Prediction of the plant β -barrel proteome: A case study of the chloroplast outer envelope

ENRICO SCHLEIFF,¹ LUTZ ANDREAS EICHACKER,¹ KERSTIN ECKART,¹
THOMAS BECKER,¹ OLIVER MIRUS,² TANJA STAHL,² AND JÜRGEN SOLL¹

¹Botanisches Institut der Ludwig-Maximilian Universität München, 80368 München, Germany

²Botanisches Institut der Christian-Albrechts Universität Kiel, 24118 Kiel, Germany

(RECEIVED October 18, 2002; FINAL REVISION January 14, 2003; ACCEPTED January 14, 2003)

Abstract

In the postgenomic era, the transformation of genetic information into biochemical meaning is required. We have analyzed the proteome of the chloroplast outer envelope membrane by an *in silico* and a proteomic approach. Based on its evolutionary relation to the outer membrane of Gram-negative bacteria, the outer envelope membrane should contain a large number of β -barrel proteins. We therefore calculated the probability for the existence of β -sheet, β -barrel, and hairpin structures among all proteins of the *Arabidopsis thaliana* genome. According to the existence of these structures, a number of candidates were selected. This protein pool was analyzed by TargetP to discard sequences with signals that would direct the protein to other organelles different from chloroplasts. In addition, the pool was manually controlled for the presence of proteins known to function outside of the chloroplast envelope. The approach developed here can be used to predict the topology of β -barrel proteins. For the proteomic approach, proteins of highly purified outer envelope membranes of chloroplasts from *Pisum sativum* were analyzed by ESI-MS/MS mass spectrometry. In addition to the known components, four new proteins of the outer envelope membranes were identified in this study.

Keywords: Genome analysis; prediction; protein topology

Supplemental material: See www.proteinscience.org.

In chloroplasts, important biochemical pathways like photosynthesis, amino acid, and fatty acid biosynthesis take place. To maintain these crucial functions, transport of metabolites and proteins is facilitated by a variety of integral membrane proteins embedded in the two envelope membranes. The Toc complex of the outer envelope is required for translocation of precursor proteins (Schleiff and Soll 2000), whereas transporter proteins regulate the exchange of metabolites (Bölter and Soll 2001). Most of the proteins destined for the outer envelope differ significantly from

proteins of the inner envelope; namely, they lack a classical transit peptide required for translocation across the envelopes (Schleiff and Klösgen 2001). So far, a transit peptide has been described only for Toc75 homologs (Tranel et al. 1995). Hence, prediction programs searching for transit peptides do not detect the majority of proteins of the outer envelope of chloroplasts. To further analyze the proteome of the outer envelope, we had to consider two types of membrane proteins, those containing hydrophobic α -helices, which can be revealed well by computer algorithms (von Heijne 1996), and those that are embedded in the membrane by monomeric or oligomeric β -barrel structures. The latter are thought to represent the majority of the outer envelope proteins, because chloroplasts are endosymbiotically derived from prokaryotic Gram-negative bacteria (Martin and Herrmann 1998). Evidence exists that the outer envelope emerged from the outer membrane of the encapsulated cyanobacterium (Cavalier-Smith 1987). Therefore, the majority

Reprint requests to: Enrico Schleiff, Department für Biologie I, Botanisches Institut, Ludwig-Maximilian Universität München, Menzinger Str. 67, 80368 München, Germany; e-mail: schleiff@botanik.biologie.uni-muenchen.de; fax: 0049-89-17861-185.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.0237503>.

of the outer envelope proteins are believed to be of bacterial origin, with the exception of those added to the outer membrane during evolution in order to adapt to metabolism in a eukaryotic cell.

The proteins of the outer membrane of Gram-negative bacteria are exclusively of β -barrel structure. This might be for several reasons. First, it is believed that no translocation machinery exists at the inner membrane, which is able to translocate hydrophobic α -helices across the inner membrane. Second, the phospholipid divergence between outer and inner leaflet of the outer membrane might favor the insertion of β -barrels over α -helices, and in addition, the formation of a β -barrel requires less amino acids to traverse a membrane than does an α -helix. Third, work on the insertion of β -barrel toxins revealed that insertion of these β -barrel proteins is not dependent on proteinaceous components, even though proteinaceous components might stimulate the insertion (Escuyer and Collier 1991).

This leads to the idea that β -barrel membrane proteins are evolutionary older. Interestingly, this observation could be confirmed for β -barrel proteins of the outer membrane from *Escherichia coli* (Kleinschmidt and Tamm 1996) and mitochondria (Schleiff et al. 1999), which are also of endosymbiotic origin. Therefore, the origin of the outer envelope from Gram-negative bacteria indicates a mainly β -barrel protein content of this membrane, including solute channels, enzymes, and receptor proteins. Unfortunately, the origin of these proteins cannot be used to identify unrecognized outer envelope proteins, because β -barrel proteins reveal a high divergence among primary sequences, even within one class of proteins. However, we conclude that the identification of β -barrel proteins within one genome should be a good indication for an identification of chloroplast outer envelope proteins in a proteome approach, because this class of proteins was not identified to be present in thylakoid membranes, plasma membranes, or endoplasmatic reticulum-derived membranes (Tamm et al. 2001).

The main structural feature of β -barrel proteins is their composition of an even number of eight to 22 membrane-spanning β -sheets with an antiparallel topology, which are connected by alternating long and short loops, forming so-called β -hairpin structures (Wimley 2002). Already, 10 amino acids are sufficient for these β -sheets to traverse the membrane bilayer. Hydrophobic residues may face the lipid environment, and amino acids with an intermediate polarity may indicate the interior of the channel. In the range of the bilayer interface, these secondary structures expose aromatic amino acid residues to the outer environment (Schulz 2000).

In the postgenomic area, it is essential to transform genetic-derived information into functional and structural protein data (Orengo et al. 1999). Chloroplast proteins of other localization than the outer envelope can be identified by the special hydrophobic features of their amino acids (von

Heijne 1996) or by analysis of their transit sequences (Emanuelsson and von Heijne 2001). However, this is not valid for the β -barrel proteins of the outer membrane of chloroplasts, because such proteins lack a cleavable transit peptide for organella targeting. A prediction of internal targeting sequences, as postulated for those proteins, is still not possible owing to the lack of experimental information (Schleiff and Soll 2000; Chacinska et al. 2002). Several methods were proposed to predict β -barrel proteins (Schirmer and Cowan 1993; von Heijne 1996), and recent progress in crystallization of the β -barrel proteins enabled a detailed statistical analysis (Ulmschneider and Sansom 2001; Wimley 2002). Wimley (2002) developed algorithms to screen the genomic derived sequences of *E. coli* and *Pseudomonas aeruginosa* for β -barrel proteins. Based on these algorithms, we developed an approach to analyze the genomic-derived amino acid sequence of β -barrel proteins from *Arabidopsis thaliana* with putative localization in the outer envelope of chloroplasts. Here, we define and discuss a pool of 891 proteins by combining a computational β -barrel, isoelectric point (pI), and TargetP calculations and manual selection called BITS (Beta barrel, pI, TargetP, manually selected). This pool provides an aid to identify outer envelope proteins after mass spectroscopy or N-terminal sequencing. Furthermore, the possibility to use these algorithms for topology prediction was tested. In the complementary proteomic approach, four new proteins could be identified. In addition to known proteins such as preprotein translocators and solute transporters, an 80-kD protein of the ABC-transporter family, an 80-kD β -glucosidase, a protein of 32 kD with homology with dehydrogenases, and a 37-kD protein of unknown function (OEP37) were identified. We present evidence that OEP37, like OEP24 and Toc75, is an outer envelope β -barrel channel.

