

Were class C iron-containing superoxide dismutases of trypanosomatid parasites initially imported into a complex plastid? A hypothesis based on analyses of their N-terminal targeting signals

A. BODYŁ^{1*} and P. MACKIEWICZ²

¹Department of Biodiversity and Evolutionary Taxonomy, Zoological Institute, University of Wrocław, ul. Przybyszewskiego 63/77, 51-148 Wrocław, Poland

²Department of Genomics, Faculty of Biotechnology, University of Wrocław, ul. Przybyszewskiego 63/77, 51-148 Wrocław, Poland

(Received 6 January 2008; revised 14 April 2008; accepted 30 April 2008; first published online 14 July 2008)

SUMMARY

Trypanosomatid parasites possess 2 distinct iron-containing superoxide dismutases (Fe-SODs) designated SODA and SODC, both of which are targeted to their mitochondria. In contrast to SODAs that carry typical mitochondrial transit peptides, SODCs have highly unusual mitochondrial targeting signals. Our analyses clearly show that these pre-sequences are bipartite possessing a signal peptide-like domain followed by a transit peptide-like domain. Consequently, they resemble N-terminal extensions of proteins targeted to multi-membrane plastids, suggesting that trypanosomatids once contained a eukaryotic alga-derived plastid. Further support for this hypothesis comes from striking similarities in length, hydropathy profile, and amino acid composition of SODC pre-sequences to those of *Euglena* and dinoflagellate plastid proteins. To account for these data, we propose that the Trypanosomatidae initially possessed a gene encoding a mitochondrial Fe-SOD with a classical mitochondrial transit peptide. Before or after plastid acquisition, a gene duplication event gave rise to SODA and SODC. In a subsequent evolutionary step a signal peptide was linked to SODC, enabling its import into the plastid. When the trypanosomatid plastid subsequently was lost, natural selection favoured adaptation of the SODC N-terminal signal as a mitochondrial transit peptide and re-targeting to the mitochondrion.

Key words: dinoflagellates, endosymbiosis, *Euglena*, *Leishmania*, plastid, protein targeting, reactive oxygen species, superoxide dismutase, *Trypanosoma*.

INTRODUCTION

Oxygen generates an array of reactive oxygen species (ROS) such as the superoxide radical (O_2^-), which are scavenged by specialized antioxidant machinery in aerobic organisms (Scandalios, 2005). The key components of this machinery are superoxide dismutases (SODs) that catalyse the disproportionation of superoxide radicals to hydrogen peroxide and oxygen (for a review see Wolfe-Simon *et al.* 2005). There are 3 main types of these enzymes: Cu/Zn-, Mn-, and Fe-SODs. While all 3 varieties are used in higher eukaryotes, such as plants and animals, Fe-SODs predominate in protozoans including parasitic lineages (see, for example, Fink and Scandalios, 2002; Brydges and Carruthers, 2003).

One of these parasitic lineages is the Trypanosomatidae, which contains medically important

species such as *Trypanosoma brucei* (the causative agent of African sleeping sickness) and *Leishmania donovani* (Kala Azar or visceral leishmaniasis) (for a recent review see Simpson *et al.* 2006). Molecular phylogenetic analyses clearly demonstrate that these parasites are closely related to bodonids and diplomonids (Simpson *et al.* 2006). Interestingly, the sister group to the collective trypanosomatid, bodonid, diplomonid lineage is the Euglenoidea, a class of protists with many photosynthetic species such as *Euglena gracilis* (Simpson *et al.* 2006). Based on these evolutionary relationships, Hannaert *et al.* (2003) suggested that trypanosomatid parasites harboured a eukaryotic alga-derived plastid in their evolutionary past (see also Martin and Borst, 2003).

T. brucei has 4 Fe-containing superoxide dismutases, designated as TbSODA, TbSODB1, TbSODB2, and TbSODC (Dufernez *et al.* 2006; Wilkinson *et al.* 2006). TbSODB1 has no targeting signal, TbSODB2 carries a C-terminal peroxisome/glycosome import sequence, while TbSODA and TbSODC have N-terminal extensions that resemble mitochondrial-targeting signals (Dufernez *et al.* 2006; Wilkinson *et al.* 2006). In agreement with

* Corresponding author: Department of Biodiversity and Evolutionary Taxonomy, Zoological Institute, University of Wrocław, ul. Przybyszewskiego 63/77, 51-148 Wrocław, Poland. Fax: +48 71 322 28 17. E-mail: bodyl@biol.uni.wroc.pl

these findings, targeting studies using green fluorescent protein (GFP) demonstrated that TbSODB1 localizes to the cytosol, TbSODB2 to glycosomes, and TbSODA and TbSODC to mitochondria (Dufernez *et al.* 2006; Wilkinson *et al.* 2006).

In this paper, we investigate the N-terminal extensions of Fe-SODCs of *Trypanosoma* and *Leishmania* in detail. We demonstrate that they carry bipartite targeting sequences composed of a signal peptide-like domain followed by a transit peptide-like domain. Since comparable pre-sequences are found in proteins targeted to multi-membrane plastids, we propose that this class of dismutases was initially imported into a plastid present in the ancestors of trypanosomatids. When plastids later were lost from this parasitic lineage, selection must have favoured re-direction of SODC into the mitochondrion.

MATERIALS AND METHODS

Sequences analysed

SOD sequences from *Trypanosoma* and *Leishmania* were identified by BLAST searches of sequence databases and genome projects, including GenBank (<http://www.ncbi.nlm.nih.gov>), GeneDB (<http://www.genedb.org>), and The Wellcome Trust Sanger Institute (<http://www.sanger.ac.uk>). A non-redundant set of 11 A and 9 C dismutases was selected for further analyses (see Table 1S in Supplementary material, in Online version only).

Based on investigations of known targeting/localization of trypanosomatid proteins we assembled a set of 48 trypanosomatid nuclear-encoded, mitochondrial matrix-targeted proteins from two databases: GenBank (<http://www.ncbi.nlm.nih.gov>) and UniProt (<http://www.expasy.uniprot.org>). Annotations and accession numbers are given in Table 2S in the Supplementary material (in Online version only).

