Arch Microbiol (2007) 187:281–296
DOI 10.1007/s00203-006-0194-5

ORIGINAL PAPER

Analysis of the targeting sequences of an iron-containing superoxide dismutase (SOD) of the dinoflagellate *Lingulodinium polyedrum* suggests function in multiple cellular compartments

Andrzej Bodył · Paweł Mackiewicz

Received: 2 October 2006 / Accepted: 6 November 2006 / Published online: 2 December 2006

Abstract One of the proteins targeted to the peridinin plastid of the dinoflagellate *Lingulodinium polyedrum* is the iron-containing superoxide dismutase (LpSOD). Like dinoflagellate plastid proteins of class II, LpSOD carries a bipartite presequence comprising a signal peptide followed by a transit peptide. Our bioinformatic studies suggest that its signal peptide is atypical, however, and that the entire presequence may function as a mitochondrial targeting signal. It is possible that LpSOD represents a new class of proteins in algae with complex plastids, which are co-targeted to the plastid and mitochondrion. In addition to the ambiguous N-terminal targeting signal, LpSOD contains a potential type-1 peroxisome-targeting signal (PTS1) located at its C-terminus. In accordance with a peroxisome localization of this dismutase, its mRNA has two in-frame AUG codons. Our bioinformatic analyses indicate that the first start codon resides in a much weaker oligonucleotide context than the second one. This suggests that synthesis of the plastid/mitochondrion-targeted and peroxisome-targeted isoforms could proceed through so-called leaky scanning. Moreover, our results show that expression of the two isoforms could be regulated by a 'hairpin' structure located between the first and second start codons.

A. Bodył (✉)
Department of Biodiversity and Evolutionary Taxonomy,
Zoological Institute, University of Wrocław,
ul. Przybyszewskiego 63/77,
51-148 Wrocław, Poland
e-mail: bodyl@biol.uni.wroc.pl

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

Keywords Alternative translation · Dinoflagellates · Peridinin plastid · Peroxisome targeting signal · Signal peptide · Superoxide dismutase

Introduction

Reactive oxygen species (ROS) such as the superoxide anion radical (O_2^-), hydroxyl radical ($HO^\bullet$) and hydrogen peroxide (H_2O_2), are normal by-products of oxidative metabolism. Although some ROS may function as important signaling molecules (Acker 2005), most of them are toxic and cause serious damage to DNA, proteins, and lipids. To minimize damaging effects of ROS, aerobic cells evolved two kinds of antioxidant defenses: non-enzymatic and enzymatic systems (for a review see Scandalios 2005). The components of the former are vitamins A and C, glutathione, flavonoids, alkaloids, and carotenoids; the latter involves enzymes such as catalases, superoxide dismutases, and peroxidases.

The first line of defense against ROS is represented by superoxide dismutases (SODs) (for reviews see Fink and Scandalios 2002; Wolfe-Simon et al. 2005). These enzymes catalyze dismutation of superoxide radicals to hydrogen peroxide and oxygen. There are three main types of SODs. The first class utilize Cu and Zn, the second Mn, and the third Fe in the active site. Cu/Zn-SODs mostly are found in the cytosol and the nucleus, Mn-SODs are characteristic of mitochondria, whereas Fe-SODs reside mainly in plastids. Experiments with SOD-deficient mutants clearly demonstrate that these enzymes are indispensable for the normal functioning of all cell types in aerobic environments (Manfredini et al. 2004; Seyler et al. 2001; Woodruff et al. 2004).

In eukaryotic cells, the largest amounts of ROS are generated by mitochondria, plastids, and peroxisomes (for reviews see Blokhina et al. 2003; del Rio et al. 2002). Although each of these organelles possesses its own antioxidant system (Blokhina et al. 2003; del Rio et al. 2002), many components are common to all three. Thus, it is reasonable to hypothesize that at least some antioxidant enzymes are encoded by single genes, but targeted to two or all three ROS-generating compartments. A good example of such an enzyme is catalase A of the yeast *Saccharomyces cerevisiae* (Petrova et al. 2004). This protein has two kinds of targeting signals, which allow it to be imported into both the mitochondrion and the peroxisome. Another dually targeted antioxidant enzyme is glutathione reductase from higher plants (Rudhe et al. 2004). This protein carries an ambiguous N-terminal targeting signal, which is responsible for its transport into both the plastid and the mitochondrion.

In this paper we present detailed bioinformatics analyses of another antioxidant enzyme, which seems to be targeted to as many as three ROS-generating compartments. This enzyme is Fe-SOD encoded in the nuclear genome of the dinoflagellate *Lingulodinium polyedrum* (formerly *Gonyaulax polyedra*). Available experimental data indicate that this enzyme is imported into the plastid (Okamoto et al. 2001), but our results provide evidence that it could be targeted to the mitochondrion and peroxisome as well.

Typical import mechanism of nuclear-encoded proteins into the dinoflagellate plastid

Dinoflagellates are a diverse group of unicellular eukaryotes that constitute one of the main components of marine and freshwater phytoplankton (Graham and Wilcox 2000). About a half of the characterized species are photosynthetic, and carry a variety of different kinds of plastids. The most widespread of these is the so-called peridinin plastid (for a review see Schnepf and Elbrächter 1999). This plastid is surrounded by three membranes and contains the accessory pigments chlorophyll c, as well as the carotenoid peridinin from which it gets its name (Schnepf and Elbrächter 1999). Evolutionary origin of the peridinin plastid remains controversial. Most authors accept its ultimate secondary origin from a rhodophyte endosymbiont (see, for example, Blanchard and Hicks 1999; Harper and Keeling 2003; Yoon et al. 2005). It has been postulated that the peridinin plastid evolved directly from the same secondary endosymbiosis that created the apicomplexan plastid, as well as those of cryptophytes, heterokonts, and haptophytes (Cavalier-Smith 2003; Harper and Keeling 2003; Yoon et al. 2005). Recent phylogenies based on plastid-located genes, however, suggest that the peridinin plastid is derived from a haptophyte alga that already contained a rhodophyte-derived plastid (Salachian-Tabrizi et al. 2006). Thus, the peridinin plastid appears more likely to be derived from a tertiary endosymbiosis, as originally postulated by Gibbs (1978). Further evidence supporting a tertiary origin of the peridinin plastid has been discussed by Bodył (2005) and Bodył and Moszczyński (2006).

The peridinin plastid has a peculiar genome composed of minicircles that encode only a handful of photosynthetic proteins (for a review see Koumandou et al. 2004). Thus, almost all dinoflagellate plastid proteins (~2,000–3,000) are encoded in the nuclear genome (Hackett et al. 2004). These proteins carry complex N-terminal targeting signals consisting of two domains (Nassoury et al. 2003; Patron et al. 2005). The first domain has characteristic features of a signal peptide enabling co-translational import into the endoplasmic reticulum (ER), whereas the second resembles transit peptides responsible for post-translational import into the primary plastids of glaucophytes, rhodophytes, and chlorophytes. It should be noted that such bi-partite presequences are also characteristic of other algae with complex plastids, e.g. euglenoids, chlorarachniophytes, cryptophytes, heterokonts, and haptophytes (for reviews see van Dooren et al. 2001; Ishida 2005). Dinoflagellates are exceptional among eukaryotic algae, however, in that they contain two classes of plastid transit peptides (Patron et al. 2005). Class I is distinguished by the presence of a transit peptide followed by a hydrophobic domain that functions as a stop-transfer sequence. Class II possesses only the transit peptide. Consequently, dinoflagellate plastid proteins of class I contain tripartite pre-sequences (signal peptide, transit peptide, and stop-transfer sequence), whereas those of class II carry bipartite presequences (signal peptide and transit peptide).

Nassoury et al. (2003) demonstrated that the co-translational translocation into ER is the first step in the import of class-I dinoflagellate plastid proteins. During this transport, the signal peptide is cleaved by a signal peptidase; each of these proteins is anchored in the membrane by the stop-transfer domain. In this membrane-bound form class-I proteins then are trafficked to the plastid via a pathway involving not only ER, but also the Golgi apparatus. Unfortunately, their further transport across the two inner membranes is still enigmatic. It is hypothesized that this final targeting step is dependent on the Toc-Tic supercomplex

(van Dooren et al. 2001); however, no homologues of Toc and Tic proteins have been found in dinoflagellates, and there is not even strong evidence for Tocs in other algae with rhodophyte-derived plastids (McFadden and van Dooren 2005). Thus, the existence of unconventional Toc–Tic-independent pathways must be considered.

In contrast to the partially characterized targeting of dinoflagellate class-I plastid proteins, targeting of class-II proteins remains entirely undescribed. These proteins undoubtedly are translocated into the ER lumen, but it is unknown whether they are trafficked directly to the plastid, or transported first into the Golgi apparatus (Patron et al. 2005). It also is unknown if their further translocation across the two inner plastid membranes is dependent on a Toc–Tic supercomplex, or on some other translocons.

Nucleus of the dinoflagellate *Lingulodinium polyedrum* encodes a Fe-containing superoxide dismutase (SOD) targeted to the plastid

The first evidence for a Fe-SOD in the peridinin plastid of *L. polyedrum* was provided by Okamoto et al. (2001). They isolated a ~32 kDa protein from plastid extracts, which reacted with heterologous Fe-SOD antibodies. It also exhibited characteristic features of an iron superoxide dismutase, such as CN-resistance and H_2O_2 sensitivity. Moreover, the authors obtained cDNA of *L. polyedrum* SOD (AF 289824). The predicted amino acid sequence indicated the presence of conserved residues responsible for Fe binding, and exhibited high overall similarity to cyanobacterial, green algal and higher plant Fe-SODs (Okamoto et al. 2001).

