

Omp85 from the Thermophilic Cyanobacterium *Thermosynechococcus elongatus* Differs from Proteobacterial Omp85 in Structure and Domain Composition^{*[5]}

Received for publication, February 9, 2010, and in revised form, March 26, 2010. Published, JBC Papers in Press, March 29, 2010, DOI 10.1074/jbc.M110.112516

Thomas Arnold, Kornelius Zeth¹, and Dirk Linke²

From Department I, Protein Evolution, Max Planck Institute for Developmental Biology, Spemannstrasse 35, 72076 Tübingen, Germany

Omp85 proteins are essential proteins located in the bacterial outer membrane. They are involved in outer membrane biogenesis and assist outer membrane protein insertion and folding by an unknown mechanism. Homologous proteins exist in eukaryotes, where they mediate outer membrane assembly in organelles of endosymbiotic origin, the mitochondria and chloroplasts. We set out to explore the homologous relationship between cyanobacteria and chloroplasts, studying the Omp85 protein from the thermophilic cyanobacterium *Thermosynechococcus elongatus*. Using state-of-the-art sequence analysis and clustering methods, we show how this protein is more closely related to its chloroplast homologue Toc75 than to proteobacterial Omp85, a finding supported by single channel conductance measurements. We have solved the structure of the periplasmic part of the protein to 1.97 Å resolution, and we demonstrate that in contrast to Omp85 from *Escherichia coli* the protein has only three, not five, polypeptide transport-associated (POTRA) domains, which recognize substrates and generally interact with other proteins in bigger complexes. We model how these POTRA domains are attached to the outer membrane, based on the relationship of Omp85 to two-partner secretion system proteins, which we show and analyze. Finally, we discuss how Omp85 proteins with different numbers of POTRA domains evolved, and evolve to this day, to accomplish an increasing number of interactions with substrates and helper proteins.

Omp85 proteins are essential, highly conserved proteins in the outer membrane of Gram-negative bacteria. They mediate the biogenesis of almost all β -barrel proteins in the bacterial outer membrane (1). The substrate proteins are recognized by a species-specific motif at the C terminus, typically ending with an aromatic residue (2). The Omp85 proteins of proteobacteria,

also referred to as BamA, form a protein complex with three to four lipoproteins (BamB–E) designated the Bam complex in *Escherichia coli* (3, 4). Homologues of Omp85 also exist in the outer membrane of chloroplasts (translocon at the outer envelope membrane of chloroplasts (Toc75)) and mitochondria (sorting and assembly machinery component of 50 kDa (Sam50) or, alternatively, topogenesis of mitochondrial outer membrane β -barrel proteins (Tob55)), where they are part of multicomponent membrane protein complexes (5–10). This homology and the conserved function of the complexes in prokaryotes and in eukaryotic organelles are major molecular evidence for the endosymbiont theory, which states that chloroplasts arose from cyanobacteria, and mitochondria from primitive α -proteobacteria, which were incorporated into the first simple eukaryotic cells (11–14). Although Sam50 is the only member of the Omp85 family in mitochondria involved in the assembly and insertion of outer membrane proteins, there exist at least two distinct proteins in the outer envelope of chloroplasts, Toc75-III and Toc75-V (Oep80 in *Arabidopsis*), and several isoforms thereof (15, 16). Toc75-III, embedded in the Toc-complex, functions as the translocation pore for chloroplast proteins that are encoded on the nuclear DNA. In mitochondria, this task is accomplished by the TOM complex, whereas in bacteria the outer membrane precursors are translocated through the cytoplasmic (inner) membrane by the Sec machinery. Toc75-V/Oep80 was shown to be essential in *Arabidopsis*; however, its role in the assembly of chloroplast outer envelope proteins still needs to be elucidated (9, 16).

Within bacteria, additional Omp85-like proteins exist, which fulfill a more specialized transport function, e.g. the type Vb secretion systems, comprising FhaC of *Bordetella pertussis* or ShlB of *Serratia marcescens*. Their genes are often found in operons with their specific transport substrates, giving rise to the alternative name of two partner secretion systems (TPSS). The transported proteins are typically adhesion or hemolysin proteins, which are either bound to the transporter on the bacterial cell surface or released into the medium (17–19).