Twelve amino acid sequences of class II nuclear-encoded, plastid-targeted proteins of *E. gracilis* were inferred from EST sequences identified by Durnford and Gray (2006). We also included 15 inferred amino acid sequences of nuclear-encoded, plastid-targeted proteins from peridinin dinoflagellates. All of these proteins belong to class II and come from corresponding gene and EST sequences (for their detailed description see Patron *et al.* 2005). A complete list of *Euglena* and dinoflagellate plastid proteins used in these studies can be found in Table 3S in the Supplementary material (in Online version only).

Prediction of targeting signals and subcellular localizations

We used 30 programs and software tools to predict targeting signals and subcellular localizations of SODAs, SODCs, mitochondrial matrix-targeted proteins of the Trypanosomatidae, and *Euglena* and

dinoflagellate plastid proteins. These tools can be divided into 4 groups: (i) programs specializing in predicting signal peptides; (ii) programs that distinguish different kinds of N-terminal targeting signals, such as signal peptides, mitochondrial transit peptides, and plastid transit peptides; (iii) programs specializing in predicting mitochondrial transit peptides or mitochondrial localizations; and (iv) programs that find different subcellular localizations of a protein. A full list of these programs and references is available in Table 4S in the Supplementary material (in Online version only).

Based on the results generated, we calculated predictabilities that each sequence contained mitochondrial transit peptide and/or mitochondrial localization as well as signal peptide and/or some localization that requires presence of this peptide (extracellular, secretory pathway, ER, plasma membrane, lysosome, the Golgi apparatus). Because all these latter localizations require at least temporary residence within the endomembrane system, we refer to them collectively as endomembrane localization. Predictabilities of the targeting signals and/or intracellular localizations were expressed as percentages of given positive predictions among total potential predictions.

Calculation of substitution rates in SODCs

SODC amino acid sequences were aligned by MSA 2.1 (Gupta *et al.* 1995) and edited in GeneDoc 2.6 (Nicholas and Nicholas, 1997). A best-fit amino acid substitution model was selected by ProtTest 1.2.7 (Abascal *et al.* 2005) according to the Akaike Information Criterion (AIC), Second-order AIC (AICc), and Bayesian Information Criterion (BIC); this turned out to be the JTT+ Γ , which was used in subsequent analyses. Site substitution rates were computed with the program CODEML from the PAML 3.15 package (Yang, 1997), given the tree inferred by PHYML 2.4.4 (Guindon and Gascuel, 2003) and a JTT+ Γ model; all parameters were estimated assuming 40 rate categories for the discrete-gamma model.

Statistical analyses

The non-parametric ANOVA Kruskal-Wallis test was used to determine the statistical significance of support for competing hypotheses. These tests and correspondence analyses were carried out using Statistica software (StatSoft, Inc., 2006). The Benjamini-Hochberg multiple comparisons procedure for controlling false discovery rate (Benjamini and Hochberg, 1995) was employed in comparisons of amino acid content of pre-sequences and predictabilities of the targeting signals and/or intracellular localizations in the R 2.4.1. package (R Development Core Team, 2006).

RESULTS

Analysis of targeting signals of SODAs and SODCs from Trypanosoma and Leishmania

Fe-SODAs. Hydropathy profiles indicate that *Trypanosoma* SODAs possess an N-terminal hydrophilic domain corresponding to the entire pre-sequence (Fig. 1), suggesting that they are, indeed, imported into the mitochondrial matrix. In support of this hypothesis, 86% (values are given as median percentages) of computational predictions found a mitochondrial transit peptide and/or mitochondrial localization (Fig. 2). In contrast, predictability of a signal peptide and/or localization within the endomembrane system (e.g. ER) was only 2% (Fig. 2).

Unlike *Trypanosoma* SODAs, *Leishmania* pre-sequences possess a hydrophobic domain (Fig. 1). Concordant with this finding, a signal peptide and/or endomembrane localization was predicted for *Leishmania* SODAs at 21% (Fig. 2). Even so, prediction of a mitochondrial transit peptide and/or mitochondrial localization was considerably higher at 57% (Fig. 2). This suggests that, although the N-terminal extensions of *Leishmania* SODAs may bear some resemblance to signal peptides, their potential for mitochondrial targeting is much stronger.

Fe-SODCs. Hydropathy profile analyses of SODCs yielded unexpected results. In contrast to SODAs, SODCs carry bipartite pre-sequences composed of a hydrophobic domain followed by a hydrophilic domain. The first domain encompasses ~20 aa, whereas the second is ~70 aa (Fig. 3). The N-terminal position of the hydrophobic domain in *Trypanosoma* and *Leishmania* SODC pre-sequences is suggestive of a signal peptide. Computational programs found a signal peptide and/or endomembrane localization for these dismutases with 52% predictability (Fig. 2). In turn, a mitochondrial transit peptide and/or mitochondrial localization was recognized with lower predictability (40%) (Fig. 2).

Further clues to the character of the SODC N-terminal extensions come from programs specialized in identifying signal peptides. These programs found a signal peptide in SODCs with the predictability of 63% (quartile range: 63%–75%), while the same prediction for SODAs was effectively 0% (quartile range: 0%–25%).

Comparison of SODA and SODC pre-sequences with N-terminal extensions of trypanosomatid mitochondrial matrix-targeted proteins

The data presented above indicate that SODAs carry typical mitochondrial targeting signals, whereas those of SODCs appear unusual for mitochondrial-targeted proteins. Therefore, we further investigated

the possibility that SODC N-terminal extensions fall within the normal diversity of trypanosomatid mitochondrial targeting signals. To test this hypothesis, we compared the pre-sequences of SODAs and SODCs with those of trypanosomatid mitochondrial matrix-targeted proteins, focusing on length, hydropathy profile, as well as predictions of targeting signals and intracellular localizations.

The N-terminal extensions of trypanosomatid mitochondrial matrix-targeted proteins are composed of 7–52 aa (Table 1). As expected, at 31–36 aa, SODA pre-sequences fall within an ideal length range for mitochondrial matrix proteins ($P=0.16$). In contrast, the N-terminal extensions of SODCs contain 90–101 aa (Table 1). This is significantly longer (almost 2×) than the longest pre-sequences of other determined matrix proteins ($P<0.0001$).