Additional support for the plastid localization of *L. polyedrum* SOD, termed further LpSOD, comes from its N-terminal extension. Okamoto et al. (2001) observed that this pre-sequence has a hydrophobic domain (presumably a signal peptide) followed by a hydrophilic stretch (presumably a transit peptide); however, they did not characterize it in great detail. For example, they did not search for an additional hydrophobic domain. We decided to perform such a detailed characterization for two reasons. First, LpSOD was not included in previous import (Nassoury et al. 2003) and bioinformatic (Patron et al. 2005) studies of dinoflagellate plastid proteins. Second, the existence of superoxide dismutases in different cell compartments suggested to us that detailed bioinformatic analyses of the N-terminal extension of LpSOD could reveal novel targeting characteristics.

Characterization of the N-terminal targeting signal of *L. polyedrum* Fe-SOD

We initially characterized the N-terminal targeting signal of LpSOD by making a hydrophobicity profile. Two hydrophobic stretches are present, which surround a hydrophilic region rich in hydroxylated (S, T, Y) and basic (H, K, R) residues (Fig. 1). This hydrophobicity profile resembles those of other proteins targeted to the peridinin plastid, and especially ones belonging to class I (Nassoury et al. 2003; Patron et al. 2005). This suggests that the first domain of the LpSOD pre-sequence functions as a signal peptide, the second as a transit peptide, and the third as a stop transfer sequence. A closer examination of each of these domains, however, revealed interesting features that qualify this N-terminal extension of LpSOD as a very peculiar targeting signal.

The signal peptide-like domain

To confirm that the first hydrophobic domain of LpSOD (Fig. 1) likely functions as a signal peptide, we used programs that predict different kinds of N-terminal targeting signals (e.g. TargetP, iPSORT, Predotar) and programs specialized in the identification of signal peptides (e.g. PrediSi, SIGFIND, SignalP-HMM)—see Table 1. Our analyses produced unexpected results. None of the six software tools used to find various kinds of N-terminal targeting signals predicted a signal peptide. In turn, programs specially designed to predict signal peptides gave inconsistent results. Three of them, DetecSig in ConPred II, PrediSi and SignalP-NN, did not find a signal peptide. Sigcleave in the EMBOS package identified a signal peptide, but with a score below the optimum of 3.5. Four other programs (e.g. SIGFIND) found a signal peptide, most with similarly low scores. Interestingly, each of these software tools predicted a cleavage site between amino acids 14 and 15. This site is preceded by the AXA motif known to be recognized by the signal peptidase (Martoglio and Dobberstein 1998).

These results are unexpected and require further explanation. Signal peptides represent a universal targeting signal for protein translocation across RER membranes in all eukaryotic organisms and it is difficult to explain the results we obtained as shortcomings in existing software tools (e.g., that they were designed and tested using sequences primarily from yeast and animals). Indeed, the first hydrophobic domains of other dinoflagellate plastid proteins, including those with short pre-sequences, were recognized as signal peptides with high scores by the programs we used

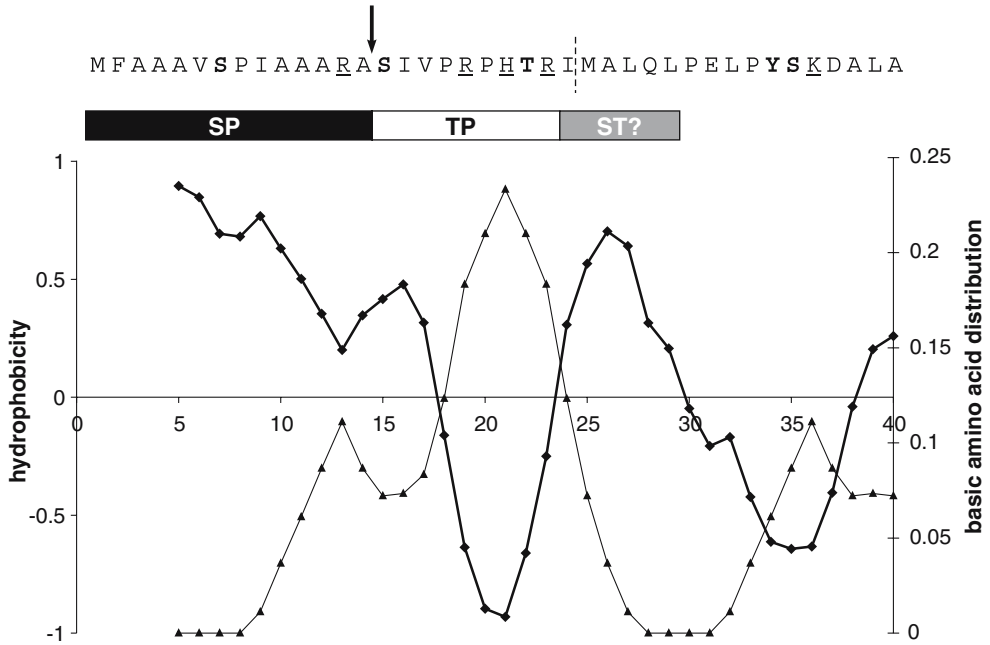

Fig. 1 Organization of the N-terminal extension of LpSOD. In this region, two targeting domains can be distinguished: the signal peptide (SP) and the transit peptide (TP). At present, it is not known whether the second hydrophobic stretch represents a vestigial stop transfer domain (ST?). The arrow indicates the cleavage site of SP, which is preceded by the characteristic AXA motif. The vertical dashed line separates the presequence from the mature protein. Basic residues (H, K, R) are underlined, while hydroxylated ones (S, T, Y) are in bold. In the positional property profile, the thick line indicates hydrophobicity according to the Kyte-Doolittle scale (Kyte and Doolittle 1982), whereas the thin line shows the distribution of basic residues. These profiles were made according to the procedure used by ProtScale (ExPASy server: www.expasy.ch): window size 9 and linear weight variation model with 10% relative weight of the window edges compared to the window center

(data not shown; see also the next section). Thus, our combined results suggest that the first hydrophobic domain of the N-terminal extension of LpSOD represents a very peculiar signal peptide.

The characteristic feature of the LpSOD signal peptide is its weak ER-targeting signal. This feature suggests that the dinoflagellate superoxide dismutase may be post- rather than co-translationally imported into the peridinin plastid. Recent analyses of protein import into higher plant plastids demonstrate that, in addition to the common post-translational targeting system, a co-translational targeting pathway also is present (Villarejo et al. 2005). Two such distinct targeting systems could also operate in the peridinin plastid. At present, this possibility cannot be excluded. An alternative hypothesis also should be considered, however, that LpSOD is imported not only to the plastid but also to another compartment. The mitochondrion is a likely candidate for that compartment. Mitochondrial transit peptides resemble plastid transit peptides in any case (Zhang and Glaser 2002), suggesting that the dinoflagellate SOD could be imported into the mitochondrion after modification or degeneration of the signal peptide. We will elaborate on this hypothesis in a later section of this article.

The transit peptide-like domain

The presence of the AXA motif indicates that the signal peptide-like domain is cleaved off during the co-translational translocation of LpSOD into the ER. Thus, the nine amino acid hydrophilic domain (residues 15–24) (Fig. 1) would be responsible for the passage of this protein across the two inner plastid membranes. If so, we would hypothesized that this domain should resemble the transit peptides responsible for protein import into the primary plastids of glaucophytes, rhodophytes, and chlorophytes. To test this prediction, we made a bioinformatic characterization after removing the putative signal peptide. Interestingly, neither of the two programs designed to predict plastid transit peptides (ChloroP and PCLR) found such a peptide (Table 1). Of six programs designed to predict different kinds of N-terminal targeting signals, only iPSORT identified a plastid transit peptide (amino acids 21–27), but this region overlaps partially with the sequence of the mature protein.

There are at least two explanations, which are not mutually exclusive. (1) Transit peptides of plastids from dinoflagellates and other algae derived ultimately rhodophyte lineage (e.g. red algae, cryptophytes,

Table 1 Prediction of targeting sequences and subcellular localization of the Fe-SOD of *L. polyedrum* by different software tools

Name and reference	Organismal model	Potential predictions	The whole sequence	The sequence without SP	The sequence without presequence
Programs predicting signal peptides					
DetecSig in ConPred II (Lao and Shimizu 2001)	Eukaryote	SP, ST	–	–	–
Phobius (Käll et al. 2004)	Eukaryote	SP, ST	SP [1–14]	–	–
PrediSi (Hiller et al. 2004)	Eukaryote	SP	–	–	–
Sigcleave in EMBOSS 3.0.0 (Rice et al. 2000)	Eukaryote	SP	SP: 3.2 [2–14]	–	–
SIGFIND 2.11 (Reczko et al. 2002)	Eukaryote	SP	SP: 0.51 [1–14]	–	–
SignalP-HMM 3.0 (Nielsen and Krogh 1998)[a]	Eukaryote	SP, ST	SP: 0.57; CS: 0.51 [1–14]	–	–
SignalP-NN 3.0 (Bendsten et al. 2004)	Eukaryote	SP	–	–	–
Programs predicting different kinds of N-terminal targeting signals					
BLSTM-LOC 1.0	Plant	SP, PT, MT, other	MT: 60%; CT: 40%	Other: 92%	Other: 100%
(139.91.72.10/blstm/blstm.html)	Non-plant	SP, MT, other	MT: 84%	Other: 100%	Other: 100%
IPSORT	Plant	SP, PT, MT	PT [6–18]	PT [21–27]	–
(Bannai et al. 2002)	Non-plant	SP, MT	MT [6–11]	–	–
Predotar 1.03	Plant	SP, PT, MT	Possibly MT: 0.31	–	Other: 0.99
(Small et al. 2004)	Animal/fungi	SP, MT	Possibly MT: 0.34	–	Other: 0.99
PredSL	Plant	SP, PT, MT, other	MT: 0.97	Other	Other
www.195.134.85.247/PredSL/index.html	Non-plant	SP, MT, other	MT: 0.62 [1–19]	Other	Other
Prowler 1.1	Plant	SP, PT, MT, other	MT: 0.38; CT: 0.23	Other: 0.67	Other: 0.91
(Boden and Hawkins 2005)	Non-plant	SP, MT, other	MT: 0.89	Other: 0.80	Other: 0.94
TargetP 1.1	Plant	SP, PT, MT	PT: 0.38 (RC = 5) [1–13]	–	Other: 0.82 (RC = 2)
(Emanuelsson et al. 2000)[b]	Non-plant	SP, MT	MT: 0.73 (RC = 3) [1–14]	–	Other: 0.88 (RC = 2)
Programs predicting plastid and mitochondrial transit peptides					
ChloroP 1.1 (Emanuelsson et al. 1999)	Plant	PT	–	–	–
PCLR 0.9 (Schein et al. 2001)	Plant	PT	–	–	–
MitoProt II 1.0a4 (Claros and Vincens 1996)	Eukaryote	MT	MT: 0.75 [1–25]	MT: 0.69 [15–6]	–
Programs predicting mitochondrial or peroxisome localization of a protein					
MITOPRED	Plant	mit	mit: 70%	mit: 60%	–
(Guda et al. 2004)	Animal/yeast	mit	mit: 99%	mit: 77 %	–
PeroxiP	Plant	per	M2, M4	M2, M4	M2, M4
(Emanuelsson et al. 2003)[c]	Non-plant	per	M2, M4	M2, M4	M2, M4
PTS1 Predictor	General	per	per: 10.97 (P = 3.2e−04)	per: 10.97 (P = 3.2e−04)	per: 10.97 (P = 3.2e−04)
(Neuberger et al. 2003a, b)[c]	Metazoa	per	per: 8.50 (P = 1.8e−04)	per: 8.50 (P = 1.8e−04)	per: 8.50 (P = 1.8e−04)
	Fungi	per	per: 9.06 (P = 2.0e−04)	per: 9.06 (P = 2.0e−04)	per: 9.06 (P = 2.0e−04)

