

Table 1S. Sequences of SODAs and SODCs from trypanosomatids (*Leishmania* and *Trypanosoma*) used in this study

(Identical sequences coming from different strains or species are described by an abbreviation.)

Abbreviation	Organism	Accession number	Source ¹
LbSODA	<i>L. braziliensis</i> M2904	LbrM08_V2.0330	GeneDB
LdcSODA	<i>L. donovani chagasi</i> Jonas	AAC38830.1	GenBank
Ld/i/mSODA	<i>L. donovani</i> 1S2D	AAQ14562.1	GenBank
Ld/i/mSODA	<i>L. infantum</i> JPCM5	LinJ08.0300	GeneDB
Ld/i/mSODA	<i>L. major</i> Friedlin	AAQ14563.1	GenBank
LmSODA	<i>L. major</i> Friedlin	CAJ02208.1	GenBank
Tb427SODA	<i>T. brucei</i> 427	AAX77683.1	GenBank
Tb927SODA	<i>T. brucei</i> TREU927	XP_845007.1	GenBank
TbgSODA	<i>T. brucei gambiense</i>	Tbgamb.9921	GeneDB
TcoSODA	<i>T. congolense</i>	congo1275h05.q1k_8	GeneDB
TcrBSODA	<i>T. cruzi</i> CL Brener	XP_812157.1	GenBank
TcrYSODA	<i>T. cruzi</i> Y	AAX84933.1	GenBank
TvSODA	<i>T. vivax</i>	Tviv1213f02.q1k	Sanger Institute
LbSODC	<i>L. braziliensis</i> M2904	LbrM32_V2.2870	GeneDB
LiSODC	<i>L. infantum</i> JPCM5	LinJ32.3130	GeneDB
LmSODC	<i>L. major</i> Friedlin	LmjF32.2630	GeneDB
Tb927SODC	<i>T. brucei</i> TREU927	XP_829639.1	GenBank
Tb427/gSODC	<i>T. brucei</i> 427	AAX77682.1	GenBank
Tb427/gSODC	<i>T. brucei gambiense</i>	Tbgamb.42456	GeneDB
TcoSODC	<i>T. congolense</i>	congo1267c10.q1k	Sanger Institute
TcrB1SODC	<i>T. cruzi</i> CL Brener	XP_804696.1	GenBank
TcrB2SODC	<i>T. cruzi</i> CL Brener	XP_815674.1	GenBank
TvSODC	<i>T. vivax</i>	Tviv1770c08.p1k_5	GeneDB

¹GeneBank: <http://www.ncbi.nlm.nih.gov>; GeneDB: <http://www.genedb.org>; The Sanger Institute: <http://www.sanger.ac.uk>.

Table 2S. Sequences of mitochondrial matrix-targeted proteins of the Trypanosomatidae used in this study¹

Accession number	Organism	Annotation	Source ²
2206467C	<i>Crithidia fasciculata</i>	histone H1-like protein p16	GenBank
AAA68599.2	<i>Crithidia fasciculata</i>	DNA polymerase beta	GenBank
AAB70749.1	<i>Crithidia fasciculata</i>	kinetoplast-associated protein p18-1	GenBank
AAC32801.1	<i>Crithidia fasciculata</i>	histone H1-like protein p21	GenBank
AAN46297.1	<i>Crithidia fasciculata</i>	histone H1-like kinetoplast DNA-binding protein p17	GenBank
AAQ88427.1	<i>Crithidia fasciculata</i>	kinetoplast DNA ligase k beta	GenBank
A4H9H8	<i>Leishmania braziliensis</i>	citrate synthase	GenBank
CAJ05692.1	<i>Leishmania major</i>	serine hydroxymethyltransferase	GenBank
CAJ07717.1	<i>Leishmania major</i>	malate dehydrogenase	GenBank
P12076	<i>Leishmania major</i>	heat shock 70-related protein 1	UniProt
P56281	<i>Leishmania tarentolae</i>	heat shock 60 protein	UniProt
Q25417	<i>Leishmania tarentolae</i>	aldehyde dehydrogenase	UniProt
AAC27101.1	<i>Trypanosoma brucei</i>	malate dehydrogenase	GenBank
AAK64278.1	<i>Trypanosoma brucei</i>	RNA-editing complex protein MP81	GenBank
AAK64279.1	<i>Trypanosoma brucei</i>	RNA-editing complex protein MP63	GenBank
AAK64280.1	<i>Trypanosoma brucei</i>	RNA-editing complex protein MP42	GenBank
AAM81963.1	<i>Trypanosoma brucei</i>	DNA polymerase I-like protein B	GenBank
AAM81964.1	<i>Trypanosoma brucei</i>	DNA polymerase I-like protein C	GenBank
AAX69289.1	<i>Trypanosoma brucei</i>	short-chain dehydrogenase	GenBank
AAX69898.1	<i>Trypanosoma brucei</i>	acyl carrier protein	GenBank
AAX79977.1	<i>Trypanosoma brucei</i>	oxidoreductase	GenBank
P82863	<i>Trypanosoma brucei</i>	RNA-editing ligase 1 MP52	UniProt
Q95W12	<i>Trypanosoma brucei</i>	RNA-editing complex protein MP18	UniProt
XP_001218997.1	<i>Trypanosoma brucei</i>	conserved hypothetical protein Tb927.1.2990	GenBank
XP_827428.1	<i>Trypanosoma brucei</i>	iron-sulfur cluster assembly protein	GenBank
XP_827446.1	<i>Trypanosoma brucei</i>	acyl transferase-like protein	GenBank
XP_827827.1	<i>Trypanosoma brucei</i>	citrate synthase	GenBank
XP_828352.1	<i>Trypanosoma brucei</i>	succinyl-CoA:3-ketoacid-coenzyme A transferase	GenBank
XP_829062.1	<i>Trypanosoma brucei</i>	2-oxoglutarate dehydrogenase E1 component	GenBank
XP_829682.1	<i>Trypanosoma brucei</i>	hypothetical protein Tb11.01.7930	GenBank
XP_843727.1	<i>Trypanosoma brucei</i>	conserved hypothetical protein Tb927.3.1330	GenBank
XP_843773.1	<i>Trypanosoma brucei</i>	pyruvate dehydrogenase E1 beta subunit	GenBank
XP_845400.1	<i>Trypanosoma brucei</i>	L-threonine 3-dehydrogenase	GenBank

