**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, and that reduces FAD to FADH$_2$ in the process (although it is capable of the reverse reaction under favorable conditions). High-resolution crystal structures of Complex II from bacterial (E. coli; [1]), avian (chicken; [2]) and mammalian (pig; [3]) 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 mito-
chondrial genomes of certain protists such as red algae and jakobid flagellates [4]. 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 [5]. A monophyletic ‘Euglenozoaa’ clade comprising *Euglena* (and related euglenids) along with two aplastidic lineages, the kinetoplastids (encompassing trypanosomatids and bodonids) and the predominantly free-living diplomemids, has been postulated principally on the basis of shared ultrastructural features, including disc-shaped mitochondrial cristae and flagellar paraxonomal rods [6]. Phylogenetic reconstructions based on small subunit ribosomal RNA (SSU rRNA; [7]) and protein [8] 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 [5] (evidenced by the presence in *Euglena* of a plastid with three surrounding membranes) and possible ephemeral, cryptic endosymbioses [9], 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 [10]. Moreover, the internal branching patterns within Euglenozoa are still not completely resolved [11], although phylogenies based on conserved protein genes seem to be consistent in placing euglenids at the base of Euglenozoa, with diplomemids and kinetoplastids forming a later diverging sister group [9].

Here we 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, we 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*.

**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 *nad6*) have been identified thus far [[12,13], 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 [14] 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 (‘TTTTTTT-TCG’) of the conserved SL sequence at the 5’ end (Figures 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, we 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 2 for partial protein alignments and additional files 1 and 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 diplomemid EST libraries available in TBESTDB [http://amoebidia.bcm.umontreal.ca/pepdb/searches/login.php?bye=true](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 identifi-
fied during sequencing of the mitochondrial genome of Diplonema papillatum [15]. Nevertheless, parsimony considerations argue that SdhB is nucleus-encoded and bipartite in diplonemids as well, given phylogenetic evidence indicating that diplonemids and trypanosomatids [8] or diplonemids and euglenids [7] 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.

Both of the deduced Euglena SdhB pieces (Figure 3) are predicted to contain mitochondrial targeting peptides (mTPs). TargetP [16] predicts a mitochondrial localization for SdhB-n and SdhB-c with 91.5% and 96.1% confidence, respectively, whereas the confidence levels with MitoProtII [17] 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 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 [18].

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. [19] and in several apicomplexan parasites [20] and dinoflagellates [21] 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 [19]. 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 [19]. 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 structural perspective, the split in the SdhB sequence in Euglena and trypanosomatids occurs in a region that might be particularly tolerant of such disruption (Figure 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 [1]. SdhB from *E. coli* is organized into two

---

**Figure 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.
Figure 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 'Additional file 4: Accession numbers and allied information'.
The crystal structure suggests that it contributes a hydrogen bond to the thiol group, important in stabilizing the [4Fe-4S] cluster ligand C152 [1]. It is thought that this H-bond maintains a higher midpoint potential in the [4Fe-4S] cluster. Interestingly, Cheng et al. [23] found a direct relationship between the midpoint potential of the [4Fe-4S] cluster and the turnover rates of succinate dehydrogenase, whereas Hudson et al. [24] 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 [23]. 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 (Figure 2B, b) and a Tyr-Ala-Ala motif (Thr-Ala-Ala in Euglena) instead of Ser-Thr-Ser (Figure 2B, c). Therefore, there exist some interesting parallels between the euglenozoan SdhB and E. coli FrdB sequences, although phylogenetic analyses (see additional file 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.

Methods
Expressed sequence tags (ESTs) from E. gracilis strain Z were prepared as described in [25]. ESTs encoding Euglena SdhB were identified by a tBLASTn [26] search of the taxonomically broad EST database (TBestDB; [27]) 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). Database accession numbers are given in additional file 4. The programs TargetP [16] and MitoProt II [17] were used to assess the probability of mitochondrial localization for Euglena and trypanosome SdhB-n and SdhB-c. When using TargetP for Euglena proteins, we selected the 'Plant' organism group in order to include the possibility of plastid-targeting, whereas we selected the 'Animal' organism group for trypanosomes, 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 [28] from a de-gapped alignment of the 30-most N-terminal residues from 107 predicted *E. gracilis* mitochondrial-targeted proteins.

Conserved domains were identified by searching the Pfam and SMART databases at the SMART server [29], using *E. gracilis* SdhB-n and SdhB-c as queries. Protein alignments were constructed using Muscle v3.6 [30] 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.

**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.

**Competing interests**

The authors declare that they have no competing interests.