Table 1 continued

Name and reference	Organismal model	Potential predictions	The whole sequence	The sequence without SP	The sequence without presequence
Programs predicting different subcellular localizations of a protein					
CELLO 2.5 (Yu et al. 2004)	Eukaryote	cyt, er, ext, ga, lys, mit, nuc, per, pm, pla, vac	mit: 1.27; ext: 1.21	mit: 1.23; cyt: 0.94	pla: 1.03; cyt: 0.94
ESLpred (Bhasin and Raghava 2004)[d]	Eukaryote	cyt, ext, mit, nuc	mit: 75% (RI = 2)	mit: 75% (RI = 2)	mit: 53% (RI = 1)
LOCSVMpsi 1.3	Eukaryote	cyt, ext, mit, nuc	mit: 93% (RI = 9)	mit: 85% (RI = 7)	cyt: 48% (RI = 2)
(Xie et al. 2005)[c]	Eukaryote	cyt, er, ext, ga, lys, mit, nuc, per, pla, pm, vac	pla: 40% (RI = 1)	mit: 43% (RI = 2)	cyt: 40% (RI = 1)
LOCTree (Nair and Rost 2005)[d]	Plant	cyt, ext, mit, nuc, pla	pla: RI = 1; mit or pla RI = 4	pla: RI = 5; mit or pla: RI = 8	pla: RI = 1; mit or pla: RI = 8
	Animal	cyt, ext, mit, nuc	mit: RI = 8	mit: RI = 5	mit: RI = 8
PLOC	Plant	cyt, er, ext, ga, mit, nuc, per, pla, pm, vac	pm	mit	cyt
(Park and Kanehisa 2003)	Fungi	cyt, er, ext, ga, mit, nuc, per, pm, vac	mit	cyt or ext	cyt or ext
	Animal	cyt, er, ext, ga, lys, mit, nuc, per, pm	mit	cyt or mit	cyt or ext
Protcomp 6.0	Plant	cyt, er, ext, mit, nuc, per, pla, pm	NN: ext: 1.1; IP: pla: 1.9	NN: mit: 1.2; IP: pla: 1.9	NN: cyt: 1.1; IP: pla: 1.9
(www.softberry.com)[e]	Animal/fungi	cyt, er, ext, ga, lys, mit, nuc, per, pm	NN: per: 3.0; IP: mit: 4.2	NN: per: 2.5; IP: mit: 4.3	NN: per: 2.4; IP: mit: 3.8
pSLIP (Sarda et al. 2005)	Eukaryote	cyt, ext, mit, nuc, pla, pm	cyt	cyt	cyt
PSORT II (Nakai and Horton 1999)	Yeast/animal	cyt, er, ext, ga, mit, nuc, per, pm, vac, ves	per: 67%	per: 78%	per: 78%
PSORT 6.4	Plant	cyt, er, ext, ga, mit, nuc, vac, per, pla, pm	per: 0.93; mit matrix: 0.67	per: 0.93	per: 0.93
(Nakai and Horton 1999)	Yeast	cyt, er, ext, ga, mit, nuc, per, pm, vac	per: 0.93; mit matrix: 0.67	per: 0.93	per: 0.93
	Animal	cyt, er, ext, ga, lys, mit, nuc, per, pm	per: 0.93; mit matrix: 0.67	per: 0.93	per: 0.93
pTARGET (Guda and Subramaniam 2005)	Fungi/metazoa	cyt, er, ext, ga, lys, mit, nuc, per, pm	mit: 100%	mit: 100%	mit: 100%
SubLoc 1.0 (Hua and Sun 2001)[d]	Eukaryote	cyt, ext, mit, nuc	cyt: 56% (RI = 1)	cyt: 74% (RI = 2)	cyt: 56% (RI = 1)
WoLF PSORT	Plant	cyt, er, ext, ga, mit, nuc, vac, per, pla, pm	per: 8	per: 9	per: 12
(Horton et al. 2006)	Fungi	cyt, er, ext, ga, mit, nuc, per, pm, vac	per: 13	per: 17	per: 18
	Animal	cyt, er, ext, ga, lys, mit, nuc, per, pm	mit: 21	mit: 17	cyt: 19.5

Score, accuracy or confidence of a given prediction is shown after colon. "–" denotes negative prediction. The range of SP, MT or PT prediction in the sequence is shown in square brackets

MT mitochondrial transit peptide, *SP* signal peptide, *ST* stop transfer domain, *PT* plastid transit peptide, *cyt* cytosol, *er* endoplasmic reticulum, *ext* extracellular space, *ga* Golgi apparatus, *lys* lysosome, *mit* mitochondrion, *nuc* nucleus, *per* peroxisome, *pm* plasma membrane, *pla* plastid, *vac* vacuole, *ves* vesicles of secretory system

[a] SP score corresponds to the signal peptide probability, CS is the cleavage site probability

[b] RC represents the reliability class—a measure of the size of the difference between the highest, winning and the second highest output scores

[c] M denotes the method of prediction: M2 (by restrictive PTS1 motif) and M4 (by permissive PTS1 motif)

[d] P represents percent of false positives

[e] RI is the reliability index; the higher the more confident prediction

[f] NN is the result of neural networks-based prediction; IP is the result of integral prediction including homology-based prediction

heterokonts) differ from those of chlorophytes, euglenoids, and chlorarachniophytes belonging to the green plastid lineage. There is evidence in support of this argument; for example, they are characterized by the distinctive FVAP motif (Kilian and Kroth 2005; Patron et al. 2005; Gould et al. 2006). Moreover, dinoflagellate transit peptides, both classes I and II, are conspicuously abundant in glutamine residues (Patron et al. 2005). Thus, programs designed primarily based on higher plant proteins (e.g. ChloroP) may overlook transit peptides of dinoflagellates and other algae with red plastids.

(2) Protein import into the primary plastids of glaucophytes, rhodophytes, and chlorophytes proceeds post-translationally (Backer et al. 2005), and some characteristics of their transit peptides could reflect adaptations to prevent mistargeting of plastid proteins into the mitochondria (Macasev et al. 2000; Bruce 2001). In the case of the co-translational transport of LpSOD, the transit peptide-like domain would reside in the lumen of the endomembrane system, where it would not be at risk of mistargeting to the mitochondrion. Thus, its plastid-targeting domain could have lost the characteristic features of typical plastid transit peptides, because they no longer were under purifying selection. It is possible, however, that these changes do not reflect accidental degeneration.

As already suggested, the dinoflagellate superoxide dismutase may be targeted not only to the plastid, but also to the mitochondrion. If this is the case, modifications to its transit peptide-like domain could be specifically selected to improve mitochondrial import. In support of this hypothesis, several programs identified the whole pre-sequence of LpSOD as a mitochondrial targeting signal (see discussion below).

The stop transfer-like domain

The second hydrophobic domain of LpSOD is composed of six amino acids only (residues 24–30), of which as many as five occur within the mature protein (Fig. 1). This clearly contrasts with the length of classical transmembrane (TM) domains, which are made of ~20 residues forming α-helix. In accordance with this observation, none of the 12 programs that predict hydrophobic domains [e.g. TMHMM (Krogh et al. 2001), TMpred (Hofmann and Stoffel 1993), TopPred (von Heijne 1992)] found such a domain in the N-terminal portion of LpSOD (data not shown).

Kuroiwa et al. (1991) demonstrated that the stop transfer domain can be composed of as few as seven amino acid residues. In such cases of short stop-transfer domains, however, all residues must be highly hydrophobic, and must be followed by positively charged amino acids to anchor the protein in the membrane (Kuroiwa et al. 1991). In the light of this, it is doubtful that the second hydrophobic region of LpSOD could function as a stop transfer sequence. This domain has only five hydrophobic residues, and they are split by a polar glutamine (Fig. 1). Moreover, only one positively charged residue occurs in their vicinity. Consequently, these overall sequence characteristics point to the conclusion that LpSOD is most similar to bipartite class-II dinoflagellate plastid proteins.

Comparison of the presequence of *L. polyedrum* Fe-SOD with those of classes I and II proteins reveals its unusual character

The presequence of LpSOD is composed of only 25 amino acids (Fig. 1). This clearly contrasts with the length of the N-terminal extensions of dinoflagellate plastid proteins belonging to both classes I and II. The average length of pre-sequences in class-I proteins is 66 residues, whereas in class-II proteins it is 71 residues (data from Patron et al. 2005). The shortest pre-sequence (49 amino acids) in class I is found in subunit C of ATP synthase (ATP H). In class II, the shortest pre-sequence (40 amino acids) is carried by beta ketoacyl reductase (bKAR).

The signal peptide-like domain of LpSOD comprises only 14 amino acids (Fig. 1), the shortest signal peptide to date among dinoflagellate plastid proteins. The mean length of signal peptides in classes I and II proteins is 24 residues (data from Patron et al. 2005). In class I, the shortest is a 16 aa-signal peptide carried by glyceraldehyde-3-phosphate dehydrogenase (GAPDH); in class II, it is the 15 aa-signal peptide found in carbonic anhydrase (CA). Although these short sequences resemble the LpSOD signal peptide in overall length, they possess strong ER-targeting signals according to programs designed to predict signal peptides as well as various kinds of N-terminal targeting signals (data not shown). These data provide additional support for our hypothesis that LpSOD has a peculiar signal peptide with an abnormal ER-targeting signal.

The inferred transit peptide-like domain of LpSOD is composed of only nine amino acids (Fig. 1). The average length of transit peptides in class-I proteins is 28 residues, and in class-II proteins it is 47 residues (data from Patron et al. 2005). The shortest class-I transit peptides (nine amino acids) are found in ATP H and LHPb, while in class-II acyl carrier protein (ACP) has

only 18 amino acids in its transit peptide. In addition to its extremely short length the LpSOD transit peptide lacks the FVAP motif found in most classes I and II proteins (Nassoury et al. 2003; Patron et al. 2005).

The stop transfer-like domain of LpSOD comprises only six amino acids (Fig. 1). In class-I proteins, this domain typically is almost three times longer (15 amino acids), the same length found in all kinds of proteins (data from Patron et al. 2005). This is additional evidence against the second LpSOD hydrophobic domain functioning as a stop transfer sequence, and further support for its similarity to class-II dinoflagellate plastid proteins.

One additional feature of the LpSOD presequence bears consideration. In dinoflagellate plastid proteins that carry a very short signal (or transit) peptide, the second domain generally is longer. For example, the signal peptide of CA comprises 15 residues, but its transit peptide is as long as 72 residues. The nine aa-transit peptides of ATP H and LHPb also are accompanied by 72 aa-signal peptides. In the case of LpSOD, both the signal peptide- and the transit peptide-like domains are extremely short.

Is the N-terminal extension of *L. polyedrum* Fe-SOD used as a mitochondrial targeting signal?

Considering the unusual character of the N-terminal extension of LpSOD, and its abnormal ER-targeting signal, we suggest that this entire pre-sequence functions as a single targeting signal, allowing import of this dismutase into the mitochondrion. This hypothesis is supported by several lines of evidence. First, an arginine occurs two positions upstream of the methionine that signals the start of the mature portion of LpSOD (Fig. 1); this sequence motif is specifically recognized by the mitochondrial processing peptidase (Glaser et al. 1998). Second, the apicomplexans *Toxoplasma gondii* and *Plasmodium falciparum*, which are closely related to dinoflagellates (for a review see Leander and Keeling 2004), contain a mitochondrion-targeted Fe-SOD that is equipped with a bipartite pre-sequence composed of a signal peptide followed by a hydrophilic domain resembling a mitochondrial targeting signal (Brydges and Carruthers 2003; Sienkiewicz et al. 2004). Third, mitochondria maintain high concentrations of superoxide dismutases (Blokhina et al. 2003), and these enzymes are targeted not only to the mitochondrial matrix, but also to the periplasmic space (Okado-Matsumoto and Fridovich 2001).

To test the hypothesis that LpSOD could be targeted to the mitochondrion, we applied programs that predict different kinds of N-terminal targeting signals, along with programs specialized in identification of mitochondrial targeting signals (e.g. MITOPRED) (Table 1). Interestingly, nearly all programs predicted a mitochondrial transit peptide and most of them gave moderate or high scores. To further test of the mitochondrial hypothesis, we used programs predicting different subcellular localizations of a protein (Table 1). Nine of these programs identified a mitochondrial localization of the LpSOD protein. Moreover, we found that even more programs predicted a mitochondrial localization when three C-terminal residues that represent a strong competing signal for peroxisome targeting (see next section) were removed (data not shown).

Additional support for targeting of LpSOD to the mitochondrion comes from the amino acid composition of its pre-sequence. In Table 2, we compared its pre-sequence to 15 dinoflagellate presequences of class II, two dinoflagellate pre-sequences with very short signal peptides, two mitochondrial bipartite pre-sequences from the apicomplexans *T. gondii* and *P. falciparum*, and classical mitochondrial pre-sequences of *P. falciparum* and other eukaryotes. Like bipartite pre-sequences of apicomplexans and those of known mitochondrial proteins from *P. falciparum* and other eukaryotes, the LpSOD pre-sequence is very rich in basic residues. It also contains no acidic residues, resembling mitochondrial pre-sequences that generally have few acidic amino acids. In contrast to the LpSOD presequence, dinoflagellate presequences with short signal peptides are very poor in basic residues, and pre-sequences of their class-II proteins have, on average, fewer basic residues than do mitochondrial presequences.

The above analyses suggest that LpSOD is targeted to the mitochondrion, but do not provide an indication as to which mitochondrial sub-compartment it is finally delivered. The presence of arginine at the −2 position upstream of the second methionine suggests that this enzyme is imported into the mitochondrial matrix, and that its N-terminal targeting signal is cleaved off by the mitochondrial processing peptidase. However, an alternative hypothesis should be considered; that is, LpSOD is targeted to the periplasmic space. It has been suggested that targeting a protein to the periplasmic space should be much simpler than to the mitochondrial matrix, because the former does not require as many targeting signals (Lucattini et al. 2004; Murcha et al. 2005).

Does *L. polyedrum* Fe-SOD represent a new protein class in algae with complex plastids?

The evidence we have advanced suggest that LpSOD represents a new protein class in dinoflagellates. The

Table 2 Percentage of acidic and basic residues in different kinds of plastid and mitochondrial presequences

Analysed presequences	Acidic (D, E)	Basic (H, K, R)
LpSOD	0.0	16.7
Dinoflagellate plastid proteins of class II[a]	4.6	11.3
ptCA of *Heterocapsa triquetra*[b]	6.9	8.0
ptGAPDH of *Heterocapsa triquetra*[c]	2.0	4.0
mtFe-SOD of *Plasmodium falciparum*[d]	5.7	14.3
mtFe-SOD of *Toxoplasma gondii*[d]	2.4	13.1
Mitochondrial proteins of *P. falciparum*[e]	2.4	25.6
Mitochondrial proteins of eukaryotes[f]	1.4	17.8

pt plastid; *mt* mitochondrion

[a] Averaged over 15 pre-sequences published by Patron et al. (2005)

[b] The pre-sequence of carbonic anhydrase (class-II protein) with a very short signal peptide published by Patron et al. (2005)

[c] The pre-sequence of glyceraldehyde-3-phosphate dehydrogenase (class-I protein) with a very short signal peptide published by Patron et al. (2005); in the case of this pre-sequence, only signal peptide and transit peptide domains without the transmembrane region were analysed

[d] The apicomplexan bipartite pre-sequence targeted to the mitochondrion

[e] Data for 40 pre-sequences published by Bender et al. (2003)

[f] Data for 282 sampled pre-sequences published by Bender et al. (2003)

characteristic feature of this class would be the presence of a combined signal-transit peptide presequence, but with a weak ER-targeting signal. This would permit them to be targeted not only to plastids, but also to mitochondria. Deviation from a standard ER-targeting sequence in the signal peptide would reduce the probability of an interaction between LpSOD and the signal recognition particle (SRP). This sequence, rather, would cause it to associate with an alternative targeting complex [e.g. nascent polypeptide-associated complex (NAC) (George et al. 1998)], which would enable delivery of the dinoflagellate superoxide dismutase to the mitochondrion (Fig. 2). Consistent with this hypothesis, NAC is capable of effective competition with SRP, and of targeting a bound protein to the mitochondrion (Fünfschilling and Rospert 1999; George et al. 2002).

In higher plants an analogous class of mitochondrion/plastid-targeted proteins has been characterized (Chew et al. 2003; Peeters and Small 2001; Rudhe et al. 2002). A diagnostic feature is the presence of an ambiguous N-terminal targeting signal, which enables their import into both plastids and mitochondria. It is possible that further representatives of this hypothetical class of mitochondrion/plastid-targeted proteins will be found as more sequences become available, not only in dinoflagellates but also in other algae with complex plastids.

Peroxisome: a third ROS-generating compartment to which *L. polyedrum* Fe-SOD is targeted?

When characterizing the targeting signals of *L. polyedrum* Fe-SOD, Okamoto et al. (2001) scrutinized only the N-terminal targeting signal; however, this protein contains one other potential targeting signal. Its C-terminus is equipped with the SKL tripeptide that represents a classical and conserved peroxisome targeting signal of type-1 (PTS1) (Gould et al. 1987, 1989), suggesting that *L. polyedrum* Fe-SOD may be targeted to the peroxisome as well. This would make sense given that peroxisomes (and their specialized forms such as glyoxysomes and glycosomes) generate large amounts of ROS (for a review see del Rio et al. 2002), which are scavenged by different kinds of superoxide dismutases including Fe-SODs (Bueno et al. 1995; Plewes et al. 2003).

Although the presence of the PTS1 tripeptide suggests that LpSOD is imported into the peroxisome, such a short sequence could originate by chance. In fact, when searching the SWISS-PROT database, Emanuelsson et al. (2003) found that among 62 proteins carrying the C-terminal SKL tripeptide, 20 did not have any annotation indicating peroxisomal location. However, Neuberger et al. (2003a, b) and Reumann (2004) demonstrated that, in peroxisome-targeted proteins, the region adjacent to the C-terminal tripeptide also shows several conserved properties that can be used to predict peroxisomal localization of a protein. Interestingly, we have found all these properties in the region upstream of PTS1 in LpSOD. The 3-mer directly upstream PTS1 lacks acidic residues but contains a basic R residue at position −5 counting from the C-terminal end. This gives the region of the last six residues a high positive net charge (+2) and a high p*I* value (11.0). There also are several strongly hydrophobic residues (A and L) in the C-terminal 18 residues as well. Finally, a proline is present further upstream (position −10), as is expected in SKL-carrying proteins.

Further support for the peroxisome localization of LpSOD is provided by a variety of computational tools, including PSORT, PSORT II, WoLF PSORT, PeroxiP, and PTS1 predictor (Table 1). The first three programs predict the subcellular localization of a given protein, whereas the last two are specialty programs for predicting peroxisome-targeted proteins. All of

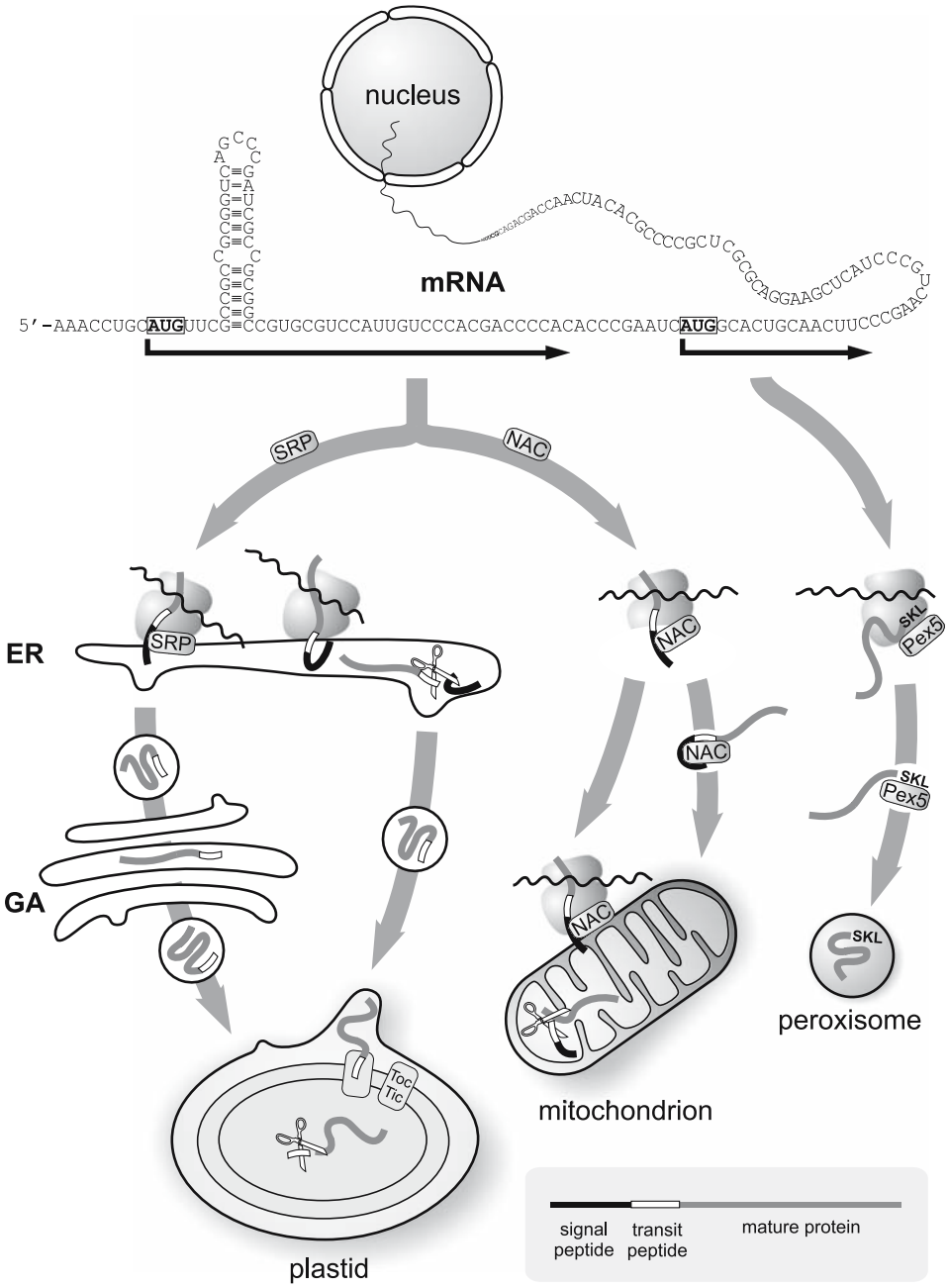

Fig. 2 A hypothetical model for Fe-SOD targeting within the *L. polyedrum* cell. The mRNA for *L. polyedrum* Fe-SOD contains two in-frame AUG start codons and a probable 'hairpin' structure located downstream of the first AUG codon. Translation initiation at the first start codon enables synthesis of the longer isoform which carries a bipartite presequence composed of the signal peptide followed by the transit peptide. Since the first domain represents a weak ER-targeting signal, there may be competition between the signal recognition particle (*SRP*) and nascent polypeptide-associated complex (*NAC*). If the newly synthesized *L. polyedrum* Fe-SOD is bound by SRP, it is imported into the plastid. The first targeting step in this pathway involves co-translational translocation into the endoplasmic reticulum (*ER*). During this transport, the signal peptide is removed by the signal peptidase. In the next vesicle-mediated targeting step, the imported protein is delivered to the plastid, but it is not known whether this pathway involves the Golgi apparatus (*GA*). After fusion of the ER (or GA)-derived vesicle with the outer plastid membrane, *L. polyedrum* Fe-SOD is translocated across the two inner membranes with the help of the Toc/Tic translocon. After reaching the stroma, the transit peptide is cleaved off by the stromal processing peptidase. If the newly synthesized *L. polyedrum* Fe-SOD is bound by NAC, it is imported into the mitochondrion. This import may occur post-translationally or co-translationally. After reaching the mitochondrial matrix, the bipartite targeting signal is removed in a single step by the mitochondrial processing peptidase. Translation initiation at the second start codon enables synthesis of the shorter isoform which is devoid of the N-terminal targeting signal, but contains the C-terminal peroxisome targeting signal of type 1 (*PTS1*). After translation in the cytosol, *L. polyedrum* Fe-SOD is bound by the Pex5 receptor. The Fe-SOD/Pex5 complex is then delivered to the peroxisome

these programs predicted peroxisome localization of LpSOD with moderate to high probability.

Transcripts of *L. polyedrum* Fe-SOD contain alternative sites for translation initiation: additional evidence for peroxisome targeting of the protein

The bioinformatic results we have obtained support the hypothesis that LpSOD is targeted to three distinct compartments within the dinoflagellate cell: the plastid, the mitochondrion, and the peroxisome. The first step in protein import into the peridinin plastid involves the co-translational transport into ER (Nassoury et al. 2003). Interestingly, an identical transport mechanism, but engaging Sec-unrelated proteins, was shown for mitochondria (Mukhopadhyay et al. 2004; Verner 1993). Thus, the existence of such a co-translational LpSOD import mechanism is compatible with import into both of these organelles, but could prevent the PTS1 signal from being available in the cytosol for recognition by the Pex5 receptor, and import of the LpSOD–Pex5 complex into the peroxisome matrix. Eukaryotic cells have evolved mechanisms, however, that enable nuclear-encoded proteins to be targeted to several compartments, each via a distinct targeting system (Silva-Filho 2003; Small et al. 1998). Alternative transcription, alternative exon-splicing, and alternative translation are especially relevant from the standpoint of possible differential targeting of LpSOD to the plastid/mitochondrion or the peroxisome. Unfortunately, at present only the cDNA of LpSOD is available (Okamoto et al. 2001), which precluded our computational analyses the first two mechanisms. To look for evidence of alternative translation, we searched for an additional ATG codon in the available cDNA sequence. Such a codon could function as an alternative start site for translation and enable synthesis of a peroxisome-targeted isoform, which lacks an N-terminal targeting pre-sequence.

According to our predictions, the mRNA for LpSOD contains two in-frame AUG codons (Fig. 2). The first, termed AUG1, occupies positions 9–11 in the cDNA sequence and encodes the methionine beginning the N-terminal targeting signal (Met-1). The second potential start codon (AUG2) is encoded by nucleotides 81–83 at the beginning the mature LpSOD protein (Met-25). The presence of AUG2 in this location could be coincidental, and not necessarily mean that it functions as an alternative start site for translation. An AUG is recognized by 40S ribosome as a start codon only when it resides in a specific oligonucleotide context (Kozak 2002). A purine (preferably A) at position −3 and a G at position +4 are the key nucleotides for providing such a translation initiation context (Joshi et al. 1997; Kozak 1987; Zhang 1998), where the +1 position refers to A in the AUG codon. Interestingly, the nucleotide sequence surrounding AUG1 indicates that it resides in a very weak context: each of the two key positions is occupied by U instead of A or G (Fig. 2). In contrast, AUG2 is positioned in a strong initiation context, with a purine in both the −3 position and the +4 position (Fig. 2). Moreover, the −3 position is occupied by A, as it is preferred for strong contexts. These data are entirely consistent with AUG2 function as a strong alternative start site for translation.

In addition to the preferred purines at positions −3 and +4, other nucleotides typically occur upstream and downstream of a strong initiation codon (Joshi et al. 1997; Kozak 1987; Zhang 1998). Therefore, we extended our analyses of the AUG1 and AUG2 contexts to positions −8 to +6, and compared them with the same positions for known initiators in different organisms, including lower plants, dicotyledons, and vertebrates (Table 3). To increase objectivity, we converted the available base fraction matrices to scoring matrices (taking logarithms from fraction/0.25) and counted the sum of log odd scores for AUG1 and AUG2 contexts. We also included in these analyses the contexts of AUG start codons found for 81 non-redundant dinoflagellate coding sequences retrieved from GenBank (Table 4). In the case of all matrices, and especially those for dinoflagellate sequences, AUG2 obtained higher scores than AUG1 (Table 3), thereby demonstrating that it contains more appropriate attributes for a translation initiator than does the first potential start codon. In agreement with these results, the ATG program (Pesole et al. 2000) ranked the AUC*AUG*GC (AUG2) 23th among 77 872 analyzed contexts, whereas UGC*AUG*UU (AUG1) was placed at a distant 766th position. Additional support for AUG2 as a good translation start site comes from the ATGpr, AUG evaluator, Geneid, HMMgene, NetStart, and TIS Miner programs (Table 5). Using all these software tools, we were able to analyze the suitability of AUG1 and AUG2 contexts in 20 model organisms. The second start codon obtained higher scores than the first in almost all cases, including in a member of the Apicomplexa (*P. falciparum*), the group thought to be most closely related to dinoflagellates.

The data presented above suggest that the mRNA for LpSOD is used for synthesis of not only a plastid/mitochondrion-targeted isoform, but also one targeted to the peroxisome (Fig. 2). The first isoform beginning

Table 3 Alignment of the contexts of AUG start codons characteristic for different organism groups with those of AUG1 and AUG2 codons of LpSOD mRNA

Group	Reference	AUG start codon context												Scores	
		−8	−7	−6	−5	−4	−3	−2	−1	**AUG**	+4	+5	+6	AUG1	AUG2
Drosophila	Cavener (1987)	a	a	a	u	C/A	A	A	A/C	AUG	a	c	c	−2.8	−0.3
Vertebrates	Salzberg (1997)	c	c	g	c	C	A	C/A	C	AUG	G/A	c	g	−4.4	2.6
Plant kingdom	Joshi et al. (1997)	a	a	a	c	a	A	A/C	a	AUG	G	C	g	−4.6	2.0
Dicotyledons	Joshi et al. (1997)	a	a	a	a	a	A	A/C	a	AUG	G	C	u	−5.4	1.8
Monocotyledons	Joshi et al. (1997)	c	g	g	c	a/c	A/G	A/C	c	AUG	G	C	G	−4.2	1.7
Non-angiosperms	Joshi et al. (1997)	u	u	g	c	a	A/G	A/C	a/c	AUG	G	C	g	−9.1	2.4
Lower plants	Joshi et al. (1997)	c	a	g	c	A/C	A	A/C	a	AUG	G	C	c/g	−5.4	2.3
Dinoflagellates	Table 4	u	c	g	c	a	G/A	C	C	AUG	G	C	g	−6.2	3.6
LpSOD contexts	AUG1	A	A	A	C	C	U	G	C	**AUG**	U	U	C		
	AUG2	C	C	C	G	A	A	U	C	**AUG**	G	C	A		

The plant kingdom includes all green plants, i.e. green algae, bryophytes, pteridophytes, gymnosperms, and angiosperms; non-angiosperms include bryophytes, pteridophytes, and gymnosperms; lower plants are mostly represented by green algae. Position +1 refers to A in the AUG codon. A score is the sum of logarithms took from fraction/0.25. Consensus sequences fulfill the 50/75 consensus rule described by Cavener (1987)

Table 4 Nucleotide frequencies at positions flanking AUG start codons found for 81 dinoflagellate coding sequences

Nucleotides	Positions											
	−8	−7	−6	−5	−4	−3	−2	−1	**AUG**	+4	+5	+6
A	8	20	18	23	36	40	23	21	AUG	10	6	16
C	30	33	24	30	35	9	53	57	AUG	5	72	23
G	19	22	35	20	14	49	14	20	AUG	67	9	36
U	43	25	23	28	15	2	10	2	AUG	19	14	25
Consensus	u	c	g	c	a	G/A	C	C	**AUG**	G	C	g

Position +1 corresponds to A in the AUG codon. The identified consensus sequence fulfils the 50/75 consensus rule described by Cavener (1987)

at AUG1 would be equipped with both the N-terminal plastid/mitochondrion targeting signal and the C-terminal peroxisome targeting signal. Synthesis of the second isoform, beginning at AUG2, would result in the presence of the C-terminal peroxisome targeting signal only.

Leaky scanning, 'hairpin' structure, and synthesis of the plastid/mitochondrion-targeted isoform of LpSOD

Synthesis of the peroxisome-targeted isoform of LpSOD would be made possible by the so-called leaky scanning (Kaipio et al. 2005; Kozak 2002; Outten and Culotta 2004). During mRNA scanning by 40S ribosome, if AUG1 were to be bypassed, translation would start at AUG2 resulting in the synthesis of a peroxisome-targeted isoform (Fig. 2). Because the first start codon resides in a very weak context (see the previous section), however, it could easily be bypassed in almost all cases, thereby making impossible synthesis of the plastid/mitochondrion-targeted isoform. Therefore, LpSOD transcripts should possess additional characteristics that encourage translation initiation from AUG1.

One way to compel the ribosome to start translation at an AUG residing in a weak context is though generation of a 'hairpin' structure downstream of such a codon (Frolov and Schlesinger 1996; Kozak 1990; Takahashi et al. 2005). This structure pauses mRNA scanning by the 40S ribosome, which increases the probability that a weak start codon will be identified. In her classical paper, Kozak (1990) demonstrated that efficiency of a 'hairpin' structure depends on its position in relation to the first start codon. The strongest effect is observed with a hairpin located 14 nucleotides from the first AUG codon (Kozak 1990). However, this localization disables synthesis of alternative isoforms, because protein synthesis starts at the first start codon in almost all cases. The further the 'hairpin' structure occurs from its optimal location, the more often the first codon is bypassed and protein synthesis begins at a second start codon. This leads to production of respectable amounts of a second isoform (Kozak 1990).

Our analyses of the nucleotides between AUG1 and AUG2 in the LpSOD mRNA indicate that they have the ability to form a stable 'hairpin' structure (Fig. 2). The palindrome program in the EMBOSS 3.0.0 package (Rice et al. 2000) located palindromic sequences (nucleotides 18–24 and 43–37). Moreover, all of the nucleotides involved are cytosine and guanine, indicating high stability of the predicted 'hairpin' structure. In agreement with this finding, the RNA structure 4.2 program (Mathews et al. 2004) and RNAfold from Vienna RNA package (Hofacker 2003) predicted stable secondary structure (nucleotides 15–43) with free energy −20.2 kcal/mol in this region (Fig. 2). Although

Table 5 Results of the evaluation of AUG1 and AUG2 contexts as translation initiation sites by different software tools

Program and its reference	Organism/group	AUG1	AUG2
ATGpr, Salamov et al. (1998)	*Homo sapiens*	0.96	0.67
AUG evaluator, Rogozin et al. (2001)	*Arabidopsis* sp.	–	79
	Aspergillus sp.	–	78
	Drosophila sp.	–	83
	Gallus sp.	–	82
	Homo sapiens	–	86
	Mus sp.	–	87
	Rattus sp.	–	86
	Saccharomyces sp.	–	75
	Xenopus sp.	–	88
	Zea sp.	–	75
Geneid 1.2, Blanco et al. (2002)	*Arabidopsis thaliana*	−3.86	2.48
	Aspergillus nidulans	−1.46	1.63
	Caenorhabditis elegans	−1.82	1.60
	Coprinus cinereus	−1.95	−1.55
	Cryptococcus neomorfans	0.29	1.74
	Dictyostelium discoideum	−3.52	3.50
	Drosophila melanogaster	−1.02	2.13
	Homo sapiens	−1.74	2.76
	Neurospora crassa	−0.22	−0.49
	Oryza sativa	0.70	−0.97
	Plasmodium falciparum	−4.08	0.53
	Tetraodon nigroviridis	3.92	7.30
	Triticum aestivum	−4.59	1.37
HMMgene 1.1a, Krogh (1997)	*Caenorhabditis elegans*	–	1.000
	Homo sapiens	0.062	0.938
NetStart, Pedersen and Nielsen (1997)	*Arabidopsis thaliana*	0.680	0.649
	Vertebrates	0.596	0.839
TIS Miner, Liu and Wong (2003)	Vertebrates	0.433	0.669

the presence and precise localization a 'hairpin' structure must be determined empirically, the palindrome program in the EMBOSS package suggests that it is located only 6 nt from AUG1. This location seems optimal for the synthesis of the two LpSOD isoforms. Were the 'hairpin' structure further away from the first AUG codon (e.g. nucleotides 10–14), it could seriously limit (or even disable) synthesis of the peroxisome-targeted isoform.

Gel electrophoresis suggests existence of several Fe-SOD isoforms in *L. polyedrum*

The detailed bioinformatic analyses presented here indicate that LpSOD transcripts encode the plastid/mitochondrion-targeted isoform, as well as a peroxisome-targeted isoform (Fig. 2). Additional support for the possibility of two isoforms comes from electrophoretic analyzes of the crude extracts of *L. polyedrum* cells made by Okamoto et al. (2001). Interestingly, on their gel (see Fig. 1 in their paper), Fe-SODs are represented not by one but by several bands. Thus, it is possible that one of them corresponds to the plastid/mitochondrion isoform, and another to the peroxisome isoform.

Perspectives

Dinoflagellates are an understudied group of protists. Many questions remain, not only regarding the import mechanism of plastid proteins of class II, but also about the targeting of plastid proteins to compartments other than plastid, such as the mitochondrion or the peroxisome. Until now, no protein had been identified with characteristics indicating it could be targeted to two or three compartments, either in dinoflagellates and in any other algae with complex plastids. LpSOD is the first candidate for such a protein. Although we do not as yet have experimental data confirming the dual (or triple) targeting of this protein, the comprehensive bioinformatic analyses presented here certainly suggest that the dinoflagellate superoxide dismutase is targeted not only to the plastid, but also to the mitochondrion and the peroxisome.