XP_845557.1	<i>Trypanosoma brucei</i>	conserved hypothetical protein Tb927.6.4400	GenBank
XP_845738.1	<i>Trypanosoma brucei</i>	andtrypanothione/tryparedoxin dependent peroxidase 3	GenBank
XP_847200.1	<i>Trypanosoma brucei</i>	isocitrate dehydrogenase [NADP]	GenBank
XP_847316.1	<i>Trypanosoma brucei</i>	conserved hypothetical protein Tb927.8.4860	GenBank
XP_847436.1	<i>Trypanosoma brucei</i>	2-amino-3-ketobutyrate coenzyme A ligase	GenBank
XP_951580.1	<i>Trypanosoma brucei</i>	hypothetical protein Tb927.2.3180	GenBank
XP_951693.1	<i>Trypanosoma brucei</i>	3-oxoacyl-(ACP) reductase	GenBank
P82864	<i>Trypanosoma brucei brucei</i>	RNA-editing ligase MP48	UniProt
Q04933	<i>Trypanosoma brucei brucei</i>	dihydrolipoyl/dihydrolipoamide dehydrogenase	GenBank
Q37683	<i>Trypanosoma brucei brucei</i>	heat shock 60 protein	UniProt
O79469	<i>Trypanosoma cruzi</i>	peroxiredoxin	UniProt
P20583	<i>Trypanosoma cruzi</i>	heat shock 70 protein	UniProt
P90597	<i>Trypanosoma cruzi</i>	dihydrolipoyl dihydrolipoamide dehydrogenase	GenBank
Q95046	<i>Trypanosoma cruzi</i>	heat shock 60 protein	UniProt
XP_811987.1	<i>Trypanosoma cruzi</i> strain CL Brener	glycerol-3-phosphate dehydrogenase	GenBank

¹The N-terminal extensions of SODAs and SODCs from *Trypanosoma* species and SODAs from *Leishmania* were complete, but SODCs from *Leishmania* were shorter than those from *Trypanosoma*. However, a detailed inspection of *Leishmania* genomic sequences upstream of *sodc* genes revealed that these regions had no in-frame stop codons for a long distance, and that their inferred amino acid sequences were similar to the N-terminal regions of *Trypanosoma* SODCs. Therefore, we extended the sequences from *Leishmania* by 43 residues in these apparently homologous regions.

²GenBank: <http://www.ncbi.nlm.nih.gov>; UniProt: <http://www.expasy.uniprot.org>

Table 3S. Sequences of nuclear-encoded plastid-targeted proteins of euglenoids and dinoflagellates used in this study

Accession number	Organism	Annotation	Reference
<i>Euglenoids</i>			
EEL00000923	<i>Euglena gracilis</i>	PEP/phosphate translocator	Durnford and Gray (2006)
EEL00002060	<i>Euglena gracilis</i>	Mg-protoporphyrin IX methyltransferase	Durnford and Gray (2006)
EEL00002416	<i>Euglena gracilis</i>	peptide chain release factor, RF 2	Durnford and Gray (2006)
EEL00002660	<i>Euglena gracilis</i>	ClpB	Durnford and Gray (2006)
EEL00003294	<i>Euglena gracilis</i>	ABC transporter	Durnford and Gray (2006)
EEL00003630	<i>Euglena gracilis</i>	photosystem II protein (PsbW)	Durnford and Gray (2006)
EEL00003784	<i>Euglena gracilis</i>	phosphoribulokinase	Durnford and Gray (2006)
EEL00003797	<i>Euglena gracilis</i>	PSI subunit IV protein (PsaE)	Durnford and Gray (2006)
EEL00004932	<i>Euglena gracilis</i>	50S ribosomal protein L9	Durnford and Gray (2006)
EEL00006808	<i>Euglena gracilis</i>	squalene and phytoene synthases	Durnford and Gray (2006)
EEL00008550	<i>Euglena gracilis</i>	Short-chain (SC) dehydrogenase	Durnford and Gray (2006)
EEL00009282	<i>Euglena gracilis</i>	MECP synthase	Durnford and Gray (2006)
<i>Dinoflagellates</i>			
CK784108	<i>Alexandrium tamarense</i>	acyl carrier protein	Patron <i>et al.</i> (2005)
CF064566	<i>Amphidinium carterae</i>	ribosomal protein L16	Patron <i>et al.</i> (2005)
CF065976	<i>Amphidinium carterae</i>	ATP synthase subunit I	Patron <i>et al.</i> (2005)
CF067008	<i>Amphidinium carterae</i>	ATP synthase subunit gamma	Patron <i>et al.</i> (2005)
CF067087	<i>Amphidinium carterae</i>	cytochrome b	Patron <i>et al.</i> (2005)
CF067099	<i>Amphidinium carterae</i>	an unknown protein	Patron <i>et al.</i> (2005)
CF067331	<i>Amphidinium carterae</i>	a membrane protein	Patron <i>et al.</i> (2005)
CF067332	<i>Amphidinium carterae</i>	photosystem II protein L	Patron <i>et al.</i> (2005)
AAW79290	<i>Heterocapsa triquerta</i>	acyl carrier protein	Patron <i>et al.</i> (2005)
AAW79296	<i>Heterocapsa triquerta</i>	ATP synthase subunit gamma	Patron <i>et al.</i> (2005)
AAW79300	<i>Heterocapsa triquerta</i>	carbonic anhydrase	Patron <i>et al.</i> (2005)
AAW79321	<i>Heterocapsa triquerta</i>	phosphoribulokinase	Patron <i>et al.</i> (2005)
AAW79330	<i>Heterocapsa triquerta</i>	beta-keto acyl reductase	Patron <i>et al.</i> (2005)
AAW79335	<i>Heterocapsa triquerta</i>	dimethyladenosine synthase	Patron <i>et al.</i> (2005)
AAW79349	<i>Heterocapsa triquerta</i>	photosystem II protein L	Patron <i>et al.</i> (2005)

