhB-n and SdhB-c in *Euglena* and trypanosomes and performed bioinformatics analyses. RMRG and MWG prepared the manuscript. Both authors read and approved the final manuscript.

4.8 Acknowledgements

RMRG was supported by a CGSD from the Natural Sciences and Engineering Research Council (NSERC) and a Predoctoral Scholarship from the Killam Trusts. MWG gratefully acknowledges salary support from the Canada Research Chairs Program as well as operating support from the Canadian Institutes of Health Research (MOP-4124).

Chapter 5 An Ancient Fission of Mitochondrial *cox1*

This chapter includes work published in **Gawryluk, R. M. R.** & Gray, M. W. (2010). An ancient fission of mitochondrial *cox1*. *Mol Biol Evol* **27**, 7-10.

5.1 Abstract

Many genes inherited from the α-proteobacterial ancestor of mitochondria have undergone evolutionary transfer to the nuclear genome in eukaryotes. In some rare cases, genes have been functionally transferred in pieces, resulting in 'split' proteins that presumably interact in *trans* within mitochondria, fulfilling the same role as the ancestral, intact protein. I describe a nucleus-encoded mitochondrial protein (here named Cox1-c) in the amoeboid protist *Acanthamoeba castellanii* that is homologous to the C-terminal portion of conventional mitochondrial Cox1, while the corresponding portion of the mitochondrion-encoded *A. castellanii* Cox1 is absent. Bioinformatic searches retrieved nucleus-encoded Cox1-c homologs in most major eukaryotic supergroups; in these cases, also, the mitochondrion-encoded Cox1 lacks the corresponding C-terminal motif. These data constitute the first report of functional re-location of a portion of *cox1* to the nucleus. This transfer event was likely ancient, with the resulting nuclear *cox1-c* being differentially activated across the eukaryotic domain.

Abbreviations: Cox1, full-length cytochrome oxidase subunit 1; Cox1-c, a nuDNA-encoded protein homologous to the C-terminal portion of ancestral Cox1; Cox1(-), Cox1 protein missing the C-terminal portion that corresponds to Cox1-c; EGT, endosymbiotic

gene transfer; mtDNA, mitochondrial DNA; mTP, mitochondrial targeting peptide; nuDNA, nuclear DNA

5.2 Introduction

Mitochondria are biochemically and ultrastructurally diverse eukaryotic organelles whose complete protein repertoire (proteome) ranges from several hundred to >1000 polypeptides [17; 19; 21; 23]. Although mitochondria have retained a relict genome (mtDNA), ≥95% of proteins that function in mitochondria are encoded in the nucleus (nuDNA) and imported post-translationally [43]. Many genes formerly encoded by mtDNA have been functionally transferred to the nucleus via endosymbiotic gene transfer (EGT; see [1]). Functional mitochondrion-to-nucleus EGT is not common in animals, but is still an active process in green algae [135] and land plants [49; 386].

In most cases, the integrity of a gene is preserved in the course of gene transfer (i.e., the gene remains a single, continuous entity). However, several cases have been reported of fission in genes coding for mitochondrial proteins, resulting in either both halves encoded in the nuDNA [135; 229] or one part encoded in the mtDNA and the other in the nuDNA [145; 146]. A notable example of mitochondrial gene fission is *cox2* in chlorophyte algae. Whereas both halves (Cox2a and Cox2b) of the typical Cox2 protein are encoded in the nuDNA in *Chlamydomonas* [135], Cox2a has been retained in the mtDNA of *Scenedesmus* [155] while *Scenedesmus* Cox2b is nucleus-encoded [146], a clear scenario for the separate functional transfer of two halves of a mitochondrial gene to the nucleus.

From an evolutionary perspective, most documented examples of fissioned mitochondrial genes are relatively recent, i.e., the events leading to them have occurred within well recognized lineages, such as chlorophyte algae [135], euglenozoans [229] or eudicots [145], and are therefore phylogenetically meaningful [229] and relatively straightforward to interpret (but *cf.* [146; 147]).

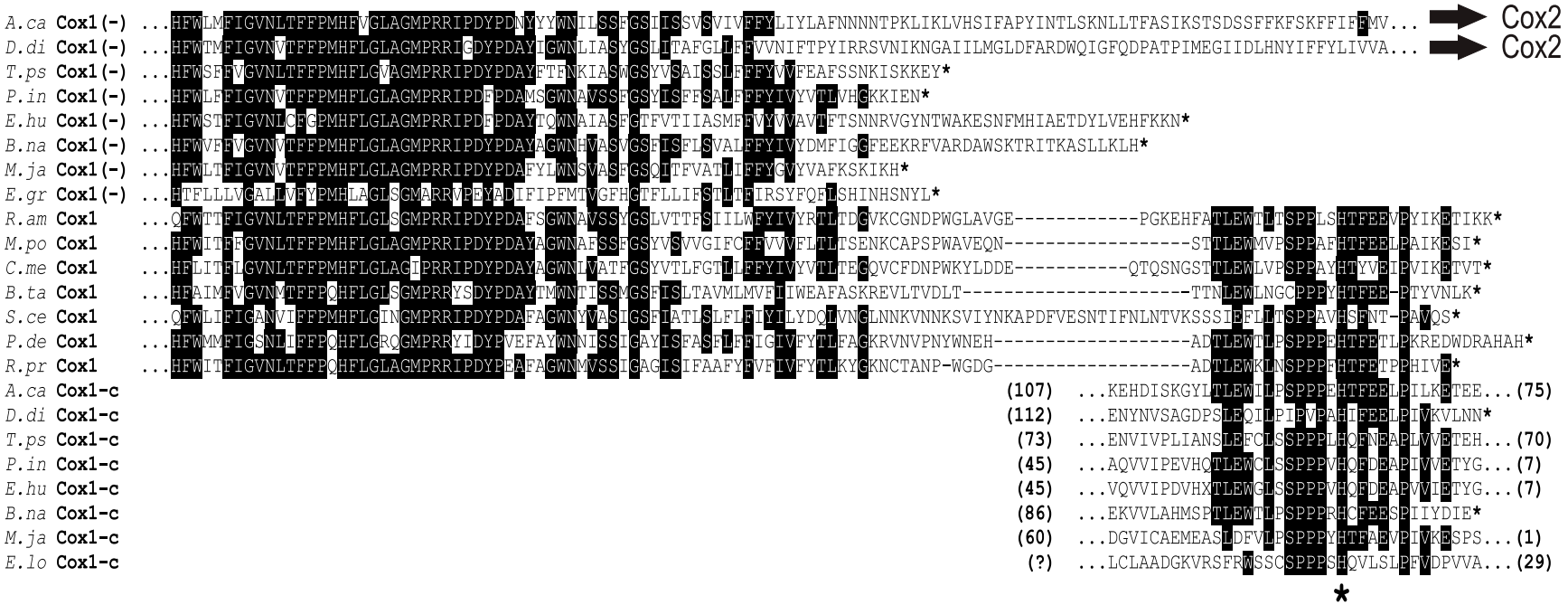
I describe here a previously undetected split in the C-terminal region of the cytochrome *c* oxidase subunit 1 gene (*cox1*), a fission that has a punctate distribution within eukaryotes. I suggest that this distribution coupled with the phylogenetic breadth covered by this unique molecular event (which is evident in at least four of six currently proposed [111] eukaryotic supergroups) reflects an ancient partial gene transfer that was differentially activated in diverse eukaryotic lineages.

5.3 Results and Discussion

During an ongoing investigation of the mitochondrial proteome of the amoeboid protozoon *Acanthamoeba castellanii*, a high-confidence peptide that matched an inferred 24.8 kDa nucleus-encoded protein (based on translated EST sequence) was identified by tandem mass spectrometry. BLASTp [189] analyses demonstrated that a short (~25 amino acid) region of this protein exhibits strong similarity to a conserved motif present within the C-terminal region of Cox1 proteins from α-proteobacteria and minimally-divergent mtDNAs (e.g., *Reclinomonas*, *Marchantia*; see Figure 5.1), whereas the remainder of the inferred protein sequence showed no strong similarity to proteins in available sequence databases. Multiple protein alignments established that the mtDNA-encoded 'fused' Cox1/2 of *Acanthamoeba* [178; 194] lacks this C-terminal portion, which I consider to be an ancestral part of a covalently continuous (i.e., conventional) Cox1 due to its presence in

this form in α-proteobacteria and *Reclinomonas* mitochondria [37]. I refer to the nuDNA-encoded *Acanthamoeba* protein homologous to the C-terminal portion of full length Cox1 as 'Cox1-c' and the truncated mtDNA-encoded Cox1 homolog as 'Cox1(-)'.

Figure 5.1 Partial alignment of Cox1, Cox1(-) and Cox1-c protein sequences.The C-terminal region of Cox1 was aligned with inferred Cox1-c protein sequences. In this alignment, '…' denotes that the sequence continues, while '*' represents the stop codon. The large * demarcates H503 of bovine Cox1, which is 100% conserved in Cox1-c. Numbers adjacent to Cox1-c sequences indicate the number of N-terminal and C-terminal amino acids omitted from the alignment. Species abbreviations: *A.ca*, *Acanthamoeba castellanii*; *D.di*, *Dictyostelium discoideum*; *T.ps*, *Thalassiosira pseudonana*; *P.in*, *Phytophthora infestans*; *E.hu*, *Emiliania huxleyi*; *B.na*, *Bigelowiella natans*; *M.ja*, *Malawimonas jakobiformis*; *E.gr*, *Euglena gracilis*; *R.am*, *Reclinomonas americana*; *M.po*, *Marchantia polymorpha*; *C.me*, *Cyanidioschyzon merolae*; *B.ta*, *Bos taurus*; *S.ce*, *Saccharomyces cerevisiae*; *P.de*, *Paracoccus denitrificans*; *R.pr*, *Rickettsia prowazekii*; *E.lo*, *Euglena longa*. GenBank accession numbers are provided in Supplemental Table 5.1.


A.ca Cox1(-) ...HFWLMFIGVNLTFFPMHFVGLAGMPRRIPDYPDNYYYWNILSSFGSLISVSVIVFFYLIYLAFNNNNTPKLIKLVHSIFAPYINTLSKNLLTFASIKSTSDSSFFKFSKFFIFFMV...
D.di Cox1(-) ...HFWTMFIGVNVTFFPMHFLGLAGMPRRIGDYPDAYIGWNLIASYGSLITAFGILFFVVNIFTPYIRRSVNIKNGAIILMGLDFARDWQIGFQDPATPIMEGIIDLHNYIFFYLIVVA...
T.ps Cox1(-) ...HFWSFFVGVNLTFFPMHFLGVAGMPRRIPDYPDAYFTFNKIASWGSYVSAISSLFFFYVVFEAFSSNKISKKEY*
P.in Cox1(-) ...HFWLFFIGVNVTFFPMHFLGLAGMPRRIPDFPDAMSGWNAVSSFGSYISFFSALFFFYIVYVTLVHGKKIEN*
E.hu Cox1(-) ...HFWSTFIGVNLCFGPMHFLGLAGMPRRIPDFPDAYIQWNAIASFGTFVTIIASMFFVYVVAVTFTSNNRVGYNTWAKESNFMHIAETDYLVEHFKKN*
B.na Cox1(-) ...HFWVFFVGVNVTFFPMHFLGLAGMPRRIPDYPDAYAGWNHVASVGSFISFLSVALFFYIVYDMFIGGFEEKRFVARDAWSKTRITKASLLKLH*
M.ja Cox1(-) ...HFWLTFIGVNVTFFPMHFLGLAGMPRRIPDYPDAFYLWNSVASFGSQITFVAILIFFYGVYVAFKSKIKH*
E.gr Cox1(-) ...HTFLLLVGALLVFYPMHLAGLSGMARRVPEYADIFIPFMTVGFHGTFLLIFSTLTFIRSYFQFLSHINHSNYL*
R.am Cox1 ...QFWTTFIGVNLTFFPMHFLGLSGMPRRIPDYPDAFSGWNAVSSYGSLVTTFSIILWFYIVYRTLTDGVKCGNDPWGLAVGE----------PGKEHFATLEWTLTSPPLSHTFEEVPYIKETIKK*
M.po Cox1 ...HFWITFFGVNLTFFPMHFLGLAGMPRRIPDYPDAYAGWNAFSSFGSYVSVVGIFCFFVVVFLTLTSENKCAPSPWAVEQN----------STTLEWMVPSPPAFHTFEELPAIKESI*
C.me Cox1 ...HFLITFLGVNLTFFPMHFLGLAGIPRRIPDYPDAYAGWNLVATFGSYVTLFGTLLFFYIVYVTLTEGQVCFDNPWKYLDDE----------QTQSNGSTTLEWLVPSPPAYHTYVEIPVIKETVT*
B.ta Cox1 ...HFAIMFVGVNMTFFPQHFLGLSGMPRRYSDYPDAYTMWNTISSYGSFISLTAVMLMVFIIWEAFASKREVLTVDLT----------TTNLEWLNGCPPPYHTFEE-PTYVNLK*
S.ce Cox1 ...QFWLIFIGANVIFFPMHFLGINGMPRRIPDYPDAFAGWNYVASIGSFIATLSLFLFIYILYDQIVNGLNNKVNNKSVIYNKAPDFVESNTIFNLNTVKSSSIEFLLTSPPAVHSENT-PAVQS*
P.de Cox1 ...HFWMMFIGSNLIFFPQHFLGRQGMPRRYIDYPVEFAYWNNISSIGAYISFASFLFFIGIVFYTLFAGKRVNVPNYWNEH----------ADTLEWTLPSPPPEHTFETLPKREDWDRAHAH*
R.pr Cox1 ...HFWITFIGVNLTFFPQHFLGLAGMPRRIPDYPEAFAGWNMVSSIGAGISIFAAFYFVFIVFYTLKYGKNCTANP-WGDG----------ADTLEWKLNSPPPFHTFETPPHIVE*
A.ca Cox1-c (107) ...KEHDISKGYLTLEWILPSPPPEHTFEELPILKETEE...(75)
D.di Cox1-c (112) ...ENYNVSAGDPSLEQILPIPVPAHIFEELPIVKVLNN*
T.ps Cox1-c (73) ...ENVIVPLIANSLEFCLSSPPPLHQFNEAPLVVETEH...(70)
P.in Cox1-c (45) ...AQVVIPEVHQTLEWCLSSPPPVHQFDEAPIVVETYG...(7)
E.hu Cox1-c (45) ...VQVVIPDVHXTLEWGLSSPPPVHQFDEAPVVIETYG...(7)
B.na Cox1-c (86) ...EKVVLAHMSPTLEWTLPSPPPRHCFEESPIIYDIE*
M.ja Cox1-c (60) ...DGVICAEMEASLDFVLPSPPPYHTFAEVPIVKESPS...(1)
E.lo Cox1-c (?) ...LCLAADGKVRSFRWSSCSPPPSHQVLSLPFVDPVVA...(29)
Cox2
Cox2
*

The *Acanthamoeba* Cox1-c sequence was used to query the proteome of a related amoebozoon, *Dictyostelium discoideum* [180] via BLASTp. This search identified a homolog that is similar to Cox1-c in the region of the conserved motif of conventional Cox1, although the degree of similarity is not as great as in the case of the *Acanthamoeba* Cox1-c. The *Dictyostelium* Cox1-c homolog had already been identified as the largest nuDNA-encoded constituent of the *Dictyostelium* cytochrome oxidase complex, and was previously named CoxIV [242]. However, the authors failed to recognize the similarity to the C-terminus of conventional Cox1 (likely because the motif is not as well conserved in *Dictyostelium*), and suggested instead that this protein might be related to subunit VI of yeast and subunit Va of human cytochrome oxidase [387]. Together, these results suggest that the novel *Acanthamoeba* Cox1-c protein is a *bona fide* component of the cytochrome oxidase complex, and that *Dictyostelium* CoxIV is homologous to the C-terminal portion of conventional Cox1.

I generated a phylogenetically-broad alignment in order to examine the C-terminal portion of mtDNA-encoded Cox1 across the diversity of eukaryotes (Figure 5.1). The alignment revealed that the ancestral C-terminal motif of Cox1 is missing from the mtDNA of a wide variety of eukaryotes (i.e., they encode Cox1(-) rather than Cox1), including species from four of the six currently proposed eukaryotic supergroups. In contrast to *Acanthamoeba* and *Dictyostelium*, where the continuous nature of the Cox1/2 ORF [178] obscures the actual C-terminus of mature Cox1(-), other inferred Cox1(-) sequences are clearly truncated.

I hypothesized that separate, nuDNA-encoded homologs of Cox1-c might be present in those eukaryotes with a mtDNA-encoded Cox1(-). Searches of EST, nuclear

genomic, and inferred protein sequence databases allowed us to identify nuDNA-encoded homologs of Cox1-c in many of the species lacking the motif in mtDNA-encoded Cox1(-) (Figure 5.1). Cox1-c homologs exist in the supergroups Amoebozoa, Chromalveolata, Rhizaria and Excavata (Table 5.1), all of which also encode a Cox1(-) in their mtDNA. Although not all Cox1-c homologs are as similar to the C-terminus of conventional Cox1 as is *Acanthamoeba* Cox1-c, a consensus motif S-P-P-P-X-H (the H 100% conserved) aids in the identification of *bona fide* Cox1-c homologs (Figure 5.1 - * denotes conserved H). As expected, the mitochondrial targeting peptide (mTP) program TargetP [90] predicts mitochondrial localization for Cox1-c homologs having complete N-terminal sequence, at a probability of $\geq$ 0.65.

Table 5.1 The structure of Cox1 across eukaryotes. Organisms are categorized according to eukaryotic supergroup membership [111]. 'Cox1' refers to intact (conventional) Cox1 in the mtDNA, while 'Cox1(-)' specifies truncated mtDNA-encoded Cox1. 'Cox1-c' is a nuDNA-encoded protein homologous to the C-terminal portion of Cox1. GenBank accession numbers are provided in Supplemental Table 5.1.