Omp85 proteins, including the TPSS proteins, share a common domain composition as follows: a β -barrel pore at the C terminus and a variable number of polypeptide transport-associated (POTRA)³ domains at the N terminus (17, 20). POTRA

* This work was supported by institutional funding from the Max Planck Society, German Science Foundation Grants SFB766/B4 (to D. L.) and ZE522/2-3 (to K. Z.), and the Landesstiftung Baden-Württemberg, Functional Nanostructures (to D. L.).

Author's Choice—Final version full access.

[5] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental material S1–S5 and additional references.

The atomic coordinates and structure factors (code 2x8x) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

¹ To whom correspondence may be addressed. Tel.: 49-7071-601323; Fax: 49-7071-601349; E-mail: kornelius.zeth@tuebingen.mpg.de.

² To whom correspondence may be addressed. Tel.: 49-7071-601357; Fax: 49-7071-601349; E-mail: dirk.linke@tuebingen.mpg.de.

³ The abbreviations used are: POTRA, polypeptide transport-associated; OMP, outer membrane protein; r.m.s.d., root mean square deviation; PSI, Paul Scherrer Institut; SeMet, selenomethionine; S, siemens.

Thermosynechococcus Omp85 Structure and Function

domains show a $\beta\alpha\alpha\beta\beta$ topology; they are localized in the bacterial periplasm, and presumably recruit substrate proteins, coordinate the complex partners, and/or mediate homo-oligomerization (21–24).

Structural data exist of full-length FhaC from *B. pertussis*, which contains two POTRA domains (17), and of the BamA POTRA domains 1–4 (of 5) from *E. coli* (25), both obtained by x-ray crystallography. No detailed structural information on the eukaryotic members of the Omp85 family, namely Sam50/Tob55 from mitochondria and Toc75 from chloroplasts or of the closely related cyanobacterial Omp85, is available to date.

In cyanobacteria, neither a Toc-like nor a Bam-like complex has been identified. Moreover, cyanobacterial Omp85 proteins, as Toc75, include only three POTRA domains, not five as in proteobacteria. These findings prompted us to further study this protein, as it seems to work independently from other complex components and has a different and apparently simpler domain composition compared with proteobacterial Omp85.

Here, we report the first three-dimensional structure of all three POTRA domains from Omp85 of the thermophilic cyanobacterium *Thermosynechococcus elongatus* (*TeOmp85*). We demonstrate that the electrophysiological characteristics of *TeOmp85* in the presence or absence of the POTRA domains correspond well to those of other Omp85 family members. Furthermore, we identified, annotated, and classified 567 POTRA domain-bearing proteins from all kingdoms of life, and we show that even though they strongly differ in their number of POTRA domains, their homologous relationships across species can be studied using state-of-the-art homology detection tools. We find that the most C-terminal POTRA domain is best conserved between species, followed by the most N-terminal POTRA domain, independent of how many POTRA domains are located in-between. From this observation, we draw conclusions on how more complex Omp85 proteins with up to seven POTRA domains may have evolved from relatively simple ancestor proteins and that these ancestor proteins must have been already present when the first endosymbiotic events occurred.

EXPERIMENTAL PROCEDURES

Molecular Biology—All constructs were obtained by PCR amplification using genomic DNA isolated from *T. elongatus* BP1 cells. The primers used are given in [supplemental material S1](#). The PCR products of *TeOmp85* and *TeOmp85-C* were digested with BsaI and ligated in the vector pASK-IBA33⁺ (C-terminal His₆ tag; IBA, Goettingen, Germany). The PCR product of *TeOmp85-N* was digested with BamHI and NdeI and ligated in the vector pET15b (N-terminal His₆ tag; EMD Biosciences, Madison, WI).

Protein Production—*E. coli* BL21 (DE3) omp8 cells (26) were transformed with the given plasmids. The cells were grown in Luria Bertani medium in shaking flasks at 37 °C, and production of the proteins was induced at an A_{578} of 0.6 by adding 0.2 $\mu\text{g}/\text{ml}$ anhydrotetracycline (pASK-IBA33⁺ vectors) or 1 mM isopropyl 1-thio- β -D-galactopyranoside (pET15b vector), respectively. After 4 h, the cells were pelleted in a centrifuge.

Production of L-Selenomethionine-labeled TeOmp85-N—*E. coli* BL21 (DE3) C41 cells (27) were transformed with the plasmid pTeOmp85-N. The cells were grown in Luria Bertani

medium to an A_{578} of 0.6, harvested using a centrifuge, washed with M9 minimal medium, and resuspended in SeMet/M9 medium for selenomethionine labeling (28). All other steps were conducted as described above.

Purification of TeOmp85 and TeOmp85-C—The cells were resuspended in 30 ml of resuspension buffer (150 mM NaCl, 10 mM MgCl₂, 10 mM MnCl₂, 20 mM Tris-HCl, pH 8.5, and a pinch of DNase I) and lysed using a French press. Because of the overall hydrophobic character of the proteins, they precipitated as inclusion bodies after translation. The inclusion bodies were harvested by spinning the cell lysate at $4000 \times g$ at 4 °C for 30 min. The pelleted inclusion bodies were resuspended in 150 mM NaCl, 20 mM Tris-HCl, pH 8.5. Lauryldimethylamine *N*-oxide was added to a final concentration of 1% (w/v), and the suspension was incubated on a shaker for 1 h at room temperature. The inclusion bodies were subsequently pelleted at $4000 \times g$, 4 °C for 30 min and washed four times with water. The pellet was finally resuspended in three times its volume water and stored at –20 °C. For fast protein liquid chromatography purification, an aliquot of inclusion body suspension was solubilized in 6 M guanidinium HCl, 500 mM NaCl, 10% glycerol, 20 mM Tris-HCl, pH 8.5, and applied on a nickel-nitrilotriacetic acid column (nickel-Sepharose, GE Healthcare) equilibrated in the same buffer. The proteins were eluted by the application of a 0–500 mM gradient of imidazole (29).

Refolding of TeOmp85 and TeOmp85-C—The purified proteins were precipitated in a final concentration of 90% (v/v) EtOH at –20 °C. The precipitate was washed with water and solubilized in 8 M urea, 20 mM Tris-HCl, pH 8.5. The proteins were refolded by fast dilution with a 20-fold excess of 1% lauryldimethylamine *N*-oxide, 20 mM Tris-HCl, pH 8.5, at room temperature. Residual urea was removed by dialysis for 24 h against the refolding buffer.

Purification of TeOmp85-N—The cells were resuspended in 30 ml of resuspension buffer (150 mM NaCl, 10 mM MgCl₂, 10 mM MnCl₂, 20 mM Tris-HCl, pH 8.5, and a pinch of DNase I) and lysed using a French pressure cell. Cell debris was removed by spinning the lysate for 30 min at $60,000 \times g$, 4 °C. For FPLC purification, the supernatant was then applied to a nickel-nitrilotriacetic acid column (nickel-Sepharose, GE Healthcare) equilibrated with 150 mM NaCl, 20 mM Tris-HCl, pH 8.5. For elution, a 0–500 mM gradient of imidazole was applied. The fractions with pure protein were pooled and concentrated using spin concentrators (Amicon, 30-kDa cutoff, Millipore). For polishing, the protein was applied to a preparative gel sizing column (Superdex 200, GE Healthcare) equilibrated with 150 mM NaCl, 20 mM Tris-HCl, pH 8.0.

Crystallization of TeOmp85-N—The protein was concentrated to 30 mg/ml as determined by the BCA assay (Thermo Fisher Scientific, Rockford, IL). Crystals of *TeOmp85-N* were obtained at 293 K by the vapor diffusion hanging drop method against 200 μl of a reservoir solution. Crystal drops were prepared by mixing 2.5 μl of protein at 30 mg/ml concentration with 2.5 μl of reservoir solution. Crystals were obtained with 8% PEG 4000, 0.8 M LiCl, 0.1 M Tris-HCl, pH 8.5, with a size of up to $300 \times 300 \times 300 \mu\text{m}$.

X-ray Data Collection—All diffraction data were collected at the Swiss Light Source beamline X10SA (PXII) at the Paul

TABLE 1
Summary of data collection, phasing, and refinement statistics

TeOmp85-N	
Data collection^a	
Wavelength	0.933 Å
Space group	I432
Resolution	40–1.97 Å (2.01 to 1.97 Å)
Cell constants	<i>a</i> = 158.18 Å
Unique reflections	21,967 (1047)
Redundancy	22 (25)
Completeness	90.9% (97.7%)
<i>R</i> _{merge-F}	4.7% (40.2%)
<i>I</i> / σ (<i>I</i>)	28.7 (3.1)
Wilson <i>B</i> -factor	52
Phasing statistics	
Figure of merit after SHARP (40 to 2.8)	0.41
Figure of merit after DM (40 to 2.3)	0.86
Refinement statistics^a	
Space group	I432
Resolution	25 to 1.97 Å (2.02–1.97 Å)
<i>R</i> _{cryst}	0.20% (0.25%)
<i>R</i> _{free}	0.23% (0.31%)
Non-hydrogen atoms	1988
Protein residues	235
Waters	157
Mean <i>B</i> -value	27.5 Å ²
r.m.s.d. of bond length	0.024 Å
r.m.s.d. of angle	2.06°
Model quality (non glycine, non proline)	
Residues in most favored region	195% (96.1%)
Residues in most allowed region	8% (3.9%)
Residues in outlier region	0% (0%)

^a Figures in parentheses refer to the highest resolution shell.