Although hydropathy profiles of the N-terminal extensions of trypanosomatid mitochondrial matrix-targeted proteins have somewhat varied patterns, their common feature is the presence of one-domain targeting signals (see Fig. 1S in Supplementary material, in Online version only). In this way, they clearly resemble SODAs. By contrast, none of the 48 trypanosomatid mitochondrial matrix-targeting sequences analysed resembles the peculiar, bipartite pre-sequences present in SODCs (Fig. 3).

Predictability of a signal peptide and/or endomembrane localization was very low for trypanosomatid mitochondrial matrix-targeted proteins, amounting to only 5% (Fig. 2). Interestingly, computational programs found almost the same (4%) predictability for SODAs. In contrast, predictability of a signal peptide and/or endomembrane localization for SODCs was more than 10 times higher (52%). This is similar to predictions of mitochondrial transit peptide and/or mitochondrial localization for mitochondrial matrix proteins and SODAs, at 63% and 72% respectively (Fig. 2). For SODCs, however, predictability of such a peptide and/or localization was only 40%.

The above comparisons demonstrate that SODA N-terminal extensions are typical for trypanosomatid mitochondrial matrix-targeted proteins, whereas SODC pre-sequences clearly deviate from the matrix targeting signals in each of the features analysed.

N-terminal extensions of Trypanosoma and Leishmania SODCs resemble pre-sequences of euglenoid and dinoflagellate plastid proteins

Hydropathy profiles and bioinformatic analyses both distinguish 2 domains in *Trypanosoma* and *Leishmania* SODC pre-sequences. The N-terminal domains are hydrophobic and resemble signal peptides, whereas the C-terminal domains are hydrophilic and possess features of transit peptides. Interestingly, this kind of bipartite targeting signal, known as a class II pre-sequence, is present on

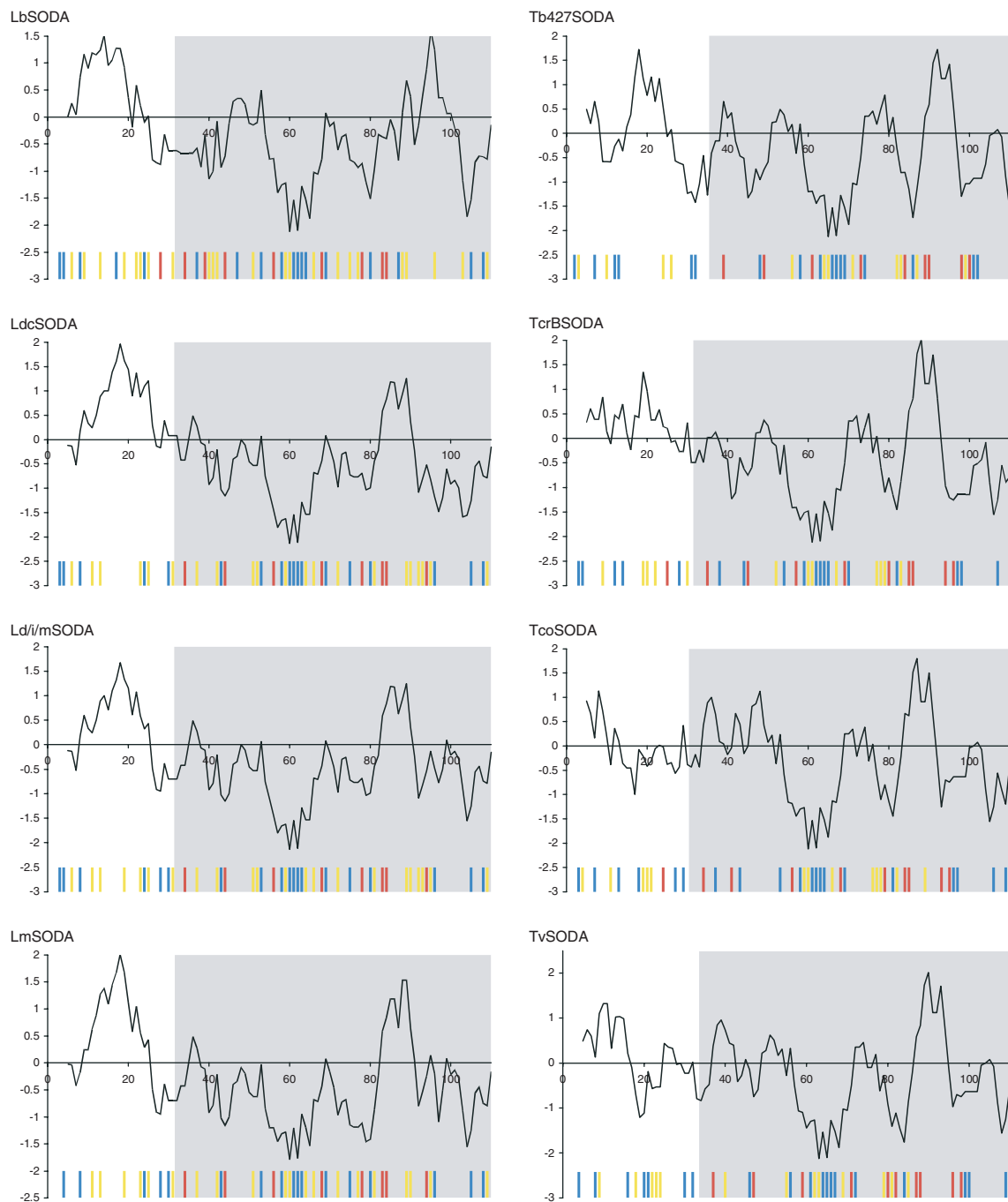


Fig. 1. Analyses of the N-terminal extensions of trypanosomatid mitochondrion-targeted SODAs, which carry single domain pre-sequences. A hydrophobic stretch is discernible in *Leishmania* pre-sequences. There are striking similarities between hydropathy profiles of SODAs and trypanosomatid mitochondrial matrix-targeted proteins (see Fig. 1S in Supplementary material, in Online version only); this is especially clear in the case of LdcSODA and AAK64280.1, Tb427SODA and O79469, TcoSODA and XP_843727.1 as well as TvSODA and XP_847200.1. The thick black landscape curve shows the hydrophobicity profiles of sequences according to the Kyte-Doolittle scale (Kyte and Doolittle, 1982) assuming a sliding window length of 9 residues. Regions present in the mature proteins are shaded in grey. Short coloured vertical lines under the profile show the distribution of particular amino acid residues grouped in 3 physicochemical classes: hydroxylated (Ser, Thr, Tyr) in yellow, basic (Arg, His, Lys) in blue, and acidic (Asp, Glu) in red.

