Chloroplast signal length requirement reflects the outer membrane and TOC complex dimension

Abstract

Background and Purpose: The evolution of an efficient preprotein targeting and translocation system was a central prerequisite for the endosymbiotic integration of α-proteobacteria and cyanobacteria as cellular organelles. Today, it is widely accepted that during evolution most (pre-)proteins destined for these two organelles were equipped with an N-terminal targeting signal for localization. While multiple modes of evolution of these extensions are currently discussed, all evolved signals serve the same function — forming a signal for targeting to the correct organelle and translocation across both membranes. We aimed to generalize the current idea for the length requirement of the N-terminal extension for efficient translocation.

Methods: To explore translocation efficiency in vivo we used protoplasts isolated from different plant sources. We compared the behavior of native and artificial precursor proteins in this plant cell system by fluorescence microscopy.

Results: We demonstrate that the minimal length of the N-terminal amino acid stretch in a loosely folded conformation of a precursor of a chloroplast protein is about 60 amino acid residues. This amino acid stretch in a loosely folded state is prerequisite that a preprotein can traverse the outer membrane in vivo.

Conclusion: We generalize the evidence that two distinct prerequisites framed the evolutionary process of development of targeting signals for chloroplast translocation. (i) The emerging signal had to be sufficiently distinct to signals existing for targeting to other cellular compartments. (ii) The N-terminal signal had to evolve with physico-chemical properties that serve both purposes: targeting and translocation. With respect to the latter, the length of the unfolded polypeptide is defined by the dimension of the translocon and the resulting distance between the cytosolically exposed receptors acting on the cis side of the membrane and the molecular machinery energizing translocation acting in trans – in the intermembrane space.

INTRODUCTION

The functional integrity of cellular compartments depends on massive protein transport through the cytosol and translocation across according membranes (1). Central for such processes are signals either encoded as amino acid extensions or as signatures within the mature part of the precursor proteins (preproteins; 2,3). For example, at least two signals have been described for targeting toward peroxisomes. One signal is positioned at the extreme C-terminus of proteins targeted to
peroxisomes and was annotated as “peroxisomal targeting signal 1” (PTS1; 4). It consists of the tripeptide Ser-Lys-Leu (SKL) or conserved variants (5). A second signal for peroxisomal targeting is formed by an N-terminal polypeptide segment annotated as the “peroxisomal targeting signal 2” (PTS2, 4). This signal is cleaved off by processing peptidases upon translocation. In line, for targeting to mitochondria and chloroplasts a multitude of different signals has evolved (6), but N-terminal polypeptide extensions are by far most dominant.

Although the transit peptides of chloroplast-targeted proteins are discussed to be important for targeting and translocation, a common nature is hard to grasp. Transit peptides are enriched in uncharged residues at the N-terminus of the signal and in hydroxylated amino acids in general (7–9). Many sequence motifs have been discussed, but the only experimentally validated and unchallenged motif is a pair of consecutive, positively charged amino acid residues required for preferential import into mature chloroplasts (10). The length of the cleavable part annotated as transit peptides is in the range of 10 to more than 150 amino acid residues, but more than 50% of the yet identified preproteins have transit peptides shorter than 50 amino acid residues (11,12).

Various scenarios have been proposed with regard to transit peptide evolution (e.g. 13). For a subset of proteins it is assumed that existing signals for secretion have been recycled. Considering that the chloroplast protein translocation pore Toc75 originated from the outer membrane β-barrel protein assembly factor Omp85 (e.g. 14) and the importance of a C-terminal phenylalanine or tryptophane residue for targeting to bacterial Omp85 (e.g. 15), mutations of such signals to include aromatic amino acids might have caused an evolutionary advantage. Indeed, signals for translocation into the primitive plastid of Cyanophora paradoxa still contain a phenylalanine at their N-terminus (16–21). In line with such scenario, the phenylalanine of a precursor protein from C. paradoxa enables the precursor protein to interact with a cyanobacterial Omp85 (22).