Scherrer Institut (PSI) in Villigen, Switzerland. Data were recorded on a mar225 CCD detector (Marreserach GmbH, Norderstedt, Germany). X-ray data of the native crystals were acquired at a wavelength of 0.933 Å and of the SeMet derivative at 0.9782 Å. For the SeMet experiments, 360 images with 1° rotation were recorded for subsequent structure determination by the single anomalous dispersion technique. Single crystals were flash-frozen in their mother liquid supplemented with 30% PEG 400, and data collection was performed at 100 K. Native crystals diffracted to 1.97 Å and SeMet derivatized crystals to 2.1 Å, respectively. The crystal system is I432 with cell constants of *a* = 156.51 Å, α = 90°.

Data Processing, Structure Determination, and Refinement—The single anomalous dispersion data of the Se-Met crystal were integrated, merged, and scaled using the program package XDS/XSCALE (30). One selenomethionine site was located, and experimental phases were calculated using the program package SHARP (31). A partial model was built automatically using ArpWarp (32); missing parts of the protein were built manually and initially refined against the derivative data set. This initial model was further refined against the native data by several cycles of model rebuilding with COOT (33) and automatic refinement using REFMAC (34). Solvent waters were added using the COOT program. The N-terminal His₆ tag and 45 residues as well as the last glycine residue of the protein domain were absent in the model. The geometry was finally checked with PROCHECK (35). The x-ray data statistics are listed in Table 1.

TeOmp85-C Modeling and Sequence Analysis—The sequence of the C-terminal domain of TeOmp85 (TeOmp85-C) was used to identify protein models with significant similarity in the Protein Data Bank data base (release date December 20, 2009) using the

program HHpred (36). The best hit was with the FhaC structure (Protein Data Bank code 2qdz; see [supplemental material S5](#)). The sequence of TeOmp85-C was modeled onto this structure using the program Modeler (37). Regions of highest conservation were identified using the FRpred program (38).

Protein Data Bank Depositions—The atomic coordinates for the crystal structures of TeOmp85-N have been deposited in the Protein Data Bank under the accession number 2x8x.

Single Channel Conductance Measurements—Single channel conductance values were recorded using a BLM workstation (Warner Instruments, Hamden, CT) with a BC-535 amplifier and LPF-8 Bessel filter connected to an Axon Digidata 1440A digitizer. Data were recorded and evaluated using the pClamp 10.0 software (Molecular Devices, Sunnyvale, CA) supplied with the digitizer. 0.5 μ l of a 1% (w/v) solution of 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine in 1:1 (v/v) methanol/chloroform was applied to a 150- μ m aperture in the 1.2-ml polysulfone cuvette (Warner Instruments). After evaporation of the solving agents, the cuvette was filled with the measurement buffer, which was 1 M KCl, 10 mM Tris-HCl, pH 8.5. 1 μ l of a 1% (w/v) solution of 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine in 9:1 (v/v) *n*-decan/butanol was painted onto the 150- μ m aperture of the cuvette. 10 μ l of protein solution (1 mg/ml) was added to the cuvette, which contained the ground electrode of the setup.

Annotation of POTRA Domains—Omp85, TpSS, FtsQ, and patatin-like protein sequences were obtained by PSI-BLAST (39) searches against selected databases of fully sequenced genomes available in November 2008. The respective NCBI taxonomy identifiers are given in [supplemental material S2](#). The gene identifiers of the query protein sequences are given in Tables 2 and 3. In every case, PSI-BLAST was used only with a single iteration to prevent corruption of the results. The secondary structure of the PSI-BLAST results was annotated using the tool Quick2D provided in the bioinformatics toolkit server of the MPI for Developmental Biology, Tübingen, Germany (40). Typically, the PSIPRED (41) result was used to determine POTRA and β -barrel domains. Subsequently, the protein sequences were ordered by the number of POTRA domains, and every POTRA domain was given a specific header indicating its origin and its position within the protein.