proteins imported into multi-membrane plastids such as those of *Euglena* (Durnford and Gray, 2006) and dinoflagellates (Patron *et al.* 2005) (see also Ishida, 2005; Hempel *et al.* 2007 for reviews on other complex plastids). The N-terminal extensions of

trypanosomatid SODCs and type II plastid pre-sequences also display striking similarities in their hydropathy profiles (see Fig. 3, and Fig. 2S in Supplementary material, in Online version only). In view of these resemblances, we performed detailed

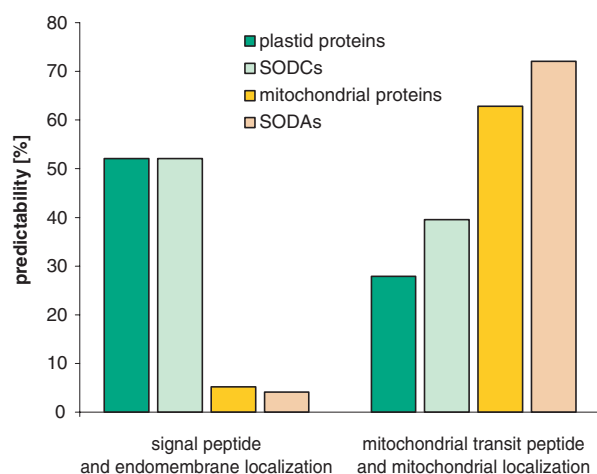


Fig. 2. Median predictability of targeting signals and subcellular localizations for *Euglena* and dinoflagellate plastid proteins, SODCs, trypanosomatid mitochondrial matrix-targeted proteins, and SODAs. Interestingly, a signal peptide and/or endomembrane localization were found with the same predictability for SODCs and plastid proteins. Moreover, the predictability of a mitochondrial transit peptide and/or mitochondrial localization for SODCs was intermediate between those for plastid proteins and mitochondrial proteins.

comparisons of the N-terminal extensions of SODCs with those of *Euglena* and dinoflagellate class II plastid proteins.

Pre-sequence lengths. The N-terminal extensions of *Euglena* and dinoflagellate plastid proteins comprise 40 to 175 aa (Table 1). Consequently, the ~100 aa pre-sequences of *Trypanosoma* and *Leishmania* SODCs fall nicely within the range of these class II plastid proteins. In fact, differences in pre-sequence length between SODCs and plastid proteins, as well as those between *Euglena* proteins and dinoflagellate proteins, are indistinguishable statistically ($P=1.0$). These striking results, along with the observation that SODC N-terminal extensions do not come close to overlapping the maximum size for mitochondrial matrix proteins and SODAs ($P<0.0001$) (see Table 1), are consistent with SODC having an alternative evolutionary history from other mitochondrial-directed proteins.

Targeting properties. Predictability of a signal peptide and/or endomembrane localization for euglenoid and dinoflagellate plastid proteins was 52% (Fig. 2). Interestingly, we found no difference in the predictabilities of such a peptide and/or localization for SODCs. This suggests that the pre-sequences of SODCs possess relatively well-preserved signal peptides. By contrast, predictability of a signal peptide and/or endomembrane localization for trypanosomatid mitochondrial matrix-targeted proteins and SODAs was more than 10 times lower (see Fig. 2).

A mitochondrial transit peptide and/or mitochondrial localization were recognized in *Euglena* and dinoflagellate plastid proteins with the predictability of 28% (Fig. 2). These results are not surprising given that these pre-sequences possess hydrophilic domains that function as N-terminal targeting signals in protein import into plastids and mitochondria (Bruce, 2001; Neupert and Herrmann, 2007). In accordance with these findings, transit peptide-like domains of many proteins imported into multi-membrane plastids have features of mitochondrial transit peptides (see, for example, Brydges and Carruthers, 2003). It is noteworthy that the 40% predictability of the mitochondrial transit peptide and/or mitochondrial localization for SODCs is intermediate between those of plastid proteins (28%) and trypanosomatid mitochondrial matrix-targeted proteins (63%) or SODAs (72%) (see Fig. 2). Thus, it is reasonable to postulate that the N-terminal extensions of SODCs originally were plastid bipartite targeting signals, which underwent a transformation into mitochondrial transit peptides.

Amino acid composition provides additional support for the plastidic nature of the N-terminal extensions of dismutases C

It is well established that targeting information of an N-terminal extension is contained in physico-chemical features of its amino acids rather than in its specific sequence order (see, for example, Bruce, 2001; Patron *et al.* 2005; Neupert and Herrmann, 2007). Therefore, we compared the contents of particular amino acids and their physico-chemical properties among pre-sequences.

Particular amino acids frequencies. The N-terminal extensions of trypanosomatid SODCs have comparable amino acid frequencies to *Euglena* and dinoflagellate class II plastid protein pre-sequences (Table 5S in Supplementary material, in Online version only). The only clear differences were for Ala ($P=0.031$) and Tyr ($P=0.011$). In contrast, the similar amino acid compositions of SODC and plastid protein pre-sequences are quite different from those of mitochondrial proteins and SODAs (see Table 5S). We identified substantial differences in Glu (E) and Asp (D) content between the N-terminal extensions of SODCs and SODAs ($P=0.015$ and $P=0.005$ for E and D, respectively) as well as those of mitochondrial proteins ($P<0.0001$ for both amino acids). In fact, pre-sequences of SODAs and other mitochondrion-targeted proteins are nearly devoid of these amino acids. Second, both SODC and plastid protein pre-sequences are characterized by reduced numbers of Met residues compared to mitochondrial proteins ($P=0.003$ and $P<0.0001$, respectively) and to SODAs ($P=0.033$