For proteins destined for the chloroplast stroma new signals had to evolve. Here, at least two scenarios are discussed: (i) a signal evolved by a process annotated as “exon shuffling” or (ii) the ancestral 5'UTR might have been used or modified to code for a signal (13). The proposal of the exon-shuffling is based on earlier observations that signal sequences of some chloroplast preproteins are encoded by three distinct exons (23–26). It is argued that in the course of evolutionary adaptation an intron loss might have occurred. The alternative scenario on 5'UTR retention was concluded from the analysis of the O-acetylserine (thiol)-lyases (27). However, it was discussed that the evolutionary mode was also dependent on the fitting accuracy of the evolving signal and the translocon (13).

Translocation of precursor proteins across the outer envelope membrane is catalyzed by the translocon of the outer chloroplast membrane (TOC; 28,29). The complex is composed of the two GTP-regulated receptors Toc34 and Toc159 and the pore-forming Toc75. In addition, it was shown that the Hsp90 receptor Toc64 dynamically associates with these proteins (30–32). Currently, at least three proteins are suggested to drive the transfer of the precursor from the TOC complex to the translocon of the inner chloroplast membrane (TIC; 28,29), namely Toc12 attached to the outer envelope in trans, Toc64 (33–35) and the intermembrane space-localized Tic22 (36–38). However, the mode of this transfer and of the subsequent translocation across the IMS is largely unknown.

In here, we analyzed the conservation of the transit peptide length requirement for protein translocation in vivo. This information is central to understand which energizing modules are involved in the process. In the past at least three different elements have been proposed: an energizing module at the cis side of the outer envelope membrane, one at the trans side of the outer envelope membrane and a third one in the stroma (e.g. 28,29). While the energizing module at the cis side is attributed to GTP hydrolysis, the other two events discussed depend on ATPases required for successful translocation (6). In here, we document that the evolutionary pressure for chloroplast signal evolution is presented by the energizing modules at the trans side of the outer envelope membrane, because the minimal length of 60 amino acid residues of a loosely folded N-terminal polypeptide stretch is only able to span the outer membrane.

**SUBJECT AND METHODS**

**Construct generation**

Constructs have been previously generated or created as previously described (12,33) and corresponding amplified cDNA-fragments were ligated into pML94-GFP and pML94-GFP-SKL via Acc65I and BcuI (12) or via EcoRI or Ncol and XhoI into pML94-titin-GFP and pML94-titin-GFP-SKL.

**Plant material, Protoplast isolation and transformation**

Sterile tobacco (Nicotiana tabacum cv. Petit Havana), tomato (Solanum lycopersicum cv. Moneymaker) and Arabidopsis thaliana (ecotype Columbia) plants were grown on soil or gelrite-solidified Murashige and Skoog (MS) medium supplemented with 30 or 20 g/L sucrose, respectively. Tomato, tobacco and A. thaliana leaf mesophyll protoplasts were isolated from 4-6 week-old plants and transformed using the polyethylene glycol-mediated transformation protocol (39,40). Generally, 10⁸ protoplasts were transformed with 20 μg plasmid-DNA, if necessary the final amount of DNA was adjusted by addition of the plasmid pRT-Neo. Cells were incubated in light for 6 h at 26°C in light.
Microscopic analyses

Intracellular localization of GFP fusion proteins in protoplasts was analyzed by fluorescence microscopy using a Leica SP5 confocal laser scanning microscope with a HCX PL APO lambda blue 63x NA 1.4 oil objective as in (12). GFP and chlorophyll fluorescence were excited at 488 nm. The emission of GFP was recorded at 495 to 540 nm and of chlorophyll at 670 to 720 nm.