Table 4S. Programs predicting targeting signals and subcellular localizations used in this study.

Program name	Reference
<i>Programs specializing in the prediction of signal peptides (SP)</i>	
DetecSig in ConPred II	Lao and Shimizu (2001)
Phobius	Käll <i>et al.</i> (2004)
PrediSi	Hiller <i>et al.</i> (2004)
Sigcleave in EMBOSS 3.0.0	Rice <i>et al.</i> (2000)
SignalP-HMM 3.0	Nielsen and Krogh (1998)
SignalP-NN 3.0	Bendsten <i>et al.</i> (2004)
SIG-Pred	bioinformatics.leeds.ac.uk/prot_analysis/Signal.html
SOSUisignal	Gomi <i>et al.</i> (2004)
<i>Programs that distinguish different kinds of N-terminal targeting signals, such as signal peptides, mitochondrial transit peptides, and plastid transit peptides</i>	
BLSTM-LOC 1.0	139.91.72.10/blstm/blstm.html
iPSORT	Bannai <i>et al.</i> (2002)
Predotar 1.03	Small <i>et al.</i> (2004)
PredSL	www.195.134.85.247/PredSL/index.html
PProwler 1.2	Bodén and Hawkins (2005)
TargetLoc	Hoglund <i>et al.</i> (2006)
TargetP 1.1	Emanuelsson <i>et al.</i> (2000)
<i>Programs specializing in the prediction of mitochondrial transit peptides or mitochondrial localizations</i>	
MITOPRED	Guda <i>et al.</i> (2004)
MitoProt II 1.0a4	Claros and Vincens (1996)
SubMito	Du and Li (2006)
<i>Programs that predict different subcellular localizations of a protein</i>	
BaCelLo	Pierleoni <i>et al.</i> (2006)
CELLO 2.5	Yu <i>et al.</i> (2004)
ESLpred	Bhasin and Raghava (2004)
LOCtree	Nair and Rost (2005)
MultiLoc	Hoglund <i>et al.</i> (2006)
PLOC	Park and Kanehisa (2003)
Protcomp 6.0	www.softberry.com
pSLIP	Sarda <i>et al.</i> (2005)
PSORT II	Nakai and Horton (1999)
PSORT 6.4	Nakai and Horton (1999)
SubLoc 1.0	Hua and Sun (2001)
WoLF PSORT	Horton <i>et al.</i> (2006)

Table 5S. The percentage composition of amino acids in SODA, SODC, *Euglena* and dinoflagellate plastid protein, and trypanosomatid mitochondrial matrix-targeted protein pre-sequences

(The table contains median and quartile range (in parentheses) of amino acids for which differences between the groups were statistically significant at $P < 0.05$.)

Amino acid	Pre-sequences			
	SODCs	Plastid proteins (<i>Euglena</i> and dinoflagellates)	SODAs	Mitochondrial proteins (trypanosomatids)
Alanine (A)	7.9 (7.2-9.9)	16.7 (14.3-21.4)	13.9 (12.5-15.6)	7.9 (0.0-12.5)
Arginine (R)	11.3 (10.9-12.4)	7.4 (4.2-8.5)	9.4 (6.3-11.8)	16.7 (14.0-23.1)
Asparagine (N)	4.0 (4.0-4.1)	1.7 (0.8-2.4)	0.0 (0.0-5.6)	0.0 (0.0-1.0)
Aspartic acid (D)	4.4 (4.0-6.2)	2.4 (1.0-3.4)	0.0 (0.0-0.0)	0.0 (0.0-0.0)
Glutamic acid (E)	4.1 (4.1-4.4)	2.1 (1.0-4.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)
Glutamine (Q)	4.0 (3.3-4.1)	3.2 (1.7-6.6)	0.0 (0.0-2.8)	0.0 (0.0-4.4)
Glycine (G)	6.9 (5.6-6.9)	6.6 (5.5-8.2)	6.3 (3.1-8.3)	0.0 (0.0-5.9)
Leucine (L)	4.4 (4.0-8.2)	9.8 (7.9-12.7)	12.5 (11.1-15.6)	11.1 (5.7-16.7)
Lysine (K)	3.0 (2.2-3.1)	1.9 (1.0-3.0)	3.1 (3.1-5.6)	0.0 (0.0-3.9)
Methionine (M)	3.3 (3.1-4.0)	3.2 (1.8-5.0)	6.5 (6.3-13.9)	8.3 (5.0-11.1)
Proline (P)	4.1 (4.0-4.4)	6.0 (4.8-9.2)	8.3 (6.3-9.4)	0.0 (0.0-5.8)
Tyrosine (Y)	7.8 (7.2-7.9)	0.9 (0.0-2.1)	3.1 (2.8-6.3)	0.0 (0.0-0.0)