Annotation of β -Barrel Domains—The β -barrel portion was considered to be the sequence immediately following the most C-terminal POTRA domain to the C terminus. Information about the number of POTRA domains was included in the sequence header.

Cluster Analysis of POTRA and β -Barrel Domains—For phylogenetic analysis, the derived POTRA sequences were clustered using CLANS (42). Singletons showing no pairwise BLAST connections better than 1×10^{-15} were discarded. Only pairwise BLAST *p* values better than 1×10^{-2} (POTRA analysis) or 1×10^{-4} (β -barrel analysis) were chosen to exert attractive forces. After formation of distinct clusters, the clustering space was reduced to two dimensions. The intensity of the connecting lines is proportional to the reciprocal of the pairwise BLAST *p* value ranging from light gray (POTRA analysis, 1×10^{-2} ; β -barrels analysis, 1×10^{-4}) to black (POTRA domain analysis, 1×10^{-20} ; β -barrel analysis, 1×10^{-40}).

TABLE 2
PSI-BLAST query sequences

POTRA domain proteins from different protein families were used.

Query sequences	Organism	gi no.	Database
<i>TeOmp85</i>	<i>T. elongatus</i>	22299332	441 genomes
FhaC	<i>B. pertussis</i>	33592931	441 genomes
Ctl1556	<i>C. tepidum</i>	21674374	441 genomes
FtsQ	<i>E. coli</i>	146031	441 genomes
DivIB	<i>B. subtilis</i>	409708	NR ^a
Thr0438	<i>T. elongatus</i>	22297981	37 cyanobacterial genomes
Fn2196	<i>Fusobacterium nucleatum</i>	148322338	NR ^b
Ttha0561	<i>T. thermophilus</i>	55980530	NR ^b
Mxan5763	<i>Myxococcus xanthus</i>	108761598	NR ^b
<i>TeOmp85</i>	<i>T. elongatus</i>	22299332	NR_euk ^c
CGI-51 (Sam50)	<i>Homo sapiens</i>	4929571	NR_euk ^c

^a Search for DivIB.

^b Search for Omp85-related proteins with 4, 6, and 7 POTRA domains; only sequences of these proteins were extracted.

^c NR_euk means nonredundant database was filtered for eukaryotic sequences.

TABLE 3
PSI-BLAST results used for further analysis

Retrieved sequences were combined, duplicates were removed, and proteins were annotated according to their domain composition.

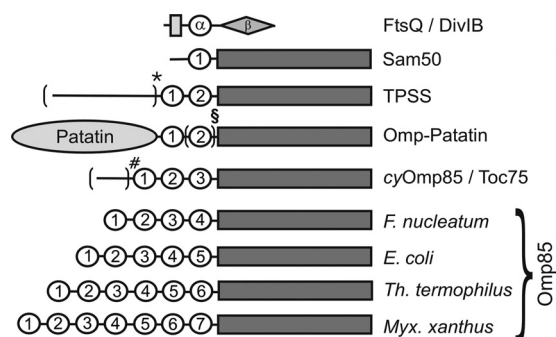
Protein family	No. of POTRA domains	Count
Omp85 and related	3	41
Omp85 and related	4	7 ^a
Omp85 and related	5	130
Omp85 and related	6	7 ^a
Omp85 and related	7	6 ^a
<i>cyOmp85</i>	3	37
Toc75	3	28
TPSS, including cyanobacterial TPSS	3	89
Omp85-like phospholipases (patatin-like)	1–2	44
Sam50	1	77
FtsQ	1	92
DivIB	1	9 ^a
Total		567

^a Note that of the results from NR searches, only selected sequences were used for further analysis.

RESULTS

Bioinformatics

Data Acquisition and Processing—POTRA domains occur in different protein families. The most prominent examples are the Omp85 and TPSS proteins in Gram-negative bacteria and the Omp85 homologues in mitochondria and chloroplasts of eukaryotes, Sam50 and Toc75, respectively. Moreover, POTRA domains are found in the highly conserved cell division protein FtsQ in Gram-negative bacteria and its homologue DivIB in Gram-positive bacteria (20). Their occurrence in a number of different essential proteins spread over all kingdoms of life makes it interesting to analyze the evolutionary history and present relationships of the POTRA domain proteins. To this end, we identified POTRA domains from all sequenced genomes but mostly excluding multiple genomes of the same species. The initial PSI-BLAST search was run against 441 fully sequenced genomes (available in November 2008, see supplemental material S2) using *TeOmp85* as the query sequence. The search result included Omp85 proteins with four to seven POTRA domains, TPSS proteins, cyanobacterial Omp85 with three POTRA domains, chloroplast Toc75, and finally three TPSS-like proteins with an additional N-terminal domain. Because some proteins, e.g. the TPSS proteins and the Omp85 proteins with 4, 6, and 7 POTRA domains, were under-