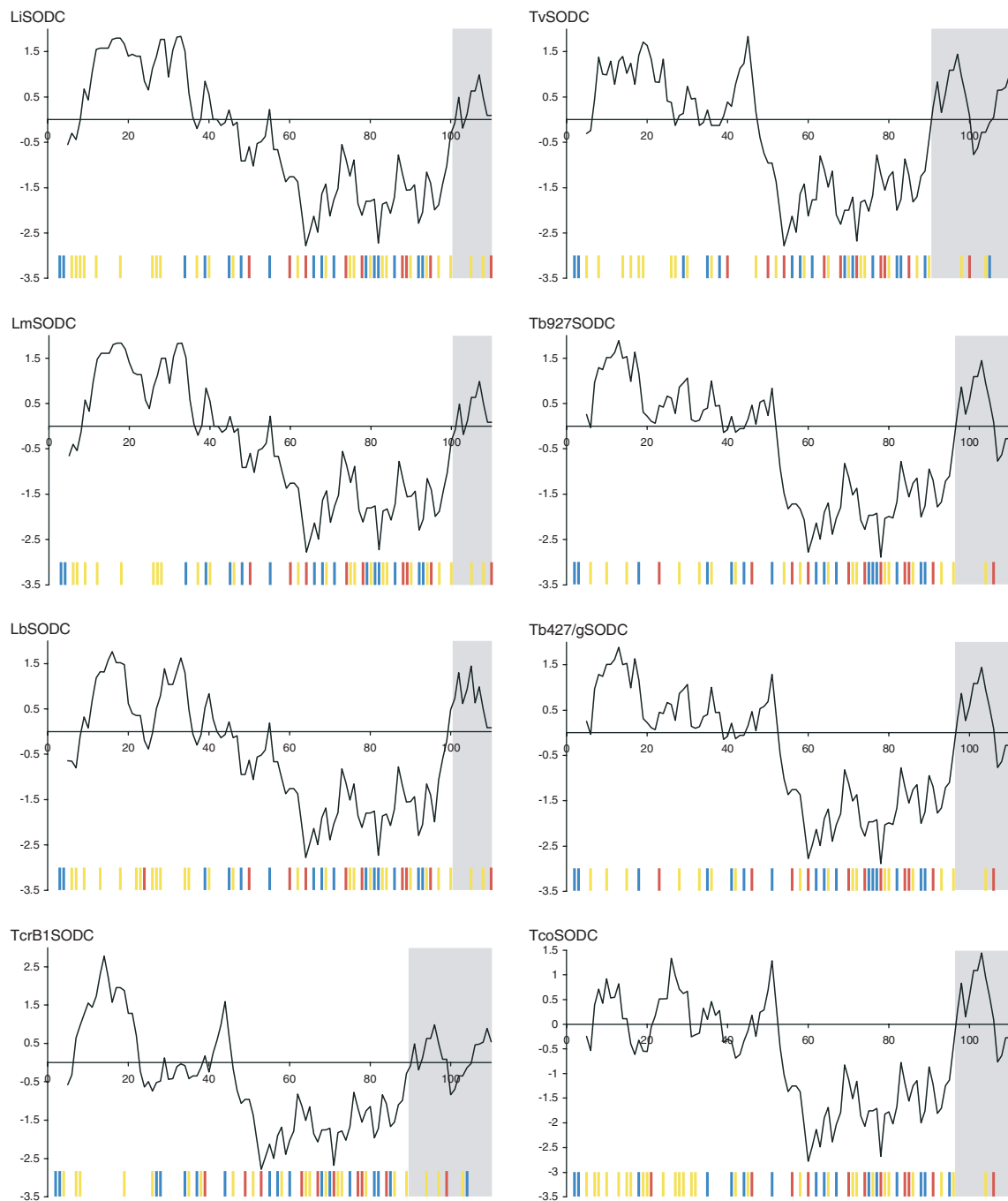


Fig. 3. Analyses of the N-terminal extensions of mitochondrion-targeted SODCs of the Trypanosomatidae, which carry bipartite pre-sequences with a hydrophobic domain followed by a hydrophilic domain. In this way, they strongly resemble N-terminal extensions of *Euglena* and dinoflagellate plastid proteins (see Fig. 2S in Supplementary material in Online version only). Please compare, for example, LiSODC with EEL00003797, LmSODC with AAW79321, TcrB1SODC with EEL00002416, and TvSODC with AAW79349.

and $P=0.006$, respectively). Third, there are significant differences in the frequencies of Tyr ($P<0.0001$) and Asp ($P=0.003$) between the N-termini of SODCs and mitochondrial proteins. Finally, SODC pre-sequences have fewer Leu residues than do SODAs ($P=0.041$).

Physico-chemical amino acids groups. Significant discrepancies in the proportions of acidic, basic, polar, and hydrophobic/non-polar amino acids also

were identified between the pre-sequences of SODCs/plastid proteins and those of mitochondrial proteins/SODAs. The differences described below are statistically significant at $P<0.01$ unless otherwise noted. First, the N-terminal extensions of SODCs and plastid proteins contain more acidic residues than those of SODAs or mitochondrial proteins (plastid proteins versus SODAs, $P=0.047$). Second, SODC pre-sequences are richer in polar residues than SODA and mitochondrial

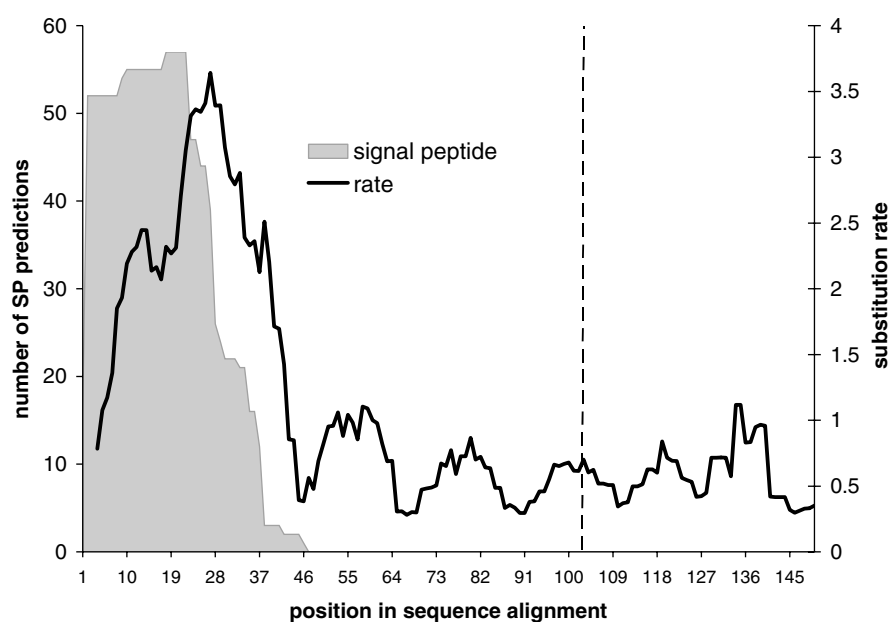


Fig. 4. Substitution rate and prediction of a signal peptide along the alignment of SODC N-terminal extensions.