RESULTS

The efficiency of short transit peptides for chloroplast import in Nicotiana tabacum

In a previous study the translocation competence of different pNTT1 (nucleotide triphosphate transporter 1) fusion proteins upon transient expression in tomato (Solanum lycopersicum cv. Moneymaker) were used to assess the minimal signal length requirement of 60 amino acid residues for chloroplast translocation (12). The constructs were based on the 27th immunoglobulin domain of the human muscle protein titin (named titin hereafter) fused to GFP, previously used to investigate the preprotein translocation mode (12,41,42). Thereby, the titin domain serves as common passenger domain that was shown to be tightly folded already during the targeting process. Unfolding of this passenger domain is prerequisite for translocation and requires pulling force of 200pN at the N-terminus (42,43,44). However, the minimal length of a signal stretch of 60 amino acids for in vivo import of a passenger protein was only documented for tomato so far. Consequently, we asked whether this rule is generally applicable.

At first we expressed the different pNTT1-based titin-GFP fusion constructs (Figure 1a) in tobacco (N. tabacum cv. Petit Havana) protoplasts (Figure 1b). The localization of the preprotein was determined by confocal laser scanning microscopy analyses of GFP fluorescence distribution using the thylakoid derived chlorophyll derived auto-fluorescence as marker of chloroplast positioning. A cytosolic distribution of the GFP signal (Figure 1b, left) was observed upon expression of a construct harboring the 21 amino acid transit peptide of pNTT1 (Figure 1a, blue) extended by the subsequent amino acids of the mature domain (Figure 1a, red) to a final length of 50 residues fused C-terminally to titin (Figure 1a, black) and GFP (Figure 1a, green; tp(50)NTT1-titin-GFP). This confirms that 50 amino acid residues are not sufficient for translocation of the protein into chloroplasts. In contrast, extending the mature section as such that 60 (tp(60)NTT1-titin-GFP) or 70 N-terminal amino acids residues of pNTT1 are present resulted in chloroplast-localized GFP fluorescence (tp(70)NTT1-titin-GFP; Figure 1b, middle and right).

To extend this observation, the first 50 amino acid residues of pRbl11 (precursor of chloroplast-localized rhomboid-like protein; 12,45) were fused to titin-GFP (tp(50)Rbl11-titin-GFP; Figure 2a). The construct was subsequently transformed into and expressed in N. tabacum protoplasts. The expression of this construct yielded a cytosolic GFP distribution as assessed by fluorescence microscopy (Figure 2b, left). In contrast, when the full-length protein is fused to titin-GFP and expressed in N. tabacum protoplasts we observed a chloroplastic GFP signal surrounding the chlorophyll fluorescence (pRbl11-titin-GFP; Figure 2b, right). The rim signal was expected, because pRbl11 has predicted envelope localization (45).

Next, we used the transit peptide of the precursor of the small subunit of RUBISCO (pSSU), which has a
length of 56 amino acid residues. Expression of the fusion protein composed of the transit peptide fused to titin-GFP (Figure 3a) in *N. tabacum* protoplasts resulted in a dual localization of the fusion protein in the cytosol and chloroplasts (Figure 3b, left). This result is consistent with the one observed after expression of this construct in tomato protoplasts (12). In contrast, when the full-length pSSU is fused to titin-GFP or GFP (pSSU-titin-GFP, pSSU-GFP) and expressed in *N. tabacum* protoplasts, the encoded protein is effectively imported into chloroplast in tobacco as judged from the overlay between GFP and auto-fluorescence signal (Figure 3b, middle right).