**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.

**Additional material**

Additional file 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 2A. This alignment includes SdhB sequences from a phylogenetically broad collection of eukaryotes. Shading of columns represents at least 70% identity.  
Click here for file  
[http://www.biomedcentral.com/content-supplementary/1756-0500-2-16-S1.pdf]

Additional file 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 2B. See additional file 1 for details.  
Click here for file  
[http://www.biomedcentral.com/content-supplementary/1756-0500-2-16-S2.pdf]

**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 Canadian Institutes of Health Research (MOP-4124).

**References**

1. Yankovskaya V, Horsefield R, Tornroth S, Luna-Chavez C, Miyoshi H, Leger C, Byrne B, Cecchini G, Iwata S. Architecture of succinate dehydrogenase and reactive oxygen species generation. Science 2003, 299:700-704.
2. Huang LS, Sun G, Cobessi D, Wang AC, Shen JT, Tung EY, Anderson VE, Berry EA: 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 2006, 281:5965-5972.
3. Sun F, Huo X, Zhai Y, Wang A, Xu J, Su D, Bartlam M, Rao Z: Crystal structure of mitochondrial respiratory membrane protein complex II. Cell 2005, 121:1043-1057.
4. Gray MV: Evolution of organellar genomes. Curr Opin Genet Dev 1999, 9:678-687.
5. Gibbs SL: The chloroplast of *Euglena* may have evolved from symbiotic green algae. Can J Bot 1978, 56:2883-2889.
6. Simpson AGB: The identity and composition of the Euglenozoa. Arch Protistenkd 1997, 148:318-328.
7. Morera D, Lopez-Garcia P, Rodriguez-Valera F: New insights into the phylogenetic position of diplomonads: G+C content bias, differences of evolutionary rate and a new environmental sequence. Int J Syst Evol Microbiol 2001, 51:2211-2219.
8. Simpson AGB, Roger AJ: Protein phylogenies robustly resolve the deep-level relationships within Euglenozoa. Mol Phylogenet Evol 2004, 30:201-212.
9. Henze K, Badr A, Wettstein M, Cerff R, Martin W: A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution. Proc Natl Acad Sci USA 1995, 92:9122-9126.
10. Ahmadinejad N, Dagan T, Martin W: Genome history in the symbiotic hybrid *Euglena gracilis*. Gene 2007, 402:35-39.
11. Simpson AGB, Lukeš J, Roger AJ: The evolutionary history of kinetoplastids and their kinetoplasts. Mol Biol Evol 2002, 19:2071-2083.

12. Yasuhira S, Simpson L: Phylogenetic affinity of mitochondria of Euglena gracilis and kinetoplastids using cytochrome oxidase I and hsp60. J Mol Evol 1997, 44:341-347.

13. Tessier LH, Speck H van der, Gualberto JM, Grienenberger JM: The coxl gene from Euglena gracilis: a protist mitochondrial gene without introns and genetic code modifications. Curr Genet 1997, 31:208-213.

14. Tessier LH, Keller M, Chan RL, Fournier R, Weil JH, Imbault P: Short leader sequences may be transferred from small RNAs to pre-mature mRNAs by trans-splicing in Euglena. EMBO J 1991, 10:2621-2625.

15. Marande W, Burger G: Mitochondrial DNA as a genomic jigsaw puzzle. Science 2007, 318:415.

16. Emanuelsson O, Nielsen H, Brunak S, von Heijne G: Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 2000, 300:1005-1016.

17. Claros MG, Vincens P: Computational method to predict mitochondrially imported proteins and their targeting sequences. Eur J Biochem 1996, 241:779-786.

18. Tasker M, Timms M, Hendriks E, Matthews K: 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 2001, 39:272-285.

19. Pérez-Martínez X, Antarramian A, Vázquez-Acevedo M, Funes S, Tolkuina E, d’Alayer J, Claros MG, Davidson E, King MP, González-Halphen D: Subunit II of cytochrome c oxidase in Chlamydomonas algae is a heterodimer encoded by two independent nuclear genes. J Biol Chem 2001, 276:11302-11309.

20. Funes S, Davidson E, Reyes-Prieto A, Magallón S, Herion P, King MP, González-Halphen D: A green algal apicoplast ancestor. Science 2002, 298:2155.

21. Waller RF, Keeling PJ: Alveolate and chlorophycean mitochondrial cox2 genes split twice independently. Gene 2006, 383:33-37.

22. Cecchini G, Schroder I, Gunsalus RP, Maklashina E: Succinate dehydrogenase and fumarate reductase from Escherichia coli. Biochim Biophys Acta 2002, 1553:140-157.

23. Cheng VWT, Pla E, Zhao Z, Rothery RA, Weiner JH: The iron-sulfur clusters in Escherichia coli succinate dehydrogenase direct electron flow. J Biol Chem 2006, 281:27662-27668.

24. Hudson JM, Heffron K, Kotlyar V, Sher Y, Maklashina E, Cecchini G, Armstrong FA: Electron transfer and catalytic control by the iron-sulfur clusters in a respiratory enzyme, E. coli fumarate reductase. J Am Chem Soc 2005, 127:6977-6989.

25. Durnford DG, Gray MW: Analysis of Euglena gracilis plastid-targeted proteins reveals different classes of transit sequences. Eukaryot Cell 2006, 5:2079-2091.

26. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997, 25:3389-3402.

27. O’Brien EA, Koski LB, Zhang Y, Yang L, Wang E, Gray MW, Burger G, Lang BF: TBestDB: a taxonomically broad database of expressed sequence tags (ESTs). Nucleic Acids Res 2007, 35:D445-D451.

28. Perez-Bercoff A, Koch J, Burglin TR: LogoBar: bar graph visualization of protein logos with gaps. Bioinformatics 2006, 22:112-114.

29. Schuler J, Milpeitz F, Bork P, Ponting CP: SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci USA 1998, 95:5857-5864.

30. Edgar RC: MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 2004, 5:113.

---

Publish with BioMed Central and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp