ELSEVIER

Biochimica et Biophysica Acta 1541 (2001) 54–63

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

www.bba-direct.com

Review

Dual targeting to mitochondria and chloroplasts

Nemo Peeters [a,*], Ian Small [b]

[a] *Department of Molecular Biology and Genetics, 321 Biotechnology Building, Cornell University, Ithaca, NY 14853-2703, USA*
[b] *Unité de Recherches en Génomique Végétale (INRA-URGV), 2, rue Gaston Crémieux, CP 5708, 91057 Evry cedex, France*

Received 13 August 2001; accepted 13 August 2001

Abstract

Plant cells contain two organelles originally derived from endosymbiotic bacteria: mitochondria and plastids. Their endosymbiotic origin explains why these organelles contain their own DNA, nonetheless only a few dozens of genes are actually encoded by these genomes. Many of the other genes originally present have been transferred to the nuclear genome of the host, the product of their expression being targeted back to the corresponding organelle. Although targeting of proteins to mitochondria and chloroplasts is generally highly specific, an increasing number of examples have been discovered where the same protein is imported into both organelles. The object of this review is to compare and discuss these examples in order to try and identify common features of dual-targeted proteins. The study helps throw some light on the factors determining organelle targeting specificity, and suggests that dual-targeted proteins may well be far more common than once thought.

Keywords: Mitochondria; Chloroplast; Protein import; Dual targeting

1. Introduction

Plant cells contain two organelles originally derived from endosymbiotic bacteria: mitochondria and plastids. Amongst many other functions, these two organelles are best known for their roles in energy metabolism, notably respiration and photosynthesis. The closest bacterial organisms to the endosymbiotic ancestors of these organelles have nearly a thousand genes (*Rickettsia* [1]) or several thousands (cyanobacteria [2]). Since the endosymbiosis, many of the genes of the endosymbiotic bacteria have been lost, leaving the organelle genomes with less than a hundred protein-coding genes each [3,4]. The vast majority of mitochondrial and plastid proteins are encoded in the nucleus, synthesized by cytosolic ribosomes and subsequently imported into the organelles via active protein transport systems. The total number of proteins present in mitochondria and chloroplasts is thought to be about 2000–3000 for each of them [5].

Mitochondria originated much earlier than plastids and thus the first plastids arose in cells that already contained an efficient system for targeting cytosolically synthesized proteins to mitochondria. One might have expected evolution to have seized this opportunity to reuse the same machinery for targeting proteins to plastids, but in fact this seems not to be the case; the two protein import systems have clearly been derived independently and do not share homology. In this situation, it is thus easy to understand why protein targeting is usually highly specific.

* Corresponding author.
E-mail address: nmp23@cornell.edu (N. Peeters).

0167-4889/01/$ – see front matter
PII: S0167-4889(01)00146-X

Nevertheless, it is becoming increasingly clear that despite the profound differences in the two import machineries, a certain number of proteins are efficiently recognized by both systems and are imported into both organelles.

The object of this review is to present and discuss the remarkable cases where indeed proteins encoded by a single gene in the nuclear genome are targeted to both mitochondria and chloroplasts.

2. Mitochondrial and chloroplast protein import systems

Our intention is not to go into the details of each protein import apparatus, since there are several good recent reviews that cover this subject [6–9] and other reviews in this issue. For proteins encoded in the nucleus and then translated in the cytosol there are many different possible final locations. In the *Arabidopsis thaliana* nuclear genome probably 20–25% of the genes encode proteins that are targeted to either mitochondria or chloroplasts [5]. In each of these organelles the proteins can then be located in the membranes, intermembrane spaces or in the matrix or stroma. The vast majority of the organellar proteins have to cross the double membrane via the protein import channel.

Mitochondria have two complexes of proteins called TOM proteins and TIM proteins, respectively located in the outer membrane and the inner membrane, which together form the protein import channel. The proteins that will be imported generally have a mitochondrial targeting sequence located at the N-terminus, although there are proteins that have internal or even C-terminal targeting signals. This latter case has been found only once, for a yeast mitochondrial helicase [10]. The N-terminal presequence cannot be described as a consensus sequence but contains conserved features that can be identified with more or less confidence. In plants, mitochondrial targeting sequences are generally longer than in other organisms (40 amino acids on average) [9], they have a net positive charge (rich in arginine and poor in acidic amino acids) and contain many aliphatic residues (mainly leucine and alanine). The structure adopted by the presequence is generally an amphiphilic α helix [11]. It can be noted also that plant mitochondrial targeting sequences are particularly rich in serine residues.