Results

Selection of outer envelope proteins by the β -barrel score and β -sheet number

Most of the outer envelope proteins do not contain classical transit peptides and cannot be identified by simple transit peptide prediction. Because we expect most proteins of the outer envelope to contain a β -barrel structure, we analyzed the genome of *A. thaliana* for the presence of β -barrel proteins as outlined in Figure 1. First, genomic-derived sequences were used to calculate the β -barrel score (BBS; Wimley 2002). Most proteins of the *A. thaliana* genome revealed a BBS of less than one (Supplemental Material, Fig. B). This observation is in line with the analysis of the *E. coli* and *P. aeruginosa* genome, which revealed a BBS between zero and one for ~87% of the proteins. Wimley (2002) demonstrated that known β -barrel proteins of these species have a BBS higher than one. However, in *A.*

Figure 1. In silico analysis of the *Arabidopsis thaliana* genome for identification of putative chloroplast proteins. Three approaches (from left to right) were used to identify proteins of chloroplastic location. (Left) Number of genes predicted by Chloro P to present all proteins localized in the inner envelope, thylakoid membranes, and chloroplastic stroma. (Middle) Numbers were taken from Ferro et al. (2002) for α -helical membrane protein prediction and include the inner envelope transporter proteins. (Right) Numbers and selection steps show the algorithms used by Wimley (2002) and our selections described here (see Materials and Methods).

thaliana, only ~2.2% of all sequences revealed a BBS higher than one (Supplemental Material, Fig. B). Hence, by the same criteria, significantly less genes were selected in *A. thaliana* in comparison to the *E. coli* and *P. aeruginosa* genome (Wimley 2002). One explanation could be that eukaryotic cells do contain β -barrel proteins only to a lower extent. However, this seems unlikely because plant cells contain two endosymbiotic-derived organelles, which mainly consist of β -barrel proteins at their cytosolic-exposed membrane. Furthermore, we observed that a BBS cut-off higher than one did not detect the homologs of the solute channel OEP21 (Bölter et al. 1999) and OEP24 (Pohlmeyer et al. 1998), which were predicted by other methods to be β -barrel proteins of the outer envelope of chloroplasts. We therefore decreased the cut-off value to 0.7. Now, 4.5% of all genomic-derived sequences were found in this gene pool selection (Fig. 1), including OEP21 and OEP24, but not the homologs of OEP16, which are predicted to form α -helical channel.

Still, some of the known outer envelope proteins with proposed β -barrel conformation, such as Toc75 (Hinnah et al. 1997; Sveshnikova et al. 2000), were not selected because the BBS selection method results in rejection of proteins containing large soluble regions. We therefore calculated the number of predicted membrane β -sheets within each protein by using the number of peaks above two of the exact β score (EBS). Most of the known channel-forming membrane proteins consist of at least 14 β -sheets (Schirmer and Cowan 1993). For our analysis, we decided to select all protein sequences with >13 predicted β -sheets (Fig. 1). We found that ~10% of the genomic-derived sequences from *A. thaliana* contain at least 14 predicted β -sheets. A combina-

tion of genes selected by either criterion, namely, a BBS cut-off 0.7 or a β -sheet content >13, revealed 3,782 sequences (13%), resulting in a β -barrel protein rate similar to that of *E. coli* and *P. aeruginosa* (Wimley 2002).

This computational selection probably allowed the detection of most membrane-located β -barrel proteins in the plant cell. To specifically select chloroplast envelope sequences, we used the observation that all described envelope membrane solute channels have a basic pI. A cut-off selection of pI > 7 reduced the number of sequences by almost one half (Fig. 1). We further analyzed this pool for the presence of targeting sequences by using TargetP. All sequences with a clearly identified targeting signal directing not to the chloroplast (of class 1 and class 2) were removed from the list (Fig. 1). Proteins with a chloroplast-targeting signal were not removed at this point, although most of the outer envelope proteins are thought to lack any targeting signal. This selection detected 1,566 sequences (41%) of the proteins selected by the BBS and β -sheet content.

To further limit the pool of selected proteins, all remaining sequences were analyzed manually according to their proposed or known function. During this selection, 675 sequences already characterized and localized to other compartments of the plant cell were removed, leaving 891 sequences (3.1%) of all genomic-derived sequences, with a possible localization in the plastid outer envelopes (Fig. 1). This pool of candidates was named the BITS pool.

Topological analysis of proteins identified in the BITS pool

The final aim of computer-supported gene analysis is the identification of structural and functional elements. Functional elements are already summarized in many databases (such as MIPS) independent of the fold of proteins. Structural information is hard to obtain, especially when no homolog proteins are available or in the case of membrane proteins. In the case of membrane proteins, the topology of helical regions is predictable as outlined (Supplemental Material). However, the topology prediction of β -barrel proteins still suffers from applicable methods. We therefore developed a protocol for such a modeling based on the BBS, the alternating hydrophobicity calculation, and the homolog comparison as described in Supplemental Material. To test this protocol, we started with two known proteins for which structural information was available.

First, the gene *At1g45170* from *A. thaliana* encoding the homologous protein of OEP24 from pea was selected from the BITS pool (Table 1). This protein was described as a high-conductance solute channel with a slight selectivity for cations transporting triphosphates, sugar, and charged amino acids (Pohlmeyer et al. 1998). Although no sequence similarity to mitochondrial porins could be identified, the protein was shown to functionally replace mitochondrial

Table 1. Features of proteins identified from the outer envelope membrane of *Pisum sativum*