REFERENCES (FOR TABLES 3 AND 4)

- Bannai, H., Tamada, Y., Maruyama, O., Nakai, K. and Miyano, S.** (2002). Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics* **18**, 298-305.
- Bendsten, J. D., Nielsen, H., von Heijne, G. and Brunak, S.** (2004). Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology* **340**, 783-795.
- Bhasin, M. and Raghava, G. P.** (2004). ESLpred: SVM-based method for subcellular localization of eukaryotic proteins using dipeptide composition and PSI-BLAST. *Nucleic Acids Research* **32** (Web Server issue), W414-W419.
- Bodén, M. and Hawkins, J.** (2005). Prediction of subcellular localization using sequence-biased recurrent networks. *Bioinformatics* **21**, 2279-2286.
- Claros, M. G. and Vincens, P.** (1996). Computational method to predict mitochondrially imported proteins and their targeting sequences. *European Journal of Biochemistry* **241**, 779-786.
- Du, P. and Li, Y.** (2006). Prediction of protein submitochondria locations by hybridizing pseudo-amino acid composition with various physicochemical features of segmented sequence. *BMC Bioinformatics* **7**, 518.
- 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.
- Emanuelsson, O., Nielsen, H., Brunak, S. and von Heijne, G.** (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology* **300**, 1005-1016.
- Gomi, M., Sonoyama, M. and Mitaku, S.** (2004). High performance system for signal peptide prediction: SOSUisignal. *Chem-Bio Informatics Journal* **4**, 142-147.
- Guda, C., Guda, P., Fahy, E. and Subramaniam, S.** (2004). MITOPRED: a web server for the prediction of mitochondrial proteins. *Nucleic Acids Research* **32**, W372-W374.
- Hiller, K., Grote, G., Scheer, M., Münch, R. and Jahn, D.** (2004). PrediSi: prediction of signal peptides and their cleavage positions. *Nucleic Acids Research* **32** (Web Server issue), W375-W379.
- Horton, P., Park, K. J., Obayashi, T. and Nakai, K.** (2006). Protein subcellular localization prediction with WoLF PSORT. In *Proceedings of the 4th Asia-Pacific Bioinformatics Conference (APBC2006)* (ed. Jiang, T. *et al.*), pp. 39-48. Imperial College Press, Taipei.
- Hoglund, A., Donnes, P., Blum, T., Adolph, H. W. and Kohlbacher, O.** (2006). MultiLoc: prediction of protein subcellular localization using N-terminal targeting sequences, sequence motifs and amino acid composition. *Bioinformatics* **22**, 1158-1165.

- Hua, S. and Sun, Z.** (2001). Support vector machine approach for protein subcellular localization prediction. *Bioinformatics* **17**, 721-728.
- Käll, L., Krogh, A. and Sonnhammer, E. L. L.** (2004). A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology* **338**, 1027-1036.
- Lao, D. M. and Shimizu, T.** (2001). A method for discriminating a signal peptide and a putative 1st transmembrane segment. In *Proceedings of the 2001 International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences (METMBS'01)* (ed. Valafar, F.), pp. 119-125. CSREA Press, USA.
- Nair, R. and Rost, B.** (2005). Mimicking cellular sorting improves prediction of subcellular localization. *Journal of Molecular Biology* **348**, 85-100.
- Nakai, K. and Horton, P.** (1999). PSORT: a program for detecting the sorting signals of proteins and predicting their subcellular localization. *Trends in Biochemical Sciences* **24**, 34-36.
- Nielsen, H. and Krogh, A.** (1998). Prediction of signal peptides and signal anchors by a hidden Markov model. *Proceedings: International Conference on Intelligent Systems for Molecular Biology* **6**, 122-130.
- Park, K. J. and Kanehisa, M.** (2003). Prediction of protein subcellular locations by support vector machines using compositions of amino acids and amino acid pairs. *Bioinformatics* **19**, 1656-1663.
- 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.
- Pierleoni, A., Martelli, P. L., Fariselli, P. and Casadio, R.** (2006). BaCelLo: a balanced subcellular localization predictor. *Bioinformatics* **22**, 408-416.
- Rice, P., Longden, I. and Bleasby, A.** (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends in Genetics* **16**, 276-277.
- Sarda, D., Chua, G. H., Li, K. B. and Krishnan, A.** (2005). pSLIP: SVM based protein subcellular localization prediction using multiple physicochemical properties. *BMC Bioinformatics* **6**, 152.
- Small, I., Peeters, N., Legeai, F. and Lurin, C.** (2004). Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* **4**, 1581-1590.
- Yu, C. S., Lin, C. J. and Hwang, J. K.** (2004). Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Science* **13**, 1402-1406.

LEGENDS TO FIGURES

Fig. 1S. Analyses of N-terminal extensions of mitochondrial matrix-targeted proteins from trypanosomatids. Although these pre-sequences show some variation in their hydropathy profiles, they are relatively short and do not have the wide hydrophobic domain followed by a long hydrophilic region characteristic of N-terminal extensions of both SODCs (see Fig. 3) and *Euglena* and dinoflagellate plastid proteins (see Fig. 2S in Supplementary material). Overall, hydropathy profiles of mitochondrial matrix-targeted proteins clearly resemble those of SODAs (see Fig. 1). Please compare AAK64280.1 with LdcSODA, O79469 with Tb427SODA, XP_843727.1 with TcoSODA, and XP_847200.1 with TvSODA. For a further description see Fig. 1.