How the translated protein is actually targeted to the mitochondria is not well understood yet. In the case of a protein targeted to the matrix of mitochondria and possessing an N-terminal presequence as described above, cytosolic protein factors interact with the presequence. These factors are generally chaperones and can require ATP. The presequence is then transferred to the mitochondrial TOM complex proteins. These proteins, namely TOM70, TOM20 and TOM22, are generally negatively charged and can thus form electrostatic interactions with the presequence. Once the presequence is engaged in the outer membrane channel, negative charges present on the inner membrane protein TIM23, along with the electrochemical gradient across the inner membrane (gradient created by the electron transport along the mitochondrial respiratory chain), allow the presequence to tow the protein through both the outer and the inner membrane. The last steps of protein import are carried out by mitochondrial chaperones, which literally pull the protein inside the matrix. The imported protein can then be cleaved from its import signal by specific proteases, and be refolded to carry out its function inside the mitochondria.

Chloroplasts also possess an outer envelope protein complex called TOC, and an inner envelope protein complex, TIC, which differ in many ways from the equivalent mitochondrial complexes. Chloroplast proteins can be located in even more compartments than mitochondrial proteins. In addition to the envelope membranes and the inter membrane space and the stroma, many important chloroplast proteins (photosynthesis-related proteins) are located in the membrane and the lumen of the thylakoids. We will focus here only on the presequence needed for the protein to cross the double membrane envelope of the chloroplast. These targeting sequences are different from their mitochondrial counterparts but do present some similarities. They are about 50 amino acids long; rich in the hydroxylated residue serine and unlike mitochondrial presequences they do not contain many positively charged residues, especially in the first ten amino acids, and do not contain many leucine residue. However, like mitochondrial targeting sequences, they contain very few acidic amino

acids. The structure of the presequence is somewhat less well defined than for mitochondria [12]. Proteins targeted to the chloroplast are probably also recognized in the cytosol by chaperone proteins [12], before interacting with the components of the import machinery. The major difference with protein import into mitochondria is that there is no comparable electrical gradient in chloroplasts. None of the proteins from the TOC and TIC complexes have homologues in the TOM or TIM machinery [13]. Protein import into chloroplasts largely depends on the subsequent action of different protein chaperones, the process requiring GTP and ATP. A large GTPase protein, TOC160, is one of the most cytosolic-accessible TOC proteins, and is involved in recognition of the presequence. The TOC and TIC protein complexes are in close contact with each other. TIC22 is the first protein from the inner membrane complex to interact with the presequence [14]. TIC110 is believed to form the canal through which the proteins are eventually imported into the stroma. It seems that TIC110 is also in close interaction with stromal chaperones, which could be the final motor for the import of the chloroplast-targeted protein. As in mitochondria, specific proteases can remove the presequence from the mature protein.

3. Specificity of import

Before the presence of chloroplasts in plant cells, the protein import machinery in mitochondria had already co-evolved with the targeting sequence permitting the import of passenger proteins. Starting with the origin of plants, approximately 800 million years ago, the mitochondrial protein import system had to cope with a new set of proteins that had to be targeted to a new semi-autonomous organelle, the chloroplast. To maintain the specificity of import, co-evolution between import systems and signals on the imported proteins must have been essential to avoid unnecessary or lethal mis-targeting between organelles. An interesting observation is that plant mitochondrial presequences differ from other eukaryotic mitochondrial presequences, in being usually longer and also richer in serine [9]. Interestingly, it seems that the outer mitochondrial membrane proteins TOM20 and TOM22 in plants have greatly diverged from their other eukaryotic counterparts. This could possibly have been a response of the mitochondrial import system to the arrival of chloroplasts, so that the two protein import systems would interfere less [15].

As mitochondria and chloroplasts are composed of several thousand proteins encoded in the nucleus and imported into the respective organelles, the first and evident observation we can make is that targeting is generally highly specific for one or the other organelle. Even though there may be some proteins that could be shared in theory by both organelles, most of the proteins have a specific function in the organelle to which they are targeted, and one can imagine that mis-targeting has been counterselected as an unnecessary loss of valuable proteins, or might even be detrimental for the organelle.

However, there are several cases where mis-targeting has been observed. A chloroplast targeting sequence from the small subunit of ribulose-1,5-bisphosphate carboxylase from the green algae *Chlamydomonas reinhardtii*, can target dihydrofolate reductase as a reporter gene into bakers' yeast mitochondria [16]. Another chloroplast protein, triosephosphate-3-phosphoglycerate phosphate translocase, from spinach, can also be targeted to fungal mitochondria [17]. These two examples are cases where mistargeting was observed in in vitro import conditions, into isolated mitochondria. Other chloroplast proteins can also be imported into isolated mitochondria in vitro (A. Smith, personal communication).The first demonstration that mistargeting can occur in vivo was with transgenic plants expressing the bakers' yeast mitochondrial cytochrome oxidase subunit Va presequence fused to chloramphenicol acetyltransferase (CAT) as a reporter gene [18]. This showed clearly that dual targeting to both organelles is a possibility and has to be taken into consideration when examining targeting in plants.