MW (kD) ^a	Acc. no. ^b	Protein name ^a	Ref. ^d	pI ^e	BBS ^f	TMβ ^g	TMα ^h	Acc. no. ⁱ	Local 10 ^j
158,676	Q9LKR1	TOC159	210, 180, 110, 85, 50	5.83	0.089	20	—	At4g02510	OE
75,057	Q43715	TOC75	210, 180, 80, 63	8.93	0.512	26	—	At3g46740	OE
—	— ^k	TOC75-V	66	7.31	0.665	17	—	At5g19620	OE
64,497	Q9MUK5	TOC64	62	8.98	0.314	—	3	At5g09420	OE
34,157	Q41009	TOC34	35	9.42	0	—	1	At5g05000	OE
				9.11	0	—	1	At1g02280	
23,482	O49929	OEP24	26	9.22	1.1244	12	—	At1g45170	OE
				9.27				At5g42960	
20,444	Q9SM57	OEP21	21	9.49	0.773	5	—	At1g76405	OE
				8.76	0.705	6	—	At1g20810	
15,537	Q41050	OEP16	14	9.16	0.259	—	2	At2g28900	OE
—	—	Putative ABC transporter	80	5.71	0	—	7	At2g01320	oE?
—	—	Putative β-glucosidase	80	8.88	0.095	—	4	At3g06510	oE?
—	—	HP37	38	9.14	1.173	13	—	At2g43950	oE?
—	—	HP32	33	8.57	0	—	—	At4g23430	iE?
110,019	O24293	TIC110	110	5.21		—		At1g06940	IE
43,649	P21727	Phosphate translocator	31, 29	9.70		—	—	At5g46110	IE
52,763	P04717	LSU	50	9.42		—	—	rbcL (plastid gene)	S

^a Molecular weight in kD of the protein in *P. sativum*.

^b Accession number of the protein of *P. sativum*.

^c Protein name of the *P. sativum* homolog.

^d Reference number according to Fig. 3B.

^e Isoelectric point (pI) for the protein in *A. thaliana*.

^f β-Barrel score for the protein in *A. thaliana*.

^g Postulated transmembrane β-segments for the protein in *A. thaliana*.

^h Postulated transmembrane α-helices for the protein in *A. thaliana*.

ⁱ Accession number of the protein of *A. thaliana*.

^j Localization of the protein: OE, previously described to be located in the outer envelope of the chloroplast; oE?, assumption to be located in the outer envelope; IE, previously described to be located in the inner envelope of the chloroplast; iE?, assumption to be located in the inner envelope; S, stroma.

^k Compare Eckart et al. (2002).

voltage-dependent anion channel (VDAC; Röhl et al. 1999). Using our computational tools, we propose a structural model of the protein encoded by the *A. thaliana* gene (Fig. 2). Nine transmembrane sheets were identified by using the BBS and additional three by using the alternating

hydrophobicity (strands 4, 6, and 8). The transmembrane probability and the exact positioning were manually controlled as outlined (Materials and Methods). The model proposed here includes 119 amino acids (56%) in the pore formation. This is in line with the observation by circular

Figure 2. A topology model of the OEP24 homolog of *Arabidopsis thaliana*. The gene product of *At1g45170* represents a β-barrel channel-forming protein, and the proposed topology is shown. The topology was modeled as outlined in Materials and Methods. A black line presents the amino acid sequences connecting the transmembrane segments.

dichroism that ~50% of the secondary structure of OEP24 was described to be in a β -sheet conformation (Pohlmeyer et al. 1998). Furthermore, the presence of only short loops is confirmed by resistance of OEP24 against proteolysis, whereas after membrane solubilization, proteolysis yields very small fragments (Fig. 3A, lanes 1,2).

Second, the gene product of *At5g19620* was identified as a 66-kD homolog of Toc75-like proteins and was annotated Toc75-V (Eckart et al. 2002). Several models were proposed for Toc75 in the past (Hinnah et al. 1997; Sveshnikova et al. 2000). However, these models were constructed from Toc75 and synToc75 only (Sveshnikova et al. 2000). Recent identification of further members of this family (Jackson-Constan and Keegstra 2001; Eckart et al. 2002) and improvement of the prediction methods suggested a reinvestigation of the topology of this protein family. The starting pool to optimize the topology model contained Toc75 of *Pisum sativum*, the *Synechocystis* homolog synToc75 (*slr1227*), and the *A. thaliana* homolog of Toc75 from chromosome III (*At3g46740*), as well as Toc75-V (*At5g19620*). The other homologs were not considered because the size distribution did not allow a direct comparison of the proteins.

By using the described algorithm, 19 transmembrane segments were predicted initially by BBS and alternating hydrophobicity. However, one segment localized earlier between the third and fourth transmembrane sheet had to be rejected after consideration of the proteolytic results ob-

served for the inserted Toc75 from *P. sativum* (Sveshnikova et al. 2000). Therefore, a model of 18 transmembrane sheets is proposed by our protocol (Fig. 3C). The topology model indicated here contains two large cytosolic exposed domains at the N terminus. In contrast, the initial model of Hinnah and coworkers (1997) suggested a cytosolic domain for the C terminus. The identification of the smaller homolog (Supplemental Material, Fig. E) and the new algorithm now clearly shows that the C-terminal part is not exposed to the cytosol but, in contrast, has to be the core of the membrane channel. Within this region, the loops remain short and indicate a dense package of the sheets. Furthermore, the proposed model is in line with the observations established by Sveshnikova and coworkers (2000), in which a protease treatment of right-side-out outer envelope vesicles resulted in degradation of the protein to a size of ~45 kD. A similar observation was made for Toc75-V (Fig. 3A, lane 5). The results from proteolysis of disrupted outer envelope vesicles further confirmed the presence of large regions exposed to the intermembrane space, because the protein becomes degraded on membrane solubilization (Fig. 3A, lane 6).

Identification of outer envelope proteins by mass spectrometry

In order to confirm our predictions and to use the BITS pool, proteins of highly purified outer envelope membranes

Figure 3. The topology of the outer envelope membrane protein Toc75. (A) The distribution of the proteolytic fragments of OEP24, OEP37, and Toc75-V are shown after treatment of right-side-out outer envelope vesicles with thermolysin in the absence (lane -) and in the presence (lane +) of 0.1% Triton X-100. (B) A purified outer envelope fraction OE (corresponding to 400 μ g protein) was separated by SDS-PAGE. The protein spots analyzed by mass spectrometry are labeled by arrowheads and are numbered according to Table 1. (C) The topological model of Toc75 was designed as outlined in the Figure 2 legend.

were analyzed by mass spectroscopy. In total, 16 proteins were identified from 117 peptide sequences by mass spectrometric analysis (Table 1; Fig. 3). According to their mode of identification, the proteins can be placed in two different groups. The first group contains proteins, which were already identified in *P. sativum*, namely, Toc159, Toc75, Toc75-V, Toc64, Toc34, OEP24, OEP21, OEP16, Tic110, the large subunit of the Rubisco (LSU), and a phosphate translocator (Schleiff and Soll 2000; Bölter and Soll 2001). With the exception of the three proteins named, all other proteins are components of the outer envelope. Previous observations explain the contaminations of the outer envelope preparation by exactly those proteins. Tic110 is found to be in association with the Toc machinery, and it was postulated that Tic 110 forms a stable, not precursor-dependent, contact site (Akita et al. 1997; Nielsen et al. 1997). The phosphate translocator was originally identified in the inner envelope (Flügge et al. 1989) and later thought to form the outer envelope translocon (Schnell et al. 1990), documenting that this protein might be tightly associated with the outer membrane. The stromal LSU is the most dominant protein of plastids and contaminates any fraction (Block et al. 1983). In addition to the three proteins described, which were only present at low concentration as determined by immunodecoration of the outer envelope fraction (Fig. 4B, lane 2), no other contaminations were detectable. This indicates that the outer envelope fractions used in this study were of high purity.