Table 1. Lengths of the pre-sequences of SODAs and Cs, *Euglena* and dinoflagellate plastid proteins, and mitochondrial matrix-targeted proteins of the Trypanosomatidae given in amino acid residues

Type of pre-sequence	Minimum	Median	Maximum
Trypanosomatid mitochondrial proteins	7	17	52
SODAs	31	32	36
SODCs	90	97	101
<i>Euglena</i> plastid proteins	62	99	175
Dinoflagellate plastid proteins	40	62	109

pre-sequences ($P=0.014$). Lastly, decreased levels of hydrophobic and non-polar amino acids distinguish SODC pre-sequences from all other pre-sequence groups (versus mitochondrial proteins, $P=0.01$).

Correspondence analysis. To better visualize differences in the amino acid composition of various pre-sequences, we performed correspondence analyses (see Fig. 3S in Supplementary material, in Online version only). In agreement with our hypothesis that SODCs are descended from plastid-targeted proteins, they occur in close proximity to plastid proteins but clearly are separated from SODAs/mitochondrial proteins.

Signal peptide-like domains of Trypanosoma and Leishmania SODCs have undergone gradual degeneration

Hydropathy profiles of *Trypanosoma* and *Leishmania* SODCs show that this character is strongly

conserved in transit peptide-like domains, but highly variable in signal peptide-like domains (Fig. 3). It is even possible to identify sequences at what appear to be various stages of evolutionary degradation. In the signal peptide of *L. infantum*, the main hydrophobic domain has begun to divide into 2 smaller domains (Fig. 3). In *L. major* this division is still incomplete, whereas in *L. braziliensis* 2 separate domains clearly are resolved. A similar fragmentation of the signal peptide-like domain is observable in the SODCs of *T. congolense* and *T. vivax* (Fig. 3).

These results suggest an ongoing degeneration of the signal peptide region of trypanosomatid SODCs, a process we argue is related to their re-direction to the mitochondrion. This hypothesis is further supported in comparisons of amino acid substitution rates in different regions of SODC (Fig. 4). Inferred substitution rates in the transit peptide are the same as in mature proteins, whereas they are several times higher in domains predicted to be signal peptides with high probability. Moreover, there are numerous insertions and deletions in these signal peptide-like regions (see Fig. 4S in Supplementary material, in Online version only). Comparable indels are absent from the transit peptide domain, and these differences characterize all sequences analysed.

DISCUSSION

Trypanosomatids encode 4 iron-containing superoxide dismutases, with 2 of them, SODA and SODC, targeted to the mitochondrion (Dufernez *et al.* 2006; Wilkinson *et al.* 2006). Although transported into the same organelle, SODA and SODC differ dramatically in the lengths of their pre-sequences (see Table 1 and Dufernez *et al.* 2006; Wilkinson

et al. 2006). Considering these discrepancies, Dufernez *et al.* (2006) suggested that the two kinds of Fe-SODs are delivered to distinct mitochondrial subcompartments, with SODA transported into the mitochondrial matrix and SODC targeted to the inter-membrane space.

Our detailed bioinformatics studies support matrix residence of SODA, with pre-sequences clearly resembling those of trypanosomatid mitochondrial matrix-targeted proteins in length, amino acid composition, and targeting properties. In contrast, our analyses cast doubt on targeting of *Trypanosoma* and *Leishmania* SODC to the mitochondrial inter-membrane space. Although their pre-sequences do carry well-distinguished hydrophobic domains (Fig. 3), these are localized at their N-termini not near their centres as is typical of inter-membrane space-targeted proteins (Neupert and Herrmann, 2007). Moreover, SODCs sharply deviate from established trypanosomatid mitochondrial inner membrane/inter-membrane space-targeted proteins. For example, cytochrome c_1 in *T. brucei* is devoid of any N-terminal targeting signal, whereas the Rieske protein from *T. brucei* carries an N-terminal extension composed of only 17 aa (Priest and Hajduk, 1996, 2003).

SODC pre-sequences possess an N-terminal signal peptide followed by a hydrophilic domain (Fig. 3), the latter similar to both mitochondrial and plastid transit peptides. In this way, they resemble the N-terminal extensions of proteins transported into multi-membrane plastids (Ishida, 2005; Patron *et al.* 2005; Durnford and Gray, 2006; Hempel *et al.* 2007), which evolved from eukaryotic algae by secondary and tertiary endosymbioses (for a review see Bodył, 2005). Moreover, there are compelling similarities between pre-sequences of SODCs and those of *Euglena* and dinoflagellate plastid proteins of class II in length, hydropathy profiles, amino acid composition, and targeting properties (see Table 1 and Figs 2 and 3). Considering these resemblances, we propose that SODC originally was imported into a eukaryotic alga-derived plastid. According to this model, the Trypanosomatidae initially possessed 1 iron-containing superoxide dismutase targeted to the mitochondrial matrix (Fig. 5). The gene encoding this dismutase underwent duplication, giving rise to SODA and SODC paralogues, both targeted to the mitochondrion; this key event could have occurred either before or after acquisition of the eukaryotic plastid. Subsequent linking of a signal peptide to SODC caused it to be re-targeting to the plastid, presumably a highly advantageous adaptation. After the trypanosomatid plastid lost its ability to photosynthesize, however, its production of ROS decreased drastically, and mutations that further re-directed SODC to the mitochondrion were favoured by selection. Although the plastid has been lost completely, evidence of its prior existence lies in the