**The efficiency of truncated transit peptides for chloroplast import in vivo**

So far the conclusion of the minimal length requirement is based on extending short transit peptides (Figures 1, 2; 12,41). Thus, we aimed to confirm that the requirement of a signal of 60 amino acids is also valid when a transit peptide of a larger size is truncated. To this end, we inspected a set of fusion constructs based on the transit peptide of pOE33 (precursor of the oxygen evolving complex subunit of 33 kDa). pOE33 contains a so-called bipartide transit peptide with a stromal signal (amino acid residues 1-30) and a thylakoid signal (amino acid residues 31-79). We fused N-terminal sections of the signal of 55, 60 and 65 amino acids, as well as the full-length transit peptide to titin-GFP (Figure 4a) to generalize the findings for pNTT1. Expression of tp(55)OE33-titin-GFP in tobacco (Figure 4b) or tomato (Figure 4c) protoplasts revealed an exclusive cytosolic distribution of GFP (Figure 4b,c; top panel). Thus, the pOE33-derived 55 amino acid residues are not sufficient to drive protein translocation into chloroplasts. While using tp(60)OE33-titin-GFP we observed a dual distribution of the GFP signal in the cytosol and in chloroplasts as judged from the GFP-fluo-
cence overlay with the auto-fluorescence (Figure 4b,c; second panel). However, comparison between tobacco (Figure 4b) and tomato (Figure 4c) protoplasts revealed that the translocation efficiency in tomato is somewhat higher than in tobacco protoplasts, which might suggest that slight species-specific variations of the transit peptide length requirement exist. Moreover, the dual localization parallels the observation for tpSSU-titin-GFP (Figure 3b), which shows that the length requirement is also slightly dependent on the signal used.

As expected, using 65 amino acid residues (tp(65) OE33-titin-GFP) or the entire transit peptide of pOE33 (tpOE33-titin-GFP) we observed an exclusive localization of the GFP signal in chloroplasts as judged from the overlay with the auto-fluorescence (Figure 4b,c; third and fourth panel). Thus, we conclude that the observed signal length requirement does not dependent on the experimental strategy.

![Diagram](https://example.com/diagram)

**Figure 4.** 60 amino acids of the transit peptide of pOE33 are sufficient for preprotein translocation into chloroplasts in vivo. (a) The constructs used are shown as scheme. The transit peptide is shown in blue, the titin domain in black and the GFP domain in green. (b,c) Intracellular distribution of pOE33 constructs in tobacco (b) and tomato (c) protoplasts. Protoplasts were transformed with pOE33-based titin-GFP fusion constructs indicated on the side and analyzed by CLSM after 6 hours of expression. The GFP-fluorescence, the chlorophyll auto-fluorescence and the overlay between GFP- and chlorophyll auto-fluorescence for representative images are shown.

**The transit peptide length requirement for chloroplast import is valid in A. thaliana**

*A. thaliana* is generally used as a model system. Consequently we extended our study to protoplasts isolated from *A. thaliana*. We particularly approached the dual localized proteins tp(60)OE33-titin-GFP and tpSSU-titin-GFP. To monitor the dual localization we created fusion proteins with C-terminal PTS1 signal (Figure 5a). This tripeptide directs proteins from the cytosol to peroxisomes (5,12). Indeed, for both proteins, tp(60)OE33-titin-GFP-SKL and tpSSU-titin-GFP-SKL, we found a chloroplast localized GFP signal as well as accumulation of GFP fluorescence in specific spots representing peroxisomes (Figure 5b, left and middle panel). On the one hand, the dual localization shows that 60 amino acids are the minimum for sufficient translocation into *A. thaliana* chloroplasts. On the other hand, this shows that the dual localization cannot be competed for by the peroxisomal signal and only proteins accumulating in the cytosol are directed to the latter organelle. In line, all precursor constructs with signals longer than 60 amino acids are efficiently translocated to chloroplasts in *A. thaliana* protoplasts as exemplified for pSSU-GFP-SKL. For this
Figure 5. Arabidopsis thaliana displays similar import characteristics like tomato and tobacco. (a) The constructs used are shown as scheme. (b) Arabidopsis thaliana protoplasts were transformed with fusion constructs indicated on top and analyzed by CLSM. The GFP-, chlorophyll auto-fluorescence and the respective overlays are shown along with the transmission channel for representative images.

construct an exclusive GFP localization in chloroplasts was observed (Figure 5b, right panel), which documents that the N-terminal signal for chloroplast targeting overwrites the information for peroxisomal targeting represented by PTS1.