Eukaryotic supergroup	**Organism**	**mtDNA Cox1**	**Cox1-c**	**Comments**
Excavata	*Reclinomonas*	Cox1	?	
	Naegleria	Cox1	No	
	Euglena	Cox1(-)	Yes	
	Trypanosoma	Cox1(-)	?	
	Malawimonas	Cox1(-)	Yes	
Amoebozoa	*Acanthamoeba*	Cox1(-)	Yes	Continuous Cox1/2
	Dictyostelium	Cox1(-)	Yes	Continuous Cox1/2
	Physarum	Cox1(-)	Yes	
	Hyperamoeba	?	Yes	
Rhizaria	*Bigelowiella*	Cox1(-)	Yes	
Chromalveolata	*Tetrahymena*	Cox1(-)	?	
	Thalassiosira	Cox1(-)	Yes	
	Guillardia	Cox1	No	
	Cafeteria	Cox1(-)	?	
	Phytophthora	Cox1(-)	Yes	
	Emiliania	Cox1(-)	Yes	
	Plasmodium	Cox1(-)	?	
Opisthokonta	*Bos*	Cox1	No	
	Monosiga	Cox1	No	
	Saccharomyces	Cox1	No	
Plantae	*Arabidopsis*	Cox1	No	
	Marchantia	Cox1	No	
	Chlamydomonas	Cox1	No	
	Porphyra	Cox1	No	

The retention of Cox1-c coding sequence in the nuDNA of many eukaryotes after it was lost from the mtDNA argues strongly for the functional importance of this region of Cox1. None of the residues required for coordination of metal clusters is located at the extreme C-terminus of Cox1; however, crystal structures of the bovine enzyme in reduced and oxidized states demonstrate that H503 (which corresponds to the 100% conserved H residue of Cox1-c) is part of an important controlling site for entrance of protons into the dioxygen reduction pathway [388]. Because proton pumping and dioxygen reduction are fundamental functions of cytochrome oxidase, these data suggest that Cox1-c (and the C-terminal portion of full-length Cox1) serves a critical function in the cytochrome oxidase complex, thereby explaining why this part of the protein is not easily lost.

Interestingly, full-length Cox1 proteins are interspersed with split Cox1 proteins within eukaryotic supergroups (Table 5.1). For instance, within Excavata, Cox1 is full length in the jakobids (*Reclinomonas*) and heteroloboseans (*Naegleria*) but split in euglenozoans (*Euglena*; *Trypanosoma* and *Diplonema* encode Cox1(-), but corresponding nuDNA-encoded proteins could not be identified). Within Chromalveolata, haptophytes such as *Emiliania* have Cox1(-) proteins and nuDNA-encoded Cox1-c, whereas related cryptophytes such as *Guillardia* have full-length mtDNA-encoded Cox1 proteins. Thus far, the only supergroups in which the described Cox1 fission has not been detected are the Opisthokonta (including fungi, animals and their unicellular ancestors) and Plantae (comprising red algae, green algae and land plants). Conversely, all amoebozoons examined so far have a nuDNA-encoded Cox1-c and mtDNA-encoded Cox1(-). From the perspective of parsimony, the widespread yet

punctuate distribution of the described *cox1* split suggests that sequence encoding the small C-terminal domain of Cox1 was transferred to the nuclear genome early in eukaryotic evolution and differentially lost/activated in different lineages. Alternatively, it is possible that the sequence encoding the C-terminal region of Cox1 (or *cox1* in its entirety) was transferred to the nucleus many times independently. Although this scenario appears less likely, our finding could reflect many parallel failed attempts at functional relocation of *cox1* to the nucleus (perhaps due to the high hydrophobicity of the intact Cox1 protein), with only a sequence encoding this small, hydrophilic C-terminal end acquiring the necessary expression and targeting signals.

5.4 Acknowledgements

RMRG was supported by a CGSD from the Natural Sciences and Engineering Research Council (NSERC) and a Predoctoral Scholarship from the Killam Trusts. This work was supported by operating grant MOP-4124 from the Canadian Institutes of Health Research.

Chapter 6 Discussion

6.1 The *A. castellanii* Mitochondrial Proteome

6.1.1 Overview

The work presented in this thesis, both biochemical and bioinformatic in nature, has expanded our knowledge of the function and evolution of mitochondria. MS/MS analyses of highly enriched *A. castellanii* whole mitochondria and mitochondrial fractions have together uncovered nearly 750 nuDNA- and mtDNA-encoded proteins, including constituents of all submitochondrial compartments. In addition, bioinformatic interrogations of *A. castellanii* sequence databases have augmented our understanding of various metabolic pathways, protein classes and protein complexes present in *A. castellanii* mitochondria, yielding, in combination with MS/MS identifications, in excess of 850 putative mitochondrial proteins. These results compare well with MS/MS-based mitochondrial proteomic investigations of other eukaryotes, such as *C. reinhardtii* (349 proteins [25]), *A. thaliana* (416 proteins [21]) and *T. thermophila* (573 proteins [23]), and even approach the coverage of animal (615 [18], 951 [24], 1117 [26] proteins) and yeast (750 [19], 749 [22] proteins) mitochondrial proteomes.

6.1.2 Limitations of this Study

Although, this study has yielded significant new information about the mitochondria of *A. castellanii*, several limitations should be noted. Firstly, the deeply-sequenced (translated) EST dataset that was queried with tandem mass spectra does not represent all nuDNA-encoded proteins in their entirety; thus, there was effectively a double bias against certain low abundance proteins. As a result, it is likely that peptides

from a number of authentic mitochondrial proteins were detected by MS/MS, but remain unassigned due to the absence of the corresponding sequence from our database. Ideally, a high-quality predicted protein dataset, based on a completely sequenced nuclear genome, should be used in conjunction with EST data; however, these tools are not yet available in the case of *A. castellanii*. Secondly, the identification of a number of contaminating, (putatively) non-mitochondrial proteins limits our confidence in the mitochondrial localization of novel 'hypothetical' proteins lacking classical mTPs. To varying extents, non-mitochondrial contaminants have been reported in every major mitochondrial proteomics effort to date. Encouragingly, however, few 'hypothetical' proteins identified in my analysis contained sequence information that suggests localization to other subcellular compartments, indicating that most contaminants represent high-abundance species that also dominate the proteomes of other organelles. Furthermore, it is quite likely that certain 'contaminants' are actually functionally associated with mitochondria [196], or dual-targeted to mitochondria and other compartments [109].

6.1.3 Conclusions

In general, my results support the emerging picture of the mitochondrial proteome as a relatively labile entity of a mosaic evolutionary origin [23; 25; 63], including many eukaryote- and lineage-specific additions, built around an α-proteobacterial 'core' consisting of predominantly bioenergetic and translational proteins. MS/MS analyses uncovered many proteins characteristic of classical mitochondria, including those of the PDH complex, the TCA cycle, the ETC (discussed below), expression and translation of mtDNA-encoded proteins, Fe-S cluster biosynthesis, amino acid and fatty acid

metabolism, vitamin and cofactor biosynthesis, and protein import, among others. In this sense, *A. castellanii* mitochondria contain many signatures of 'classical' mitochondria derived from studies of model eukaryotes.

Intriguingly, however, various other non-canonical features of *A. castellanii* mitochondria demonstrate functional and evolutionary versatility, thereby highlighting the value of exploring and understanding protist mitochondrial proteomes. For instance, the DNA polymerase found in *A. castellanii* mitochondria is related to bacterial PolI, rather than the phage-like Polγ that replicates mtDNA in Opisthokonta [281], underscoring the impact of non-orthologous gene replacement in mitochondrial proteome evolution. Moreover, from a functional perspective, peptides from all PDH subunits, along with PFO [203], HydF and a putative ASCT, were detected by MS/MS, suggesting that mitochondria derived from aerobically grown *A. castellanii* contain proteins associated with both aerobic and anaerobic pyruvate metabolic pathways. It may be that *A. castellanii* constitutively expresses these anaerobiosis-related proteins in order to be 'ready' for a shift to low-oxygen conditions, as suggested for *Euglena* [389]. In this sense, the functions of *A. castellanii* mitochondria may more closely reflect those from ecologically similar organisms (e.g., *E. gracilis* [390] and *C. reinhardtii* [25; 84] mitochondria employ multiple pyruvate metabolic pathways) than closer evolutionary relatives, such as *D. discoideum*.

6.1.4 Future Directions

Comparative mitochondrial proteomics is still in its early stages; although consistent functional and evolutionary trends are already emerging, comprehensive mitochondrial proteomic investigations remain limited to very few organisms. Indeed,

mitochondrial proteome surveys need to be undertaken for a large number of protists, if we wish to achieve the same sorts of robust and nuanced conclusions that comparative genomics has provided with respect to the origin and evolution of mtDNA [1; 15]. It will be particularly important to understand the mitochondrial proteomes of protists occupying important ecological and evolutionary niches. For instance, proteomic investigations of organisms like *M. brevicollis* and *R. americana* might permit further understanding of the ancestral animal and eukaryotic mitochondrial proteomes, respectively. Additionally, MS/MS characterization of *E. gracilis* and *N. gruberi* mitochondria would surely provide a further appreciation of the role mitochondria play in adaptation to life in both aerobic and anaerobic conditions.

Along with wholesale evolutionary comparisons of mitochondrial protein compendiums from diverse eukaryotes, it is imperative to use proteomic data as a starting point for more targeted investigations of organellar function. Mitochondrial proteomic investigations present a bewildering number of such platforms, and it can be quite difficult to choose among them. One particularly intriguing avenue of study, with respect to *A. castellanii* mitochondria, relates to the proteomic consequences of shifting from aerobic to anaerobic conditions (and *vice versa*). Although the study of this transition will require the development of methods for purification of mitochondria from *A. castellanii* exposed to low O_2 conditions, various proteomic strategies, including comparisons of mitochondrial protein profiles from aerobically and anaerobically grown cells after 2D IEF/SDS-PAGE, and quantitative proteomics using isotope-coded affinity tags, will provide a bird's-eye view of how the mitochondrial proteome responds to fluctuating environmental conditions.

6.2 The *A. castellanii* Mitochondrial ETC

6.2.1 Overview

I have carried out an analysis of the composition, structure and evolution of the *A. castellanii* mitochondrial ETC using a combination of BN-PAGE [140] coupled to MS/MS, as well as sensitive bioinformatic searches. This investigation has yielded multiple novel insights into the evolution of the ETC both in *A. castellanii* and eukaryotes in general.

All 5 ETC complexes that participate in coupled oxidative phosphorylation (CI-CV) were isolated via BN-PAGE; 4 of these were identified by in-gel enzyme activity assays [142], which permitted some appreciation of their size and higher-order structure. In general, most complexes are similar in size to their counterparts in other eukaryotes; however, an especially abundant, large and stable multimeric CV was also demonstrated.

6.2.2 Limitations of BN-PAGE

BN-PAGE has proven to be a useful tool in elucidating the structure and composition of the *A. castellanii* ETC; however, it is limited in several ways. For instance, certain complexes are not as free from contamination as they would be using certain other approaches, including affinity- and immunological-based approaches, making results somewhat more difficult to interpret. For instance, the low abundance of CI has obscured the detection of certain components by MS/MS; higher abundance proteins (e.g., from partially dissociated CV) are frequently detected, thereby limiting the coverage of CI and potentially failing to detect novel CI proteins. Similarly, it was not possible to determine whether any lineage-specific CII proteins exist, as a large number of other (presumably multimeric) mitochondrial proteins co-migrate with CII due to its

small size. BN-PAGE was found to be most useful for determining the composition of large, abundant and well-separated complexes, especially CIII and CV.

6.2.3 Composition and Evolution of the A. castellanii ETC

Based on combined MS/MS and bioinformatic analyses, I suggest that ≥ 42, 4, 10, ~10 and 18 subunits constitute CI-CV, respectively. Most ETC subunits conserved in other eukaryotic groups were identified, including all of the ancestral components presumably inherited from the eubacterial ancestor of mitochondria and a large proportion of conserved supernumerary subunits [75]. Importantly, however, my analyses provided novel perspectives as well. Notably, I detected two γ-type carbonic anhydrases (γCA) [182], AcCa1 and AcCa2, via MS/MS analysis of CI. As described above, the presence of γCAs in CI from land plants and green algae, along with their complete absence in animals and fungi, led other investigators to conclude that they were relatively recent additions to CI in Plantae [75; 127]. In contrast, I demonstrated that γCAs are CI subunits in *A. castellanii*. Furthermore, the detection of three such proteins in a *T. thermophila* ETC supercomplex composed of CI and CIII (and comigrating with CV_2; my unpublished results, [139]) and the presence of N-terminal mTPs in homologs from many eukaryotes, argue strongly that these proteins are ancestral components of mitochondrial CI that were lost in opisthokonts. These results emphasize the value of incorporating data from a broad phylogenetic sampling prior to drawing far-reaching conclusions.

Whereas the sequences of γCAs are relatively well conserved, more sensitive bioinformatic analyses, especially PSI-BLAST [189] and HMMER [191], were crucial in the 'search' for proteins that are widely conserved in eukaryotes, but unidentifiable via

standard BLAST searches in *A. castellanii* sequence databases (e.g., CI subunits MWFE and ASHI, as well as the CV subunit Atpd) and for inferences of previously unappreciated homologies (e.g., yeast Atpi/j and a 6-kDa plant CV subunit). It is important to note that the availability of databases containing homologs from a wide variety of eukaryotic groups has made these analyses feasible. Specifically, employing protist-encoded homologs of various ETC proteins from animals, fungi and plants in these searches has helped 'bridge the gap' between these various groups: e.g., while the relationship between a plant and fungal protein might not be evident, each may be somewhat more similar to a given protist homolog, thereby functionally and evolutionarily linking the plant and fungal proteins. Additionally, the MS/MS-based identification of a protein as a component of a given ETC complex greatly increases the confidence associated with conclusions drawn using tools such as HMMER. On their own, HMMER and tools like it may not be suitable for the *de novo* identification and annotation of very divergent 'missing' proteins from a sequence database, but they can be extremely powerful when used in conjunction with the knowledge that the top-scoring protein match from the sequence database is an authentic component of the complex that is 'missing' the subunit in question.

In general, my results support and extend previous analyses that suggest many of the ETC complex proteins lacking bacterial homologs are broadly distributed across the domain Eucarya [75]. As noted above, however, my analyses indicate that the ancestral mitochondrial CI contained even more subunits than previously inferred [75]. Moreover, all *A. castellanii* CII, CIII and CIV proteins and most CV proteins identified have homologs in other eukaryotic groups, although it should be noted that this trend likely does not hold

for more 'divergent' organisms [137-139]. My conclusions have important implications for the evolution of mitochondria; I infer that the vast majority of ETC components were present in the last common ancestor of eukaryotes, suggesting that the considerable increase in complexity of the mitochondrial ETC relative to bacteria happened relatively quickly, prior to the divergence of the major eukaryotic groups. Furthermore, my results strongly suggest that a further proportion of *A. castellanii* lineage-specific mitochondrial proteins were early eukaryotic inventions, but are simply too divergent to confidently annotate using standard bioinformatic techniques. The isolation and characterization of various other mitochondrial protein complexes, such as mitoribosomes and the protein import machinery, from a broad sampling of eukaryotic diversity will be important and may reveal further hitherto unappreciated homologies.

6.2.5 Future Directions

Whereas this work has characterized the composition of *A. castellanii* ETC complexes extensively, further work is required to understand the specific roles of various proteins. For instance, we still do not understand how CV achieves such a stable dimeric/multimeric structure; further dissection of CV into distinct subcomplexes, as done in *C. reinhardtii* [106; 255] and *A. thaliana* [108], would provide more precise protein-protein interaction data. Furthermore, the use of single-particle electron microscopy (EM) [115; 117] could provide information about the relative association of CV monomers and whether this interaction is mediated by the F_1 or F_o sector (or both). Linear sucrose density gradient centrifugation effectively separates *A. castellanii* CV from other mitochondrial complexes, so these experiments are quite feasible. Similarly, single-particle EM studies of CI structure will be very informative, although methods for

improving yields for CI will have to be developed. Although γCAs are now established components of *A. castellanii* CI, and we expect that they serve approximately the same function as in land plants and green algae, it is not known whether they form a distinct CI domain [115].

6.3 Split Mitochondrial Proteins

6.3.1 Overview

Relatively few examples of fissioned genes encoding split mitochondrial proteins have been reported in the scientific literature; in this thesis, I have presented two novel cases. The split proteins that I have described are: 1) SdhB [229] (CII; consisting of SdhB-n and SdhB-c), present in *E. gracilis* and trypanosomatids and 2) Cox1 [177] (CIV; consisting of Cox1(-) and Cox1-c), an arrangement present in a punctate fashion in wide variety of eukaryotes. In the case of euglenozoan SdhB-n and SdhB-c, both halves are encoded in nuDNA, translated on cytosolic ribosomes, and imported into mitochondria. In contrast, Cox1-c is nuDNA-encoded whereas Cox1(-) is invariably encoded in mtDNA, as is full-length Cox1. These examples have provided insights into the structure of split proteins and have posed further evolutionary questions about the process of EGT and the utility of rare genomic changes as phylogenetic tools.

Although my investigations did not provide direct biochemical evidence that split SdhB-n/SdhB-c remain as separate polypeptides *in vivo* (as opposed to being post-translationally fused), an independent and simultaneous investigation reported distinct SdhB-n and SdhB-c proteins from purified *T. cruzi* CII [137]. Similarly, I have not shown directly that *A. castellanii* Cox1-c and Cox1(-) remain separate; however, MS/MS analyses of SDS-PAGE gel slices indicates that Cox1-c and Cox1(-) migrate at markedly

different molecular weights. In addition, mature *D. discoideum* Cox1-c and Cox1(-) exist as separate proteins [242].

6.3.2 Protein Structure and Function

One of the mysteries pertaining to split proteins relates to the mechanism by which they interact in *trans* such that function is not perturbed excessively. As noted above, in certain other cases, such as Nad11 in stramenopiles [156; 157], gene fissions have occurred in between distinct protein folding domains. Thus, it may be that gene fission is relatively well tolerated, as long as the protein domains are intact (i.e., proteins are able to fold properly) and able to interact with each other. In agreement with this expectation, the fission of euglenozoan *sdhB* has occurred in a way that preserves the two folding domains of this protein, including an N-terminal portion (containing the [2Fe-2S] cluster) that forms a fold similar to plant-like ferredoxins and a C-terminal portion (containing [3Fe-4S] and [4Fe-4S] clusters) reminiscent of bacterial ferredoxins [229]. In the case of *cox1-c*, it is not clear whether the fission occurred within a conserved domain (the conserved domain database depicts the entire Cox1 protein as a single domain), although the region of Cox1 corresponding to Cox1-c is far removed from any residues involved in binding heme or copper.

As in the case of split *cox2* (encoding CoxII a and CoxII b) in green algae [135] the C-terminal region of *E. gracilis* SdhB-n bears an ~35-amino acid extension relative to homologs in other species; similarly, the N-terminal region of SdhB-c has an N-terminal extension, although its exact length in the mature protein is obscured by the presence of a cleavable mTP. In the case of Cox1-c homologs, there appears to be an N-terminal extension in general, although, once again, the presence of an mTP obfuscates the nature

of the mature N-termini. In contrast, the C-termini of various Cox1(-) proteins tend to be truncated in relation to Cox1 proteins due to the absence of the Cox1-c region, although the mature C-terminus of *A. castellanii* Cox1(-) has not been definitively established, as Cox1(-) is encoded as part of a dicistronic ORF that includes the Cox2 coding region [178; 194]. While there are still no direct biochemical or bioinformatic data supporting the assertion that protein extensions aid in the interaction of split proteins, their existence in most split proteins suggests that they are somehow involved. Alternatively, in some cases, it might be suggested that other proteins in a complex mediate the interaction of protein halves. Interestingly, purified *T. cruzi* CII consists of an unusually large number of distinct polypeptides [137], 12, in comparison to 4 in most organisms studied to date. Could it be that the accumulation of supernumerary proteins that bind and stabilize SdhB in trypanosomatids somehow permitted the subsequent fission of *sdhB*?