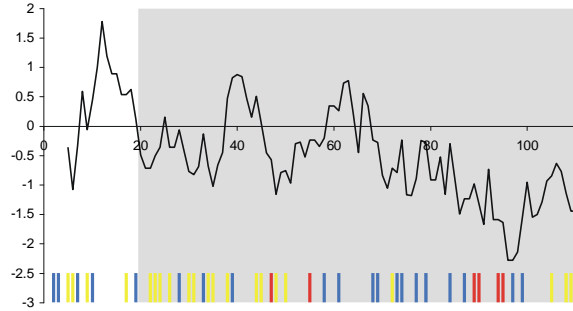
Fig. 2S. Analyses of the N-terminal extensions of proteins targeted to multimembrane plastids, represented here by *Euglena* (left panel) and dinoflagellate (right panel) proteins. Their pre-sequences have a bipartite nature with a distinct hydrophobic domain followed by a long hydrophilic domain. Interestingly, a very similar pre-sequence architecture is characteristic of SODCs (see Fig. 3). Please compare, for example, EEL00003797 with LiSODC, AAW79321 with LmSODC, EEL00002416 with TcrB1SODC, or AAW79349 with TvSODC. For a further description see Fig. 1.

Fig. 3S. Correspondence analyses of the amino acid compositions of pre-sequences of SODAs, trypanosomatid mitochondrial matrix-targeted proteins, SODCs, as well as *Euglena* and dinoflagellate plastid proteins. The analyses were performed on four physicochemical classes of amino acids: acidic (D, E), basic (H, K, R), polar (N, Q, S, T, Y) and hydrophobic or nonpolar (A, C, F, G, I, L, M, P, V, W). The first two dimensions explain in sum 82.3 % of the variance. The acidic class constitutes 54 % of the first dimension and polar class almost 49 % in the second dimension. The ellipsoids shown indicate 85 % confidence.

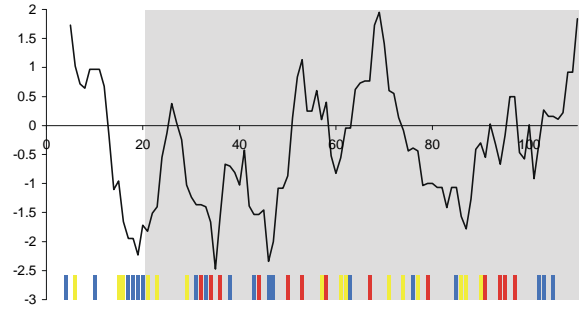
Fig. 4S. Alignment of the N-terminal extensions of SODCs. Particular residues are shaded according to their levels of conservation. The band below the alignment shows predictability of a signal peptide. The intensity of shading is proportional to the fraction of software that predicted the given site to be a signal peptide, as shown by the scale bar.

Fig. 1S

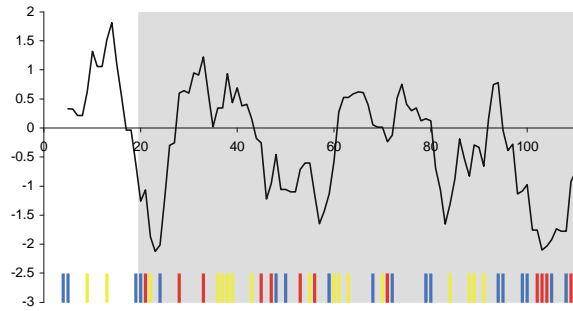
T. brucei - RNA-editing complex protein MP42 (AAK64280.1)



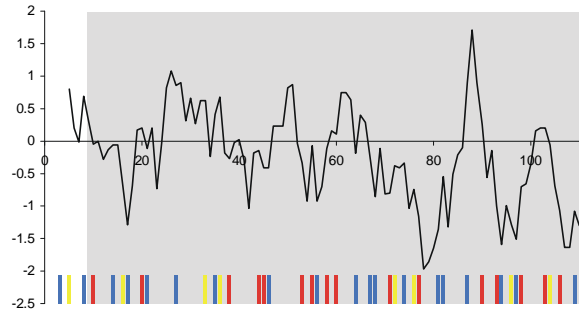
L. major - serine hydroxymethyltransferase (CAJ05692.1)



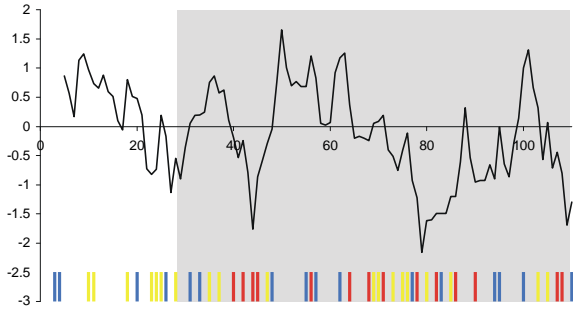
L. major - heat shock 70-related protein 1 (P12076)