The second group contains proteins previously not identified in *P. sativum*. For their identification, the *A. thaliana* database (http://mips.gsf.de/proj/thal/db/search/search_frame.html) was screened with blast searches for proteins homologous to the de novo sequenced peptides. One protein was identified as a homologous ABC transporter protein by four peptide sequences (*At2g01320*; data not shown). ABC proteins are involved in transport processes. In particular, it was shown that *atABC1* is required for transport and correct distribution of protoporphyrin IX, which may act as a light-specific signaling factor involved in coordinating intercompartmental communication between plastids and the nucleus (Moller et al. 2001). We conclude that the identified protein encoded by *At2g01320* and the identified homologous protein in *P. sativum* are located in the outer chloroplast membrane. Interestingly, additional peptides isolated from the same protein band (Fig. 3B, cf. No. 80) revealed the existence of a protein homologous to a putative β glucosidase of *A. thaliana* encoded by *At3g06510*. Previously, it was observed that a β -glucosidase is associated with chloroplast membranes in *Beta vulgaris* (Zakharova and Petrova 2000). Two further proteins of 37 kD (HP37) and 32 kD (HP32) were identified with homology with the gene products of *At2g43950* and *At4g23430*, respectively. In contrast to the two proteins mentioned above, no information about these proteins could be found in the literature. The 32-kD protein

was previously identified while sequencing inner envelope proteins by a similar approach (Ferro et al. 2002). However, because Ferro and coworkers (2002) had to deal with cross-contaminations with outer envelope and other membranes, it remains unclear where this protein might be located. Peptides generated from the protein with the reference No. 23 (Fig. 3, cf. No. 23) did not show any relation to known proteins (Supplemental Material, Fig. A) and therefore could not be identified.

Six of the sequenced proteins could directly be identified as integral β -barrel proteins, namely, Toc75 and Toc75-V, Toc159, OEP24, OEP21, and HP37. For OEP24 and the Toc75 homolog, representative topology models were predicted here (Figs. 2, 3). For OEP21, an eight β -barrel structure was proposed previously (Bölter et al. 1999), and HP37 will be discussed in detail (see below). The protein translocon subunit Toc159, which is believed to have a dual localization (i.e., in the outer envelope and in the cytosol; Hiltbrunner et al. 2001), was not selected in the BITS pool because of the pI value below 7. Closer analysis after identification of the protein within the outer envelope (Table 1) revealed that ~14 of the putative transmembrane segments are located in the C-terminal 52-kD M-domain (data not shown), which was found to be resistant against protease treatment of intact chloroplasts (Chen et al. 2000). This might open the question whether the C-terminal domain forms a β -barrel fold within the intermembrane space or is inserted into the outer envelope by this β -barrel conformation. The latter would be in line with earlier observations that the C-terminal 25 kD of Toc159 inserts into the outer envelope (Muckel and Soll 1996).

The other six proteins—namely, Toc64, Toc34, OEP16, the putative ABC transporter, the putative β -glucosidase, and HP32 (Table 1)—do not belong to the class of β -barrel proteins. In order to investigate the presence and localization of helical transmembrane segments, prediction programs were used as outlined. Only transmembrane helices predicted by at least three different algorithms were selected (Supplemental Material, Table A). Interestingly, for five proteins, helical segments were proposed (Table 1). One member of this class with an apparent molecular weight of 14 kD, was identified as OEP16 (Pohlmeyer et al. 1997), an amino acid-selective channel of the outer envelope. By sequence analysis, we found three homolog sequences within the *A. thaliana* genome, namely, *At2g28900*, *At4g16160*, and *At3g62880*. Furthermore, this protein is closely related to Tim23, Tim22, and Tim17, two translocator proteins of the mitochondrial inner membrane (Rassow et al. 1999). In OEP16, three of the four transmembrane helices present in Tim22 and Tim23 were predicted. In *At2g28900* the identified transmembrane region at amino acid 102–118 represents helix 3 and the region at amino acid 125–142 helix 4. According to Rassow and coworkers (1999), helix 2 would be located between amino acid 70 and 91. In line with

Figure 4. OEP37, a newly identified protein of the outer envelope membrane. (A) The alignment of amino acid sequences of protein homologous in *Arabidopsis thaliana* and *Pisum sativum* (accession nos. At2g43950 and AJ243759) is shown. In the third line, the location of sequenced peptides is presented. The gray bar on the top indicates the position of the transit peptide proposed by ChloroP. (B) The subcellular localization of OEP37 was investigated. Chloroplast proteins (Ch, lane 1), proteins of isolated outer envelope (OE, lane 2), proteins of isolated inner envelope (IE, lane 3), stromal proteins (St, lane 4), and thylakoid proteins (Th, lane 5) were subjected to SDS-PAGE followed by immunodecoration with antisera against the indicated proteins of the outer (Toc75, Toc34, OEP24, OEP16) and the inner envelope (Tic110, Tic55), the stroma (cpn60), and the thylakoids (LHCP). (C) The topological model of OEP37 was created as outlined in the Figure 2 legend.

earlier investigations, the helix 1 present in Tim17 was not identified (Pohlmeyer et al. 1997). Our data show that almost all of the sequenced proteins represent membrane-inserted proteins. Only for At4g23430,

can no membrane segment be proposed with high reliability. This protein was recently also identified while sequencing inner envelope proteins (Ferro et al. 2002) and reveals remarkable homology with dehydrogenases, specifi-

cally to short-chain oxidoreductases. Because Tic110 was the only contamination with inner envelope proteins in our outer envelope membrane preparation, we suggest that according to our computational prediction the 32-kD protein is associated with the membrane surface by protein–protein or protein–membrane interaction, and most likely faces the intermembrane space.

Analysis of HP37

The previously not identified HP37 was investigated in more detail. Peptide sequences derived from a protein migrating at ~37 kD on SDS-PAGE (Table 1, HP37) were used to isolate a full-length cDNA clone from *P. sativum* (accession no. AJ243759; Fig. 4). The calculated molecular weight of the protein from *P. sativum* is 37,361 Da. The protein has a pI of 8.2 and a BBS of 1.42. The homologous protein in *A. thaliana* is encoded by the gene *At2g43950* and has a molecular mass of 39 kD. The proteins from *P. sativum* and *A. thaliana* share 63% identity and 75% homology over their entire length. To further prove the accuracy of our outer envelope purification method, a chloroplast fractionation was carried out. Chloroplasts were fractionated (Fig. 4B, lane 1) into outer (lane 2) and inner envelope (lane 3), stroma (lane 4), and thylakoids (lane 5). Clearly, HP37 is colocalized with the proteins of the outer envelope like Toc75, Toc34, OEP24, and OEP16. We therefore conclude that HP37 is a protein of the outer envelope membrane (Fig. 4B), and we will refer to the protein as OEP37.