Although the molecular functions of SdhB and Cox1 have been probed in great detail, my discoveries of these gene fissions have provided some novel and intriguing insights. The SdhB-n and SdhB-c proteins encoded by *E. gracilis* and trypanosomatids are perhaps the most divergent SdhB homologs reported yet, lacking a number of otherwise universally-conserved residues. All Cys residues required for coordination of Fe-S clusters are conserved, however, indicating that the proteins are functional; this conclusion is supported by the detection of SdhB-n and SdhB-c in *T. cruzi* CII [137]. While the identification of split SdhB has not suggested novel roles for SdhB *per se*, certain features of *E. gracilis* SdhB reminiscent of the homologous, Fe-S subunit of the anaerobically functioning *E. coli* fumarate reductase lead me to speculate that CII may also be employed as a fumarate reductase (Frd) in *E. gracilis*. This suggestion is

consistent with previous reports of 1) elevated levels of the low-potential electron carrier rhodoquinone under anaerobic conditions [389; 391], 2) a requirement for fumarate reductase activity in mitochondrial anaerobic wax ester fermentation [392], 3) insensitivity of *E. gracilis* fumarate reductase to prokaryotic-type Frd inhibitors (but not Sdh inhibitors) [391], and 4) the absence of a distinct fumarate reductase sequence from deeply-sequenced *E. gracilis* EST libraries. Ultimately, the fumarate reductase activity of *E. gracilis* mitochondria will have to be identified via biochemical assays, but, in combination with previous reports, these data suggest that CII may carry this activity in *E. gracilis*.

Cox1-c does not bind any of the metal cofactors directly involved in the catalytic mechanism of CIV. Muramoto *et al.* [388] have reported that His503 of bovine Cox1 – corresponding to the 100% conserved His residue of Cox1-c – forms a coupling site known as the H pathway that is utilized in the control of proton uptake into the dioxygen reduction pathway. Although other investigators have questioned whether the H pathway is a general feature of Cox1 [393], as it is apparently lacking in bacteria, my discovery that Cox1-c is virtually always encoded in nuDNA wherever mtDNA-encoded Cox1 is truncated argues strongly that this portion of Cox1 is plays an important role in the function of CIV. Intriguingly, a recent investigation of C-terminally truncated yeast Cox1 has uncovered another potential function role for the C-terminus of Cox1, controlling the synthesis of Cox1 itself [394]. Briefly, a yeast-specific protein, Mss51, serves two distinct functions with respect to Cox1 synthesis: 1) it binds to the 5′ UTR of *cox1* mRNA, serving as a translational activator, and 2) it interacts with the C-terminal region of the Cox1 polypeptide during early stages of CIV assembly. The sequestration of Mss51 by the C-terminal region of Cox1 in partially assembled CIV limits further

translational activation of *cox1* mRNA until Mss51 is released later in the assembly process, thereby ensuring that the levels of unassembled Cox1 remain relatively low [394]. Maintaining the correct stoichiometry of Cox1 is important, as accumulation of incompletely assembled CIV results in cellular peroxide sensitivity due to the pro-oxidant nature of the heme *a*-Cox1 intermediate [395]. It is unclear what effects the translocation of the C-terminal region of Cox1 (Cox1-c) to nuDNA might have, but it is tempting to speculate that Cox1-c might somehow serve a role in the coordination of nuDNA- and mtDNA-encoded CIV protein levels, at least in some lineages.

6.3.3 Split Proteins and Eukaryotic Phylogeny

It has been suggested that supposedly rare genomic changes [159], including gene fissions [146] and fusions [160; 161], provide invaluable molecular landmarks capable of providing deep evolutionary insights that are independent of the limitations of phylogenetic models. As described above, the use of split mitochondrial proteins for these purposes has already yielded highly controversial conclusions [146; 147; 165]. Whereas the fission of *sdhB* in *E. gracilis* and trypanosomatids [137; 229] – both members of the highly supported phylum Euglenozoa [373] – may represent a relatively recent event that specifically unites these groups, my discovery of fissioned *cox1* in numerous eukaryotic groups (Section 5.3) [177] has provided strong evidence that gene fissions should **not** be used as independent metrics of deep eukaryotic evolution.

Nucleus-encoded Cox1-c, present invariably along with mtDNA-encoded Cox1(-), exists in a wide variety of eukaryotes. Importantly, the distribution of Cox1-c does not agree with established evolutionary relationships among eukaryotic supergroups. For instance, the relation between Amoebozoa and Opisthokonta is strongly supported (they

are often considered as one supergroup, Unikonta), while both are only distantly related to Plantae [111]; Cox1 is bipartite in all amoebozoan species examined thus far, whereas it is intact in opisthokonts and members of Plantae. In sum, the fact that the co-occurrence of Cox1-c and Cox1(-) does not accurately mirror *established* phylogenetic relationships argues that it cannot be diagnostic of tenuous ancient evolutionary relationships, even though the location of the gene fission occurs at the same precise location of *cox1*.

The mechanism(s) by which the same *cox1* gene fission can occur in distantly related groups is poorly established, although at least two strong possibilities exist. Firstly, as indicated by the title of our publication [177] ("An ancient fission of mitochondrial *cox1*"), it is possible that an ancient gene duplication and transfer event happened in the ancestor of extant eukaryotes, with differential activation of *cox1-c* in various eukaryotic lineages after the radiation of the major groups. In this scenario, *cox1-c* was lost in certain lineages (and the ancestral, intact Cox1 was retained), whereas in other groups, Cox1-c functionally replaced the C-terminus of intact Cox1 and the portion of *cox1* encoding the C-terminal region was eventually jettisoned. Another possibility is that the fission of *cox1* has occurred in the same fashion numerous times in distinct eukaryotic lineages. Parsimony arguments would suggest that this is less likely, although the existence of strong biases could produce this result. To this end, I hypothesized that the highly hydrophobic nature of the Cox1 polypeptide might have prevented the functional transfer of intact *cox1* to the nucleus many times [177]. Protein hydrophobicity has been demonstrated to be a major impediment to mitochondrial protein import [396; 397] and has been proposed as a rationale for the retention of mtDNA in general [398]. Cox1-c, in contrast, corresponds to a relatively hydrophilic portion of Cox1, which represents a

much more amenable substrate for the protein import machinery. In this scenario, multiple independent transfers of complete or partial *cox1* duplicates to the nucleus in various lineages may have *converged* on a similar solution, the activation of only the 3′ portion of *cox1* that encodes a protein (Cox1-c) suitable for import into mitochondria. Importantly, the finding that Nad11 appears to have split multiple times in stramenopiles [156-158] suggests that strikingly similar gene fissions can occur multiple times. Although it is not possible to determine which of these scenarios is responsible for the distribution of Cox1-c proteins in modern-day eukaryotes, my recently discovered, unpublished result of a split Cox1 *within* Fungi favors the latter (convergent) hypothesis.

Regardless of the mechanism, the point remains that the co-occurrence of fissioned genes in disparate lineages should not be considered as reliable evolutionary markers, even though some probably are unique. In a similar vain, gene fusion events may also be quite susceptible to artifact, as analyses of MS-ICL (and ICL-MS) proteins suggest that fusions can occur independently, as well. Overall, my results suggest that 'rare genomic changes' may not be very rare at all and that the most parsimonious interpretations of gene fission events may hinder our attempts at elucidating the ancient affiliations of eukaryotic lineages.

Appendix A Copyright Agreement

OXFORD UNIVERSITY PRESS LICENSE
TERMS AND CONDITIONS

Jul 01, 2011

This is a License Agreement between Ryan MR Gawryluk ("You") and Oxford University Press ("Oxford University Press") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Oxford University Press, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	2676100958687
License date	May 25, 2011
Licensed content publisher	Oxford University Press
Licensed content publication	Molecular Biology and Evolution
Licensed content title	An Ancient Fission of Mitochondrial cox1:
Licensed content author	Ryan M.R. Gawryluk, Michael W. Gray
Licensed content date	01/01/2010
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	Comparative proteomics: studies on the composition and evolution of the mitochondrial proteome in eukaryotic microbes (protists)
Publisher of your work	n/a
Expected publication date	Aug 2011
Permissions cost	0.00 USD
Value added tax	0.00 USD
Total	0.00 USD
Total	0.00 USD
Terms and Conditions	

STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL

1. Use of the material is restricted to the type of use specified in your order details.

2. This permission covers the use of the material in the English language in the following territory: world. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.

3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it apply to quotations, images, artistic works etc that have been reproduced from other sources which may be part of the material to be used.

4. No alteration, omission or addition is made to the material without our written consent. Permission must be re-cleared with Oxford University Press if/when you decide to reprint.

5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society must be included in the credit line.

6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should there be a problem clearing these rights, please contact journals.permissions@oxfordjournals.org

7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.

8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press reserves the right to take any and all action to protect its copyright in the materials.

9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person without Oxford University Press's written permission.

10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing

Appendix B Supplemental Material

The manuscripts presented in this thesis refer to a large number of Supplemental Files, Tables and Figures. Due to the lengthy nature of certain files (especially FASTA files containing sequences of proteins identified by MS/MS), the data are not appended to the thesis itself (they will later be publically available on DalSpace). I present below titles and descriptions for the Supplemental Material referred to in the text.

Supplemental Figure 2.1 The dichotomy of Core1/MPP β in mitochondria. A ML reconstruction of Core1 (CIII) and MPP β homologs from various eukaryotes indicates that gene duplication events *within* established groups may account for the presence of distinct Core1 and MPP β proteins. MPPB, MPP β; MPPB/Core1, putative dual functional MPP β and Core1; Sce, *S. cerevisiae*; Ncr, *N. crassa*; Hsa, *H. sapiens*; Tth, *T. thermophila*; Ath, *A. thaliana*; Cre, *C. reinhardtii*; Aca, *A. castellanii*. Note that for *T. thermophila*, Core1 has been identified biochemically (my unpublished results), but the putative MPP β homolog is inferred from proteomic data and absence from isolated CIII. Bootstrap values (from 100 replicates) are presented as a measure of statistical support.

Supplemental Figure 2.2 In-gel cytochrome *c* oxidase and ATPase enzyme assays of *A. castellanii* mitochondria solubilized with different amounts of DDM (final concentrations depicted for each lane by "%"). ETC complexes were separated on a 4-12% linear gradient using a minigel setup and the entire gel was soaked in either cytochrome *c* oxidase or ATPase enzyme assay solution. Two distinct isoforms of CIV can be visualized over a wide range of DMM concentrations. Dimeric *A. castellanii* CV is extremely stable, but dissociates into a smaller subcomplex when treated with high detergent concentrations.

Supplemental Table 2.1 *A. castellanii* CV may contain a malate dehydrogenase isoform. DDM-solubilized *A. castellanii* mitochondrial protein complexes were separated by centrifugation on a linear, 10-40% sucrose gradient (described above); aliquots from various fractions were analyzed by BN-PAGE and several isoforms of CV (including putative tetrameric, dimeric, 820-kDa, 520-kDa and F_1 isoforms) were subjected to MS/MS analysis. In each case, aside from the F_1 isoform, malate dehydrogenase was confidently detected.

Supplemental File 2.1 Aca_mtDNA-encoded_MSMS. A FASTA file containing sequences of mtDNA-encoded proteins identified by MS/MS.

Supplemental File 2.2 Contaminants_MSMS. A FASTA file containing sequences of putative non-mitochondrial proteins identified by MS/MS.

Supplemental File 2.3 Aca_nuDNA-encoded_MSMS. A FASTA file containing sequences of mitochondrial proteins identified by MS/MS.

Supplemental File 2.4 Aca_nuDNA_mtDNA_mito_bioinf. A FASTA file containing sequences of putative mitochondrial proteins identified by bioninformatic interrogations.

Supplemental File 2.5 Aca_pyruvate_and_TCA_cycle. A FASTA file containing sequences of enzymes associated with pyruvate metabolism and the TCA cycle. Proteins identified by MS/MS and bioinformatics are presented.

Supplemental File 2.6 Aca_CI_and_associated. A FASTA file containing sequences of CI proteins and CI assembly proteins. Proteins identified by MS/MS and bioinformatics are presented.

Supplemental File 2.7 Aca_CII. A FASTA file containing sequences of CII proteins. Proteins identified by MS/MS are presented.

Supplemental File 2.8 Aca_CIII_and_associated. A FASTA file containing sequences of CIII proteins and CIII assembly proteins. Proteins identified by MS/MS are presented.

Supplemental File 2.9 Aca_CIV_and_associated. A FASTA file containing sequences of CIV proteins and CIV assembly proteins. Proteins identified by MS/MS and bioinformatics are presented.

Supplemental File 2.10 Aca_CV_and_associated. A FASTA file containing sequences of CV proteins and CV assembly proteins. Proteins identified by MS/MS are presented.

Supplemental File 2.11 Aca_branched_ETC. A FASTA file containing sequences of enzymes associated with the branched respiratory chain. Proteins identified by MS/MS are presented.

Supplemental File 2.12 Aca_lysine biosynthetic proteins. A FASTA file containing sequences of lysine biosynthetic proteins. Proteins identified by MS/MS and bioinformatics are presented.

Supplemental File 2.13 Aca_PPR repeat. A FASTA file containing sequences of pentatricopeptide repeat proteins. Proteins identified by MS/MS and bioinformatics are presented.

Supplemental File 2.14 Aca_translation-associated. A FASTA file containing sequences of translation-associated proteins. Proteins identified by MS/MS and bioinformatics are presented.

Supplemental File 2.15 Aca-mito-RP-MS-peptides. A FASTA file containing sequences of mitoribosomal proteins, with MS/MS-identified peptides mapped onto the sequences. Proteins identified by MS/MS and bioinformatics are presented.

Supplemental File 2.16 Aca_protein import. A FASTA file containing sequences of proteins associated with mitochondrial protein import. Proteins identified by MS/MS and bioinformatics are presented.

Supplemental File 2.17 Aca_protein folding and degradation. A FASTA file containing sequences of mitochondrial proteases. Proteins identified by MS/MS and bioinformatics are presented.

Supplemental File 2.18 Aca_FeS proteins. A FASTA file containing sequences of mitochondrial FeS cluster biosynthetic proteins. Proteins identified by MS/MS and bioinformatics are presented.

Supplemental Figure 3.1. Phylogenetically broad alignment of eukaryotic and prokaryotic γCAs. A multiple alignment of eukaryotic and prokaryotic γCAs including complete and partial sequences from diverse eukaryotic groups.

Supplemental Table 3.1. Table of CI and CI* proteins identified by tandem mass spectrometry. The table lists proteins identified in the 940 kDa (CI) and 820 kDa (CI*) complexes via MASCOT searches of a six-frame translation of a 454 EST database (N) and predicted proteins from the *Acanthamoeba* mitochondrial genome (M). Proteins with ion scores of 37 and above are reported from the 454 EST database while ion scores of 29 or higher are reported for inferred mtDNA-encoded proteins. Bolded entries highlighted in color indicate that the protein in question was identified in both CI and CI* complexes. The two categories designated 'Other Proteins' are assumed to be contaminants; in particular, CV components (ATP synthase subunits) appear to be exceptionally abundant mitochondrial proteins, as judged by the results of MS/MS analysis of whole mitochondria.

Supplemental Table 3.2. Table of accession numbers. The table compiles accession numbers and other relevant information for γCA homologs available in public databases.

Supplemental Table 3.3. Table of mitochondrial targeting probabilities. The table reports the probabilities of mitochondrial localization for γCA proteins (according to TargetP and MitoProtII). Bolded entries indicate proteins predicted to be mitochondrial with ≥ 70% probability.

Supplemental File 3.1. MS/MS peptide coverage. The peptides identified by MS/MS analysis in AcCa1 and AcCa2 are mapped onto the corresponding protein sequences. Each peptide is colour-coded according to whether it was identified only in the 940 kDa CI, only the 820 kDa CI*, or in both samples.

Supplemental File 3.2. Sequences of γCA proteins. The inferred protein sequences of γCA proteins are reported in fasta format. Proteins are designated as indicated in Supplemental Table 3.2.

Supplemental Figure 4.1. Phylogenetically broad alignment of the N-terminal portion of SdhB. The figure displays more extensive protein alignments of the N-terminal half of SdhB-n than are presented in Figure 4.2A. This alignment includes SdhB sequences from a phylogenetically broad collection of eukaryotes. Shading of columns represents at least 70% identity.

Supplemental Figure 4.2. Phylogenetically broad alignment of the C-terminal portion of SdhB. The figure displays more extensive protein alignments of the C-terminal half of SdhB-c than are presented in Figure 4.2B. See Supplemental Figure 4.1 for details.

Supplemental Figure 4.3. Maximum likelihood phylogenetic tree of concatenated SdhB-n and SdhB-c proteins. This maximum likelihood phylogenetic tree reconstruction demonstrates that the euglenozoan SdhB-n and SdhB-c proteins are orthologs of mitochondrial SdhB (as opposed to FrdB). Euglenozoan SdhB-n and SdhB-c protein sequences were concatenated and aligned with SdhB and FrdB sequences from other eukaryotes and prokaryotes. The alignments were edited and PHYML was used to reconstruct the phylogeny. The WAG amino acid substitution model was used, with no invariable sites, 8 substitution rate categories and an estimated Γ distribution parameter. Nonparametric bootstrap analyses (100) were performed.

Supplemental Table 4.1. Database accession numbers, annotation, gene location, protein size. GenBank accession numbers, TBestDB entry information, associated annotations, gene locations (nuclear or mitochondrial DNA), sizes of inferred proteins.

Supplemental Table 5.1. Cox1, Cox1(-) and Cox1-c accession numbers. The table compiles accession numbers for Cox1, Cox1(-) and Cox1-c homologs reported in Table 5.1 and Figure 5.1.

References

1. Gray, M. W., Burger, G. & Lang, B. F. (1999). Mitochondrial evolution. *Science* **283**, 1476-1481.

2. Roger, A. J. (1999). Reconstructing early events in eukaryotic evolution. *Am Nat* **154**, S146-S163.

3. Cavalier-Smith, T. (1983). A 6-kingdom classification and unified phylogeny. In *Endocytobiology II. Intracellular Space as Oligongenetic Ecosystem* (Schenk, H. E. A. & Schwemmler, W., eds.), pp. 1027-1034. De Gruyter, Berlin.

4. Keeling, P. J., Luker, M. A. & Palmer, J. D. (2000). Evidence from beta-tubulin phylogeny that microsporidia evolved from within the fungi. *Mol Biol Evol* **17**, 23-31.