**DISCUSSION**

**The signal length requirement**

It is accepted that the majority of nuclear encoded chloroplast proteins carry N-terminal transit peptides, which are cleaved off during or after translocation (2,3,6). The signal itself is ill-defined and several reports indicated that elements downstream of the transit peptides can influence the translocation process. For example, import of pFNR (precursor of ferredoxin–NADP+ reductase) is accelerated upon deletion of the first 24 amino acid residues of the mature domain (46), C-terminal truncations of pSSU drastically reduced its import competence (47), while C-terminal deletions of the NADPH/protoclorophyllide oxidoreductase influenced the sorting of the protein inside the chloroplast (48). On the example of the tRNA-NT it was demonstrated that the stability of the mature domain affects the final localization of the dual targeted preprotein as well (49). Further, the transit peptide was found to be essential but not generally sufficient for protein translocation on the examples of pSSU, pFNR, pNTT1, plastocyanin and pOE33 (Figures 1-5; 12,41).

In *in vitro* and *in vivo* studies (Figures 1-5; 12,27,41) document that a length requirement of about 60 amino acid residues exists for translocation into chloroplasts. While small variations might exist between species (e.g. Figure 4) and transit peptides (e.g. Figures 3,4), the concept can be generalized as exemplified for different preproteins and while using protoplasts from different sources (Figures 1-5). Thus, we have to conclude that in the native context a ‘signal’ of sufficient length appears to be present, irrespective of the position of the processing site. It can be assumed that the preceding region is prone to unfolding if not already unfolded in the preprotein context. Based on the titin fusions, which are not associated with Hsp70 type chaperones in *in vitro* (12,41), the results suggest that an initial intraorganellar ‘motor’ provides a ‘pulling’ force of at least 350 pN, because a tightly folded titin mutant was imported as efficiently as wild-type titin used herein (41,43,44).

The rationality of the signal length requirement

The observed length of 60 amino acids raises the question why particularly such length is required for efficient translocation. If one considers that the translocation channel of the chloroplast outer membrane Toc75 belongs to the class of outer membrane proteins annotated as Omp85 (14) the structure of the protein can be approximated from existing structures of the *E. coli* protein of the same family, BamA (50). Moreover, the structures of Toc33/34 (51,52) and of the Hsp70 docking domain of Toc12 have been modeled (35). In addition, the existence of an intermembrane space Hsp70 involved in preprotein translocation has been postulated for a long time (35,53-55) and structural information has been provided for the Hsp70-like DnaK (56,57). The structural information of all described components can be placed into a model to obtain a first estimate on the dimensions during protein translocation (Figure 6). It needs to be mentioned that at stage no structural information exists for Toc159, which contains a 52kDa domain in the intermembrane space (58) and thus, the dimension of this domain cannot be considered at stage. However, if one places an unfolded polypeptide from receptor side across the translocation channel of the model generated, it becomes obvious that 60 amino acids are just enough to reach from receptor to the Hsp70 docking side (Figure 6). Thus, could this be the stromal chaperone as either Hsp70 (59-61) or Hsp90 (62) are discussed as molecular “motor” in the stroma?
Preproteins contain an Hsp70 binding site within the first 20 amino acids (63), which requires the release of an amino acid stretch of this length from the translocon as shown in the model (Figure 6). Additionally, about 20 amino acids are required to traverse the translocation pore present in the membrane (Figure 6). Thus, if one excludes an intermembrane space molecular “motor” for translocation, more than 40 amino acids would be required to span both, outer and inner membrane in case the two translocation machines tightly interact with each other. However, for the latter no experimental evidence exists. Moreover, the intermembrane space localized 52kDa domain of Toc159 (58) and the intermembrane space localized C-terminus of Toc33 (note, the Toc33 structure in the model represents the 28kDa cytosolic domain of the protein; Figure 6) are not shown in the model, because structural information on these domains are not available. The presence of these domains would require additional amino acids of the emerging N-terminus of the preproteins. In addition, the TOC structure observed by electron microscopy (39,64) revealed a dense rim of the TOC complex at the cytosolic side. This high density of domains at the cytosolic side of the TOC complex again argues for the requirement of a certain number of amino acids at the N-terminus of the preproteins before the signal enters the outer membrane pore.