As already mentioned, the newly identified OEP37 belongs to the class of β -barrel proteins (Table 1). So far nothing except the protein sequences for the homologs of *P. sativum* and *A. thaliana* is known about the protein. Therefore, topology prediction might already lead to a possible function of this uncharacterized protein. Modeling of the

channel reveals again a 12 β -sheet structure, as before for the OEP24 homolog (Fig. 5). However, in contrast to OEP24, OEP37 seems to contain large soluble domains at both sides of the membrane. The overall distribution of charged amino acids indicates that OEP37 forms a voltage-gated channel. The fragments observed after protease treatment also indicates for this protein that large loops are exposed to the intermembrane space, because the protein becomes accessible to proteolysis after vesicle disruption (Fig. 3A, lanes 3,4).

Discussion

We have presented an algorithm for the prediction of outer envelope β -barrel proteins. A combination of our method with the general prediction of chloroplast proteins by Chloro P and the method described by Ferro and coworkers (2002) provides a basis to cover the proteome of the plastid compartment. However, while analyzing the genome of *Arabidopsis thaliana*, we observed that even known β -barrel proteins such as Toc75 have a BBS smaller than 0.7 (Table 1). This observation indicated that the polypeptide loop regions of β -barrel membrane proteins are longer in eukaryotes than in prokaryotes, because the BBS is strongly dependent on the assumption of a loop with a length between five and 25 amino acids. Hence, although the β -barrel protein content of the outer membrane of mitochondria and chloroplasts originated from the prokaryotic ancestors, the loop and exposed regions may have been altered in length and possibly in function during evolution. These changes might have been necessary, especially for mitochondrial and chloroplast proteins to adopt new functions or higher selectivity than required in the outer membrane of bacteria.

We have further analyzed the selected sequences according to their pI. The majority of the selected proteins have a pI between 7.5 and 10 (Supplemental Material, Fig. C) with

Figure 5. Distribution of the selected proteins according to size and isoelectric point (pI). (A) The pI of each protein was calculated as described (Materials and Methods) and plotted against the sequence length of the gene product (in amino acids). Each open circle represents one protein. The gray curve shows the gaussian distribution of the isoelectric point values, and the black curve represents the gaussian distribution of the amino acid numbers. (B) The β -barrel score (BBS) for Toc75, Toc75V, and a transcription factor (control) is shown as a function of the distance between the two scanned areas (loop length) for the hairpin score (HS) calculation. (C) The BBS for Toc75 and Toc75V is shown as a function of the position of the 350-amino-acid area for which the BBS was calculated. The center of the protein was defined as 0 position.

a maximum of the gaussian distribution at a pI of 8.8 (Fig. 5). This is in line with earlier observations from two-dimensional mapping of the envelope proteins (J. Soll, unpubl.) and results presented for inner envelope proteins (Ferro et al. 2002). Analysis of the sequence length prior to manual selection yielded two clusters with a gaussian distribution ~500 (cluster 1) and 1,200 (cluster 2) amino acids (Supplemental Material, Fig. D). Remarkably, some proteins with up to 3,800 amino acids were also selected. The final sequence pool revealed a distribution, which can be described by a single gaussian distribution (Supplemental Material, Fig. D), with a peak at 500 amino acids and a comparable sigma as the first peak before selection (Fig. 2). We believe that the second cluster represents large soluble proteins with a hydrophobic β -sheet interior, resulting in their misidentification as membrane spanning β -sheets. This is in line with the observation that the second peak represents proteins selected by β -sheet number and not by BBS (data not shown).

Both observations together, namely, the existence of longer loop regions and the insufficiency of the selection by β -sheet number, indicate that a new criterion has to be developed to identify β -barrel proteins in higher plants. Two possibilities exist. In line with the model of OEP24 and OEP37 (Figs. 2, 4), we suggest to increase the loop size acceptance for the hairpin score (HS) calculation. So far, taking the suggestion of Wimley (2002) into account, we are working with a loop gap of five amino acids. To test whether the calculated score depends on this assumption, we increased the loop gap from 5 to 15 amino acids (Fig. 5B). For Toc75 and Toc75-V, we observed a maximum of the BBS at a loop size of 7 for Toc75-V and of 13 for Toc75 (Fig. 5B), strongly indicating that loop regions of eukaryotic proteins are larger than those of prokaryotic (Wimley 2002). In contrast, for a soluble test protein, the transcription factor encoded by *At1g21740* containing a β -sheet structure and a calculated BBS of 0.662 by using the standard algorithm, we observed a constant decrease of the BBS when the loop gap was increased (Fig. 5B). Based on the model created, we observed that certain large regions are not involved in channel formation in Toc75-V (Fig. 3). Therefore, a low BBS is observed because the score is calculated by dividing the HS summation by the number of amino acids of the whole sequence. This hypothesis is supported by the identification of homologous proteins with shorter sequence than that of Toc75 (Jackson-Constan and Keegstra 2001) and Toc75-V (Eckart et al. 2002). The alignment (Supplemental Material, Fig. E) between Toc75-V and *At3g48620* indicates that the majority of the topological elements have to be located at the C terminus of Toc75-V. In line with the model of Toc75-V presented here, the shorter homolog also accounts for a β -barrel channel, because 13 transmembrane segments of Toc75-V are present within the aligned region. Together, this would indicate that the HS should only be calculated for a region accounting for a pore-forming seg-

ment of the polypeptide. We therefore used a window of 350 amino acids, reflecting the average size of many porins, and calculated the highest HS for a protein by selecting the highest score for such a window sliding over the sequence. While doing that, we observed that for Toc75 and Toc75-V, the highest BBS was observed for the C-terminal 350 amino acids (Fig. 5C), as expected from the topological model. However, for Toc75 a second peak appeared in the center of the protein. We therefore suggest that analysis of the BBS distribution allows the selection of the core channel regions as well.

The protocol proposed for topological predictions was tested, and the models constructed are in line with experimental observations. Earlier modeling was based on alternating hydrophobicity prediction (von Heijne 1996) or a combination of hydrophobicity and amino acid localization (Sveshnikova et al. 2000). Our results indicate that a combination of the analysis of alternating hydrophobicity and BBS is required to build more reliable topological models. Positioning of the sheet ends still requires the consideration of the amino acid distribution in the capping regions as outlined. By using such models, observations can be explained, and a hypothesis for channel behavior can be formulated. For example, Hinnah and coworkers postulated that the cation selectivity of Toc75 from *P. sativum* is defined by a high negative charge at the "mouth" of the channel (Hinnah et al. 2002). Because Toc75-V belongs to the same protein family, similar properties are expected. A close look at the model of Toc75-V is in line with this interpretation. When only the first three amino acids of the channel pointing toward the cytosol were analyzed for charge distribution, a ratio of seven negatively charged amino acids over one positively charged amino acid was found. Furthermore, the negative charges are present in sheets 2, 3, 6, 7, 10, 14, and 17 and support the idea of the formation of a charged "mouth-like structure."