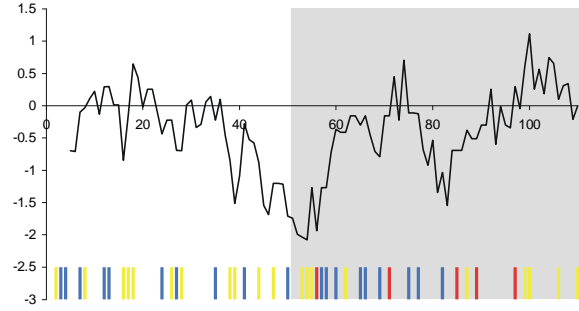
L. tarentolae - p51 Aldehyde dehydrogenase (Q25417)



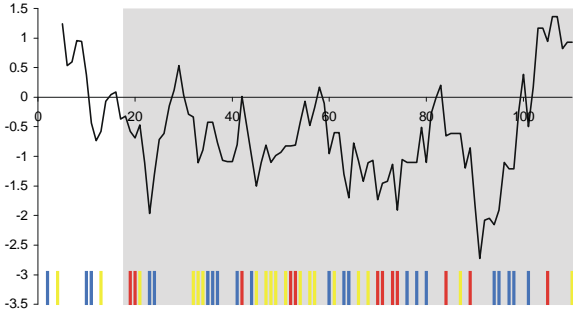
T. brucei - isocitrate dehydrogenase [NADP] (XP_847200.1)



T. brucei - RNA-editing complex protein MP63 (AAK64279.1)



T. brucei - conserved hypothetical protein (XP_843727.1)



T. cruzi - peroxiredoxin (O79469)

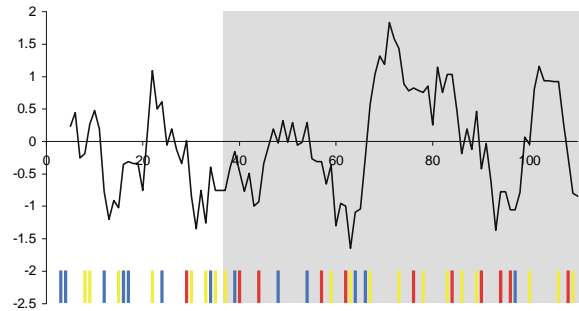
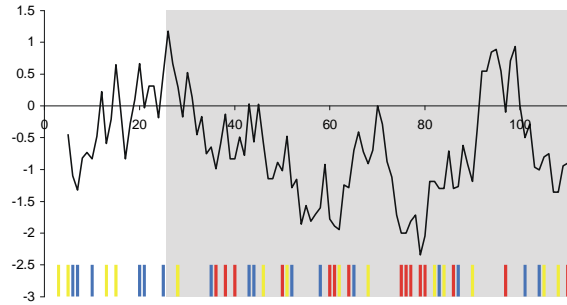
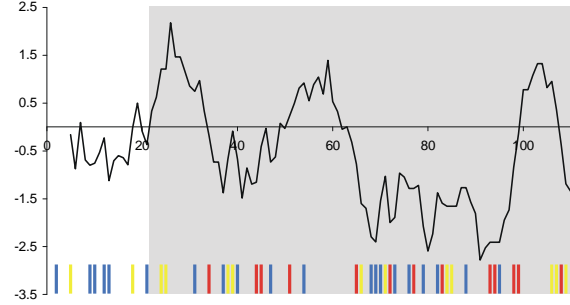


Fig. 1S continued

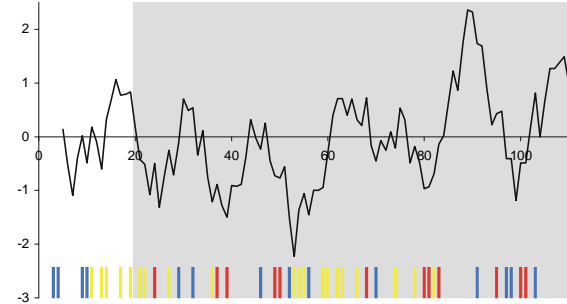
T. brucei - conserved hypothetical protein (XP_001218997.1)



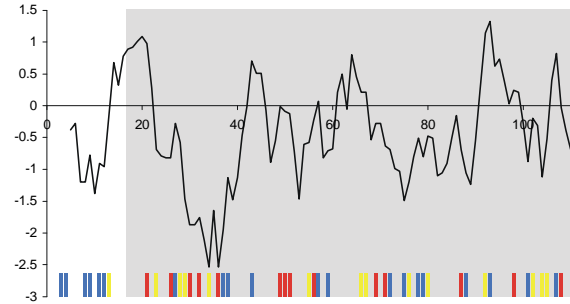
T. brucei - DNA polymerase I-like protein B (AAM81963.1)



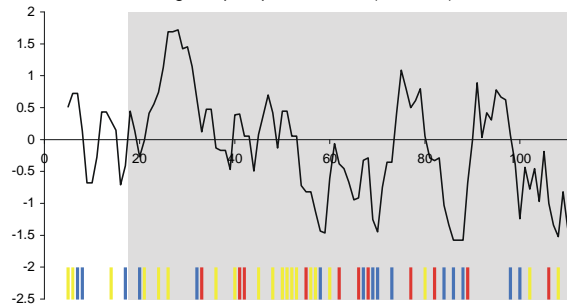
T. brucei - conserved hypothetical protein (XP_845557.1)



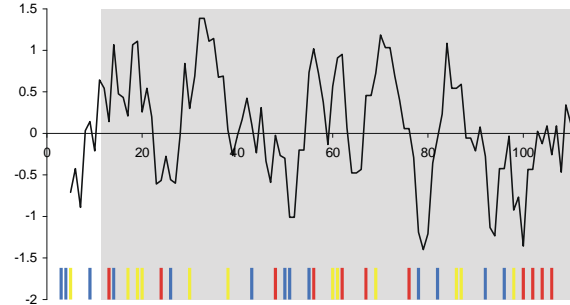
T. brucei - RNA-editing ligase TbMP48 (P82864)