Thus, as exemplified in the model, the observed requirement of 60 amino acids cannot be explained by a direct contact to the energizing machinery in the stroma, but fits very well with the dimensions assumed for the TOC translocon (Figure 6). This conclusion based on structural estimations is supported by experimental evidence, as it was documented that an intermembrane space-localized ATPase is involved in protein transloca-

Figure 6. The structural arrangement of domains involved in preprotein translocation. An illustrative model of the components likely to interact with the transit peptide during translocation across the outer envelope is shown. The central translocation pore Toc75 is represented by the crystal structure of BamA including the most C-terminal POTRA domain (blue, 4C4V, 50). The b-barrel was embedded in a membrane of phosphatidylethanolamine lipids. The G domain of Toc33 facing the C-terminus toward the membrane (cyan; 3BB3; 51) was placed on the cis side of the translocation pore in close vicinity of the POTRA domain as Toc34-POTRA interaction was observed before (67). The substrate binding domain (SBD) of a model of E. coli Hsp70 bound to DnaJ (green, 2KHO, 56,57) was structurally aligned with the crystal structure of an SBD-peptide complex (1DKZ, 68) to include a substrate peptide. In addition, DnaJ was replaced with a homology model of Toc12 (yellow; 35) and positioned with the N-terminus at the trans side of the membrane. The tp(60)NTT1-titin is represented by the N-terminal 60mer peptide in loose conformation (red) fused to titin (orange, 2RQ8; 69). Note, a multiple sequence alignment of a non-redundant set of pNTT1 homolog revealed a conserved putative Hsp70-binding motif in the N-terminal region (NP_178146, res 8-16, RGGLSLPTK), which contains three leucines flanked by basic amino acids. All molecular modeling was done with YASARA (www.yasara.org) and molecular graphics were rendered with Povray (www.povray.org).
tion (65), and the stromal chaperone system does not provide an unfolding capacity (66) that would be required to translocate preproteins with the tightly folded titin passenger (Figures 1–6, 12, 41). Summing up, the identified requirement of 60 amino acids in a loosely folded state for efficient import in vivo can only be explained by the existence of an ATP-consuming molecular motor, most likely presented by the long time postulated intermembrane space Hsp70.

Acknowledgements: A model of E. coli Hsp70 bound to DnaJ was kindly provided by Erik Zuiderweg (University Michigan). This project was funded by the Cluster of Excellence Frankfurt “Macromolecular Complexes” and by the Deutsche Forschungsgemeinschaft in project DFG SCHL 585-3 and in the frame of SFB807/IP17.

REFERENCES

Cytosolic Hsp70s in the IMS of chloroplasts

B. Tillmann et al.


47. BETYLESEN EB., CHANG L., GESTWICKI JE, ZUIDERWEG ER 2009 Solution conformation of wild-type E. coli Hsp70 (DnaK) chaperone complexed with ADP and substrate. Proc Natl Acad Sci USA 106: 8461-8466.


64. SCHLEIFF E, SOLL J, KÜCHLER M, KÜHLBRANDT W, HARRER R. 2003 Characterisation of the translocon of the outer envelope of chloroplasts. *J Cell Biol* 160: 541-551. [http://dx.doi.org/10.1083/jcb.200210060](http://dx.doi.org/10.1083/jcb.200210060)