4. Dual targeting

4.1. Demonstrating dual targeting

Numerous techniques are available for studying targeting of proteins to organelles, not all of which are ideal when dual targeting is a possibility. One can

Table 1
Dual targeted proteins

Protein	Accession no.	Type of dual targeting		Donor species	Host species	Experimental evidence for dual targeting		References
		Ambiguous signal	Twin presequence			In vitro	In vivo	
CoxVa	NP_014346	✓		Yeast	Tobacco		CAT fusion	[18]
Triose phosphate 3-phosphoglycerate phosphate translocator	P11869	✓		Spinach	Spinach, bean	✓		[35][a]
Glutathione reductase	P27456	✓		Pea	Tobacco		Enzyme activity, PAT fusion	[30]
Ferrochelatase-I	P42043	✓		*Arabidopsis*	Pea	✓		[31][a]
Methionyl-tRNA synthetase	O23761	✓		*Arabidopsis*	Pea, tobacco	✓	GFP fusion, enzyme activity	[24]
Histidyl-tRNA synthetase	AF020715	✓		*Arabidopsis*	Tobacco		GFP fusion	[25]
Cysteinyl-tRNA synthetase	AC005311	✓		*Arabidopsis*	Tobacco		GFP fusion	[26]
Asparaginyl-tRNA synthetase	AJ222644	✓		*Arabidopsis*	Tobacco		GFP fusion	[26]
RNA polymerase T2	CAC17120	✓		*Arabidopsis*	*Arabidopsis*, tobacco		GFP fusion	[28]
Mercaptopyruvate sulfurtransferase	BAA85149	✓		*Arabidopsis*	*Arabidopsis*		GFP fusion, partial immunoblot	[19][b]
Methionine amino-peptidase (MAP1C)	AAG33976	✓		*Arabidopsis*	Onion		GFP fusion	[29]
Methionine amino-peptidase (MAP1D)	AAG33977	✓		*Arabidopsis*	Onion		GFP fusion	[29]
Peptide deformylase (PDF1B)	AAG33980	✓		*Arabidopsis*	Onion		GFP fusion	[29]
Peptide deformylase (PDF1B)	AAG33972	✓		Tomato	Onion		GFP fusion	[29]
Glycyl-tRNA synthetase	O23150	✓		*Arabidopsis*	Tobacco, *Nicotiana benthamiana*, potato	✓	GFP fusion, enzyme activity	[27]
Lysyl-tRNA synthetase unpublished	AP000603	✓		*Arabidopsis*	Tobacco		GFP fusion	Peeters et al.,
Tryptophanyl-tRNA synthetase unpublished	AF058914	✓		*Arabidopsis*	Tobacco		GFP fusion	Peeters et al.,
Pseudouridine synthase unpublished	O22928	✓		*Arabidopsis*	Tobacco		GFP fusion	Peeters et al.,
RNA binding protein unpublished	BAB03001	✓		*Arabidopsis*	Tobacco		GFP fusion	Gualberto et al.,
Phosphoribosyl aminoimidazole (AIR) synthase	AAC14578	✓		Cowpea	Cowpea		Immunoblot	[32]

Table 1
Dual targeted proteins

Protein	Accession no.	Type of dual targeting		Donor species	Host species	Experimental evidence for dual targeting		References
		Ambiguous signal	Twin presequence			In vitro	In vivo	
Protoporphyrinogen oxidase-II	AAC97124		✓	Spinach	Spinach		Immunoblot, GFP fusions	[22]
THI1	AB046993		✓	*Arabidopsis*	Tobacco		GUS fusion, immunogold	[23]

The origin of the protein, and the experimental evidence for dual targeting, are indicated. The first entry in this table is a yeast mitochondrial targeting sequence expressed in tobacco. The other 21 entries are recently discovered examples of plant dual targeting presequences.
[a]In vivo results disprove in vitro observations.

purify the organelles from plant tissue and monitor the presence of a specific protein in the organellar fractions by Western blot or enzyme activity, but this approach can suffer from problems with cross-contamination and insufficient specificity of the antibodies or the reaction being followed. Alternatively, one can study import of in vitro translated polypeptides into isolated organelles, but this approach potentially suffers from the necessarily nonphysiological conditions employed and the lack of the correct cellular context (cytoskeleton, cytosolic chaperones, other organelles, etc.). Finally, one can study targeting in situ by immunofluorescence techniques or expressing fusion proteins containing a visible marker (usually green fluorescent protein, GFP). The latter techniques are probably the best adapted for the study of dual-targeted proteins and GFP fusions in particular are being increasingly employed for this purpose. There remains a doubt about to what extent the targeting is influenced by the marker protein or by the fact that the fusion protein is often greatly overexpressed with respect to the corresponding natural protein. In our experience, we have never observed dual targeting of control GFP fusions known to be specifically targeted to one or other organelle, whatever the degree of overexpression. In rare cases, with some fusions, the import apparatus appears to saturate and some fusion protein can build up in the cytosol, but again the targeted protein remains specific to one organelle. The only published example we are aware of where GFP fusions have given conflicting results to other techniques (in this case immunoblot detection in subcellular fractions) is for *Arabidopsis* mercaptopyruvate sulfurtransferase [19]. In any case, the best studies confirm targeting by comparing two or three complementary techniques, but this is not always a viable option.

4.2. Examples of dual targeting

There are many examples in eukaryotes where one gene produces products located in different parts of the cell. A recent review describes some of the different possibilities in plant cells [20]. In this review, we will concentrate on cases where the product(s) of a single gene is (are) located in mitochondria and chloroplasts. Although this was first discovered [18] as a possibility in a heterologous system (a yeast mitochondrial presequence dual targeting a passenger protein in tobacco cells), in the last few years an increasing number of natural examples have been discovered (Table 1).

There are two basic ways in which a single gene can provide a product to both organelles (Fig. 1).

4.2.1. Twin presequences

It can have 'twin' targeting sequences, represented by a mitochondrial and a chloroplast targeting sequence in tandem at the N-terminus of the protein. By having alternative transcription starts (Fig. 1Ai), alternative translation starts (Fig. 1Aii), alternative exon splicing (Fig. 1Aiii), or a combination of the above, two proteins can be made from the same gene. Each of the two proteins has a different presequence located at its N-terminus. The same mature protein can then be targeted to mitochondria and

chloroplasts by different targeting sequences. The longer protein has two targeting sequences directly following each other. As has been shown previously by experiment, in this case, it is the most N-terminal presequence that dictates the destination of the protein [21]. It is only recently that it has been proved that nature can adopt this way of dual targeting a protein to both organelles. In spinach, an enzyme necessary for the biosynthesis of chlorophyll (chloroplasts) and haem (chloroplasts and mitochondria) has been found to be dual-targeted. This protein, protoporphyrinogen oxidase II (protox-II) has two in-frame initiation codons, and two different proteins are made by alternative translation, the longer protein being imported into chloroplasts and the shorter one into mitochondria [22]. Dual targeting of Arabidopsis THI1 protein (an enzyme of the thiamine biosynthesis pathway) also appears to use two alternative translation starts (see Fig. 1Ai), again the longer protein being targeted to chloroplasts and the shorter one to mitochondria [23].

4.2.2. *Ambiguous presequences*

The second way of obtaining dual targeting is to have a targeting presequence, which we will call 'ambiguous' (Fig. 1B), that is recognized as an import signal by both mitochondria and chloroplasts (although it may well not be the same regions of

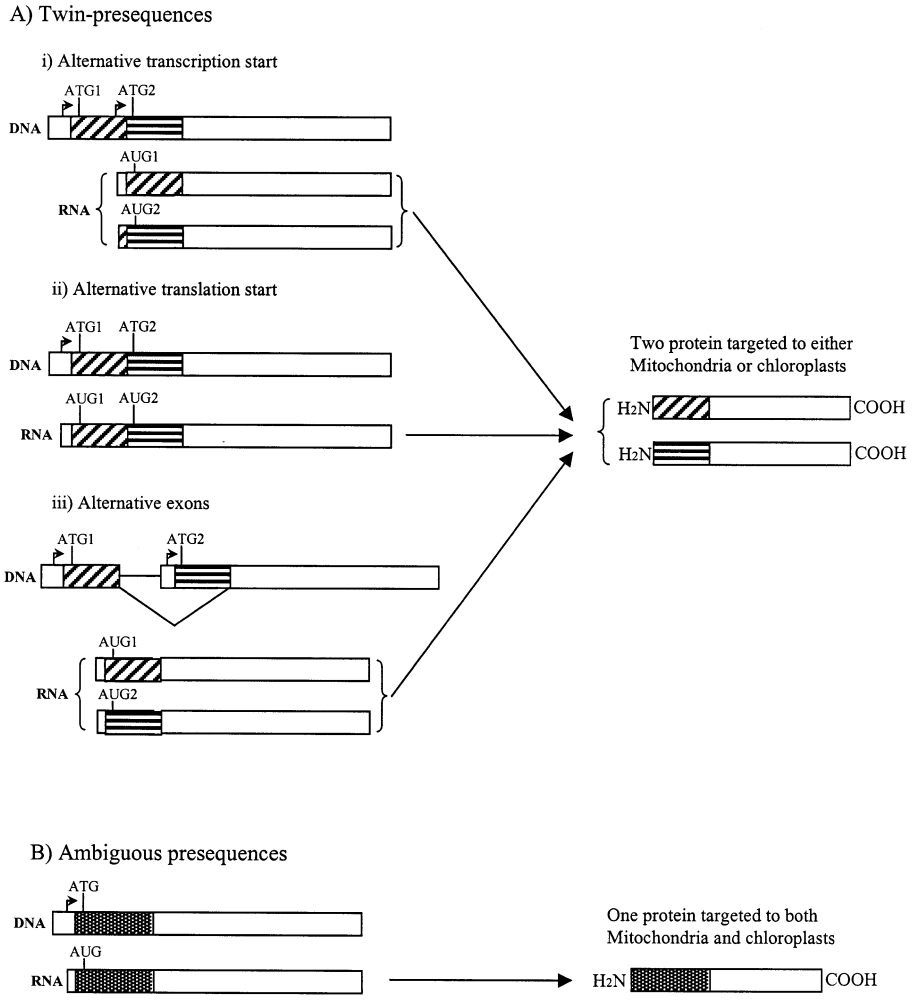

Fig. 1. One nuclear gene for two organellar proteins. One gene can encode for two different proteins by alternative transcription (Ai), alternative translation starts (Aii), alternative exon splicing (Aiii), or a combination of the above. The longer protein can be targeted to one organelle and the shorter one to the other. A single gene can also code for a single polypeptide that has an ambiguous targeting signal, allowing the protein to be targeted to both mitochondria and chloroplasts (B).

the targeting sequence that are recognized). To date we are aware of 19 proposed examples of dual targeting of proteins to both mitochondria and chloroplast by means of an ambiguous targeting signal (Table 1). A large majority of them are proteins that are involved in gene expression in the organelles: in this group we find aminoacyl-tRNA synthetases (MetRS [24], HisRS [25], CysRS, AspRS [26], GlyRS [27], LysRS, TrpRS (Peeters et al., unpublished), but also an RNA polymerase [28], two methionine aminopeptidases that remove the first methionine from translated proteins, and a peptidyl deformylase, that removes the formyl group from the initiator methionine [29]. Another dual targeted enzyme indirectly related to gene expression in organelles is a pseudouridine synthase, which probably modifies bases on organelle tRNA (Peeters et al., unpublished). Other examples of dual targeting are a RNA binding protein (Gualberto et al., unpublished), glutathione reductase [30] and ferrochelatase-I [31] both probably related to protection against oxidative stress, mercaptopyruvate sulfurtransferase [19] (physiological role unknown), and phosphoribosyl aminoimidazol synthase, involved in purine synthesis [32], although the evidence for dual targeting for the latter three enzymes is not conclusive.

To date, little detailed experimental work has been carried out on these ambiguous targeting sequences to discover why they lack the high specificity of most organelle targeting sequences. Formally, one can imagine two hypotheses to explain ambiguous targeting sequences. One possibility is that they contain a new type of signal, different from those studied at present, recognized by a specific receptor or import pathway shared by both organelles. The second possibility is that they are imported via the same route as specifically targeted proteins and that therefore they must contain a mixture of signals capable of being recognized by the two different import pathways on mitochondria and plastids.

The evidence from in vitro work suggests that the import pathway used by these proteins is indistinguishable from that taken by other imported precursors. This is supported by a cursory analysis of the composition of these targeting sequences (Table 2). The 39 residues following the initiator methionine of each of the 19 ambiguous sequences were analyzed and compared to the amino acid composition obtained with the 39 residues at the same position from 138 mitochondrial, 428 chloroplast targeting sequences and 290 839 residues of plant cytosolic proteins. Even though this analysis has been done on only a small group of dual targeted proteins, it is sufficient to reveal some general features. The ambiguous targeting signals are poor in negatively charged residues (aspartate, D and glutamate, E) just like classical mitochondrial or chloroplast targeting signals. Again, like classical targeting organelle targeting signals, they are enriched in arginine and serine, the values being intermediate between mito-

Table 2
The most significant amino acid composition differences between targeting presequences

Amino acid	Targeting sequence			Cytoplasmic proteins
	Ambiguous	Mitochondrial	Chloroplastic	
A	**7.4**	12.0	12.9	**7.9**
D	**0.9**	**1.8**	**0.8**	5.5
E	**0.6**	**1.9**	**0.9**	6.9
F	8.1	**3.4**	**4.9**	**3.9**
L	14.0	11.6	**8.6**	**8.7**
R	9.0	11.3	**5.8**	**5.0**
S	18.1	16.2	19.5	**6.9**

The percentage of different amino acids was calculated on the first 39 amino acids after the initiator methionine. These percentages were calculated on 19 ambiguous targeting presequences (see Table 1 for the list of the 19 sequences), 138 mitochondrial targeting presequences and 428 chloroplast targeting presequences. These percentages have been compared to the composition of plant cytosolic proteins (calculated from 290 839 residues). The protein sequences used are available from the authors on request. Mitochondrial and Chloroplastic percentages were compared to the cytosolic composition, and ambiguous targeting sequence composition was compared to Mitochondrial, Chloroplastic and Cytosolic data. A chi-square test of one degree of freedom was performed, and entries in the table found to be significantly different from each other ($P < 0.001$) are shown in bold.

chondrial and plastid targeting sequences. Two remarkable features are the low content in alanine residues (only 7.4%, as opposed to the 12% in regular organelle targeting presequences), and the high content in phenylalanine and leucine (22%, as opposed to 14%), suggesting that the dual-targeted presequences may be more hydrophobic on average. Secondary structure prediction of poorly conserved sequences is never easy and the results should be interpreted with caution, but it seems less easy to demonstrate potential N-terminal α helices in these ambiguous targeting sequences than in classical mitochondrial targeting sequences. All in all, one gets the impression that these sequences are intermediary in character between mitochondrial and plastid targeting sequences and that they contain features from both. An inter-

Table 3
Prediction of targeting by TargetP and Predotar for the dual-targeted proteins

Protein	Sequence	Predotar			Target P						Predicted locations
		cp score	mit score	Loc.	cTP	mTP	SP	Other	Loc.	RC	
COXVa	CoxVa/CAT	0.002	0.986	M	0.036	0.769	0.058	0.191	M	3	M
TPT	TPT	0.165	0.480		0.153	0.293	0.028	0.143		5	
Glutathione reductase	Glutathione reductase	0.725	0.000	C	0.917	0.080	0.007	0.061	C	1	C
Ferrochelatase-I	Ferrochelatase-I	0.157	0.683	M	0.869	0.141	0.023	0.073	C	2	M/C
Methionyl-tRNA synthetase	SYMO_ARATH	0.999	0.010	C	0.840	0.182	0.016	0.056	C	2	C
Histidyl-tRNA synthetase	SYHO_ARATH	0.007	0.002		0.220	0.594	0.061	0.060		4	
Cysteinyl-tRNA synthetase	SYCO_ARATH	0.730	0.672	O	0.804	0.398	0.017	0.016	C	3	O/C
Asparaginyl-tRNA synthetase	SYNO_ARATH	0.976	0.159	C	0.829	0.353	0.004	0.025	C	3	C
RNA polymerase	rpoT2	0.006	0.000		0.720	0.046	0.003	0.234	C	3	
Mercaptopyruvate sulfurtransferase	AtMST1-GFP	0.303	0.681	M	0.711	0.295	0.014	0.018	C	3	M/C
Methionine aminopeptidase (MAP1C)	MAP1C	0.905	0.003	C	0.509	0.024	0.017	0.293		4	C
Methionine aminopeptidase (MAP1D)	MAP1D	0.000	0.034		0.311	0.532	0.010	0.032		4	
Peptide deformylase (PDF1B)	PDF1B_ARATH	0.875	0.150	C	0.739	0.547	0.012	0.002	C	5	C
Peptide deformylase (PDF1B)	PDF1B_LYCES	0.968	0.008	C	0.858	0.379	0.006	0.005	C	3	C
Glycyl-tRNA synthetase	SYGO_ARATH	0.989	0.082	C	0.820	0.222	0.021	0.007	C	3	C
Lysyl-tRNA synthetase	SYKO_ARATH	0.332	0.262		0.487	0.514	0.001	0.011		5	
Tryptophanyl-tRNA synthetase	SYWO_ARATH	0.568	0.643	O	0.854	0.271	0.013	0.007	C	3	O/C
Pseudouridine synthase	PUSH_ARATH	0.218	0.935	M	0.490	0.638	0.018	0.009		6	M
RNA binding protein	BAB03001	0.920	0.069	C	0.477	0.504	0.013	0.055		5	C
Phosphoribosyl aminoimidazole (AIR) synthase	AIR synthase	0.955	0.022	C	0.970	0.030	0.065	0.012	C	1	C
Protoporphyrinogen oxidase-II	ProtoxII-Long	0.624	0.026	C	0.811	0.023	0.090	0.086	C	2	C
Protoporphyrinogen oxidase-II	ProtoxII-Short	0.001	0.193		0.123	0.504	0.048	0.400		5	
THI1	THI1-Long	0.964	0.023	C	0.925	0.067	0.021	0.091	C	1	C
THI1	THI1-Short	0.230	0.000		0.013	0.077	0.378	0.837		3	

TargetP [34] and Predotar (http://www.inra.fr/Internet/Produits/Predotar/) are programs that are designed to predict both mitochondrial and chloroplast targeting sequences. This table shows the performance of these prediction tools on proteins that have been shown experimentally to be dual-targeted. The column Loc. indicates the predicted location. M stands for mitochondrial, C for chloroplast and O for organelle (both mitochondria and chloroplasts). Where no clear prediction was found a blank space is left in the column. For four dual-targeted proteins, Predotar predicts mitochondrial targeting and TargetP chloroplast targeting. It may be possible to use such prediction disagreements as an indicator of potential dual-targeted proteins. The cutoffs used were 0.5 for Predotar (both organelles) and 0.62 (mitochondria) or 0.76 (chloroplast) for TargetP (the latter cutoffs corresponding to 90% specificity [34]).

esting experiment to carry out would be to make point mutations or deletions in one of these ambiguous targeting sequences to see whether it is possible to render it specific to one or other organelle.

4.3. Predicting dual-targeted proteins

Different programs have been designed to predict the presence of mitochondrial or plastid targeting presequences such as Mitoprot [33], TargetP [34] and Predotar (http://www.inra.fr/Internet/Produits/Predotar/). These programs are generally quite good, with more than 80% of proteins correctly predicted. However, these programs fare very badly with dual-targeted proteins (Table 3). Mitoprot does not give separate mitochondrial and plastid scores so cannot be easily used for predicting dual-targeted proteins. TargetP and Predotar do give separate scores, and so can be used for predicting dual-targeted proteins, but in practice the predictions are not reliable. For the two twin-presequences (Protox II and THI1, Table 1), the longer protein is predicted to be targeted to the chloroplast (Table 3) as would be expected [22,23]. Nonetheless, the shorter version of these proteins is not predicted to be a mitochondrial targeting presequence as one could expect from the experimental data [22,23].

For the 19 plant ambiguous targeting sequences studied for this review, TargetP predicts some of them to be either mitochondrial or plastid targeting sequences (12 predicted chloroplastic and only one mitochondrial), but none of them as both. Predotar predicts two dual-targeted proteins correctly (the Arabidopsis CysRS and TrpRS sequences) but fails to recognize a number of others as organellar proteins at all (fails for five presequences). Why are the predictions for dual-targeted proteins so bad? The major reason must be the paucity of known examples which makes it impossible to include dual-targeted proteins explicitly in the data sets used to train the neural networks used to develop TargetP and Predotar. Inspection of the networks used in Predotar and some experimentation using different training sets have revealed that the principal difficulty is interference between the overlapping signals in the ambiguous targeting sequences (I. Small, unpublished). During training of the network to recognize mitochondrial targeting sequences, plastid targeting sequences are used as examples of what mitochondrial targeting sequences do NOT look like. When presented with a dual-targeted sequence, the network recognizes its plastid targeting features and this represses the mitochondrial prediction. Exactly the same thing happens in reverse when predicting plastid targeting. The result is that Predotar predictions are often low when presented with double-targeted sequences. Although TargetP functions rather differently, it too must suffer from similar difficulties. It is impossible therefore at the current time to use these programs to estimate what proportion of proteins are likely to be dual-targeted; more experimental data is needed, coupled with improvements to the algorithms used for predictions.

5. Conclusion

In the meantime, probably the best way to spot potential dual-targeted proteins is to examine the complete *Arabidopsis* genome sequence for examples of single genes whose products are strongly expected to be required in both mitochondria and chloroplasts. Examples of such products include the aminoacyl-tRNA synthetases (aaRSs), absolutely essential for translation in the cytosol, mitochondria and plastids. One would therefore expect three genes for each of the 20 aaRSs, but in fact this is very rarely the case, most aaRSs are only represented by two genes, implying one of the gene products is shared between two compartments. Considering the likely genetic origins of the different aaRS genes and the tRNAs present in the organelles, we think it is highly likely that Arabidopsis AspRS, GluRS, PheRS, ProRS, ArgRS, SerRS and TyrRS are dual-targeted to mitochondria and plastids in addition to CysRS, HisRS, LysRS, MetRS, AsnRS and TrpRS which have already been shown to be dual-targeted. For this family of proteins at least, dual targeting is apparently the norm, not something extraordinary. It remains to be seen just how widespread this phenomenon is, but as scientific interest moves beyond organelle-specific metabolism (respiration and photosynthesis) and on to more basic metabolism and the study of organelle gene expression, we confidently predict that the current small group of dual-targeted proteins will turn out to be just the tip of a large iceberg.

References

[1] S.G. Andersson, A. Zomorodipour, J.O. Andersson, T. Sicheritz-Ponten, U.C. Alsmark, R.M. Podowski, A.K. Naslund, A.S. Eriksson, H.H. Winkler, C.G. Kurland, Nature 396 (1998) 133–140.

[2] T. Kaneko, S. Sato, H. Kotani, A. Tanaka, E. Asamizu, Y. Nakamura, N. Miyajima, M. Hirosawa, M. Sugiura, S. Sasamoto, T. Kimura, T. Hosouchi, A. Matsuno, A. Muraki, N. Nakazaki, K. Naruo, S. Okumura, S. Shimpo, C. Takeuchi, T. Wada, A. Watanabe, M. Yamada, M. Yasuda, S. Tabata, DNA Res. 3 (1996) 109–136.