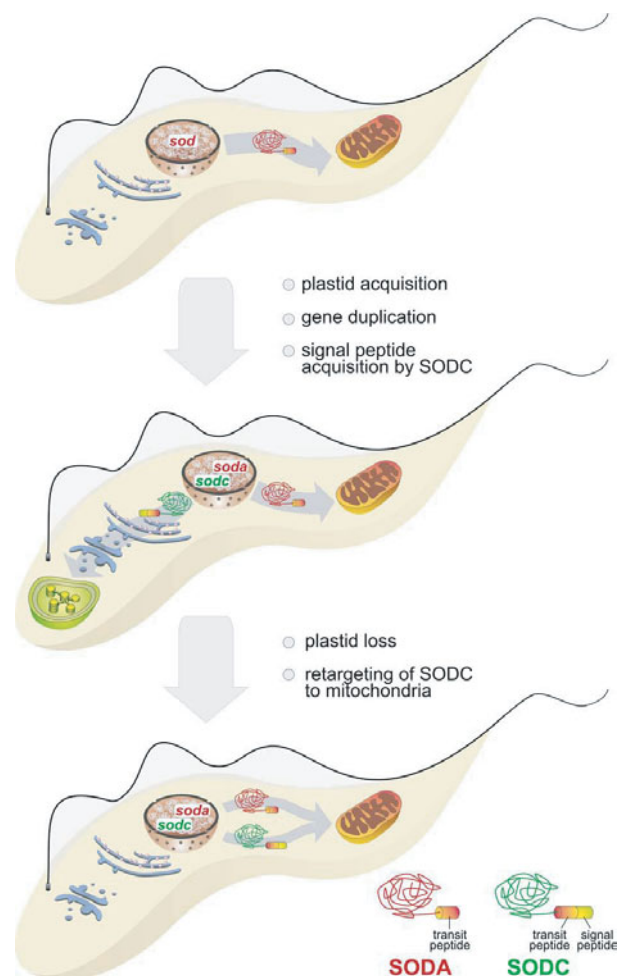


Fig. 5. Hypothetical evolutionary pathway of trypanosomatid SODs. An ancestor of trypanosomatids initially possessed a gene encoding a Fe-SOD, equipped with a classical mitochondrial transit peptide and targeted to the mitochondrion. After acquisition of a eukaryotic alga-derived plastid, this gene underwent duplication, giving rise to SODA and SODC. In a next evolutionary step, selection favoured acquisition of a signal peptide by SODC and its import into the plastid. The subsequent loss of the plastid resulted in adaptation of the N-terminal targeting signal of SODC as a mitochondrial transit peptide, and its re-direction to the mitochondrion.

peculiar, plastid-like pre-sequences of SODC proteins.

The above evolutionary scenario is most consistent with the framework provided by phylogenetic analyses of Fe-SODs, which demonstrate that SODA and SODC are paralogues that were duplicated early in trypanosomatid evolution (see Figs 5 and 6 in Dufernez *et al.* 2006). Interestingly, apicomplexan parasites possess a Fe-SOD that carries a SODC-like pre-sequence composed of a signal peptide followed by a transit peptide (Brydges and Carruthers, 2003). In accordance with our evolutionary hypothesis for SODCs, the apicomplexan

dismutase is dually targeted to the apicomplexan mitochondrion and its complex plastid surrounded by 3 or 4 membranes (Pino *et al.* 2007).

Alternative evolutionary scenarios for SODCs such as vertical inheritance from an ancient cyanobacterium-derived plastid (see Nozaki *et al.* 2003), or horizontal gene transfer from an alga with complex plastids (e.g. cryptophyte, heterokont or chlorarachniophyte; Ishida, 2005; Hempel *et al.* 2007) are inconsistent with the well-supported phylogenetic position of SODC sequences that branch with trypanosomatid SODA sequences and next with sequences of other excavates represented by trichomonads (Dufernez *et al.* 2006). These trees indicate rather a horizontal transfer of an ϵ -proteobacterial *sod* gene to the ancestor of Excavata (see Archibald *et al.* 2003) and a successive vertical inheritance of the acquired dismutase gene by all excavate taxa including trypanosomatids (Dufernez *et al.* 2006). Moreover, it could be postulated that SODCs were first targeted to the endomembrane system (e.g. ER) with the help of a signal peptide, and later re-directed to the mitochondrion through acquisition of a hydrophilic domain. However, comparison of SODs from ϵ -purple bacteria with those of eukaryotic proteins demonstrates that the sequences of bacterial proteins are strikingly similar to the trichomonad dismutases that are devoid of any pre-sequences (data not shown). Consequently, the regions encoding pre-sequences of SODAs and SODCs must have been added only in the trypanosomatid lineage, and likely functioned from the beginning as mitochondrial/plastid targeting signals. In addition, no scenario of a *de novo* origin of the SODC pre-sequence for mitochondrial targeting explains why the appended hydrophilic domain would resemble comparable domains of *Euglena* and dinoflagellate plastid proteins in length, hydrophathy profile, and amino acid composition. Taken together, available data are most consistent with the hypothesis that bipartite pre-sequences of SODCs evolved initially to function as plastid targeting signals.

Our proposed evolutionary pathway for SODC can be tested through isolation and sequencing of the SOD genes from algae with complex plastids. If they do not show a phylogenetic affinity with SODCs, this will be further support for the argument that trypanosomatid dismutases evolved vertically, as postulated by our model. Characterization of additional Fe-SODs from all other main lineages from the excavate superassemblage could supply further support of our hypothesis, by showing that they form a monophyletic clade with SODCs and SODAs, as already has been shown for trichomonad dismutases. Finally, GFP studies should be undertaken to determine whether the signal peptide-like domain in the pre-sequences of SODCs is still able to target a protein to complex plastids.

We are very grateful to Dr D. G. Durnford for supplying EST clusters of *E. gracilis* and helpful discussions about their annotation, to Dr J. W. Stiller for critical comments on a first draft of this paper and English editing, to K. Moszczyński for help with figures, and to two anonymous referees for their helpful comments and suggestions. This work was supported by funds from the research grant BW/2020/2007 to A. B.