5. Clark, C. G. & Roger, A. J. (1995). Direct evidence for secondary loss of mitochondria in *Entamoeba histolytica*. *Proc Natl Acad Sci U S A* **92**, 6518-6521.

6. Roger, A. J., Clark, C. G. & Doolittle, W. F. (1996). A possible mitochondrial gene in the early-branching amitochondriate protist *Trichomonas vaginalis*. *Proc Natl Acad Sci USA* **93**, 14618-14622.

7. Roger, A. J., Svard, S. G., Tovar, J., Clark, C. G., Smith, M. W., Gillin, F. D. & Sogin, M. L. (1998). A mitochondrial-like chaperonin 60 gene in *Giardia lamblia*: evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. *Proc Natl Acad Sci USA* **95**, 229-34.

8. Lindmark, D. G. & Müller, M. (1973). Hydrogenosome, a cytoplasmic organelle of the anaerobic flagellate *Tritrichomonas foetus*, and its role in pyruvate metabolism. *J Biol Chem* **248**, 7724-7728.

9. Müller, M. (1993). The hydrogenosome. *J Gen Microbiol* **139 (Pt 12)**, 2879-2889.

10. Akhmanova, A., Voncken, F., van Alen, T., van Hoek, A., Boxma, B., Vogels, G., Veenhuis, M. & Hackstein, J. H. P. (1998). A hydrogenosome with a genome. *Nature* **396**, 527-528.

11. Boxma, B., de Graaf, R. M., van der Staay, G. W. M., van Alen, T. A., Ricard, G., Gabaldón, T., van Hoek, A. H. A. M., Moon-van der Staay, S. Y., Koopman, W. J. H., van Hellemond, J. J., Tielens, A. G. M., Friedrich, T., Veenhuis, M., Huynen, M. A. & Hackstein, J. H. P. (2005). An anaerobic mitochondrion that produces hydrogen. *Nature* **434**, 74-79.

12. Mai, Z., Ghosh, S., Frisardi, M., Rosenthal, B., Rogers, R. & Samuelson, J. (1999). Hsp60 is targeted to a cryptic mitochondrion-derived organelle ("crypton") in the microaerophilic protozoan parasite *Entamoeba histolytica*. *Mol Cell Biol* **19**, 2198-205.

13. Tovar, J., Fischer, A. & Clark, C. G. (1999). The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*. *Mol Microbiol* **32**, 1013-21.

14. Martin, W. & Müller, M. (1998). The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37-41.

15. Gray, M. W., Lang, B. F. & Burger, G. (2004). Mitochondria of protists. *Annu Rev Genet* **38**, 477-524.

16. Lane, N. & Martin, W. (2010). The energetics of genome complexity. *Nature* **467**, 929-934.

17. Richly, E., Chinnery, P. F. & Leister, D. (2003). Evolutionary diversification of mitochondrial proteomes: implications for human disease. *Trends Genet* **19**, 356-362.

18. Taylor, S. W., Fahy, E., Zhang, B., Glenn, G. M., Warnock, D. E., Wiley, S., Murphy, A. N., Gaucher, S. P., Capaldi, R. A., Gibson, B. W. & Ghosh, S. S. (2003). Characterization of the human heart mitochondrial proteome. *Nat Biotechnol* **21**, 281-286.

19. Sickmann, A., Reinders, J., Wagner, Y., Joppich, C., Zahedi, R., Meyer, H. E., Schönfisch, B., Perschil, I., Chacinska, A., Guiard, B., Rehling, P., Pfanner, N. & Meisinger, C. (2003). The proteome of *Saccharomyces cerevisiae* mitochondria. *Proc Natl Acad Sci USA* **100**, 13207-13212.

20. Mootha, V. K., Bunkenborg, J., Olsen, J. V., Hjerrild, M., Wisniewski, J. R., Stahl, E., Bolouri, M. S., Ray, H. N., Sihag, S., Kamal, M., Patterson, N., Lander, E. S. & Mann, M. (2003). Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell* **115**, 629-640.

21. 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.

22. 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.

23. Smith, D. G. S., Gawryluk, R. M. R., Spencer, D. F., Pearlman, R. E., Siu, K. W. M. & Gray, M. W. (2007). Exploring the mitochondrial proteome of the ciliate protozoon *Tetrahymena thermophila*: direct analysis by tandem mass spectrometry. *J Mol Biol* **374**, 837-863.

24. Pagliarini, D. J., Calvo, S. E., Chang, B., Sheth, S. A., Vafai, S. B., Ong, S.-E., Walford, G. A., Sugiana, C., Boneh, A., Chen, W. K., Hill, D. E., Vidal, M., Evans, J. G., Thorburn, D. R., Carr, S. A. & Mootha, V. K. (2008). A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **134**, 112-123.

25. Atteia, A., Adrait, A., Brugière, S., Tardif, M., van Lis, R., Deusch, O., Dagan, T., Kuhn, L., Gontero, B., Martin, W., Garin, J., Joyard, J. & Rolland, N. (2009). A proteomic survey of *Chlamydomonas reinhardtii* mitochondria sheds new light on the metabolic plasticity of the organelle and on the nature of the α-proteobacterial mitochondrial ancestor. *Mol Biol Evol* **26**, 1533-1548.

26. Li, J., Cai, T., Wu, P., Cui, Z., Chen, X., Hou, J., Xie, Z., Xue, P., Shi, L., Liu, P., Yates, J. R. & Yang, F. (2009). Proteomic analysis of mitochondria from *Caenorhabditis elegans*. *Proteomics* **9**, 4539-4553.

27. Kurland, C. G. & Andersson, S. G. E. (2000). Origin and evolution of the mitochondrial proteome. *Microbiol Mol Biol Rev* **64**, 786-820.

28. Gabaldón, T. & Huynen, M. A. (2004). Shaping the mitochondrial proteome. *BBA-Bioenergetics* **1659**, 212-220.

29. Szklarczyk, R. & Huynen, M. (2010). Mosaic origin of the mitochondrial proteome. *Proteomics* **10**, 4012-4024.

30. Burger, G., Gray, M. W. & Lang, B. F. (2003). Mitochondrial genomes: anything goes. *Trends Genet* **19**, 709-716.

31. Gray, M. W. (1998). Rickettsia, typhus and the mitochondrial connection. *Nature* **396**, 109-110.

32. Sicheritz-Pontén, T., Kurland, C. G. & Andersson, S. G. E. (1998). A phylogenetic analysis of the cytochrome *b* and cytochrome *c* oxidase I genes supports an origin of mitochondria from within the Rickettsiaceae. *Biochim Biophys Acta* **1365**, 545-51.

33. Emelyanov, V. V. (2003). Common evolutionary origin of mitochondrial and rickettsial respiratory chains. *Arch Biochem Biophys* **420**, 130-41.

34. Fitzpatrick, D. A., Creevey, C. J. & McInerney, J. O. (2006). Genome phylogeneies indicate a meaningful a-proteobacterial phylogeny and support a grouping of the mitochondria with the Rickettsiales. *Mol Biol Evol* **23**, 74-85.

35. Thrash, J. C., Boyd, A., Huggett, M. J., Grote, J., Carini, P., Yoder, R. J., Robbertse, B., Spatafora, J. W., Rappé, M. S. & Giovannoni, S. J. (2011). Phylogenomic evidence for a common ancestor of mitochondria and the SAR11 clade. *Sci Rep* **1**.

36. Gray, M. W., Lang, B. F., Cedergren, R., Golding, G. B., Lemieux, C., Sankoff, D., Turmel, M., Brossard, N., Delage, E., Littlejohn, T. G., Plante, I., Rioux, P., Saint-Louis, D., Zhu, Y. & Burger, G. (1998). Genome structure and gene content in protist mitochondrial DNAs. *Nucleic Acids Res* **26**, 865-878.

37. Lang, B. F., Burger, G., O'Kelly, C. J., Cedergren, R., Golding, G. B., Lemieux, C., Sankoff, D., Turmel, M. & Gray, M. W. (1997). An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* **387**, 493-497.

38. Lang, B. F., Seif, E., Gray, M. W., O'Kelly, C. J. & Burger, G. (1999). A comparative genomics approach to the evolution of eukaryotes and their mitochondria. *J Eukaryot Microbiol* **46**, 320-6.

39. Feagin, J. E. (1992). The 6-kb element of *Plasmodium falciparum* encodes mitochondrial cytochrome genes. *Mol Biochem Parasitol* **52**, 145-8.

40. Andersson, S. G. E., Zomorodipour, A., Andersson, J. O., Sicheritz-Pontén, T., Alsmark, U. C. M., Podowski, R. M., Näslund, A. K., Eriksson, A.-S., Winkler, H. H. & Kurland, C. G. (1998). The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**, 133-140.

41. Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M. & Kawashima, K. (2002). Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA research* **9**, 189.

42. Herrmann, J. M. (2003). Converting bacteria to organelles: evolution of mitochondrial protein sorting. *Trends Microbiol* **11**, 74-9.

43. Neupert, W. & Herrmann, J. M. (2007). Translocation of proteins into mitochondria. *Annu Rev Biochem* **76**, 723-749.

44. Gabaldón, T. & Huynen, M. A. (2003). Reconstruction of the proto-mitochondrial metabolism. *Science* **302**, 609.

45. Gabaldón, T. & Huynen, M. A. (2007). From endosymbiont to host-controlled organelle: the hijacking of mitochondrial protein synthesis and metabolism. *PLoS Comput Biol* **3**, 2209-2218.

46. Esser, C., Martin, W. & Dagan, T. (2007). The origin of mitochondria in light of a fluid prokaryotic chromosome model. *Biol Lett* **3**, 180-184.

47. Gabaldón, T., Snel, B., Zimmeren, F., Hemrika, W., Tabak, H. & Huynen, M. A. (2006). Origin and evolution of the peroxisomal proteome. *Biology Direct* **1**, 8.

48. Hoepfner, D., Schildknegt, D., Braakman, I., Philippsen, P. & Tabak, H. F. (2005). Contribution of the endoplasmic reticulum to peroxisome formation. *Cell* **122**, 85-95.

49. Adams, K. L., Daley, D. O., Qiu, Y.-L., Whelan, J. & Palmer, J. D. (2000). Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature* **408**, 354-357.

50. Brennicke, A., Grohmann, L., Hiesel, R., Knoop, V. & Schuster, W. (1993). The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants. *FEBS Lett* **325**, 140-5.

51. Thorsness, P. E. & Fox, T. D. (1990). Escape of DNA from mitochondria to the nucleus in *Saccharomyces cerevisiae*. *Nature* **346**, 376-9.

52. Ricchetti, M., Fairhead, C. & Dujon, B. (1999). Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. *Nature* **402**, 96-100.

53. Stegemann, S. & Bock, R. (2006). Experimental Reconstruction of Functional Gene Transfer from the Tobacco Plastid Genome to the Nucleus. *Plant Cell* **18**, 2869-2878.

54. Hammen, P. K. & Weiner, H. (1998). Mitochondrial leader sequences: Structural similarities and sequence differences. *J Exp Zool* **282**, 280-283.

55. von Heijne, G. (1986). Mitochondrial targeting sequences may form amphiphilic helices. *EMBO J* **5**, 1335-1342.

56. Abe, Y., Shodai, T., Muto, T., Mihara, K., Torii, H., Nishikawa, S., Endo, T. & Kohda, D. (2000). Structural basis of presequence recognition by the mitochondrial protein import receptor Tom20. *Cell* **100**, 551-560.

57. Allison, D. S. & Schatz, G. (1986). Artificial mitochondrial presequences. *Proc Natl Acad Sci USA* **83**, 9011-9015.

58. Lucattini, R., Likić, V. A. & Lithgow, T. (2004). Bacterial proteins predisposed for targeting to mitochondria. *Mol Biol Evol* **21**, 652-8.

59. Ueda, M., Fujimoto, M., Arimura, S., Tsutsumi, N. & Kadowaki, K. (2008). Presence of a latent mitochondrial targeting signal in gene on mitochondrial genome. *Mol Biol Evol* **25**, 1791-1793.

60. Dolezal, P., Likic, V., Tachezy, J. & Lithgow, T. (2006). Evolution of the molecular machines for protein import into mitochondria. *Science* **313**, 314-318.

61. Hewitt, V., Alcock, F. & Lithgow, T. (2011). Minor modifications and major adaptations: The evolution of molecular machines driving mitochondrial protein import. *BBA-Biomembranes* **1808**, 947-954.

62. Gray, M. W., Burger, G. & Lang, B. F. (2001). The origin and early evolution of mitochondria. *Genome Biol* **2**, reviews1018.1-1018.5.

63. Karlberg, O., Canbäck, B., Kurland, C. G. & Andersson, S. G. E. (2000). The dual origin of the yeast mitochondrial proteome. *Yeast* **17**, 170-187.

64. Heazlewood, J. L., Howell, K. A., Whelan, J. & Millar, A. H. (2003). Towards an analysis of the rice mitochondrial proteome. *Plant Physiol* **132**, 230-242.

65. Vögtle, F. N., Wortelkamp, S., Zahedi, R. P., Becker, D., Leidhold, C., Gevaert, K., Kellermann, J., Voos, W., Sickmann, A., Pfanner, N. & Meisinger, C. (2009). Global analysis of the mitochondrial N-proteome identifies a processing peptidase critical for protein stability. *Cell* **139**, 428-439.

66. 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**, 868-83.

67. Zhaxybayeva, O. & Doolittle, W. F. (2011). Lateral gene transfer. *Curr Biol* **21**, R242-R246.

68. Henze, K., Badr, A., Wettern, M., Cerff, R. & Martin, W. (1995). A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution. *Proc Natl Acad Sci USA* **92**, 9122 - 9126.

69. Natsoulis, G., Hilger, F. & Fink, G. R. (1986). The HTS1 gene encodes both the cytoplasmic and mitochondrial histidine tRNA synthetases of *S. cerevisiae*. *Cell* **46**, 235-243.

70. Chatton, B., Walter, P., Ebel, J. P., Lacroute, F. & Fasiolo, F. (1988). The yeast VAS1 gene encodes both mitochondrial and cytoplasmic valyl-tRNA synthetases. *J Biol Chem* **263**, 52.

71. Shutt, T. E. & Gray, M. W. (2006). Twinkle, the mitochondrial replicative DNA helicase, is widespread in the eukaryotic radiation and may also be the mitochondrial DNA primase in most eukaryotes. *J Mol Evol* **62**, 588-599.

72. Shutt, T. E. & Gray, M. W. (2006). Bacteriophage origins of mitochondrial replication and transcription proteins. *Trends Genet* **22**, 90-95.

73. Jain, R., Rivera, M. C. & Lake, J. A. (1999). Horizontal gene transfer among genomes: The complexity hypothesis. *Proc Natl Acad Sci U S A* **96**, 3801-6.

74. Adams, K. L., Daley, D. O., Whelan, J. & Palmer, J. D. (2002). Genes for two mitochondrial ribosomal proteins in flowering plants are derived from their chloroplast or cytosolic counterparts. *Plant Cell* **14**, 931-943.

75. 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.

76. Brandt, U. (2006). Energy converting NADH:quinone oxidoreductase (complex I). *Annu Rev Biochem* **75**, 69-92.

77. 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 proteome. *Nucleic Acids Res* **35**, 4686-4703.

78. Szklarczyk, R. & Huynen, M. A. (2009). Expansion of the human mitochondrial proteome by intra-and inter-compartmental protein duplication. *Genome Biol* **10**, R135.

79. Panigrahi, A. K., Ogata, Y., Zíková, A., Anupama, A., Dalley, R. A., Acestor, N., Myler, P. J. & Stuart, K. D. (2009). A comprehensive analysis of *Trypanosoma brucei* mitochondrial proteome. *Proteomics* **9**, 434-450.

80. Mi-ichi, F., Yousuf, M. A., Nakada-Tsukui, K. & Nozaki, T. (2009). Mitosomes in *Entamoeba histolytica* contain a sulfate activation pathway. *Proc Natl Acad Sci USA* **106**, 21731-21736.

81. Jedelský, P. L., Doležal, P., Rada, P., Pyrih, J., Šmíd, O., Hrdý, I., Šedinová, M., Marcinčiková, M., Voleman, L., Perry, A. J., Beltrán, N. C., Lithgow, T. & Tachezy, J. (2011). The minimal proteome in the reduced mitochondrion of the parasitic protist *Giardia intestinalis*. *PLoS ONE* **6**, e17285.

82. 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* **20**, 1609-1620.

83. Brandina, I., Graham, J., Lemaitre-Guillier, C., Entelis, N., Krasheninnikov, I., Sweetlove, L., Tarassov, I. & Martin, R. P. (2006). Enolase takes part in a macromolecular complex associated to mitochondria in yeast. *BBA-Bioenergetics* **1757**, 1217-1228.

84. Atteia, A., van Lis, R., Gelius-Dietrich, G., Adrait, A., Garin, J., Joyard, J., Rolland, N. & Martin, W. (2006). Pyruvate formate-lyase and a novel route of eukaryotic ATP synthesis in *Chlamydomonas* mitochondria. *J Biol Chem* **281**, 9909-9918.

85. Meisinger, C., Sickmann, A. & Pfanner, N. (2008). The mitochondrial proteome: from inventory to function. *Cell* **134**, 22-24.

86. Gaston, D., Tsaousis, A. D. & Roger, A. J. (2009). Predicting proteomes of mitochondria and related organelles from genomic and expressed sequence tag data. *Method Enzymol* **457**, 21-47.

87. Marcotte, E. M., Xenarios, I., van Der Bliek, A. M. & Eisenberg, D. (2000). Localizing proteins in the cell from their phylogenetic profiles. *Proc Natl Acad Sci U S A* **97**, 12115-20.

88. Claros, M. & Vincens, P. (1996). Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur J Biochem* **241**, 779 - 786.

89. Guda, C., Fahy, E. & Subramaniam, S. (2004). MITOPRED: a genome-scale method for prediction of nucleus-encoded mitochondrial proteins. *Bioinformatics* **20**, 1785.

90. 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.

91. 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.

92. Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protoc* **2**, 953-971.

93. Bolender, N., Sickmann, A., Wagner, R., Meisinger, C. & Pfanner, N. (2008). Multiple pathways for sorting mitochondrial precursor proteins. *EMBO Rep* **9**, 42-49.

94. Ross-Macdonald, P., Coelho, P. S. R., Roemer, T., Agarwal, S., Kumar, A., Jansen, R., Cheung, K. H., Sheehan, A., Symoniatis, D., Umansky, L., Heidtman, M., Nelson, F. K., Iwasaki, H., Hager, K., Gerstein, M., Miller, P., Roeder, G. S. & Snyder, M. (1999). Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* **402**, 413-417.

95. Kumar, A., Agarwal, S., Heyman, J. A., Matson, S., Heidtman, M., Piccirillo, S., Umansky, L., Drawid, A., Jansen, R., Liu, Y., Cheung, K.-H., Miller, P., Gerstein, M., Roeder, G. S. & Snyder, M. (2002). Subcellular localization of the yeast proteome. *Genes Dev* **16**, 707-19.