In general, the interior of the channel contains 14 negatively charged and 10 positively charged amino acids. Furthermore, four histidines are present, which become charged at pH 5.5. Therefore reduction of the pH from 7 to 5 drastically alters the ratio of negative- and positive-charged amino acids in the channel to a value of 14 : 14. It was observed that cation selectivity was reduced by a factor of two when the pH was changed from 7 to 5 (Hinnah et al. 2002). This is in line with the conclusion from the model, because the histidines are mainly located close to the intermembrane space side not altering the charge of the "mouth" region. A further reduction of the pH to 4.0 would discharge glutamic acids, and the resulting overall ratio would increase to a value between 9 to 14 (negative to positive amino acids). The majority of the aspartic acids in our model are present within the "mouth zone" of the channel, leaving a 5 : 2 ratio of negative to positive charges at the channel entrance. This would explain why the cation selec-

tivity remained at the mouth region, although drastically reduced (Hinnah et al. 2002), possibly owing to the increase in positive charges at the intermembrane space side.

With respect to the OEP24 channel, believed to form a dimer (Pohlmeyer et al. 1998), we made an unexpected observation that was of interest. Four aspartic acids are located at the fourth position of the channel, forming β-sheets from the proposed cytosolic side (TMS 3, 4, 6, 9). This highly negatively charged ring is complemented by a glutamic acid at position 5 in TMS 11. In addition, the proposed topological model shows a 10 : 9 ratio of negative- to positive-charged amino acids, again explaining the weak cation selectivity of the channel. Furthermore, the comparison P(K+)/P(Cl-) ratio observed for OEP24 at physiological conditions (3.7; Pohlmeyer et al. 1998) lies between the ratio observed for Toc75 at pH 5 and pH 4 (6.3 and 1.9, respectively). Therefore, the model is able to explain the low ion selectivity of the channel and opens the question of the function of the acidic ring at position four.

In contrast to OEP24, the model of OEP37 shows large soluble domains, which might even penetrate the channel. The highly charged third loop with the proposed localization at the intermembrane space especially indicates an influence on the channel behavior, similar to the loop found in porins (for example, see Weiss and Schulz 1992). If so, we would expect a rapid gating behavior as the result of the high negative net charge. Furthermore, we observed a ratio of 9 : 15 (negative to positive amino acids) within the channel region of OEP37, indicating an anionic behavior of this channel on the first view. However, the large negatively charged loop region accounts for a certain cation-selectivity as well. Interestingly, within a “mouth” region the charges are balanced at 4 : 4, whereas the core domain shows a clear positive net charge and might function as the filter of the channel. The proteomic approach, together with the topology modeling, indicates that we are able to present a new member of the family of ion selective channels of the outer envelope membrane.

The analysis of the proteins selected in our BITS pool revealed that 38% of all selected proteins are computer-predicted hypothetical proteins, and 47% are proteins with identified cDNA but with unknown or predicted function. Among the known proteins are nitrate, H⁺, UDP-galactose, sulfate, sugar, potassium, peptide, and other transporter proteins. Interestingly, our pool contains for example HP44, HP59, HPSOT, and HPTLC, identified by the earlier screen of Ferro and coworkers (2002). The localization of these proteins is not certain yet; although a chloroplast-targeting signal was identified. HPSOT and HPTLC were only selected by their BBN of 14 and 15, close to the threshold, and may therefore still be helical channels as suggested by Ferro and coworkers (2002). However, HP44 and HP59 have a BBS of 1.17 and 1.37, respectively, indicating a β-barrel conformation and therefore an outer envelope localization.

In general, the procedure described here can be used to analyze other plant genomes for the presence of the β-barrel proteins. Furthermore, even though the methodology was developed to identify putative chloroplast β-barrel proteins, proteins with other localizations can be predicted by the methods presented here. The selection using Target P can be modified to selectively identify putative mitochondrial or secreted β-barrel proteins. Also, further information about general properties such as the pI or the size distribution of proteins from the proteome of the target membrane increases the reliability of the prediction and can be included in the algorithms to select candidates with a specific cellular localization, as performed with the pI selection for outer envelope proteins.

Materials and methods

Protein extraction from outer envelopes

Outer envelope membranes were isolated as described (Seedorf et al. 1995). Copurified stromal, thylakoid, inner envelope, and outer envelope fractions were used for the investigation of subcellular localization and topology analysis. All procedures for gel electrophoresis, in-gel digestion, and preparation for the following mass spectrometric analysis were carried out as described before (Eckart et al. 2002).

cDNA cloning, overexpression, and antisera production

The peptide sequences of the 37-kD protein were used to create a random primed digoxigenin-labeled probe. The subsequent screening, cloning, heterologous expression, and antisera production was performed as previously described (Pohlmeyer et al. 1997, 1998).

Statistical analysis of the genomic-derived sequences

The geneomic-derived sequences were recovered from the FTP server from The Institute for Genomic Research and were analyzed by using ChloroP and TargetP as described in Supplemental Material. For the calculation of the EBS, we followed the statistical analysis described by Wimley (2002). The values for the probability of amino acids to face the lipid bilayer (A_{ex}) or the interior of the channel (A_{in}) were further specified for the lipid core (c, defined as the phase 0.65 nm from the membrane center) and the head-group region (h, defined as the phase between 0.65 and 1.35 nm from the membrane center; Wimley 2002). The score was then calculated by using a window of 10 amino acids, which statistically most likely are able to traverse the membrane (Wimley 2002). Therefore, four amino acids are located in the core region and six in the interphase or head-group region. The score was calculated for the case that the first amino acid faces the bilayer (B₁; E₁) and that the first amino acid faces the interior of the channel (B₂; E₂). The higher of both values was then defined as the global β-strand score of this amino acid.

$$B_{k,1} = \sum_{i=0}^1 (\ln(A_{ex,c})_{k+2i+4} + \ln(A_{in,c})_{k+2i+3} + \ln(A_{ex,h})_{k+2i} + \ln(A_{in,h})_{k+2i+7} + \ln(A_{ex,h})_{k+8} + \ln(A_{in,h})_{k+1}) \quad (E1)$$

$$B_{k,2} = \sum_{i=0}^1 (\ln(A_{in,c})_{k+2i+4} + \ln(A_{ex,c})_{k+2i+3} + \ln(A_{in,h})_{k+2i} + \ln(A_{ex,h})_{k+2i+7}) + \ln(A_{in,h})_{k+8} + \ln(A_{ex,h})_{k+1} \quad (E2)$$