FIGURE 1. Domain composition of POTRA domain proteins investigated in this study (not to scale). POTRA domains are represented by open circles and are numbered starting from the N terminus. FtsQ and DivIB include an N-terminal transmembrane helix, which anchors the protein in the cytoplasmic membrane. All other proteins are outer membrane proteins with a β -barrel domain illustrated by a dark gray box. *, TPSS of *T. erythraeum* is fused to its substrate and has a N-terminal hemagglutinin-like domain. \S , most Omp-Patatin include only one POTRA domain. #, most cyanobacterial Omp85s have a proline-rich (10–40%) region of variable length and N-terminal to the first POTRA domain.

represented in the result, follow-up searches were performed using the specific sequences found in the initial search as query sequences to increase the size of the data base (see Tables 2 and 3). In total, 567 POTRA domain proteins were included in the further analysis. At least one Omp85 protein was found in every genome, except in genomes of the archaea and the firmicutes, chloroflexi, mollicutes, and actinobacteria, which are commonly referred to as Gram-positive bacteria. Some cyanobacterial strains like *Nostoc* sp., *Anabaena* sp., or *Trichodesmium* sp. have up to three isoforms. Pathogenic proteobacteria like *Pseudomonas* sp., *Burkholderia* sp., *Ralstonia* sp., *Bartonella* sp., *Yersinia* sp., *Haemophilus* sp. and also *Fusobacterium* sp. contain several isoforms of the TPSS proteins.

The resulting sequences were analyzed using the tool Quick2D as described above, annotating POTRA domains by locating the $\beta\alpha\alpha\beta$ secondary structure topology motif. Every POTRA domain sequence was annotated with a header, including information about the position of the domain in the protein and the gi number of the protein, e.g. in *E. coli*, BamA, the N-terminal POTRA domain was annotated as POTRA1. The barrel domains were exclusively located at the C terminus of the protein and always followed directly after the last POTRA domain. The domain composition of the proteins is illustrated in Fig. 1; for the complete list of annotated POTRA and β -barrel domains see supplemental material S3.

Cluster Analysis of POTRA Domains of Omp85 Homologues—We clustered the extracted POTRA and the β -barrel domains separately using CLANS (42) to visualize their sequence relationships (Fig. 2). CLANS is a Java™ applet, which compares protein sequences using pairwise BLAST *E*-values. Protein sequences (represented by a dot) are seeded into a three-dimensional space and are mutually attracted due to a virtual force proportional to their pairwise *p*-value, which is the log BLAST *E*-value and indirectly indicates the degree of sequence similarity. Per default, every sequence dot is equipped with a mild repulsion force to prevent the complete collapse of the system. Highly similar sequences cluster together, and after some rounds the clustering space is reduced from three to two