96. Huh, W. K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S. & O'Shea, E. K. (2003). Global analysis of protein localization in budding yeast. *Nature* **425**, 686-691.

97. Aebersold, R. & Mann, M. (2003). Mass spectrometry-based proteomics. *Nature* **422**, 198-207.

98. Domon, B. & Aebersold, R. (2006). Mass spectrometry and protein analysis. *Science* **312**, 212-217.

99. Taylor, S. W., Fahy, E. & Ghosh, S. S. (2003). Global organellar proteomics. *Trends Biotechnol* **21**, 82-88.

100. Da Cruz, S., Xenarios, I., Langridge, J., Vilbois, F., Parone, P. A. & Martinou, J. C. (2003). Proteomic analysis of the mouse liver mitochondrial inner membrane. *J Biol Chem* **278**, 41566-41571.

101. Schmitt, S., Prokisch, H., Schlunck, T., Camp, D. G., Ahting, U., Waizenegger, T., Scharfe, C., Meitinger, T., Imhof, A., Neupert, W., Oefner, P. J. & Rapaport, D. (2006). Proteome analysis of mitochondrial outer membrane from *Neurospora crassa*. *Proteomics* **6**, 72-80.

102. Zahedi, R. P., Sickmann, A., Boehm, A. M., Winkler, C., Zufall, N., Schönfisch, B., Guiard, B., Pfanner, N. & Meisinger, C. (2006). Proteomic analysis of the yeast mitochondrial outer membrane reveals accumulation of a subclass of preproteins. *Mol Biol Cell* **17**, 1436-1450.

103. Abdrakhmanova, A., Zickermann, V., Bostina, M., Radermacher, M., Schägger, H., Kerscher, S. & Brandt, U. (2004). Subunit composition of mitochondrial complex I from the yeast *Yarrowia lipolytica*. *Biochim Biophys Acta* **1658**, 148-156.

104. Cardol, P., Vanrobaeys, F., Devreese, B., Van Beeumen, J., Matagne, R. F. & Remacle, C. (2004). Higher plant-like subunit composition of mitochondrial complex I from *Chlamydomonas reinhardtii*: 31 conserved components among eukaryotes. *Biochim Biophys Acta* **1658**, 212-224.

105. Millar, A. H., Eubel, H., Jänsch, L., Kruft, V., Heazlewood, J. L. & Braun, H.-P. (2004). Mitochondrial cytochrome *c* oxidase and succinate dehydrogenase complexes contain plant specific subunits. *Plant Mol Biol* **56**, 77-90.

106. Vázquez-Acevedo, M., Cardol, P., Cano-Estrada, A., Lapaille, M., Remacle, C. & González-Halphen, D. (2006). The mitochondrial ATP synthase of chlorophycean algae contains eight subunits of unknown origin involved in the formation of an atypical stator-stalk and in the dimerization of the complex. *J Bioenerg Biomembr* **38**, 271-282.

107. Carroll, J., Fearnley, I. M., Skehel, J. M., Shannon, R. J., Hirst, J. & Walker, J. E. (2006). Bovine complex I is a complex of 45 different subunits. *J Biol Chem* **281**, 32724-32727.

108. Klodmann, J., Sunderhaus, S., Nimtz, M., Jänsch, L. & Braun, H.-P. (2010). Internal architecture of mitochondrial complex I from *Arabidopsis thaliana*. *Plant Cell* **22**, 797-810.

109. Yogev, O. & Pines, O. (2010). Dual targeting of mitochondrial proteins: Mechanism, regulation and function. *BBA-Biomembranes* **1808**, 1012-1020.

110. Dinur-Mills, M., Tal, M. & Pines, O. (2008). Dual targeted mitochondrial proteins are characterized by lower MTS parameters and total net charge. *PLoS ONE* **3**, e2161.

111. Keeling, P. J., Burger, G., Durnford, D. G., Lang, B. F., Lee, R. W., Pearlman, R. E., Roger, A. J. & Gray, M. W. (2005). The tree of eukaryotes. *Trends Ecol Evol* **20**, 670-676.

112. Schmitz-Linneweber, C. & Small, I. (2008). Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci* **13**, 663-670.

113. Runswick, M. J., Fearnley, I. M., Skehel, J. M. & Walker, J. E. (1991). Presence of an acyl carrier protein in NADH: ubiquinone oxidoreductase from bovine heart mitochondria. *FEBS Lett* **286**, 121-124.

114. Braun, H.-P. & Schmitz, U. K. (1995). Are the 'core' proteins of the mitochondrial bc_1 complex evolutionary relics of a processing protease? *Trends Biochem Sci* **20**, 171-175.

115. Sunderhaus, S., Dudkina, N. V., Jänsch, L., Klodmann, J., Heinemeyer, J., Perales, M., Zabaleta, E., Boekema, E. J. & Braun, H.-P. (2006). Carbonic anhydrase subunits form a matrix-exposed domain attached to the membrane arm of mitochondrial complex I in plants. *J Biol Chem* **281**, 6482-6488.

116. Cruciat, C. M., Brunner, S., Baumann, F., Neupert, W. & Stuart, R. A. (2000). The cytochrome bc_1 and cytochrome *c* oxidase complexes associate to form a single supracomplex in yeast mitochondria. *J Biol Chem* **275**, 18093-18098.

117. Dudkina, N. V., Eubel, H., Keegstra, W., Boekema, E. J. & Braun, H.-P. (2005). Structure of a mitochondrial supercomplex formed by respiratory-chain complexes I and III. *Proc Natl Acad Sci USA* **102**, 3225-3229.

118. Dudkina, N. V., Heinemeyer, J., Sunderhaus, S., Boekema, E. J. & Braun, H.-P. (2006). Respiratory chain supercomplexes in the plant mitochondrial membrane. *Trends Plant Sci* **11**, 232-240.

119. Nübel, E., Wittig, I., Kerscher, S., Brandt, U. & Schägger, H. (2009). Two dimensional native electrophoretic analysis of respiratory supercomplexes from *Yarrowia lipolytica*. *Proteomics* **9**, 2408-2418.

120. Dudkina, N. V., Kouřil, R., Peters, K., Braun, H. P. & Boekema, E. J. (2010). Structure and function of mitochondrial supercomplexes. *BBA-Bioenergetics* **1797**, 664-670.

121. Berry, S. (2003). Endosymbiosis and the design of eukaryotic electron transport. *Biochim Biophys Acta* **1606**, 57-72.

122. Castresana, J. (2001). Comparative genomics and bioenergetics. *BBA-Bioenergetics* **1506**, 147-162.

123. Joseph-Horne, T., Hollomon, D. W. & Wood, P. M. (2001). Fungal respiration: a fusion of standard and alternative components. *Biochim Biophys Acta* **1504**, 179-195.

124. Rasmusson, A. G., Soole, K. L. & Elthon, T. E. (2004). Alternative NAD(P)H dehydrogenases of plant mitochondria. *Annu Rev Plant Biol* **55**, 23-39.

125. Ackerman, S. H. & Tzagoloff, A. (2005). Function, structure, and biogenesis of mitochondrial ATP synthase. In *Prog Nucleic Acid Res Mol Biol* (Kivie, M., ed.), Vol. Volume 80, pp. 95-133. Academic Press.

126. Yip, C.-y., Harbour, M. E., Jayawardena, K., Fearnley, I. M. & Sazanov, L. A. (2011). Evolution of respiratory complex I. "Supernumerary" subunits are present in the α-proteobacterial enzyme. *J Biol Chem* **286**, 5023-5033.

127. Braun, H.-P. & Zabaleta, E. (2007). Carbonic anhydrase subunits of the mitochondrial NADH dehydrogenase complex (complex I) in plants. *Physiol Plantarum* **129**, 114-122.

128. Aggeler, R., Coons, J., Taylor, S. W., Ghosh, S. S., García, J. J., Capaldi, R. A. & Marusich, M. F. (2002). A functionally active human F1F0 ATPase can be purified by immunocapture from heart tissue and fibroblast cell lines. Subunit structure and activity studies. *J Biol Chem* **277**, 33906-33912.

129. Taanman, J. W. & Capaldi, R. A. (1992). Purification of yeast cytochrome *c* oxidase with a subunit composition resembling the mammalian enzyme. *J Biol Chem* **267**, 22481-22485.

130. Arnold, I., Pfeiffer, K., Neupert, W., Stuart, R. A. & Schägger, H. (1998). Yeast mitochondrial F_1F_0-ATP synthase exists as a dimer: identification of three dimer-specific subunits. *EMBO J* **17**, 7170-7178.

131. Braun, H.-P., Emmermann, M., Kruft, V., Bödicker, M. & Schmitz, U. K. (1995). The general mitochondrial processing peptidase from wheat is integrated into the cytochrome bc_1-complex of the respiratory chain. *Planta* **195**, 396-402.

132. Heazlewood, J. L., Whelan, J. & Millar, A. H. (2003). The products of the mitochondrial *orf25* and *orfB* genes are F_O components in the plant F_1F_O ATP synthase. *FEBS Lett* **540**, 201-205.

133. Marx, S., Baumgartner, M., Kunnan, S., Braun, H. P., Lang, B. F. & Burger, G. (2003). Structure of the *bc1* complex from *Seculamonas ecuadoriensis*, a jakobid flagellate with an ancestral mitochondrial genome. *Mol Biol Evol* **20**, 145-53.

134. Burger, G., Lang, B. F., Braun, H.-P. & Marx, S. (2003). The enigmatic mitochondrial ORF *ymf39* codes for ATP synthase chain b. *Nucleic Acids Res* **31**, 2353-2360.

135. Pérez-Martínez, X., Antaramian, A., Vázquez-Acevedo, M., Funes, S., Tolkunova, E., d'Alayer, J., Claros, M. G., Davidson, E., King, M. P. & González-Halphen, D. (2001). Subunit II of cytochrome *c* oxidase in chlamydomonad algae is a heterodimer encoded by two independent nuclear genes. *J Biol Chem* **276**, 11302-9.

136. Panigrahi, A. K., Zíková, A., Dalley, R. A., Acestor, N., Ogata, Y., Anupama, A., Myler, P. J. & Stuart, K. D. (2008). Mitochondrial Complexes in *Trypanosoma brucei*. *Mol Cell Proteomics* **7**, 534-545.

137. Morales, J., Mogi, T., Mineki, S., Takashima, E., Mineki, R., Hirawake, H., Sakamoto, K., Ōmura, S. & Kita, K. (2009). Novel mitochondrial complex II isolated from *Trypanosoma cruzi* is composed of 12 peptides including a heterodimeric Ip subunit. *J Biol Chem* **284**, 7255-7263.

138. Zíková, A., Schnaufer, A., Dalley, R. A., Panigrahi, A. K. & Stuart, K. D. (2009). The F_oF_1-ATP synthase complex contains novel subunits and is essential for procyclic *Trypanosoma brucei*. *PLoS Pathog* **5**, e1000436.

139. Nina, P. B., Dudkina, N. V., Kane, L. A., van Eyk, J. E., Boekema, E. J., Mather, M. W. & Vaidya, A. B. (2010). Highly divergent mitochondrial ATP synthase complexes in *Tetrahymena thermophila*. *PLoS Biol* **8**, 1-5 [e1000418].

140. Schägger, H. & von Jagow, G. (1991). Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal Biochem* **199**, 223-231.

141. Schägger, H., Cramer, W. & von Jagow, G. (1994). Analysis of molecular masses and oligomeric states of protein complexes by blue native electrophoresis and isolation of membrane protein complexes by two-dimensional native electrophoresis. *Anal Biochem* **217**, 220-230.

142. Zerbetto, E., Vergani, L. & Dabbeni-Sala, F. (1997). Quantification of muscle mitochondrial oxidative phosphorylation enzymes *via* histochemical staining of blue native polyacrylamide gels. *Electrophoresis* **18**, 2059-2064.

143. 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**, 365-80.

144. Edqvist, J., Burger, G. & Gray, M. W. (2000). Expression of mitochondrial protein-coding genes in *Tetrahymena pyriformis*. *J Mol Biol* **297**, 381-93.

145. Adams, K. L., Ong, H. C. & Palmer, J. D. (2001). Mitochondrial gene transfer in pieces: fission of the ribosomal protein gene *rpl2* and partial or complete gene transfer to the nucleus. *Mol Biol Evol* **18**, 2289-2297.

146. Funes, S., Davidson, E., Reyes-Prieto, A., Magallón, S., Herion, P., King, M. P. & González-Halphen, D. (2002). A green algal apicoplast ancestor. *Science* **298**, 2155.

147. Waller, R. F. & Keeling, P. J. (2006). Alveolate and chlorophycean mitochondrial *cox2* genes split twice independently *Gene* **383**, 33-37.

148. Bonen, L. (1993). Trans-splicing of pre-mRNA in plants, animals, and protists. *Faseb J* **7**, 40-6.

149. Gogarten, J. P., Senejani, A. G., Zhaxybayeva, O., Olendzenski, L. & Hilario, E. (2002). Inteins: structure, function, and evolution. *Annu Rev Microbiol* **56**, 263-287.

150. Schnare, M. N., Heinonen, T. Y., Young, P. G. & Gray, M. W. (1986). A discontinuous small subunit ribosomal RNA in *Tetrahymena pyriformis* mitochondria. *J Biol Chem* **261**, 5187-93.

151. Boer, P. H. & Gray, M. W. (1988). Scrambled ribosomal RNA gene pieces in *Chlamydomonas reinhardtii* mitochondrial DNA. *Cell* **55**, 399-411.

152. Ullmann, A., Jacob, F. & Monod, J. (1967). Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the beta-galactosidase structural gene of *Escherichia coli*. *J Mol Biol* **24**, 339-343.

153. Handa, H., Bonnard, G. & Grienenberger, J. M. (1996). The rapeseed mitochondrial gene encoding a homologue of the bacterial protein Ccl1 is divided into two independently transcribed reading frames. *Mol Gen Genet* **252**, 292-302.

154. Unseld, M., Marienfeld, J. R., Brandt, P. & Brennicke, A. (1997). The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat Genet* **15**, 57-61.

155. Nedelcu, A. M., Lee, R. W., Lemieux, C., Gray, M. W. & Burger, G. (2000). The complete mitochondrial DNA sequence of *Scenedesmus obliquus* reflects an intermediate stage in the evolution of the green algal mitochondrial genome. *Genome Res* **10**, 819-31.

156. Oudot, M.-P., Kloareg, B. & Loiseaux-de Goër, S. (1999). The mitochondrial *Pylaiella littoralis nad11* gene contains only the N-terminal FeS-binding domain. *Gene* **235**, 131-7.

157. Oudot-Le Secq, M.-P., Fontaine, J. M., Rousvoal, S., Kloareg, B. & Loiseaux-de Goër, S. (2001). The complete sequence of a brown algal mitochondrial genome, the ectocarpale *Pylaiella littoralis* (L.) Kjellm. *J Mol Evol* **53**, 80-8.

158. Oudot-Le Secq, M.-P. & Green, B. R. (2011). Complex repeat structures and novel features in the mitochondrial genomes of the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*. *Gene* **476**, 20-26.

159. Rokas, A. & Holland, P. W. H. (2000). Rare genomic changes as a tool for phylogenetics. *Trends Ecol Evol* **15**, 454-459.

160. Stechmann, A. & Cavalier-Smith, T. (2002). Rooting the eukaryote tree by using a derived gene fusion. *Science* **297**, 89-91.

161. Stechmann, A. & Cavalier-Smith, T. (2003). The root of the eukaryote tree pinpointed. *Curr Biol* **13**, R665-6.

162. Gribaldo, S. & Philippe, H. (2002). Ancient phylogenetic relationships. *Theor Popul Biol* **61**, 391-408.

163. Fast, N. M., Kissinger, J. C., Roos, D. S. & Keeling, P. J. (2001). Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol Biol Evol* **18**, 418-26.

164. Janouškovec, J., Horák, A., Oborník, M., Lukeš, J. & Keeling, P. J. (2010). A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proc Natl Acad Sci USA* **107**, 10949-10954.

165. Waller, R. F., Keeling, P. J., van Dooren, G. G. & McFadden, G. I. (2003). Comment on "A green algal apicoplast ancestor". *Science* **301**, 49a.

166. Keeping, A., DeAbreu, D., DiBernardo, M. & Collins, R. A. (2011). Gel-based mass spectrometric and computational approaches to the mitochondrial proteome of *Neurospora*. *Fungal Genet Biol* **48**, 526-536.

167. Huang, S., Taylor, N. L., Narsai, R., Eubel, H., Whelan, J. & Millar, A. H. (2009). Experimental analysis of the rice mitochondrial proteome, its biogenesis, and heterogeneity. *Plant Physiol* **149**, 719-734.

168. Jacoby, R. P., Millar, A. H. & Taylor, N. L. (2010). Wheat mitochondrial proteomes provide new links between antioxidant defense and plant salinity tolerance. *J Proteome Res* **9**, 6595-6604.

169. Millar, A. H., Heazlewood, J. L., Kristensen, B. K., Braun, H.-P. & Møller, I. M. (2005). The plant mitochondrial proteome. *Trends Plant Sci.* **10**, 36-43.

170. Calvo, S. E. & Mootha, V. K. (2010). The mitochondrial proteome and human disease. *Annu Rev Genomics Hum Genet* **11**, 25-44.

171. Anderson, I. J., Watkins, R. F., Samuelson, J., Spencer, D. F., Majoros, W. H., Gray, M. W. & Loftus, B. J. (2005). Gene discovery in the *Acanthamoeba castellanii* genome. *Protist* **156**, 203-214.

172. Marciano-Cabral, F. & Cabral, G. (2003). *Acanthamoeba* spp. as agents of disease in humans. *Clin Microbiol Rev* **16**, 273-307.

173. Khan, N. A. (2007). *Acanthamoeba* invasion of the central nervous system. *Int J Parasitol* **37**, 131-138.

174. Schuster, F. L. (2002). Cultivation of pathogenic and opportunistic free-living amebas. *Clin Microbiol Rev* **15**, 342-354.

175. Salah, I. B., Ghigo, E. & Drancourt, M. (2009). Free-living amoebae, a training field for macrophage resistance of mycobacteria. *Clin Microbiol Infect* **15**, 894-905.

176. Lonergan, K. M. & Gray, M. W. (1993). Editing of transfer RNAs in *Acanthamoeba castellanii* mitochondria. *Science* **259**, 812-816.

177. Gawryluk, R. M. R. & Gray, M. W. (2010). An ancient fission of mitochondrial *cox1*. *Mol Biol Evol* **27**, 7-10.

178. Burger, G., Plante, I., Lonergan, K. M. & Gray, M. W. (1995). The mitochondrial DNA of the amoeboid protozoon, *Acanthamoeba castellanii*: complete sequence, gene content and genome organization. *J Mol Biol* **245**, 522-537.