These data provide an important starting point for future experimental studies. The first step in such experiments should be empirical analyses of the structure and expression of the LpSOD gene and protein. It will be interesting to determine whether, in addition to the two start codons, it contains other mechanisms for regulating isoform expression. Experiments, such as primer extension analyses and RNase protection

assays, should demonstrate how many transcript types are generated from the LpSOD gene. In turn, in vitro translation of the LpSOD mRNA should show whether it is used to synthesize two proteins with distinct molecular masses and different targeting properties. Key tests for the targeting hypotheses presented in this paper would be provided by import experiments with Green Fluorescent Protein (GFP). It would be important to determine whether (1) the bipartite presequence of LpSOD is able to render GFP targeting to mitochondria; (2) how efficient the signal peptide-like domain is in the import of GFP into the ER, and (3) whether the C-terminal domain of LpSOD is able to target GFP to peroxisomes. Moreover, modification and/or deletion of the putative 'hairpin' structure in the LpSOD mRNA should demonstrate its influence on the synthesis of LpSOD isoforms. In turn, mutations of the nucleotides surrounding AUG1 and AUG2 codons could be used to examine how these contexts influence the synthesis of the plastid/mitochondrion-targeted and peroxisome-targeted isoform. Thus, our extensive computational investigation provides a clear roadmap for exciting new avenues of empirical research into protein trafficking in these important, but enigmatic and poorly understood organisms.

Acknowledgments We are grateful to Dr M. Wołoszyńska for helpful discussions, K. Moszczyński for his help with Figure 2, and Dr J.W. Stiller for English editing. This work was supported by the Poland State Committee for Scientific Research.

References

Acker H (2005) The oxygen sensing signal cascade under the influence of reactive oxygen species. Philos Trans R Soc Lond B Biol Sci 360:2201–2210

Bannai H, Tamada Y, Maruyama O, Nakai K, Miyano S (2002) Extensive feature detection of N-terminal protein sorting signals. Bioinformatics 18:298–305

Becker T, Qbadou S, Jelic M, Schleiff E (2005) Let's talk about ... chloroplast import. Plant Biol 7:1–14

Bender A, van Dooren GG, Ralph SA, McFadden GI, Schneider G (2003) Properties and prediction of mitochondrial transit peptides from *Plasmodium falciparum*. Mol Biochem Parasitol 132:59–66

Bendsten JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340:783–795

Bhasin M, Raghava GP (2004) ESLpred: SVM-based method for subcellular localization of eukaryotic proteins using dipeptide composition and PSI-BLAST. Nucleic Acids Res 32(Web Server issue):W414–W419

Blanchard JL, Hicks JS (1999) The non-photosynthetic plastid in malarial parasites and other apicomplexans is derived from outside the green plastid lineage. J Eukaryot Microbiol 46:367–375

Blanco E, Parra G, Guigó R (2002) Using geneid to identify genes. In: Baxevanis AD, Davison DB (eds) Current protocols in bioinformatics, vol 1. Wiley, New York (Unit 4.3)

Blokhina O, Virolainen E, Fagerstedt KV (2003) Antioxidants, oxidative damage and oxygen deprivation stress: a review. Ann Bot 91:179–194

Bodén M, Hawkins J (2005) Prediction of subcellular localization using sequence-biased recurrent networks. Bioinformatics 21:2279–2286

Bodył A (2005) Do plastid-related characters support the chromalveolate hypothesis? J Phycol 41:712–719

Bodył A, Moszczyński K (2006) Did the peridinin plastid evolve through tertiary endosymbiosis? A hypothesis. Eur J Phycol 41:437–450

Bruce BD (2001) The paradox of plastid transit peptides: conservation of function despite divergence in primary structure. Biochim Biophys Acta 1541:2–21

Brydges SD, Carruthers VB (2003) Mutation of an unusual mitochondrial targeting sequence of SODB2 produces multiple targeting fates in *Toxoplasma gondii*. J Cell Sci 116:4675–4685

Bueno P, Varela J, Gimenez GG, Del Rio LA (1995) Peroxisomal copper, zinc superoxide dismutase: characterization of the isoenzyme from watermelon cotyledons. Plant Physiol 108:1151–1160

Cavalier-Smith T (2003) Genomic reduction and evolution of novel genetic membranes and protein-targeting machinery in eukaryote-eukaryote chimaeras (meta-algae). Philos Trans R Soc Lond B Biol Sci 359:109–134

Cavener DR (1987) Comparison of the consensus sequence flanking translational start sites in *Drosophila* and vertebrates. Nucleic Acids Res 15:1353–1361

Chew O, Rudhe C, Glaser E, Whelan J (2003) Characterization of the targeting signal of dual-targeted pea glutathione reductase. Plant Mol Biol 53:341–356

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

del Rio LA, Corpas FJ, Scandalio LM, Palma JM, Gomez M, Barroso JB (2002) Reactive oxygen species, antioxidant systems and nitric oxide in peroxisomes. J Exp Bot 53:1255–1272

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

Emanuelsson O, Nielsen H, von Heijne G (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci 8:978–984

Emanuelsson O, Elofsson A, von Heijne G, Cristobal S (2003) In silico prediction of the peroxisomal proteome in fungi, plants and animals. J Mol Biol 330:443–456

Fink RC, Scandalios JG (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. Arch Biochem Biophys 399:19–36

Frolov I, Schlesinger S (1996) Translation of sindbis virus mRNA: analysis of sequences downstream of the initiating AUG codon that enhance translation. J Virol 70:1182–1190

Fünfschilling U, Rospert S (1999) Nascent polypeptide-associated complex stimulates protein import into yeast mitochondria. Mol Biol Cell 10:3289–3299

George R, Beddoe T, Landl K, Lithgow T (1998) The yeast nascent polypeptide-associated complex initiates protein targeting to mitochondria *in vivo*. Proc Natl Acad Sci USA 95:2296–2301

George R, Walsh P, Beddoe T, Lithgow T (2002) The nascent polypeptide-associated complex (NAC) promotes interaction of ribosomes with the mitochondrial surface *in vivo*. FEBS Lett 516:213–216

Gibbs SP (1978) The chloroplasts of *Euglena* may have evolved from symbiotic green algae. Can J Bot 56:2883–2889

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
Gould SB, Sommer MS, Hadfi K, Zauner S, Kroth PG, Maier U-G (2006) Protein targeting into the complex plastid of cryptophytes. J Mol Evol 62:674–681
Gould SG, Keller GA, Subramani S (1987) Identification of a peroxisomal targeting signal at the carboxy terminus of firefly luciferase. J Cell Biol 105:2923–2931
Gould SJ, Keller GA, Hosken N, Wilkinson J, Subramani S (1989) A conserved tripeptide sorts proteins to peroxisomes. J Cell Biol 108:1657–1664
Graham LE, Wilcox LW (2000) Algae. Prentice Hall, UK
Guda C, Guda P, Fahy E, Subramaniam S (2004) MITOPRED: a web server for the prediction of mitochondrial proteins. Nucleic Acids Res 32:W372–W374
Guda C, Subramaniam S (2005) TARGET: a new method for predicting protein subcellular localization in eukaryotes. Bioinformatics 21:3963–3969
Hackett JD, Yoon HS, Soares MB, Bonaldo MF, Casavant TL, Scheetz TE, Nosenko T, Bhattacharya D (2004) Migration of the plastid genome to the the nucleus in a peridinin dinoflagellate. Curr Biol 14:213–218
Harper JT, Keeling PJ (2003) Nucleus-encoded, plastid-targeted glyceraldehydes-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. Mol Biol Evol 20:1730–1735
Hiller K, Grote G, Scheer M, Münch R, Jahn D (2004) PrediSi: prediction of signal peptides and their cleavage positions. Nucleic Acids Res 32(Web Server issue):W375–W379
Hofacker IL (2003) Vienna RNA secondary structure server. Nucleic Acids Res 31:3429–3431
Hofmann K, Stoffel W (1993) TMbase — a database of membrane spanning proteins segments. Biol Chem 374:166
Horton P, Park KJ, Obayashi T, Nakai K (2006) Protein subcellular localization prediction with WoLF PSORT. In: Jiang T et al. (eds) Proceedings of the 4th Asia-Pacific Bioinformatics Conference (APBC2006). Imperial College Press, Taibei, pp39–48
Hua S, Sun Z (2001) Support vector machine approach for protein subcellular localization prediction. Bioinformatics 17:721–728
Ishida K (2005) Protein targeting into plastids: a key to understanding the symbiogenetic acquisitions of plastids. J Plant Res 118:237–245
Joshi CP, Zhou H, Huang X, Chiang VL (1997) Context sequences of translation initiation codon in plants. Plant Mol Biol 35:993–1001
Kaipio K, Kallio J, Pesonen U (2005) Mitochondrial targeting signal in human neuropeptide Y gene. Biochem Biophys Res Commun 35:633–640
Käll L, Krogh A, Sonnhammer ELL (2004) A combined transmembrane topology and signal peptide prediction method. J Mol Biol 338:1027–1036
Kilian O, Kroth PG (2005) Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. Plant J 41:175–183
Koumandou VL, Nisbet RER, Barbrook AC, Howe CJ (2004) Dinoflagellate chloroplasts — where have all the genes gone? Trends Genet 20:261–267
Kozak M (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res 15:8125–8132
Kozak M (1990) Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. Proc Natl Acad Sci USA 87:8301–8305
Kozak M (2002) Pushing the limits of the scanning mechanism for initiation of translation. Gene 299:1–34
Krogh A (1997) Two methods for improving performance of an HMM and their application for gene finding. Proc Int Conf Intell Syst Mol Biol 5:179–186
Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580
Kuroiwa T, Sakaguchi M, Mihara K, Omura T (1991) Systematic analysis of stop-transfer sequence for microsomal membrane. J Biol Chem 266:9251–9255
Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of protein. J Mol Biol 157:105–132
Lao DM, Shimizu T (2001) A method for discriminating a signal peptide and a putative 1st transmembrane segment. In: Valafar F (ed) Proceedings of the 2001 International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences (METMBS'01). CSREA Press, USA, pp119–125
Leander BS, Keeling PJ (2004) Early evolutionary history of dinoflagellates and apicomplexans (Alveolata) as inferred from HSP90 and actin phylogenies. J Phycol 40:341–350
Liu H, Wong L (2003) Data mining tools for biological sequences. J Bioinform Comput Biol 1:139–168
Lucattini R, Likic VA, Lithgow T (2004) Bacterial proteins predisposed for targeting to mitochondria. Mol Biol Evol 21:652–658
Macasev D, Newbigin E, Whelan J, Lithgow T (2000) How do plant mitochondria avoid importing chloroplast proteins? Components of the import apparatus To20 and Tom22 from *Arabidopsis* differ from their fungal counterparts. Plant Physiol 123:811–816
Manfredini V, Roehrs R, Peralba MC, Henriques JA, Saffi J, Ramos AL, Benfato MS (2004) Glutathione peroxides induction protects *Saccharomyces cerevisiae sod1Δsod2Δ* double mutants against oxidative damage. Braz J Med Biol Res 37:159–165
Martoglio B, Dobberstein B (1998) Signal sequences: more than just greasy peptides. Trends Cell Biol 8:410–415
Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc Natl Acad Sci USA 101:7287–7292
McFadden GI, van Dooren GG (2004) Red algal genome affirms a common origin of all plastids. Curr Biol 14:514–516
Mukhopadhyay A, Ni L, Weiner H (2004) A co-translational model to explain the in vivo import of proteins into HeLa cell mitochondria. Biochem J 382:385–392
Murcha MW, Rudhe C, Elhafez D, Adams KL, Daley DO, Whelan J (2005) Adaptations required for mitochondrial import following mitochondrial to nucleus gene transfer of ribosomal protein S10. Plant Physiol 138:2134–2144
Nair R, Rost B (2005) Mimicking cellular sorting improves prediction of subcellular localization. J Mol Biol 348:85–100
Nakai K, Horton P (1999) PSORT: a program for detecting the sorting signals of proteins and predicting their subcellular localization. Trends Biochem Sci 24:34–36
Nassoury N, Cappadocia M, Morse D (2006) Plastid ultrastructure defines the protein import pathway in dinoflagellates. J Cell Sci 116:2867–2874
Neuberger G, Maurer-Stroh S, Eisenhaber B, Hartig A, Eisenhaber F (2003a) Motif refinement of the peroxisomal targeting signal 1 and evaluation of taxon-specific differences. J Mol Biol 328:567–579
Neuberger G, Maurer-Stroh S, Eisenhaber B, Hartig A, Eisenhaber F (2003b) Prediction of peroxisomal targeting signal 1 containing proteins from amino acid sequence. J Mol Biol 328:581–592
Nielsen H, Krogh A (1998) Prediction of signal peptides and signal anchors by a hidden Markov model. Proc Int Conf Intell Syst Mol Biol 6:122–130