The EBS representing the number of membrane spanning regions was calculated by counting of all separate regions with an EBS above two. The HS was calculated within a window of 25 amino acids by selecting the highest EBS in a window of 10 amino acids, starting at amino acid one and starting at amino acid 15 (Wimley 2002). The HS was then defined as the summation of both values. All HSs above a cut-off of six were added and divided by the number of amino acids of the sequence, giving the β -barrel value. The pI value for further selection was then numerically calculated by using the following values: pKai Arg = 12.48; Lys = 10.53; His = 6.00; N-term = 9.69, pKaj Asp = 3.86; Glu = 4.25; Cys = 8.33; Tyr = 10.07; C-term = 2.20, and the following equation:

$$\text{Net charge} = \sum_{i=1}^4 N_i * \{1 - 10^{-pK_{ai}} / (10^{-pH} + 10^{-pK_{ai}})\} - \sum_{j=1}^5 N_j * \{10^{-pK_{aj}} / (10^{-pH} + 10^{-pK_{aj}})\} \quad (E3)$$

where N represents the number of similar amino acids. The pH value for the absolute net charge value smaller than 0.01 was set as isoelectric point of the protein.

Topology modeling

β -Sheet analysis was performed as outlined above. Subsequent identification of transmembrane helices for nonselected proteins was performed by using the DAS (Cserzo et al. 1997), HMMTOP (Tusnady and Simon 2001), SOSUI (<http://sosui.proteome.bio.tuat.ac.jp/sosui/frame0.html>), TMHMM (Krogh et al. 2001), TMAPred (Hofmann and Stoffel 1993), and TopPred (von Heijne 1992). The results are given in Supplemental Material (Table 1).

The localization of the β -sheets was determined according to the regions with an EBS above two. Furthermore, the alternating hydrophobicity specific for membrane β -sheets (von Heijne 1996) was calculated by using a 10-amino-acid window typical for the size of a transmembrane sheet (Wimley 2002) for the EGS scale (Engelman et al. 1986), the POPC scale representing the lipid interface (White and Wimley 1998), and the octanol scale representing the lipid core (White and Wimley 1998). All three hydrophobicity results and the BBS were compared to build an initial model (step 1). Then the secondary structure was predicted. It was demonstrated earlier that the proposed loop regions are often found to be surface exposed in the channel formation when secondary structure prediction programs for cytosolic proteins are applied to membrane proteins. Here, secondary structure prediction was performed by using the profile-based neural network method PHD (<http://dodo.cpmc.columbia.edu/prediction/>; Rost and Sander 1993). The structure of exposed regions was modeled by using secondary structure profiles and accessibility data from PHD. The initial topology model was then corrected (step 2). The same procedure was performed for close homologs, and the structural models will be aligned according to the amino acid alignment observed with Clustal W (Jeanmougin et al. 1998). Only transmembrane segments existing in all related proteins can be considered, because alterations of the amino acid composition are frequent, whereas alteration in transmembrane segment number is rare among one

class (step 3). Then, the position of the transmembrane β -sheet was corrected on the base of the hydrophobicity, the β -barrel probability values (Wimley 2002), and the positioning of the aromatic amino acids because they are often found to cap the transmembrane region (step 4; Sveshnikova et al. 2000). Last, the localization of the loops in the cytosol or in the intermembrane space was postulated by the positive inside rule (step 5; von Heijne 1996).

Electronic supplemental material

The supplemental material contains a detailed description of the program units used for determination of the BITS pool. Additional figures of the statistical analysis of the pool are provided. In addition, the sequences of an unidentified protein are given.

Acknowledgments

We would like to thank Dr. R. Grimm of Hewlett Packard for peptide sequences of OEP37. This work was supported by grants from the Deutsche Forschungsgemeinschaft to J.S. and to L.A.E.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Akita, M., Nielsen, E., and Keegstra, K. 1997. Identification of protein transport complexes in the chloroplastic envelope membranes via chemical cross-linking. *J. Cell Biol.* **136**: 983–994.
- Block, M.A., Dorne, A.J., Joyard, J., and Douce, R. 1983. Preparation and characterization of membrane fractions enriched in outer and inner envelope membranes from spinach chloroplasts. I: Electrophoretic and immunochemical analyses. *J. Biol. Chem.* **258**: 13273–13280.
- Bölter, B. and Soll, J. 2001. Ion channels in the outer membranes of chloroplasts and mitochondria: Open doors or regulated gates? *EMBO J.* **20**: 935–940.
- Bölter, B., Soll, J., Hill, K., Hemmler, R., and Wagner, R. 1999. A rectifying ATP-regulated solute channel in the chloroplastic outer envelope from pea. *EMBO J.* **18**: 5505–5516.
- Cavalier-Smith, T. 1987. The simultaneous symbiotic origin of mitochondria, chloroplasts, and microbodies. *Ann. N.Y. Acad. Sci.* **503**: 55–71.
- Chacinska, A., Pfanner, N., and Meisinger, C. 2002. How mitochondria import hydrophilic and hydrophobic proteins. *Trends Cell Biol.* **12**: 299–303.
- Chen, K., Chen, X., and Schnell, D.J. 2000. Initial binding of preproteins involving the Toc159 receptor can be bypassed during protein import into chloroplasts. *Plant Physiol.* **122**: 813–822.
- Cserzo, M., Wallin, E., Simon, I., von Heijne, G., and Eklöfsson, A. 1997. Prediction of transmembrane α -helices in prokaryotic membrane proteins: The dense alignment surface method. *Protein Eng.* **10**: 673–676.
- Eckart, K., Eichacker, L., Sohr, K., Schleiff, E., Heins, L., and Soll, J. 2002. A Toc75-like protein import channel is abundant in chloroplasts. *EMBO Rep.* **3**: 557–562.
- Emanuelsson, O. and von Heijne, G. 2001. Prediction of organellar targeting signals. *Biochim. Biophys. Acta* **1541**: 114–119.
- Engelman, D.M., Steitz, T.A., and Goldman, A. 1986. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu. Rev. Biophys. Chem.* **15**: 321–353.
- Escuyer, V. and Collier, R.J. 1991. Anthrax protective antigen interacts with a specific receptor on the surface of CHO-K1 cells. *Infect Immun.* **59**: 3381–3386.
- Ferro, M., Salvi, D., Riviere-Rolland, H., Vermet, T., Seigneurin-Berny, D., Grunwald, D., Garin, J., Joyard, J., and Rolland, N. 2002. Integral membrane proteins of the chloroplast envelope: Identification and subcellular localization of new transporters. *Proc. Natl. Acad. Sci.* **99**: 11487–11492.
- Flügge, U.I., Fischer, K., Gross, A., Sebald, W., Lottspeich, F., and Eckerskorn, C. 1989. The triose phosphate-3-phosphoglycerate-phosphate translocator from spinach chloroplasts: Nucleotide sequence of a full-length cDNA