179. Lohan, A. J. & Gray, M. W. (2007). Analysis of 5'- or 3'-terminal tRNA editing: mitochondrial 5' tRNA editing in *Acanthamoeba castellanii* as the exemplar. In *Methods Enzymol* (Jonatha, M. G., ed.), Vol. 424, pp. 223-242. Academic Press.

180. Eichinger, L., Pachebat, J. A., Glockner, G., Rajandream, M. A., Sucgang, R., Berriman, M., Song, J., Olsen, R., Szafranski, K., Xu, Q., Tunggal, B., Kummerfeld, S., Madera, M., Konfortov, B. A., Rivero, F., Bankier, A. T., Lehmann, R., Hamlin, N., Davies, R., Gaudet, P., Fey, P., Pilcher, K., Chen, G., Saunders, D., Sodergren, E., Davis, P., Kerhornou, A., Nie, X., Hall, N., Anjard, C., Hemphill, L., Bason, N., Farbrother, P., Desany, B., Just, E., Morio, T., Rost, R., Churcher, C., Cooper, J., Haydock, S., van Driessche, N., Cronin, A., Goodhead, I., Muzny, D., Mourier, T., Pain, A., Lu, M., Harper, D., Lindsay, R., Hauser, H., James, K., Quiles, M., Madan Babu, M., Saito, T., Buchrieser, C., Wardroper, A., Felder, M., Thangavelu, M., Johnson, D., Knights, A., Loulseged, H., Mungall, K., Oliver, K., Price, C., Quail, M. A., Urushihara, H., Hernandez, J., Rabbinowitsch, E., Steffen, D., Sanders, M., Ma, J., Kohara, Y., Sharp, S., Simmonds, M., Spiegler, S., Tivey, A., Sugano, S., White, B., Walker, D., Woodward, J., Winckler, T., Tanaka, Y., Shaulsky, G., Schleicher, M., Weinstock, G., Rosenthal, A., Cox, E. C., Chisholm, R. L., Gibbs, R., Loomis, W. F., Platzer, M., Kay, R. R., Williams, J., Dear, P. H., Noegel, A. A., Barrell, B. & Kuspa, A. (2005). The genome of the social amoeba *Dictyostelium discoideum*. *Nature* **435**, 43-57.

181. Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

182. Gawryluk, R. M. R. & Gray, M. W. (2010). Evidence for an early evolutionary emergence of γ-type carbonic anhydrases as components of mitochondrial respiratory complex I. *BMC Evol Biol* **348**, 857-870.

183. Bullerwell, C. E., Schnare, M. N. & Gray, M. W. (2003). Discovery and characterization of *Acanthamoeba castellanii* mitochondrial 5S rRNA. *RNA* **9**, 287-292.

184. Peattie, D. A. (1979). Direct chemical method for sequencing RNA. *Proc Natl Acad Sci USA* **76**.

185. O'Brien, E. A., Koski, L. B., Zhang, Y., Yang, L., Wang, E., Gray, M. W., Burger, G. & Lang, B. F. (2007). TBestDB: a taxonomically broad database of expressed sequence tags (ESTs). *Nucleic Acids Res* **35**, D445-D451.

186. Stanke, M. & Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**, ii215-ii225.

187. Chevreux, B., Pfisterer, T., Drescher, B., Driesel, A. J., Müller, W. E. G., Wetter, T. & Suhai, S. (2004). Using the miraEST Assembler for Reliable and Automated mRNA Transcript Assembly and SNP Detection in Sequenced ESTs. *Genome Res* **14**, 1147-1159.

188. Kreppel, L., Fey, P., Gaudet, P., Just, E., Kibbe, W. A., Chisholm, R. L. & Kimmel, A. R. (2004). dictyBase: a new *Dictyostelium discoideum* genome database. *Nucleic Acids Res* **32**, D332-D333.

189. Altschul, S. F., Madden, T. L., Schäffer, 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. *Nucleic Acids Res* **25**, 3389-3402.

190. Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797.

191. Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics* **14**, 755-763.

192. Pérez-Bercoff, A., Koch, J. & Bürglin, T. R. (2006). LogoBar: bar graph visualization of protein logos with gaps. *Bioinformatics* **22**, 112 - 114.

193. Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690.

194. Lonergan, K. M. & Gray, M. W. (1996). Expression of a continuous open reading frame encoding subunits 1 and 2 of cytochrome *c* oxidase in the mitochondrial DNA of *Acanthamoeba castellanii*. *J Mol Biol* **257**, 1019-30.

195. Wong, T. W. & Clayton, D. A. (1986). DNA primase of human mitochondria is associated with structural RNA that is essential for enzymatic activity. *Cell* **45**, 817-825.

196. Boldogh, I. R. & Pon, L. A. (2007). Mitochondria on the move. *Trends Cell Biol* **17**, 502-510.

197. Suissa, M. & Schatz, G. (1982). Import of proteins 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.

198. 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.

199. MacKenzie, J. A. & Payne, R. M. (2004). Ribosomes specifically bind to mammalian mitochondria via protease-sensitive proteins on the outer membrane. *J Biol Chem* **279**, 9803-9810.

200. Kornmann, B., Currie, E., Collins, S. R., Schuldiner, M., Nunnari, J., Weissman, J. S. & Walter, P. (2009). An ER-mitochondria tethering complex revealed by a synthetic biology screen. *Science* **325**, 477-481.

201. Emelyanov, V. V. & Goldberg, A. V. (2011). Fermentation enzymes of *Giardia lamblia*, pyruvate:ferredoxin oxidoreductase and hydrogenase, do not localize to its mitosomes. *Microbiology*, mic.0.044784-0.

202. Mus, F., Dubini, A., Seibert, M., Posewitz, M. C. & Grossman, A. R. (2007). Anaerobic acclimation in *Chlamydomonas reinhardtii*. Anoxic gene expression, hydrogenase induction, and metabolic pathways *J Biol Chem* **282**, 25475-25486.

203. Hug, L. A., Stechmann, A. & Roger, A. J. (2010). Phylogenetic distributions and histories of proteins involved in anaerobic pyruvate metabolism in eukaryotes. *Mol Biol Evol* **27**, 311-324.

204. van Grinsven, K. W. A., van Hellemond, J. J. & Tielens, A. G. M. (2009). Acetate:succinate CoA-transferase in the anaerobic mitochondria of *Fasciola hepatica*. *Mol Biochem Parasitol* **164**, 74-79.

205. Zhang, W., Wong, K. K., Magliozzo, R. S. & Kozarich, J. W. (2001). Inactivation of pyruvate formate-lyase by dioxygen: Defining the mechanistic interplay of glycine 734 and cysteine 419 by rapid freeze-quench EPR. *Biochemistry* **40**, 4123-4130.

206. Asanuma, N., Yoshii, T. & Hino, T. (2004). Molecular characteristics and transcription of the gene encoding a multifunctional alcohol dehydrogenase in relation to the deactivation of pyruvate formate-lyase in the ruminal bacterium *Streptococcus bovis*. *Arch Microbiol* **181**, 122-128.

207. Fritz-Laylin, L. K., Prochnik, S. E., Ginger, M. L., Dacks, J. B., Carpenter, M. L., Field, M. C., Kuo, A., Paredez, A., Chapman, J., Pham, J., Shu, S., Neupane, R., Cipriano, M., Mancuso, J., Tu, H., Salamov, A., Lindquist, E., Shapiro, H., Lucas, S., Grigoriev, I. V., Cande, W. Z., Fulton, C., Rokhsar, D. S. & Dawson, S. C. (2010). The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. *Cell* **140**, 631-642.

208. Millar, A. H., Hill, S. A. & Leaver, C. J. (1999). Plant mitochondrial 2-oxoglutarate dehydrogenase complex: purification and characterization in potato. *Biochem J* **343**, 327-334.

209. Mitchell, P. & Moyle, J. (1967). Chemiosmotic hypothesis of oxidative phosphorylation. *Nature* **213**, 137-139.

210. Hunte, C., Zickermann, V. & Brandt, U. (2010). Functional modules and structural basis of conformational coupling in mitochondrial complex I. *Science* **329**, 448-451.

211. Angell, J. E., Lindner, D. J., Shapiro, P. S., Hofmann, E. R. & Kalvakolanu, D. V. (2000). Identification of GRIM-19, a novel cell death-regulatory gene induced by the interferon-β and retinoic acid combination, using a genetic approach. *J Biol Chem* **275**, 33416-33426.

212. Fearnley, I. M., Carroll, J., Shannon, R. J., Runswick, M. J., Walker, J. E. & Hirst, J. (2001). GRIM-19, a cell death regulatory gene product, is a subunit of bovine mitochondrial NADH:ubiquinone oxidoreductase (complex I). *J Biol Chem* **276**, 38345-38348.

213. Navet, R., Jarmuszkiewicz, W., Douette, P., Sluse-Goffart, C. M. & Sluse, F. E. (2004). Mitochondrial respiratory chain complex patterns from *Acanthamoeba castellanii* and *Lycopersicon esculentum*: Comparative analysis by BN-PAGE and evidence of protein–protein interaction between alternative oxidase and complex III. *Journal Bioenerg Biomembr* **36**, 471-479.

214. Meyer, E. H., Heazlewood, J. L. & Millar, A. H. (2007). Mitochondrial acyl carrier proteins in *Arabidopsis thaliana* are predominantly soluble matrix proteins and none can be confirmed as subunits of respiratory complex I. *Plant Mol Biol* **64**, 319-327.

215. Bridges, H. R., Fearnley, I. M. & Hirst, J. (2010). The subunit composition of mitochondrial NADH:ubiquinone oxidoreductase (complex I) from *Pichia pastoris*. *Mol Cell Proteomics* **9**, 2318-2326.

216. Ogata, K. & Volini, M. (1990). Mitochondrial rhodanese: membrane-bound and complexed activity. *J Biol Chem* **265**, 8087-8093.

217. Jin, C., Barrientos, A. & Tzagoloff, A. (2003). Yeast dihydroxybutanone phosphate synthase, an enzyme of the riboflavin biosynthetic pathway, has a second unrelated function in expression of mitochondrial respiration. *J Biol Chem* **278**, 14698-14703.

218. Torkko, J. M., Koivuranta, K. T., Miinalainen, I. J., Yagi, A. I., Schmitz, W., Kastaniotis, A. J., Airenne, T. T., Gurvitz, A. & Hiltunen, K. J. (2001). *Candida tropicalis* Etr1p and *Saccharomyces cerevisiae* Ybr026p (Mrf1'p), 2-enoyl thioester reductases essential for mitochondrial respiratory competence. *Mol Cell Biol* **21**, 6243-6253.

219. Mckenzie, M. & Ryan, M. T. (2010). Assembly factors of human mitochondrial complex I and their defects in disease. *IUBMB Life* **62**, 497-502.

220. Bych, K., Kerscher, S., Netz, D. J. A., Pierik, A. J., Zwicker, K., Huynen, M. A., Lill, R., Brandt, U. & Balk, J. (2008). The iron-sulphur protein Ind1 is required for effective complex I assembly. *EMBO J* **27**, 1736-1746.

221. Sheftel, A. D., Stehling, O., Pierik, A. J., Netz, D. J. A., Kerscher, S., Elsässer, H. P., Wittig, I., Balk, J., Brandt, U. & Lill, R. (2009). Human Ind1, an iron-sulfur cluster assembly factor for respiratory complex I. *Mol Cell Biol* **29**, 6059-6073.

222. Sugiana, C., Pagliarini, D. J., McKenzie, M., Kirby, D. M., Salemi, R., Abu-Amero, K. K., Dahl, H. H. M., Hutchison, W. M., Vascotto, K. A. & Smith, S. M. (2008). Mutation of *C20orf7* disrupts complex I assembly and causes lethal neonatal mitochondrial disease. *Am J Hum Genet* **83**, 468-478.

223. Carilla-Latorre, S., Gallardo, M. E., Annesley, S. J., Calvo-Garrido, J., Graña, O., Accari, S. L., Smith, P. K., Valencia, A., Garesse, R. & Fisher, P. R. (2010). MidA is a putative methyltransferase that is required for mitochondrial complex I function. *J Cell Sci* **123**, 1674-1683.

224. Vogel, R. O., Janssen, R. J. R. J., Ugalde, C., Grovenstein, M., Huijbens, R. J., Visch, H. J., Van Den Heuvel, L. P., Willems, P. H., Zeviani, M. & Smeitink, J. A. M. (2005). Human mitochondrial complex I assembly is mediated by NDUFAF1. *FEBS J* **272**, 5317-5326.

225. Dunning, C. J. R., McKenzie, M., Sugiana, C., Lazarou, M., Silke, J., Connelly, A., Fletcher, J., Kirby, D. M., Thorburn, D. & Ryan, M. T. (2007). Human CIA30 is involved in the early assembly of mitochondrial complex I and mutations in its gene cause disease. *EMBO J* **26**, 3227-3237.

226. Calvo, S. E., Tucker, E. J., Compton, A. G., Kirby, D. M., Crawford, G., Burtt, N. P., Rivas, M., Guiducci, C., Bruno, D. L. & Goldberger, O. A. (2010). High-throughput, pooled sequencing identifies mutations in *NUBPL* and *FOXRED1* in human complex I deficiency. *Nat Genet* **42**, 851-858.

227. Fassone, E., Duncan, A. J., Taanman, J. W., Pagnamenta, A. T., Sadowski, M. I., Holand, T., Qasim, W., Rutland, P., Calvo, S. E. & Mootha, V. K. (2010). *FOXRED1*, encoding an FAD-dependent oxidoreductase complex-I-specific molecular chaperone, is mutated in infantile-onset mitochondrial encephalopathy. *Hum Mol Genet* **19**, 4837-4847.

228. Ogilvie, I., Kennaway, N. G. & Shoubridge, E. A. (2005). A molecular chaperone for mitochondrial complex I assembly is mutated in a progressive encephalopathy. *J Clin Invest* **115**, 2784-2792.

229. Gawryluk, R. M. R. & Gray, M. W. (2009). A split and rearranged nuclear gene encoding the iron-sulfur subunit of mitochondrial succinate dehydrogenase in Euglenozoa. *BMC Res Notes* **2**, 16.

230. Cecchini, G. (2003). Function and structure of complex II of the respiratory chain. *Annu Rev Biochem* **72**, 77-109.

231. Crofts, A. R. (2004). The cytochrome bc_1 complex: function in the context of structure. *Annu Rev Physiol* **66**, 689-733.

232. Iwata, S., Lee, J. W., Okada, K., Lee, J. K., Iwata, M., Rasmussen, B., Link, T. A., Ramaswamy, S. & Jap, B. K. (1998). Complete structure of the 11-subunit bovine mitochondrial cytochrome bc_1 complex. *Science* **281**, 64-71.

233. Lange, C. & Hunte, C. (2002). Crystal structure of the yeast cytochrome bc_1 complex with its bound substrate cytochrome *c*. *Proc Natl Acad Sci USA* **99**, 2800-2805.

234. Glaser, E. & Dessi, P. (1999). Integration of the mitochondrial-processing peptidase into the cytochrome bc_1 complex in plants. *J Bioenerg Biomembr* **31**, 259-274.

235. Yang, M., Jensen, R. E., Yaffe, M. P., Oppliger, W. & Schatz, G. (1988). Import of proteins into yeast mitochondria: the purified matrix processing protease contains two subunits which are encoded by the nuclear *MAS1* and *MAS2* genes. *EMBO J* **7**, 2605-2612.

236. Ou, W.-J., Ito, A., Okazaki, H. & Omura, T. (1989). Purification and characterization of a processing protease from rat liver mitochondria. *EMBO J* **8**, 2605-2612.

237. Hawlitschek, G., Schneider, H., Schmidt, B., Tropschug, M., Hartl, F.-U. & Neupert, W. (1988). Mitochondrial protein import: Identification of processing peptidase and of PEP, a processing enhancing protein. *Cell* **53**, 795-806.

238. Nagayama, K., Itono, S., Yoshida, T., Ishiguro, S.-i., Ochiai, H. & Ohmachi, T. (2008). Antisense RNA Inhibition of the b subunit of the *Dictyostelium discoideum* mitochondrial processing peptidase induces the expression of mitochondrial proteins. *Biosci Biotechnol Biochem* **72**, 1836-1846.

239. Zara, V., Conte, L. & Trumpower, B. L. (2009). Biogenesis of the yeast cytochrome bc1 complex. *BBA-Mol Cell Res* **1793**, 89-96.

240. Cruciat, C. M., Hell, K., Fölsch, H., Neupert, W. & Stuart, R. A. (1999). Bcs1p, an AAA-family member, is a chaperone for the assembly of the cytochrome bc1 complex. *EMBO J* **18**, 5226-5233.

241. Wu, M. & Tzagoloff, A. (1989). Identification and characterization of a new gene (CBP3) required for the expression of yeast coenzyme QH2-cytochrome *c* reductase. *J Biol Chem* **264**, 11122-11130.

242. Bisson, R. & Schiavo, G. (1986). Two different forms of cytochrome *c* oxidase can be purified from the slime mold *Dictyostelium discoideum*. *J Biol Chem* **261**, 4373-4376.

243. Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R. & Yoshikawa, S. (1996). The whole structure of the 13-subunit oxidized cytochrome *c* oxidase at 2.8 Å. *Science* **272**, 1136-1144.

244. Cole, R. A. & Williams, K. L. (1994). The *Dictyostelium discoideum* mitochondrial genome: a primordial system using the universal code and encoding hydrophilic proteins atypical of metazoan mitochondrial DNA. *J Mol Evol* **39**, 579-588.

245. Pellizzari, R., Anjard, C. & Bisson, R. (1997). Subunits I and II of *Dictyostelium* cytochrome *c* oxidase are specified by a single open reading frame transcribed into a large polycistronic RNA. *Biochim Biophys Acta* **1320**, 1-7.

246. Eubel, H., Jänsch, L. & Braun, H.-P. (2003). New insights into the respiratory chain of plant mitochondria. Supercomplexes and a unique composition of complex II. *Plant Physiol* **133**, 274-286.

247. Fontanesi, F., Soto, I. C. & Barrientos, A. (2008). Cytochrome *c* oxidase biogenesis: new levels of regulation. *IUBMB Life* **60**, 557-568.

248. Mick, D. U., Fox, T. D. & Rehling, P. (2011). Inventory control: cytochrome *c* oxidase assembly regulates mitochondrial translation. *Nature Rev Mol Cell Biol* **12**, 14-20.

249. Boyer, P. D. (1997). The ATP synthase—a splendid machine. *Annu Rev Biochem* **66**, 717-749.

250. Yoshida, M., Muneyuki, E. & Hisabori, T. (2001). ATP synthase — a marvellous rotary engine of the cell. *Nat Rev Mol Cell Biol* **2**, 669-677.