Okado-Matsumoto A, Fridovich I (2001) Subcellular distribution of superoxide dismutases (SOD) in rat liver: Cu, Zn-SOD in mitochondria. J Biol Chem 276:38388–38393

Okamoto OK, Robertson DL, Fagan TF, Hastings JW, Colepicolo P (2001) Different regulatory mechanisms modulate the expression of a dinoflagellate iron-superoxide dismutase. J Biol Chem 276:19989–19993

Outten CE, Culotta VC (2004) Alternative start sites in the *Saccharomyces cerevisiae GLR1* gene are responsible for mitochondrial and cytosolic isoforms of glutathione reductase. J Biol Chem 279:7785–7791

Park KJ, 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 NJ, Waller RF, Archibald JM, Keeling PJ (2005) Complex protein targeting to dinoflagellate plastids. J Mol Biol 348:1015–1024

Pedersen AG, Nielsen H (1997) Neural network prediction of translation initiation sites in eukaryotes: perspectives for EST and genome analysis. Proc Int Conf Intell Syst Mol Biol 5:226–233

Peeters N, Small I (2001) Dual targeting to mitochondria and chloroplasts. Biochim Biophys Acta 1541:54–63

Pesole G, Gissi C, Grillo G, Licciulli F, Liuni S, Saccone C (2000) Analysis of oligonucleotide AUG start codon context in eukariotic mRNAs. Gene 261:85–91

Petrova V, Drescher D, Kujumdzieva AV, Schmitt MJ (2004) Dual targeting of yeast catalase A to peroxisomes and mitochondria. Biochem J 380:393–400

Plewes KA, Barr SD, Gedamu L (2003) Iron superoxide dismutases targeted to the glycosomes of *Leishmania chagasi* are important for survival. Infect Immun 71:5910–5920

Reczko M, Staub E, Fiziev P, Hatzigeorgiou A (2002) Finding signal peptides in human protein sequences using recurrent neural networks. Lect Notes Comput Sci 2452:60–67

Reumann S (2004) Specification of the peroxisome targeting signals type 1 and type 2 of plant peroxisomes by bioinformatics analyses. Plant Physiol 135:783–800

Rice P, Longden I, Bleasby A (2000) EMBOSS: the European molecular biology open software suite. Trends Genet 16:276–277

Rogozin IB, Kochetov AV, Kondrashov FA, Koonin EV, Milanesi L (2001) Presence of ATG triplets in 5' untranslated regions of eukaryotic cDNAs correlates with a 'weak' context of the start codon. Bioinformatics 17:890–900

Rudhe C, Clifton R, Whelan J, Glaser E (2002) N-terminal domain of the dual-targeted pea glutathione reductase signal peptide controls organellar targeting efficiency. J Mol Biol 324:577–585

Rudhe C, Clifton R, Chew O, Zemam K, Richter S, Lamppa G, Whelan J, Glaser E (2004) Processing of the dual targeted precursor protein of glutathione reductase in mitochondria and chloroplasts. J Mol Biol 343:639–647

Shalchian-Tabrizi K, Skånseng M, Ronquist F, Klaveness D, Bachvaroff TR, Delwiche CF, Botnen A, Tengs T, Jakobsen KS (2006) Heterotachy processes in rhodophyte-derived secondhand plastid genes: implications for addressing the origin and evolution of dinoflagellate plastids. Mol Biol Evol 23:1504–1515

Salamov AA, Nishikawa T, Swindells MB (1998) Assessing protein coding region integrity in cDNA sequencing projects. Bioinformatics 14:384–390

Salzberg SL (1997) A method for identifying splice sites and translation start sites in eukaryotic mRNA. Comput Appl Biosci 13:365–376

Sarda D, Chua GH, Li KB, Krishnan A (2005) pSLIP: SVM based protein subcellular localization prediction using multiple physicochemical properties. BMC Bioinformatics 6:152

Scandalios JG (2005) Oxidative stress: molecular perception and transduction of signals triggering antioxidant gene defences. Braz J Med Biol Res 38:995–1014

Schein AI, Kissinger JC, Ungar LH (2001) Chloroplast transit peptide prediction: a peek inside the black box. Nucleic Acids Res 29:E82

Schnepf E, Elbrächter M (1999) Dinophyte chloroplasts and phylogeny — a review. Grana 38:81–97

Seyler RW Jr, Olson JW, Maier RJ (2001) Superoxide dismutase-deficient mutants of *Helicobacter pylori* are hypersensitive to oxidative stress and defective in host colonization. Infect Immun 69:4034–4040

Sienkiewicz N, Daher W, Dive D, Wrenger C, Viscogliosi E, Wintjens R, Jouin H, Capron M, Müller S, Khalife J (2004) Identification of a mitochondrial superoxide dismutase with an unusual targeting sequence in *Plasmodium falciparum*. Mol Biochem Parasitol 137:121–132

Silva-Filho MC (2003) One ticket for multiple destinations: dual targeting of proteins to distinct subcellular locations. Curr Opin Plant Biol 6:589–595

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

Small I, Wintz H, Akashi K, Mireau H (1998) Two birds with one stone: genes that encode products targeted to two or more compartments. Plant Mol Biol 38:265–277

Takahashi K, Maruyama M, Tokuzawa Y, Murakami M, Oda Y, Yoshikane N, Makabe KW, Ichisaka T, Yamanaka S (2005) Evolutionary conserved non-AUG translation initiation in *NAT1/p97/DAP5* (*EIF4G2*). Genomics 85:360–371

van Dooren GG, Schwartzbach SD, Osafune T, McFadden GI (2001) Translocation of proteins across the multiple membranes of complex plastids. Biochim Biophys Acta 1541:34–53

von Heijne G (1992) Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. J Mol Biol 225:487–494

Verner K (1993) Co-translational protein import into mitochondria: an alternative view. Trends Biochem Sci 18:366–371

Villarejo A, Buren S, Larsson S, Dejardin A, Monne M, Rudhe C, Karlsson J, Jansson S, Lerouge P, Rolland N, von Heijne G, Grebe M, Bako L, Samuelsson G (2005) Evidence for a protein transported through the secretory pathway *en route* to the higher plant chloroplast. Nat Cell Biol 7:1224–1231

Wolfe-Simon F, Grzebyk D, Schofield O, Falkowski PG (2005) The role and evolution of superoxide dismutases in algae. J Phycol 41:453–465

Woodruff RC, Phillips JP, Hilliker AJ (2004) Increased spontaneous DNA damage in Cu/Zn superoxide dismutase (SOD1) deficient *Drosophila*. Genome 47:1029–1035

Xie D, Li A, Wang M, Fan Z, Feng H (2005) LOCSVMPSI: a web server for subcellular localization for eukaryotic proteins using SVM and profile of PSI-BLAST. Nucleic Acids Res 33(Web Sever issue):W105–W110

Yoon HS, Hackett JD, Van Dolah FM, Nosenko T, Lidie KL, Bhattacharya D (2005) Tertiary endosymbiosis driven genome evolution in dinoflagellate algae. Mol Biol Evol 22:1299–1308

Yu CS, Lin CJ, Hwang JK (2004) Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein Sci 13:1402–1406

Zhang MQ (1998) Statistical features of human exons and their flanking regions. Hum Mol Genet 7:919–932

Zhang XP, Glaser E (2002) Interaction of plant mitochondrial and chloroplast signal peptides with the Hsp70 molecular chaperone. Trends Plant Sci 7:14–21