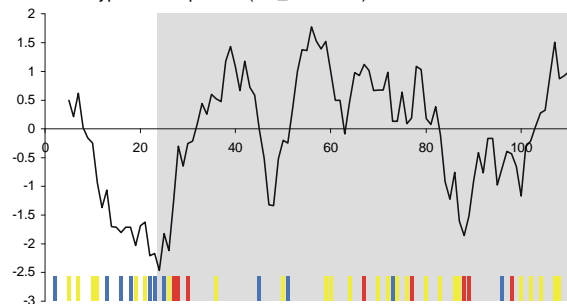
T. brucei - RNA-editing complex protein MP18 (Q95W12)



T. brucei - succinyl-coA:3-ketoacid-coenzyme A transferase (XP_828352.1)



T. brucei - hypothetical protein (XP_829682.1)



T. brucei - DNA polymerase I-like protein C (AAM81964.1)

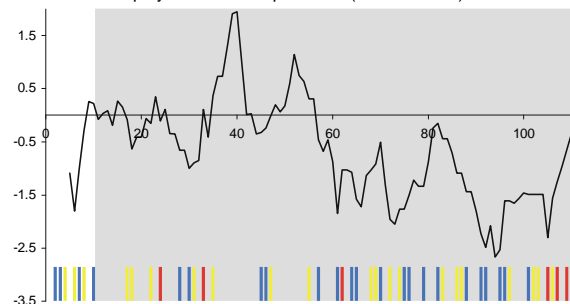
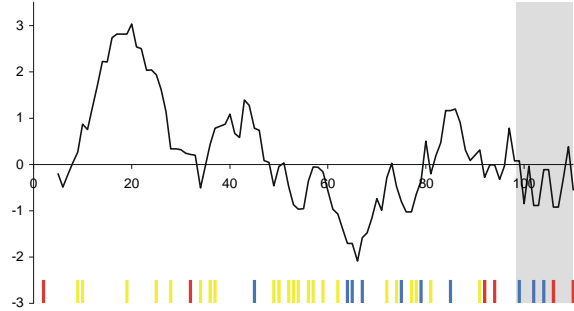
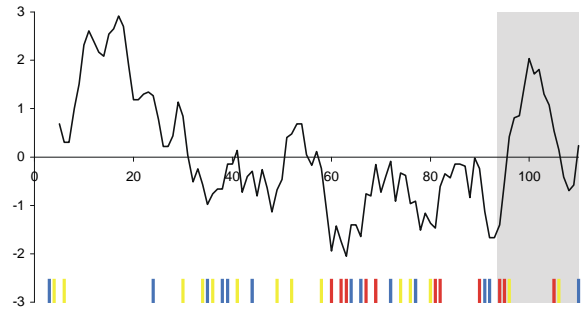


Fig. 2S

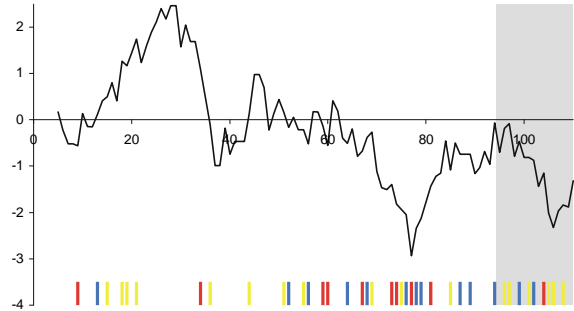
Euglena gracilis - peptide chain release factor 2 (EEL00002416)



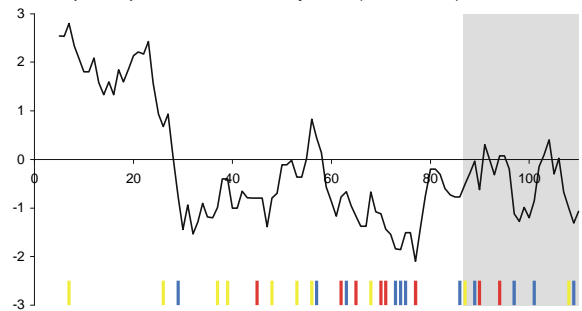
Heterocapsa triquetra - phosphoribulokinase (AAW79321)



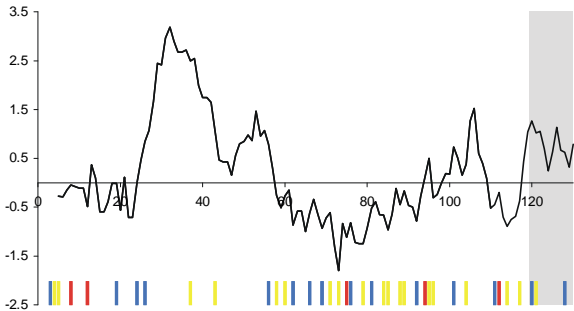
Euglena gracilis - PSI subunit IV protein, PsaE (EEL00003797)



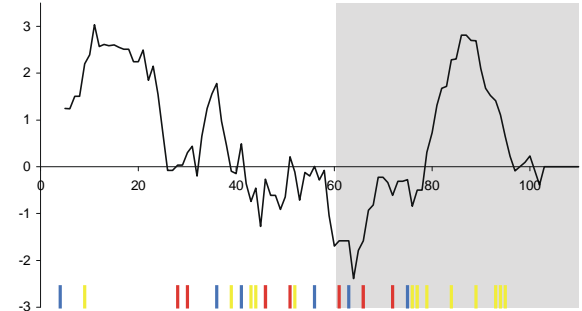
Heterocapsa triquetra - carbonic anhydrase (AAW79300)