251. Senior, A. E., Nadanaciva, S. & Weber, J. (2002). The molecular mechanism of ATP synthesis by F_1F_0-ATP synthase. *Biochim Biophys Acta* **1553**, 188-211.

252. Minauro-Sanmiguel, F., Wilkens, S. & García, J. J. (2005). Structure of dimeric mitochondrial ATP synthase: Novel F_0 bridging features and the structural basis of mitochondrial cristae biogenesis. *Proc Natl Acad Sci USA* **102**, 12356-12358.

253. Wagner, K., Perschil, I., Fichter, C. D. & van der Laan, M. (2010). Stepwise assembly of dimeric F_1F_o-ATP synthase in mitochondria involves the small F_o-subunits k and i. *Mol Biol Cell* **21**, 1494-1504.

254. Villavicencio-Queijeiro, A., Vázquez-Acevedo, M., Cano-Estrada, A., Zarco-Zavala, M., Tuena de Gómez, M., Mignaco, J., Freire, M., Scofano, H., Foguel, D., Cardol, P., Remacle, C. & González-Halphen, D. (2009). The fully-active and structurally-stable form of the mitochondrial ATP synthase of *Polytomella* sp. is dimeric. *J Bioenerg Biomembr* **41**, 1-13.

255. Cano-Estrada, A., Vázquez-Acevedo, M., Villavicencio-Queijeiro, A., Figueroa-Martínez, F., Miranda-Astudillo, H., Cordeiro, Y., Mignaco, J. A., Foguel, D., Cardol, P., Lapaille, M., Remacle, C., Wilkens, S. & González-Halphen, D. (2010). Subunit-subunit interactions and overall topology of the dimeric mitochondrial ATP synthase of *Polytomella* sp. *Biochim Biophys Acta* **1797**, 1439-1448.

256. Sonnhammer, E. L. L., von Heijne, G. & Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **6**, 175-182.

257. Liu, S. & Zhang, X. (2004). Expression and purification of a novel rice (*Oryza sativa L.*) mitochondrial ATP synthase small subunit in *Escherichia coli*. *Protein Expr Purif* **37**, 306-310.

258. Paumard, P., Vaillier, J., Coulary, B., Schaeffer, J., Soubannier, V., Mueller, D. M., Brèthes, D., di Rago, J.-P. & Velours, J. (2002). The ATP synthase is involved in generating mitochondrial cristae morphology. *EMBO J* **21**, 221-230.

259. Camougrand, N., Pélissier, P., Velours, G. & Guérin, M. (1995). *NCA2*, a second nuclear gene required for the control of mitochondrial synthesis of subunits 6 and 8 of ATP synthase in *Saccharomyces cerevisiae*. *J Mol Biol* **247**, 588-596.

260. Campanella, M., Parker, N., Tan, C. H., Hall, A. M. & Duchen, M. R. (2009). IF_1: setting the pace of the F_1F_o-ATP synthase. *Trends Biochem Sci* **34**, 343-350.

261. Ackerman, S. H. & Tzagoloff, A. (1990). *ATP10*, a yeast nuclear gene required for the assembly of the mitochondrial F_1-F_0 complex. *J Biol Chem* **265**, 9952-9959.

262. Zeng, X., Neupert, W. & Tzagoloff, A. (2007). The metalloprotease encoded by *ATP23* has a dual function in processing and assembly of subunit 6 of mitochondrial ATPase. *Mol Biol Cell* **18**, 617626.

263. Ackerman, S. H. (2002). Atp11p and Atp12p are chaperones for F_1-ATPase biogenesis in mitochondria. *BBA-Bioenergetics* **1555**, 101-105.

264. Kerscher, S. J. (2000). Diversity and origin of alternative NADH:ubiquinone oxidoreductases. *Biochim Biophys Acta* **1459**, 274-283.

265. Chaudhuri, M., Ott, R. D. & Hill, G. C. (2006). Trypanosome alternative oxidase: from molecule to function. *Trends Parasitol* **22**, 484-491.

266. Jarmuszkiewicz, W., Sluse, F. E., Hryniewiecka, L. & Sluse-Goffart, C. M. (2002). Interactions between the cytochrome pathway and the alternative oxidase in isolated *Acanthamoeba castellanii* mitochondria. *J Bioenerg Biomembr* **34**, 31-40.

267. Jarmuszkiewicz, W., Wagner, A. M., Wagner, M. J. & Hryniewiecka, L. (1997). Immunological identification of the alternative oxidase of *Acanthamoeba castellanii* mitochondria. *FEBS Lett* **411**, 110-114.

268. Henriquez, F. L., McBride, J., Campbell, S. J., Ramos, T., Ingram, P. R., Roberts, F., Tinney, S. & Roberts, C. W. (2009). *Acanthamoeba* alternative oxidase genes: Identification, characterisation and potential as antimicrobial targets. *Int J Parasitol* **39**, 1417-1424.

269. Eastmond, P. J. & Graham, I. A. (2001). Re-examining the role of the glyoxylate cycle in oilseeds. *Trends Plant Sci* **6**, 72-78.

270. Maeting, I., Schmidt, G., Sahm, H., Revuelta, J. L., Stierhof, Y.-D. & Stahmann, K. P. (1999). Isocitrate lyase of *Ashbya gossypii* - transcriptional regulation and peroxisomal localization. *FEBS Lett* **444**, 15-21.

271. Kunze, M., Kragler, F., Binder, M., Hartig, A. & Gurvitz, A. (2002). Targeting of malate synthase 1 to the peroxisomes of *Saccharomyces cerevisiae* cells depends on growth on oleic acid medium. *Eur J Biochem* **269**, 915-922.

272. Lorenz, M. C. & Fink, G. R. (2001). The glyoxylate cycle is required for fungal virulence. *Nature* **412**, 83-86.

273. Nakazawa, M., Nishimura, M., Inoue, K., Ueda, M., Inui, H., Nakano, Y. & Miyatake, K. (2011). Characterization of a bifunctional glyoxylate cycle enzyme, malate synthase/isocitrate lyase, of *Euglena gracilis*. *J Eukaryot Microbiol* **58**, 128-133.

274. Ono, K., Kondo, M., Osafune, T., Miyatake, K., Inui, H., Kitaoka, S., Nishimura, M. & Nakano, Y. (2003). Presence of glyoxylate cycle enzymes in the mitochondria of *Euglena gracilis*. *J Eukaryot Microbiol* **50**, 92-96.

275. Liu, F., Thatcher, J. D., Barral, J. M. & Epstein, H. F. (1995). Bifunctional glyoxylate cycle protein of *Caenorhabditis elegans*: A developmentally regulated protein of intestine and muscle. *Dev Biol* **169**, 399-414.

276. Nakazawa, M., Minami, T., Teramura, K., Kumamoto, S., Hanato, S., Takenaka, S., Ueda, M., Inui, H., Nakano, Y. & Miyatake, K. (2005). Molecular characterization of a bifunctional glyoxylate cycle enzyme, malate synthase/isocitrate lyase, in *Euglena gracilis*. *Comp Biochem Physiol B Biochem Mol Biol* **141**, 445-452.

277. Kondrashov, F. A., Koonin, E. V., Morgunov, I. G., Finogenova, T. V. & Kondrashova, M. N. (2006). Evolution of glyoxylate cycle enzymes in Metazoa: evidence of multiple horizontal transfer events and pseudogene formation. *Biol Direct* **1**, 31.

278. Xu, H., Andi, B., Qian, J., West, A. H. & Cook, P. F. (2006). The α-aminoadipate pathway for lysine biosynthesis in fungi. *Cell Biochem Biophys* **46**, 43-64.

279. Torruella, G., Suga, H., Riutort, M., Peretó, J. & Ruiz-Trillo, I. (2009). The evolutionary history of lysine biosynthesis pathways within eukaryotes. *J Mol Evol* **69**, 240-248.

280. Torrents, E., Trevisiol, C., Rotte, C., Hellman, U., Martin, W. & Reichard, P. (2006). *Euglena gracilis* ribonucleotide reductase. The eukaryote class II enzyme and the possible antiquity of eukaryote B12 dependence. *J Biol Chem* **281**, 5604-5611.

281. Kaguni, L. S. (2004). DNA polymerase γ, the mitochondrial replicase. *Annu Rev Biochem* **73**, 293-320.

282. Elo, A., Lyznik, A., Gonzalez, D. O., Kachman, S. D. & Mackenzie, S. A. (2003). Nuclear genes that encode mitochondrial proteins for DNA and RNA metabolism are clustered in the *Arabidopsis* genome. *Plant Cell* **15**, 1619-1631.

283. Moriyama, T., Terasawa, K., Fujiwara, M. & Sato, N. (2008). Purification and characterization of organellar DNA polymerases in the red alga *Cyanidioschyzon merolae*. *FEBS J* **275**, 2899-2918.

284. Moriyama, T., Terasawa, K. & Sato, N. (2011). Conservation of POPs, the plant organellar DNA polymerases, in eukaryotes. *Protist* **162**, 177-187.

285. Shutt, T. E. & Gray, M. W. (2006). Homologs of mitochondrial transcription factor B, sparsely distributed within the eukaryotic radiation, are likely derived from the dimethyladenosine methyltransferase of the mitochondrial endosymbiont. *Mol Biol Evol* **23**, 1169-1179.

286. Small, I. D. & Peeters, N. (2000). The PPR motif - a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem Sci* **25**, 45-47.

287. Prikryl, J., Rojas, M., Schuster, G. & Barkan, A. (2011). Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proc Natl Acad Sci USA* **108**, 415-420.

288. Knoop, V. & Rüdinger, M. (2010). DYW-type PPR proteins in a heterolobosean protist: Plant RNA editing factors involved in an ancient horizontal gene transfer? *FEBS Lett* **584**, 4287-4291.

289. Pusnik, M., Small, I., Read, L. K., Fabbro, T. & Schneider, A. (2007). Pentatricopeptide repeat proteins in *Trypanosoma brucei* function in mitochondrial ribosomes. *Mol Cell Biol* **27**, 6876-6888.

290. Cock, J. M., Sterck, L., Rouzé, P., Scornet, D., Allen, A. E., Amoutzias, G., Anthouard, V., Artiguenave, F., Aury, J. M. & Badger, J. H. (2010). The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature* **465**, 617-621.

291. Salone, V., Rüdinger, M., Polsakiewicz, M., Hoffmann, B., Groth-Malonek, M., Szurek, B., Small, I., Knoop, V. & Lurin, C. (2007). A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Lett* **581**, 4132-4138.

292. O'Brien, T. W. (2003). Properties of human mitochondrial ribosomes. *IUBMB Life* **55**, 505-13.

293. de la Cruz, V. F., Lake, J. A., Simpson, A. M. & Simpson, L. (1985). A minimal ribosomal RNA: sequence and secondary structure of the 9S kinetoplast ribosomal RNA from *Leishmania tarentolae*. *Proc Natl Acad Sci USA* **82**, 1401-1405.

294. Smirnov, A., Comte, C., Mager-Heckel, A.-M., Addis, V., Krasheninnikov, I. A., Martin, R. P., Entelis, N. & Tarassov, I. (2010). Mitochondrial enzyme rhodanese is essential for 5 S ribosomal RNA import into human mitochondria. *J Biol Chem* **285**, 30792-30803.

295. Gan, X., Kitakawa, M., Yoshino, K.-i., Oshiro, N., Yonezawa, K. & Isono, K. (2002). Tag-mediated isolation of yeast mitochondrial ribosome and mass spectrometric identification of its new components. *Eur J Biochem* **269**, 5203-5214.

296. Matthews, D. E., Hessler, R. A., Denslow, N. D., Edwards, J. S. & O'Brien, T. W. (1982). Protein composition of the bovine mitochondrial ribosome. *J Biol Chem* **257**, 8788-8794.

297. Zíková, A., Panigrahi, A. K., Dalley, R. A., Acestor, N., Anupama, A., Ogata, Y., Myler, P. J. & Stuart, K. (2008). *Trypanosoma brucei* mitochondrial ribosomes. Affinity purification and component identification by mass spectrometry. *Mol Cell Proteomics* **7**, 1286-1296.

298. Graack, H.-R. & Wittmann-Liebold, B. (1998). Mitochondrial ribosomal proteins (MRPs) of yeast. *Biochem J* **329**, 433-448.

299. Sharma, M. R., Booth, T. M., Simpson, L., Maslov, D. A. & Agrawal, R. K. (2009). Structure of a mitochondrial ribosome with minimal RNA. *Proc Natl Acad Sci USA* **106**, 9637-9642.

300. Blaha, G., Stanley, R. E. & Steitz, T. A. (2009). Formation of the first peptide bond: the structure of EF-P bound to the 70S ribosome. *Science* **325**, 966.

301. Micklinghoff, J. C., Schmidt, M., Geffers, R., Tegge, W. & Bange, F. C. (2010). Analysis of expression and regulatory functions of the ribosome-binding protein TypA in *Mycobacterium tuberculosis* under stress conditions. *Arch Microbiol* **192**, 499-504.

302. Bauerschmitt, H., Funes, S. & Herrmann, J. M. (2008). The membrane-bound GTPase Guf1 promotes mitochondrial protein synthesis under suboptimal conditions. *J Biol Chem* **283**, 17139-17146.

303. Caldon, C. E., Yoong, P. & March, P. E. (2001). Evolution of a molecular switch: universal bacterial GTPases regulate ribosome function. *Mol Microbiol* **41**, 289-297.

304. Tsuboi, M., Morita, H., Nozaki, Y., Akama, K., Ueda, T., Ito, K., Nierhaus, K. H. & Takeuchi, N. (2009). EF-G2mt is an exclusive recycling factor in mammalian mitochondrial protein synthesis. *Mol Cell* **35**, 502-510.

305. Curnow, A. W., Hong, K., Yuan, R., Kim, S., Martins, O., Winkler, W., Henkin, T. M. & Söll, D. (1997). Glu-tRNA^{Gln} amidotransferase: a novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. *Proc Natl Acad Sci USA* **94**, 11819.

306. Nagao, A., Suzuki, T., Katoh, T., Sakaguchi, Y. & Suzuki, T. (2009). Biogenesis of glutaminyl-mt tRNA^{Gln} in human mitochondria. *Proc Natl Acad Sci USA* **106**, 16209-16214.

307. Frechin, M., Senger, B., Brayé, M., Kern, D., Martin, R. P. & Becker, H. D. (2009). Yeast mitochondrial Gln-tRNA^{Gln} is generated by a GatFAB-mediated transamidation pathway involving Arc1p-controlled subcellular sorting of cytosolic GluRS. *Genes Dev* **23**, 1119-1130.

308. Pujol, C., Bailly, M., Kern, D., Maréchal-Drouard, L., Becker, H. & Duchêne, A.-M. (2008). Dual-targeted tRNA-dependent amidotransferase ensures both mitochondrial and chloroplastic Gln-tRNA^{Gln} synthesis in plants. *Proc Natl Acad Sci USA* **105**, 6481-6485.

309. Lonergan, K. M. & Gray, M. W. (1994). The ribosomal RNA gene region in *Acanthamoeba castellanii* mitochondrial DNA. A case of evolutionary transfer of introns between mitochondria and plastids? *J Mol Biol* **239**, 476-499.

310. Lu, M. & Steitz, T. A. (2000). Structure of *Escherichia coli* ribosomal protein L25 complexed with a 5S rRNA fragment at 1.8-Å resolution. *Proc Natl Acad Sci USA* **97**, 2023-2028.

311. Bullerwell, C. E., Burger, G., Gott, J. M., Kourennaia, O., Schnare, M. N. & Gray, M. W. (2010). Abundant 5S rRNA-like transcripts encoded by the mitochondrial genome in Amoebozoa. *Eukaryot Cell* **9**, 762-773.

312. Spencer, D. F., Bonen, L. & Gray, M. W. (1981). Primary sequence of wheat mitochondrial 5S ribosomal ribonucleic acid: functional and evolutionary implications. *Biochemistry* **20**, 4022-4029.

313. Wolff, G., Plante, I., Lang, B. F., Kück, U. & Burger, G. (1994). Complete sequence of the mitochondrial DNA of the chlorophyte alga *Prototheca wickerhamii*. Gene content and genome organization. *J Mol Biol* **237**, 75-86.

314. Ohta, N., Sato, N. & Kuroiwa, T. (1998). Structure and organization of the mitochondrial genome of the unicellular red alga *Cyanidioschyzon merolae* deduced from the complete nucleotide sequence. *Nucleic Acids Res* **26**, 5190-298.

315. Chacinska, A., Koehler, C. M., Milenkovic, D., Lithgow, T. & Pfanner, N. (2009). Importing mitochondrial proteins: machineries and mechanisms. *Cell* **138**, 628-644.

316. Tsaousis, A. D., Gaston, D., Stechmann, A., Walker, P. B., Lithgow, T. & Roger, A. J. (2011). A functional Tom70 in the human parasite *Blastocystis* sp.: implications for the evolution of the mitochondrial import apparatus. *Mol Biol Evol* **28**, 781-791.

317. Neupert, W. (1997). Protein import into mitochondria. *Annu Rev Biochem* **66**, 863-917.

318. Glaser, E., Sjöling, S., Tanudji, M. & Whelan, J. (1998). Mitochondrial protein import in plants. Signals, sorting, targeting, processing and regulation. *Plant Mol Biol* **38**, 311-338.

319. Moberg, P., Nilsson, S., Ståhl, A., Eriksson, A.-C., Glaser, E. & Mäler, L. (2004). NMR solution structure of the mitochondrial $F_1\beta$ presequence from *Nicotiana plumbaginifolia*. *J Mol Biol* **336**, 1129-1140.

320. Maćašev, D., Whelan, J., Newbigin, E., Silva-Filho, M. C., Mulhern, T. D. & Lithgow, T. (2004). Tom22′, an 8 kDa *trans*-site receptor in plants and protozoans, is a conserved feature of the TOM complex that appeared early in the evolution of eukaryotes. *Mol Biol Evol* **21**, 1557-1564.

321. Wojtkowska, M., Szczech, N., Stobienia, O., Jarmuszkiewicz, W., Budzinska, M. & Kmita, H. (2005). An inception report on the TOM complex of the amoeba *Acanthamoeba castellanii*, a simple model protozoan in mitochondria studies. *J Bioenerg Biomembr* **37**, 261-268.

322. Kozjak, V., Wiedemann, N., Milenkovic, D., Lohaus, C., Meyer, H. E., Guiard, B., Meisinger, C. & Pfanner, N. (2003). An essential role of Sam50 in the protein sorting and assembly machinery of the mitochondrial outer membrane. *J Biol Chem* **278**, 48520-48523.

323. Paschen, S. A., Waizenegger, T., Stan, T., Preuss, M., Cyrklaff, M., Hell, K., Rapaport, D. & Neupert, W. (2003). Evolutionary conservation of biogenesis of β-barrel membrane proteins. *Nature* **426**, 862-866.