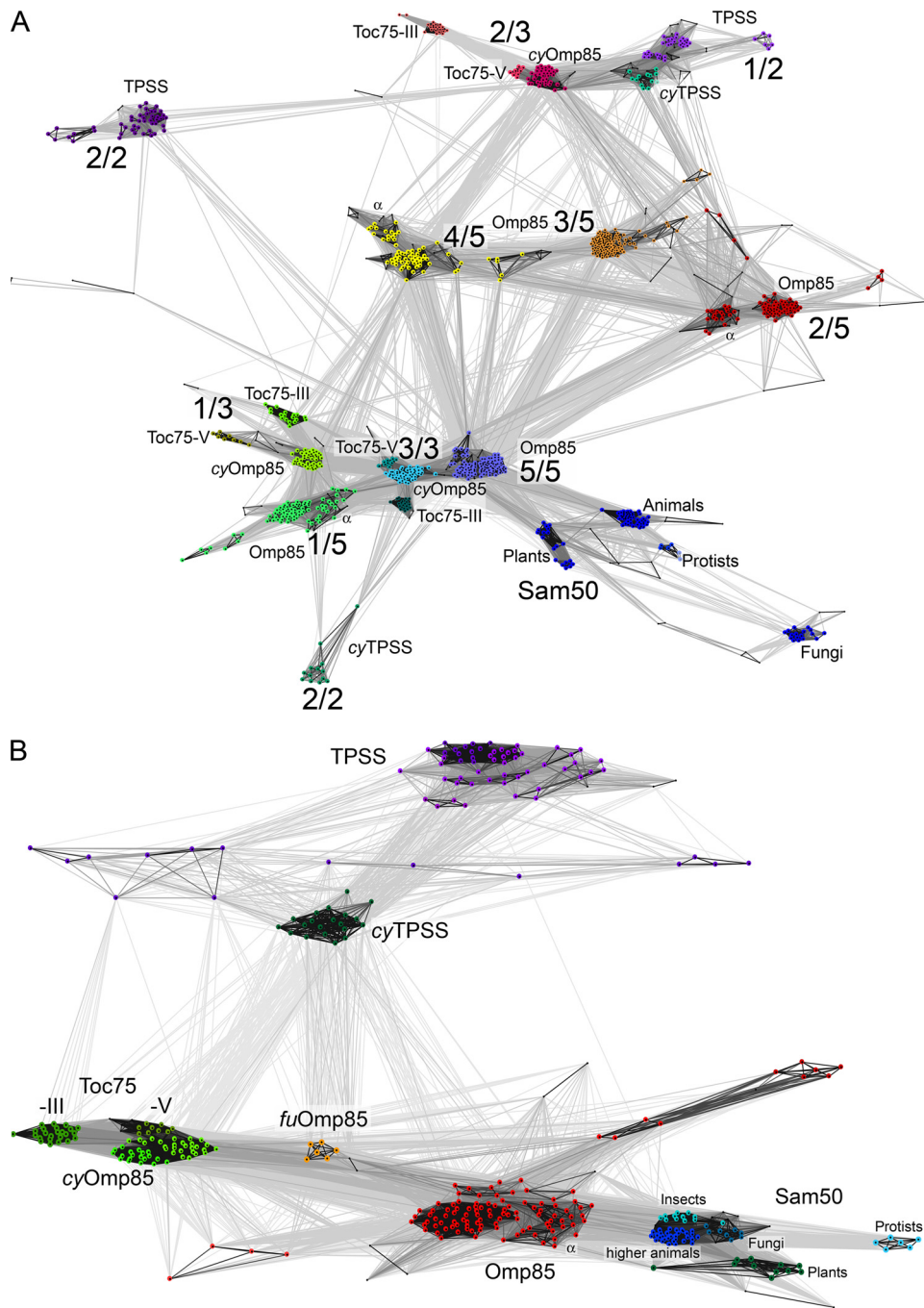


FIGURE 2. *A*, cluster map of the POTRA domains of Sam50, Toc75-III, and Toc75-V, cyanobacterial Omp85 (cyOmp85), proteobacterial Omp85, cyanobacterial (cyTPSS), and proteobacterial TPSS proteins. Each dot represents a single sequence. The intensity of the connecting lines is proportional to the quality of the pairwise BLAST p -value ranging from light gray (1×10^{-2}) to black (1×10^{-20}). The POTRA domains cluster according to their position in related proteins. The N-terminal POTRA domains of cyOmp85 (1 of 3; 1/3), Toc75 (1/3) and Omp85 (1/5) cluster close together. Likewise, the C-terminal POTRA domains (3/3 or 5/5, respectively) of these proteins, together with the single POTRA domain of Sam50, cluster close together. Note, however, that there is no direct connection between the chloroplast and cyanobacterial clusters on the one hand, and the mitochondrial clusters on the other hand, illustrating the huge evolutionary distance between them. Sam50 sequences disperse in a number of subclusters that represent the different kingdoms of eukaryotic life. The central POTRA domains of Omp85 (2/5, 3/5, and 4/5) and cyOmp85 (2/3) cluster further away; for some POTRA domains, the sequences from the α -Proteobacteria form discrete subclusters marked with α . The N-terminal TPSS and cyTPSS POTRAs (1/2) clusters in close proximity to the central POTRA (2/3) of cyOmp85. The C-terminal POTRA domain of cyTPSS and proteobacterial TPSS cluster apart from each other, indicating sequence divergence, most likely because of optimization on the specific substrates. *B*, cluster map of the β -barrel domains of Sam50, Toc75-III, and -V, cyanobacterial Omp85 (cyOmp85), *Fusobacterium* sp. Omp85 (*fu*Omp85), Omp85, cyanobacterial (cyTPSS), and proteobacterial TPSS proteins. Each dot represents a single sequence. The intensity of the connecting lines is proportional to the quality of the pairwise BLAST p -value ranging from light gray 1×10^{-4} to black (1×10^{-40}). The sequences cluster in functional groups, with Sam50, Omp85, and TPSS showing satellite subclusters. Toc75-III, the translocation channel in the chloroplast outer envelope, forms a separate cluster apart from Toc75-V and cyOmp85. Together, these clusters do not directly connect to the mitochondrial (Sam50) clusters, comparable with what is seen for the POTRA domains, above. Proteobacterial TPSS are almost exclusively linked with cyTPSS. The β -barrels of *Fusobacterium* sp. and the Clostridiales *Veillonella parvula* and *Selenomonas flueggii* cluster halfway between cyOmp85 and Omp85, due to possible recombination after horizontal gene transfer. Overall, the p -values indicate a high degree of homology between the mitochondrial Sam50 and Omp85 as well as chloroplast Toc75, cyOmp85, *fu*Omp85, and Omp85 β -barrels, providing strong support for the endosymbiont theory.