REFERENCES

- Abascal, F., Zardoya, R. and Posada, D.** (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104–2105.
- Archibald, J. M., Longet, D., Pawlowski, J. and Keeling, P. J.** (2003). A novel polyubiquitin structure in Cercozoa and Foraminifera: evidence for a new eukaryotic supergroup. *Molecular Biology and Evolution* **20**, 62–66.
- Benjamini, Y. and Hochberg, Y.** (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **57**, 289–300.
- Bodyl, A.** (2005). Do plastid-related characters support the chromalveolate hypothesis? *Journal of Phycology* **41**, 712–719.
- Bruce, B. D.** (2001). The paradox of plastid transit peptides: conservation of function despite divergence in primary structure. *Biochimica et Biophysica Acta* **1541**, 2–21.
- Brydges, S. D. and Carruthers, V. B.** (2003). Mutation of an unusual mitochondrial targeting sequence of SODB2 produces multiple targeting fates in *Toxoplasma gondii*. *Journal of Cell Science* **116**, 4675–4685.
- Dufernez, F., Yernaux, C., Gerbod, D., Noel, C., Chauvenet, M., Wintjens, R., Edgcomb, V. P., Capron, M., Oppendoes, F. R. and Viscogliosi, E.** (2006). The presence of four iron-containing superoxide dismutase isozymes in trypanosomatidae: characterization, subcellular localization, and phylogenetic origin in *Trypanosoma brucei*. *Free Radical Biology & Medicine* **40**, 210–225.
- Durnford, D. G. and Gray, M. W.** (2006). Analysis of *Euglena gracilis* plastid-targeted proteins reveals different classes of transit sequences. *Eukaryotic Cell* **5**, 2079–2091.
- Fink, R. C. and Scandalios, J. G.** (2002). Molecular evolution and structure-function relationships of the superoxide dismutase gene families in angiosperms and their relationships to other eukaryotic and prokaryotic superoxide dismutases. *Archives of Biochemistry and Biophysics* **399**, 19–36.
- Guindon, S. and Gascuel, O.** (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**, 696–704.
- Gupta, S. K., Kececioğlu, J. and Schaffer, A. A.** (1995). Improving the practical space and time efficiency of the shortest-paths approach to sum-of-pairs multiple sequence alignment. *Journal of Computational Biology* **2**, 459–472.
- Hannaert, V., Saavedra, E., Duffieux, F., Szikora, J. P., Rigden, D. J., Michels, P. A. and Oppendoes, V.**

- F. R.** (2003). Plant-like traits associated with metabolism of *Trypanosoma* parasites. *Proceedings of the National Academy of Sciences, USA* **100**, 1067–1071.
- Hempel, F., Bozarth, A., Sommer, M. S., Zauner, S., Przyborski, J. M. and Maier, U. G.** (2007). Transport of nuclear-encoded proteins into secondarily evolved plastids. *Biological Chemistry* **388**, 899–906.
- Ishida, K.** (2005). Protein targeting into plastids: a key to understanding the symbiogenetic acquisitions of plastids. *Journal of Plant Research* **118**, 237–245.
- Kyte, J. and Doolittle, R. F.** (1982). A simple method for displaying the hydropathic character of protein. *Journal of Molecular Biology* **157**, 105–132.
- Martin, W. and Borst, P.** (2003). Secondary loss of chloroplasts in trypanosomes. *Proceedings of the National Academy of Sciences, USA* **100**, 765–767.
- Neupert, W. and Herrmann, J. M.** (2007). Translocation of proteins into mitochondria. *Annual Review of Biochemistry* **76**, 723–749.
- Nicholas, K. B. and Nicholas, H. B. Jr.** (1997). GeneDoc: a tool for editing and annotating multiple sequence alignments. *Distributed by the authors*.
- Nozaki, H., Matsuzaki, M., Takahara, M., Misumi, O., Kuroiwa, H., Hasegawa, M., Shin-i, T., Kohara, Y., Ogasawara, N. and Kuroiwa, T.** (2003). The phylogenetic position of red algae revealed by multiple nuclear genes from mitochondria-containing eukaryotes and an alternative hypothesis on the origin of plastids. *Journal of Molecular Evolution* **56**, 485–497.
- Patron, N. J., Waller, R. F., Archibald, J. M. and Keeling, P. J.** (2005). Complex protein targeting to dinoflagellate plastids. *Journal of Molecular Biology* **348**, 1015–1024.
- Pino, P., Foth, B. J., Kwok, L. Y., Sheiner, L., Schepers, R., Soldati, T. and Soldati-Favre, D.** (2007). Dual targeting of antioxidant and metabolic enzymes to the mitochondrion and the apicoplast of *Toxoplasma gondii*. *PLoS Pathogens* **3**, e115.
- Priest, J. W. and Hajduk, S. L.** (1996). *In vitro* import of the Rieske iron-sulfur protein by trypanosome mitochondria. *Journal of Biological Chemistry* **271**, 20060–20069.
- Priest, J. W. and Hajduk, S. L.** (2003). *Trypanosoma brucei* cytochrome c_1 is imported into mitochondria along an unusual pathway. *Journal of Biological Chemistry* **278**, 15084–15094.
- R Development Core Team.** (2006). *R: a Language Environment for Statistical Computing*. Technical report. R Foundation for Statistical Computing, Vienna, Austria [available online at: <http://www.r-projekt.org>].
- Scandalios, J. G.** (2005). Oxidative stress: molecular perception and transduction of signals triggering antioxidant gene defences. *Brazilian Journal of Medical and Biological Research* **38**, 995–1014.
- Simpson, A. G., Stevens, J. R. and Lukeš, J.** (2006) The evolution and diversity of kinetoplastid flagellates. *Trends in Parasitology* **22**, 168–174.
- StatSoft, Inc.** (2006). *STATISTICA (Data Analysis Software System), Version 7.1*. [available online at: <http://www.statsoft.com>].
- Wilkinson, S. R., Prathalingam, S. R., Taylor, M. C., Ahmed, A., Horn, D. and Kelly, J. M.** (2006). Functional characterisation of the iron superoxide dismutase gene repertoire in *Trypanosoma brucei*. *Free Radical Biology & Medicine* **40**, 198–209.
- Wolfe-Simon, F., Grzebyk, D., Schofield, O. and Falkowski, P. G.** (2005). The role and evolution of superoxide dismutases in algae. *Journal of Phycology* **41**, 453–465.
- Yang, Z.** (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the BioSciences* **13**, 555–556.