- clone and import of the in vitro synthesized precursor protein into chloroplasts. *EMBO J.* **8**: 39–46.
- Hiltbrunner, A., Bauer, J., Vidi, P.A., Infanger, S., Weibel, P., Hohwy, M., and Kessler, F. 2001. Targeting of an abundant cytosolic form of the protein import receptor at Toc159 to the outer chloroplast membrane. *J. Cell Biol.* **154**: 309–316.
- Hinnah, S.C., Hill, K., Wagner, R., Schlicher, T., and Soll, J. 1997. Reconstitution of a chloroplast protein import channel. *EMBO J.* **16**: 7351–7360.
- Hinnah, S.C., Wagner, R., Sveshnikova, N., Harrer, R., and Soll, J. 2002. The chloroplast protein import channel toc75: Pore properties and interaction with transit peptides. *Biophys. J.* **83**: 899–911.
- Hofmann, K. and Stoffel, W. 1993. TMbase: A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler* **374**: 166.
- Jackson-Constan, D. and Keegstra, K. 2001. Arabidopsis genes encoding components of the chloroplastic protein import apparatus. *Plant Physiol.* **125**: 1567–1576.
- Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G., and Gibson, T.J. 1998. Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* **23**: 403–405.
- Kleinschmidt, J.H. and Tamm, L.K. 1996. Folding intermediates of a β -barrel membrane protein: Kinetic evidence for a multi-step membrane insertion mechanism. *Biochemistry* **35**: 12993–13000.
- Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305**: 567–580.
- Martin, W. and Herrmann, R.G. 1998. Gene transfer from organelles to the nucleus: How much, what happens, and why? *Plant Physiol.* **118**: 9–17.
- Moller, S.G., Kunkel, T., and Chua, N.H. 2001. A plastidic ABC protein involved in intercompartmental communication of light signaling. *Genes & Dev.* **15**: 90–103.
- Muckel, E. and Soll, J. 1996. A protein import receptor of chloroplasts is inserted into the outer envelope membrane by a novel pathway. *J. Biol. Chem.* **271**: 23846–23852.
- Nielsen, E., Akita, M., Davila-Aponte, J., and Keegstra, K. 1997. Stable association of chloroplastic precursors with protein translocation complexes that contain proteins from both envelope membranes and a stromal Hsp100 molecular chaperone. *EMBO J.* **16**: 935–946.
- Orengo, C.A., Todd, A.E., and Thornton, J.M. 1999. From protein structure to function. *Curr. Opin. Struct. Biol.* **9**: 374–382.
- Pohlmeier, K., Soll, J., Steinkamp, T., Hinnah, S., and Wagner, R. 1997. Isolation and characterization of an amino acid-selective channel protein present in the chloroplastic outer envelope membrane. *Proc. Natl. Acad. Sci.* **94**: 9504–9509.
- Pohlmeier, K., Soll, J., Grimm, R., Hill, K., and Wagner, R. 1998. A high-conductance solute channel in the chloroplastic outer envelope from pea. *Plant Cell* **10**: 1207–1216.
- Rassow, J., Dekker, P.J., van Wilpe, S., Meijer, M., and Soll, J. 1999. The preprotein translocase of the mitochondrial inner membrane: Function and evolution. *J. Mol. Biol.* **286**: 105–120.
- Röhl, T., Motzkus, M., and Soll, J. 1989. The outer envelope protein OEP24 from pea chloroplasts can functionally replace the mitochondrial VDAC in yeast. *FEBS Lett.* **460**: 491–494.
- Rost, B. and Sander, C. 1993. Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.* **232**: 584–599.
- Schirmer, T. and Cowan, S.W. 1993. Prediction of membrane-spanning β -strands and its application to maltoporin. *Protein Sci.* **2**: 1361–1363.
- Schleiff, E. and Klösken, R.B. 2001. Without a little help from “my” friends: Direct insertion of proteins into chloroplast membranes? *Biochim. Biophys. Acta* **1541**: 22–33.
- Schleiff, E. and Soll, J. 2000. Travelling of proteins through membranes: Translocation into chloroplasts. *Planta* **211**: 449–456.
- Schleiff, E., Silviu, J.R., and Shore, G.C. 1999. Direct membrane insertion of voltage-dependent anion-selective channel protein catalyzed by mitochondrial Tom20. *J. Cell Biol.* **145**: 973–978.
- Schnell, D.J., Blobel, G., and Pain, D. 1990. The chloroplast import receptor is an integral membrane protein of chloroplast envelope contact sites. *J. Cell. Biol.* **111**: 1825–1838.
- Schulz, G.E. 2000. β -Barrel membrane proteins. *Curr. Opin. Struct. Biol.* **10**: 443–447.
- Seedorf, M., Waegemann, K., and Soll, J. 1995. A constituent of the chloroplast import complex represents a new type of GTP-binding protein. *Plant J.* **7**: 401–411.
- Sveshnikova, N., Grimm, R., Soll, J., and Schleiff, E. 2000. Topology studies of the chloroplast protein import channel Toc75. *Biol. Chem.* **381**: 687–693.
- Tamm, L.K., Arora, A., and Kleinschmidt, J.H. 2001. Structure and assembly of β -barrel membrane proteins. *J. Biol. Chem.* **276**: 32399–32402.
- Tranel, P.J., Froehlich, J., Goyal, A., and Keegstra, K. 1995. A component of the chloroplastic protein import apparatus is targeted to the outer envelope membrane via a novel pathway. *EMBO J.* **14**: 2436–2446.
- Tusnady, G.E. and Simon, I. 2001. The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**: 849–850.
- Ulmschneider, M.B. and Sansom, M.S. 2001. Amino acid distributions in integral membrane protein structures. *Biochim. Biophys. Acta* **1512**: 1–14.
- von Heijne, G. 1992. Membrane protein structure prediction: Hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.* **225**: 487–494.
- . 1996. Prediction of transmembrane protein topology. In *Protein structure prediction: A practical approach*, 1st ed. (ed. M.J.E. Sternberg), pp 101–110. IRL Press, New York.
- Weiss, M.S. and Schulz, G.E. 1992. Structure of porin refined at 1.8 Å resolution. *J. Mol. Biol.* **227**: 493–509.
- White, S.H. and Wimley, W.C. 1998. Hydrophobic interactions of peptides with membrane interfaces. *Biochim. Biophys. Acta* **1376**: 339–352.
- Wimley, W.C. 2002. Toward genomic identification of β -barrel membrane proteins: Composition and architecture of known structures. *Protein Sci.* **11**: 301–312.
- Zakharova, N.S. and Petrova, T.A. 2000. β -Glucosidase of leaves and roots of the common beet *Beta vulgaris*. *Prikl. Biokhim. Mikrobiol.* **36**: 458–461.