324. Chan, N. C. & Lithgow, T. (2008). The peripheral membrane subunits of the SAM complex function codependently in mitochondrial outer membrane biogenesis. *Mol Biol Cell* **19**, 126-136.

325. Armstrong, L. C., Komiya, T., Bergman, B. E., Mihara, K. & Bornstein, P. (1997). Metaxin is a component of a preprotein import complex in the outer membrane of the mammalian mitochondrion. *J Biol Chem* **272**, 6510-6518.

326. Lister, R., Chew, O., Lee, M.-N., Heazlewood, J. L., Clifton, R., Parker, K. L., Millar, A. H. & Whelan, J. (2004). A transcriptomic and proteomic characterization of the *Arabidopsis* mitochondrial protein import apparatus and its response to mitochondrial dysfunction. *Plant Physiol* **134**, 777-789.

327. Sogo, L. F. & Yaffe, M. P. (1994). Regulation of mitochondrial morphology and inheritance by Mdm10p, a protein of the mitochondrial outer membrane. *J Cell Biol* **126**, 1361-1373.

328. Meisinger, C., Rissler, M., Chacinska, A., Szklarz, L. K. S., Milenkovic, D., Kozjak, V., Schönfisch, B., Lohaus, C., Meyer, H. E., Yaffe, M. P., Guiard, B., Wiedemann, N. & Pfanner, N. (2004). The mitochondrial morphology protein Mdm10 functions in assembly of the preprotein translocase of the outer membrane. *Dev Cell* **7**, 61-71.

329. Boldogh, I. R., Nowakowski, D. W., Yang, H.-C., Chung, H., Karmon, S., Royes, P. & Pon, L. A. (2003). A protein complex containing Mdm10p, Mdm12p, and Mmm1p links mitochondrial membranes and DNA to the cytoskeleton-based segregation machinery. *Mol Biol Cell* **14**, 4618-4627.

330. Kornmann, B. & Walter, P. (2010). ERMES-mediated ER-mitochondria contacts: molecular hubs for the regulation of mitochondrial biology. *J Cell Sci* **123**, 1389-1393.

331. Mesecke, N., Terziyska, N., Kozany, C., Baumann, F., Neupert, W., Hell, K. & Herrmann, J. M. (2005). A disulfide relay system in the intermembrane space of mitochondria that mediates protein import. *Cell* **121**, 1059-1069.

332. Allen, J. W. A., Ferguson, S. J. & Ginger, M. L. (2008). Distinctive biochemistry in the trypanosome mitochondrial intermembrane space suggests a model for stepwise evolution of the MIA pathway for import of cysteine-rich proteins. *FEBS Lett* **582**, 2817-2825.

333. Hell, K., Neupert, W. & Stuart, R. A. (2001). Oxa1p acts as a general membrane insertion machinery for proteins encoded by mitochondrial DNA. *EMBO J* **20**, 1281-1288.

334. Preuss, M., Ott, M., Funes, S., Luirink, J. & Herrmann, J. M. (2005). Evolution of mitochondrial Oxa proteins from bacterial YidC. Inherited and acquired functions of a conserved protein insertion machinery. *J Biol Chem* **280**, 13004-13011.

335. Neupert, W. & Herrmann, J. M. (2007). Translocation of proteins into mitochondria. *Annu Rev Biochem* **76**, 723 - 749.

336. Branda, S. S. & Isaya, G. (1995). Prediction and identification of new natural substrates of the yeast mitochondrial intermediate peptidase. *J Biol Chem* **270**, 27366-27373.

337. Pratje, E., Mannhaupt, G., Michaelis, G. & Beyreuther, K. (1983). A nuclear mutation prevents processing of a mitochondrially encoded membrane protein in *Saccharomyces cerevisiae*. *EMBO J* **2**, 1049-1054.

338. Koppen, M. & Langer, T. (2007). Protein degradation within mitochondria: versatile activities of AAA proteases and other peptidases. *Crit Rev Biochem Mol Biol* **42**, 221-242.

339. 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 the ATP-dependent degradation of inner membrane proteins in mitochondria. *EMBO J* **15**, 4218–4229. .

340. Rohrwild, M., Coux, O., Huang, H. C., Moerschell, R. P., Yoo, S. J., Seol, J. H., Chung, C. H. & Goldberg, A. L. (1996). HslV-HslU: A novel ATP-dependent protease complex in *Escherichia coli* related to the eukaryotic proteasome. *Proc Natl Acad Sci USA* **93**, 5808-5813.

341. Kanemori, M., Nishihara, K., Yanagi, H. & Yura, T. (1997). Synergistic roles of HslVU and other ATP-dependent proteases in controlling in vivo turnover of s^{32} and abnormal proteins in *Escherichia coli*. *J Bacteriol* **179**, 7219-7225.

342. Couvreur, B., Wattiez, R., Bollen, A., Falmagne, P., Le Ray, D. & Dujardin, J.-C. (2002). Eubacterial HslV and HslU subunits homologs in primordial eukaryotes. *Mol Biol Evol* **19**, 2110-2117.

343. Ruiz-González, M. & Marín, I. (2006). Proteasome-related *HslU* and *HslV* genes typical of eubacteria are widespread in eukaryotes. *J Mol Evol* **63**, 504-512.

344. Li, Z., Lindsay, M. E., Motyka, S. A., Englund, P. T. & Wang, C. C. (2008). Identification of a bacterial-like HslVU protease in the mitochondria of *Trypanosoma brucei* and its role in mitochondrial DNA replication. *PLoS Pathog* **4**, e1000048.

345. Groll, M., Bochtler, M., Brandstetter, H., Clausen, T. & Huber, R. (2005). Molecular machines for protein degradation. *ChemBioChem* **6**, 222-256.

346. Naamati, A., Regev-Rudzki, N., Galperin, S., Lill, R. & Pines, O. (2009). Dual targeting of Nfs1 and discovery of its novel processing enzyme, Icp55. *J Biol Chem* **284**, 30200-30208.

347. Lill, R. & Mühlenhoff, U. (2008). Maturation of iron-sulfur proteins in eukaryotes: mechanisms, connected processes, and diseases. *Annu Rev Biochem* **77**, 669-700.

348. Lill, R. (2009). Function and biogenesis of iron–sulphur proteins. *Nature* **460**, 831-838.

349. Sutak, R., Dolezal, P., Fiumera, H. L., Hrdy, I., Dancis, A., Delgadillo-Correa, M., Johnson, P. J., Müller, M. & Tachezy, J. (2004). Mitochondrial-type assembly of FeS centers in the hydrogenosomes of the amitochondriate eukaryote *Trichomonas vaginalis*. *Proc Natl Acad Sci USA* **101**, 10368-73.

350. Tovar, J., León-Avila, G., Sánchez, L. B., Sutak, R., Tachezy, J., van Der Giezen, M., Hernández, M., Müller, M. & Lucocq, J. M. (2003). Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation. *Nature* **426**, 172-176.

351. Satre, M., Mattei, S., Aubry, L., Gaudet, P., Pelosi, L., Brandolin, G. & Klein, G. (2007). Mitochondrial carrier family: Repertoire and peculiarities of the cellular slime mould *Dictyostelium discoideum*. *Biochimie* **89**, 1058-1069.

352. Weidner, U., Geier, S., Ptock, A., Friedrich, T., Leif, H. & Weiss, H. (1993). The gene locus of the proton-translocating NADH: ubiquinone oxidoreductase in *Escherichia coli*. Organization of the 14 genes and relationship between the derived proteins and subunits of mitochondrial complex I. *J Mol Biol* **233**, 109-22.

353. Marques, I., Duarte, M., Assuncão, J., Ushakova, A. & Videira, A. (2005). Composition of complex I from *Neurospora crassa* and disruption of two "accessory" subunits. *Biochim Biophys Acta* **1707**, 211-220.

354. 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**, 159-69.

355. Parisi, G., Perales, M., Fornasari, M. S., Colaneri, A., González-Schain, N., Gómez-Casati, D., Zimmermann, S., Brennicke, A., Araya, A., Ferry, J. G., Echave, J. & Zabaleta, E. (2004). Gamma carbonic anhydrases in plant mitochondria. *Plant Mol Biol* **55**, 193-207.

356. Alber, B. & Ferry, J. (1994). A carbonic anhydrase from the archaeon *Methanosarcina thermophila*. *Proc Natl Acad Sci USA* **91**, 6909-6913.

357. Ferry, J. (2009). The γ class of carbonic anhydrases. *Biochim Biophys Acta* **1804**, 374-381.

358. 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.

359. Martin, V., Villarreal, F., Miras, I., Navaza, A., Haouz, A., González-Lebrero, R., Kaufman, S. & Zabaleta, E. (2009). Recombinant plant gamma carbonic anhydrase homotrimers bind inorganic carbon. *FEBS Lett* **583**, 3425-3430.

360. Soto, A., Zheng, H., Shoemaker, D., Rodriguez, J., Read, B. & Wahlund, T. (2006). Identification and preliminary characterization of two cDNAs encoding unique carbonic anhydrases from the marine alga *Emiliania huxleyi*. *Appl Environ Microbiol* **72**, 5500-5511.

361. Loftus, B., Anderson, I., Davies, R., Alsmark, U. C. M., Samuelson, J., Amedeo, P., Roncaglia, P., Berriman, M., Hirt, R., Mann, B., Nozaki, T., Suh, B., Pop, M., Duchene, M., Ackers, J., Tannich, E., Leippe, M., Hofer, M., Bruchhaus, I., Willhoeft, U., Bhattacharya, A., Chillingworth, T., Churcher, C., Hance, Z., Harris, B., Harris, D., Jagels, K., Moule, S., Mungall, K., Ormond, D., Squares, R., Whitehead, S., Quail, M., Rabbinowitsch, E., Norbertczak, H., Price, C., Wang, Z., Guillen, N., Gilchrist, C., Stroup, S., Bhattacharya, S., Lohia, A., Foster, P., Sicheritz-Ponten, T., Weber, C., Singh, U., Mukherjee, C., El-Sayed, N., Petri, W., Clark, C., Embley, T., Barrell, B., Fraser, C. & Hall, N. (2005). The genome of the protist parasite *Entamoeba histolytica*. *Nature* **433**, 865-868.

362. Perales, M., Parisi, G., Fornasari, M., Colaneri, A., Villarreal, F., González-Schain, N., Echave, J., Gómez-Casati, D., Braun, H.-P., Araya, A. & Zabaleta, E. (2004). Gamma carbonic anhydrase like complex interact with plant mitochondrial complex I. *Plant Mol Biol* **56**, 947 - 957.

363. Zimmerman, S. & Ferry, J. (2006). Proposal for a hydrogen bond network in the active site of the prototypic γ-class carbonic anhydrase. *Biochemistry* **45**, 5149-5157.

364. Tripp, B., Bell, C., Cruz, F., Krebs, C. & Ferry, J. (2004). A role for iron in an ancient carbonic anhydrase. *J Biol Chem* **279**, 6683-6687.

365. Perkins, D., Pappin, D., Creasy, D. & Cottrell, J. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567.

366. Yankovskaya, V., Horsefield, R., Törnroth, S., Luna-Chavez, C., Miyoshi, H., Léger, C., Byrne, B., Cecchini, G. & Iwata, S. (2003). Architecture of succinate dehydrogenase and reactive oxygen species generation. *Science* **299**, 700 - 704.

367. Huang, L. S., Sun, G., Cobessi, D., Wang, A. C., Shen, J. T., Tung, E. Y., Anderson, V. E. & Berry, E. A. (2006). 3-nitropropionic acid is a suicide inhibitor of mitochondrial respiration that, upon oxidation by complex II, forms a covalent adduct with a catalytic base arginine in the active site of the enzyme. *J Biol Chem* **281**, 5965 - 5972.

368. Sun, F., Huo, X., Zhai, Y., Wang, A., Xu, J., Su, D., Bartlam, M. & Rao, Z. (2005). Crystal structure of mitochondrial respiratory membrane protein complex II. *Cell* **121**, 1043 - 1057.

369. Gray, M. W. (1999). Evolution of organellar genomes. *Curr Opin Genet Dev* **9**, 678-687.

370. Gibbs, S. L. (1978). The chloroplast of *Euglena* may have evolved from symbiotic green algae. *Can J Bot* **56**, 2883 - 2889.

371. Simpson, A. G. B. (1997). The identity and composition of the Euglenozoa. *Arch Protistenkd* **148**, 318-328.

372. Moreira, D., López-García, P. & Rodríguez-Valera, F. (2001). New insights into the phylogenetic position of diplonemids: G+C content bias, differences of evolutionary rate and a new environmental sequence. *Int J Syst Evol Microbiol* **51**, 2211 - 2219.

373. Simpson, A. G. & Roger, A. J. (2004). Protein phylogenies robustly resolve the deep-level relationships within Euglenozoa. *Mol Phylogenet Evol* **30**, 201-12.

374. Ahmadinejad, N., Dagan, T. & Martin, W. (2007). Genome history in the symbiotic hybrid *Euglena gracilis*. *Gene* **402**, 35 - 39.

375. Simpson, A. G. B., Lukeš, J. & Roger, A. J. (2002). The evolutionary history of kinetoplastids and their kinetoplasts. *Mol Biol Evol* **19**, 2071 - 2083.

376. Yasuhira, S. & Simpson, L. (1997). Phylogenetic affinity of mitochondria of *Euglena gracilis* and kinetoplastids using cytochrome oxidase I and hsp60. *J Mol Evol* **44**, 341 - 347.

377. Tessier, L. H., Speck, H., Gualberto, J. M. & Grienenberger, J. M. (1997). The *cox1* gene from *Euglena gracilis*: a protist mitochondrial gene without introns and genetic code modifications. *Curr Genet* **31**, 208 - 213.

378. Tessier, L. H., Keller, M., Chan, R. L., Fournier, R., Weil, J. H. & Imbault, P. (1991). Short leader sequences may be transferred from small RNAs to pre-mature mRNAs by *trans*-splicing in *Euglena*. *EMBO J* **10**, 2621 - 2625.

379. Marande, W. & Burger, G. (2007). Mitochondrial DNA as a genomic jigsaw puzzle. *Science* **318**, 415.

380. Tasker, M., Timms, M., Hendriks, E. & Matthews, K. (2001). Cytochrome oxidase subunit VI of *Trypanosoma brucei* is imported without a cleaved presequence and is developmentally regulated at both RNA and protein levels. *Mol Microbiol* **39**, 272 - 285.

381. Cecchini, G., Schröder, I., Gunsalus, R. P. & Maklashina, E. (2002). Succinate dehydrogenase and fumarate reductase from *Escherichia coli*. *Biochim Biophys Acta* **1553**, 140 - 157.

382. Cheng, V. W. T., Ma, E., Zhao, Z., Rothery, R. A. & Weiner, J. H. (2006). The iron-sulfur clusters in *Escherichia coli* succinate dehydrogenase direct electron flow. *J Biol Chem* **281**, 27662 - 27668.

383. Hudson, J. M., Heffron, K., Kotlyar, V., Sher, Y., Maklashina, E., Cecchini, G. & Armstrong, F. A. (2005). Electron transfer and catalytic control by the iron-sulfur clusters in a respiratory enzyme, *E. coli* fumarate reductase. *J Am Chem Soc* **127**, 6977 - 6989.

384. Durnford, D. G. & Gray, M. W. (2006). Analysis of *Euglena gracilis* plastid-targeted proteins reveals different classes of transit sequences. *Eukaryot Cell* **5**, 2079 - 2091.

385. Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci USA* **95**, 5857 - 5864.

386. Adams, K. L. & Palmer, J. D. (2003). Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol Phylogenet Evol* **29**, 380-395.

387. Rizzuto, R., Sandonà, D., Capaldi, R. A. & Bisson, R. (1991). Nucleotide sequence of a cDNA coding for the mitochondrial precursor protein of cytochrome *c* oxidase subunit IV from the slime mold *Dictyostelium discoideum*. *Biochim Biophys Acta* **1090**, 125-128.

388. Muramoto, K., Hirata, K., Shinzawa-Itoh, K., Yoko-o, S., Yamashita, E., Aoyama, H., Tsukihara, T. & Yoshikawa, S. (2007). A histidine residue acting as a controlling site for dioxygen reduction and proton pumping by cytochrome *c* oxidase. *Proc Natl Acad Sci USA* **104**, 7881-7886.

389. Hoffmeister, M., van der Klei, A., Rotte, C., van Grinsven, K. W. A., van Hellemond, J. J., Henze, K., Tielens, A. G. M. & Martin, W. (2004). *Euglena gracilis* rhodoquinone:ubiquinone ratio and mitochondrial proteome differ under aerobic and anaerobic conditions. *J Biol Chem* **279**, 22422-9.

390. Rotte, C., Stejskal, F., Zhu, G., Keithly, J. S. & Martin, W. (2001). Pyruvate:NADP+ oxidoreductase from the mitochondrion of *Euglena gracilis* and from the apicomplexan *Cryptosporidium parvum*: a biochemical relic linking pyruvate metabolism in mitochondriate and amitochondriate protists. *Mol Biol Evol* **18**, 710-20.

391. Castro-Guerrero, N. A., Jasso-Chávez, R. & Moreno-Sánchez, R. (2005). Physiological role of rhodoquinone in *Euglena gracilis* mitochondria. *BBA-Bioenergetics* **1710**, 113-121.

392. Hoffmeister, M., Piotrowski, M., Nowitzki, U. & Martin, W. (2005). Mitochondrial *trans*-2-enoyl-CoA reductase of wax ester fermentation from Euglena gracilis defines a new family of enzymes involved in lipid synthesis. *J Biol Chem* **280**, 4329-4338.

393. Brzezinski, P. & Gennis, R. B. (2008). Cytochrome *c* oxidase: exciting progress and remaining mysteries. *J Bioenerg Biomembr* **40**, 521-531.

394. Shingú-Vázquez, M., Camacho-Villasana, Y., Sandoval-Romero, L., Butler, C. A., Fox, T. D. & Pérez-Martínez, X. (2010). The carboxyl-terminal end of Cox1 is required for feedback assembly regulation of Cox1 synthesis in *Saccharomyces cerevisiae* mitochondria. *J Biol Chem* **285**, 34382-34389.

395. Khalimonchuk, O., Bird, A. & Winge, D. R. (2007). Evidence for a pro-oxidant intermediate in the assembly of cytochrome oxidase. *J Biol Chem* **282**, 17442-17449.

396. Claros, M. G., Perea, J., Shu, Y., Samatey, F. A., Popot, J. L. & Jacq, C. (1995). Limitations to *in vivo* import of hydrophobic proteins into yeast mitochondria. *Eur J Biochem* **228**, 762-771.

397. Oca-Cossio, J., Kenyon, L., Hao, H. & Moraes, C. T. (2003). Limitations of allotopic expression of mitochondrial genes in mammalian cells. *Genetics* **165**, 707-720.

398. von Heijne, G. (1986). Why mitochondria need a genome. *FEBS Lett* **198**, 1-4.