[3] S. Sato, Y. Nakamura, T. Kaneko, E. Asamizu, S. Tabata, DNA Res. 6 (1999) 283–290.

[4] M. Unseld, J.R. Marienfeld, P. Brandt, A. Brennicke, Nat. Genet. 15 (1997) 57–61.

[5] AGI, Nature 408 (2000) 796-815.

[6] K. Chen, X. Chen, D.J. Schnell, Biochem. Soc. Trans. 28 (2000) 485–491.

[7] K. Keegstra, K. Cline, Plant Cell 11 (1999) 557–570.

[8] C.J. Danpure, Trends Cell Biol. 5 (1995) 230–238.

[9] E. Glaser, S. Sjoling, M. Tanudji, J. Whelan, Plant Mol. Biol. 38 (1998) 311–338.

[10] C.M. Lee, J. Sedman, W. Neupert, R.A. Stuart, J. Biol. Chem. 274 (1999) 20937–20942.

[11] G. von Heijne, J. Steppuhn, R.G. Herrmann, Eur. J. Biochem. 180 (1989) 535–545.

[12] G. von Heijne, K. Nishikawa, FEBS Lett. 278 (1991) 1–3.

[13] A. Caliebe, J. Soill, Plant Mol. Biol. 39 (1999) 641–645.

[14] A. Kouranov, X. Chen, B. Fuks, D.J. Schnell, J. Cell Biol. 143 (1998) 991–1002.

[15] D. Macasev, E. Newbigin, J. Whelan, T. Lithgow, Plant Physiol. 123 (2000) 811–816.

[16] E.C. Hurt, M. Goldschmidt-Clermont, B. Pesold-Hurt, J.D. Rochaix, G. Schatz, J. Biol. Chem. 261 (1986) 11440–11443.

[17] S. Brink, U.I. Flugge, F. Chaumont, M. Boutry, M. Emmermann, U. Schmitz, K. Becker, N. Pfanner, J. Biol. Chem. 269 (1994) 16478–16485.

[18] J. Huang, E. Hack, R.W. Thornburg, A.M. Myers, Plant Cell 2 (1990) 1249–1260.

[19] T. Nakamura, Y. Yamaguchi, H. Sano, Eur. J. Biochem. 267 (2000) 5621–5630.

[20] I. Small, H. Wintz, K. Akashi, H. Mireau, Plant Mol. Biol. 38 (1998) 265–277.

[21] M. de Castro Silva Filho, F. Chaumont, S. Leterme, M. Boutry, Plant Mol. Biol. 30 (1996) 769–780.

[22] N. Watanabe, F.S. Che, M. Iwano, S. Takayama, S. Yoshida, A. Isogai, J. Biol. Chem. 23 (2001) 23.

[23] S.M. Chabregas, D.D. Luche, L.P. Farias, A.F. Ribeiro, M.A. Van Sluys, C.F. Menck, M.C. Silva-Filho, Plant Mol. Biol. 6 (2001) 639–650.

[24] B. Menand, L. Marechal-Drouard, W. Sakamoto, A. Dietrich, H. Wintz, Proc. Natl. Acad. Sci. USA 95 (1998) 11014–11019.

[25] K. Akashi, O. Grandjean, I. Small, FEBS Lett. 431 (1998) 39–44.

[26] N.M. Peeters, A. Chapron, A. Giritch, O. Grandjean, D. Lancelin, T. Lhomme, A. Vivrel, I. Small, J. Mol. Evol. 50 (2000) 413–423.

[27] A.M. Duchene, N. Peeters, A. Dietrich, A. Cosset, I.D. Small, H. Wintz, J. Biol. Chem. 276 (2001) 15275–15283.

[28] B. Hedtke, T. Borner, A. Weihe, EMBO Rep. 1 (2000) 435–440.

[29] C. Giglione, A. Serero, M. Pierre, B. Boisson, T. Meinnel, EMBO J. 19 (2000) 5916–5929.

[30] G. Creissen, H. Reynolds, Y. Xue, P. Mullineaux, Plant J. 8 (1995) 167–175.

[31] K.S. Chow, D.P. Singh, J.M. Roper, A.G. Smith, J. Biol. Chem. 272 (1997) 27565–27571.

[32] P.M. Smith, A.J. Mann, D.E. Goggin, C.A. Atkins, Plant Mol. Biol. 36 (1998) 811–820.

[33] M.G. Claros, P. Vincens, Eur. J. Biochem. 241 (1996) 779–786.

[34] O. Emanuelsson, H. Nielsen, S. Brunak, G. von Heijne, J. Mol. Biol. 300 (2000) 1005–1016.

[35] M.D. Silva-Filho, M.C. Wieers, U.I. Flugge, F. Chaumont, M. Boutry, J. Biol Chem. 272 (1997) 15264–15269.