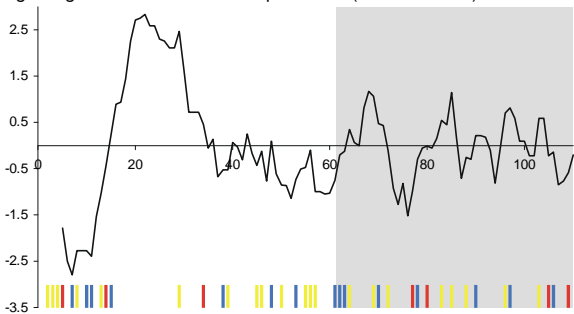
Euglena gracilis - short-chain dehydrogenase (EEL00008550)



Heterocapsa triquetra - photosystem II protein L (AAW79349)



Euglena gracilis - 50S ribosomal protein L9 (EEL00004932)



Amphidinium carterae - a membrane protein (CF067331)

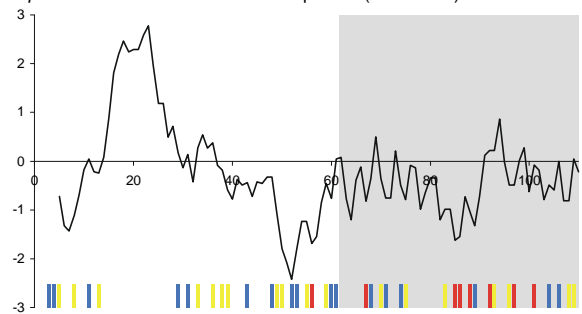


Fig. 3S

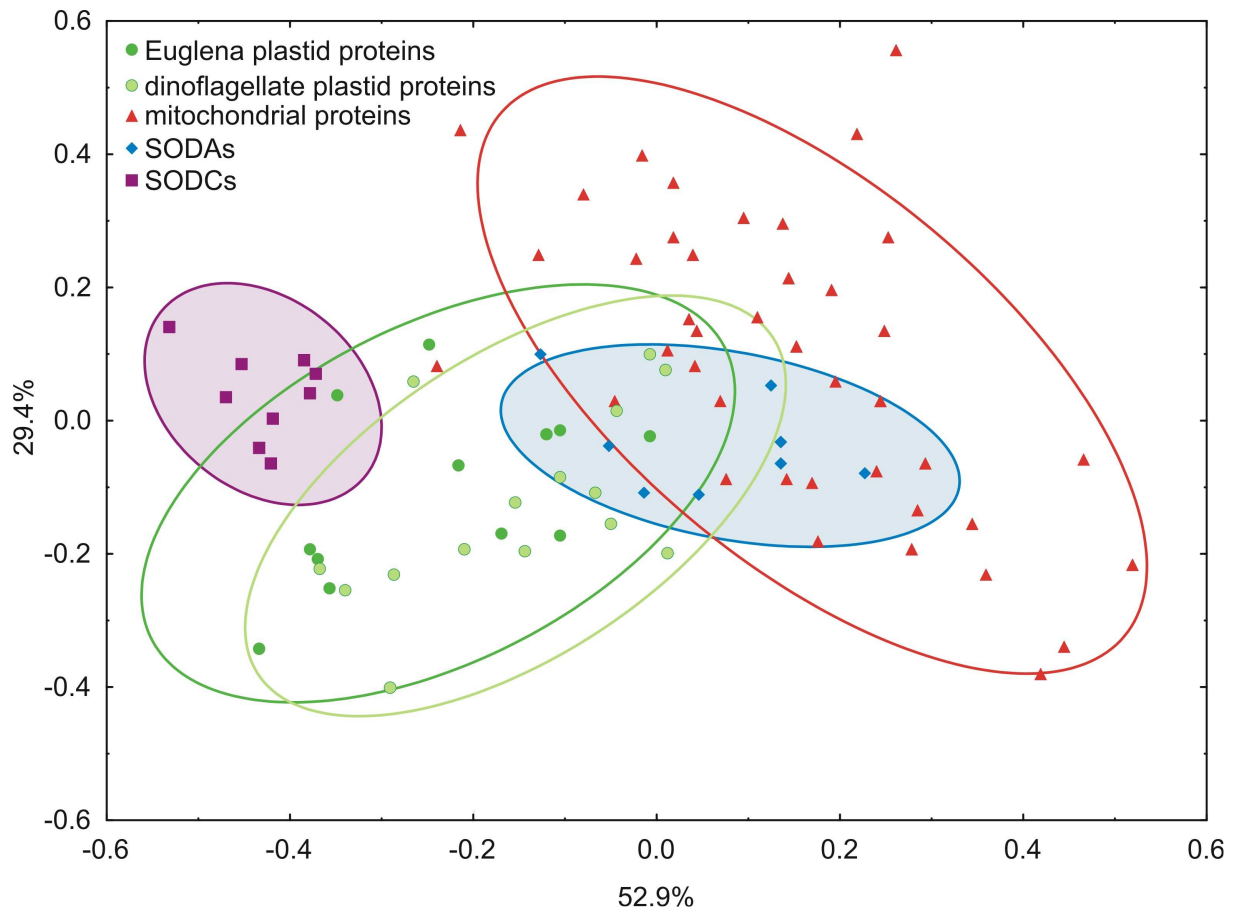


Fig. 4S