Thermosynechococcus Omp85 Structure and Function

dimensions. The intensity of the lines connecting the sequence dots is proportional to the reciprocals of the pairwise BLAST p values. Only connections with p values below 1×10^{-2} were chosen to exert an attractive force on the sequences and are shown in the map.

The cluster map of the POTRA domains is shown in Fig. 2A. For reasons of clarity, the figure shows only the analysis of Omp85 with five POTRA domains and cyanobacterial Omp85 proteins (*cyOmp85*) with three POTRA domains, Toc75, Sam50, and TPSS proteins. At first glance, every POTRA domain forms clusters with others that have the same position in closely related proteins. *cyOmp85*, Toc75-III, and Toc75-V POTRA domains cluster very close together (see also alignment in [supplemental material S4](#)). The clusters of the N- and C-terminal POTRA domains of Toc75, *cyOmp85*, and Omp85 are in close proximity, joined by the clusters of the single POTRA domain of the Sam50 proteins. However, at the p value cutoff chosen for the clustering, there are no connections between the mitochondrial (Sam50) and the chloroplast/cyanobacterial (Toc75 and *cyOmp85*) clusters, illustrating the huge evolutionary distance between them. The Sam50 proteins form several subclusters for sequences derived from fungi, protozoa, yeasts, plants, and animals. The sequences of the yeast Sam50 POTRA domains show the strongest divergence and therefore cluster in the periphery of the map.

Clusters of the central Omp85 POTRA2, -3, and -4 form a triangle-like network quite distant from the N- and C-terminal POTRA domains. The *cyOmp85*/Toc75 central POTRA2 cluster is located even further away, with strong BLAST connections to POTRA2 and -4 of Omp85. The clusters of the C-terminal POTRA domain of Omp85 and *cyOmp85*/Toc75 are the most compact ones indicating a highly conserved sequence and, most likely, function of this domain, which is closest to the transmembrane β -barrel.

Cluster Analysis of POTRA Domains of TPSS Proteins—TPSS proteins are related to Omp85, but include only two POTRA domains and a C-terminal β -barrel (see Fig. 1). The N-terminal POTRA1 (of 2) domains of cyanobacterial and proteobacterial TPSS cluster together and with the central POTRA2 (of 3) domain of *cyOmp85* and Toc75.

The clusters of the C-terminal POTRA2 domains of TPSS proteins from cyanobacteria and proteobacteria are distant from each other and are located in the periphery of the map, only weakly connected with other clusters. This is in strong contrast to the C-terminal POTRA domains of *cyOmp85*, Toc75, Omp85, and the POTRA of Sam50. They are highly conserved and cluster closely together. This indicates a different function for the C-terminal POTRA domain of TPSS and Omp85 proteins and probably reflects the difference in specific (TPSS) versus unspecific (Omp85) substrates. To date, no functional data exist about the cyanobacterial TPSS proteins. The proteobacterial TPSS proteins have been shown to secrete a number of large exoproteins like hemagglutinins, hemolysins, or adhesins (17, 19, 43–45). *Serratia* ShlB activates its hemolysin upon secretion. Mutants that secrete inactive hemolysin had insertions in the POTRA2 domain (46), indicating a possible adaptation in this POTRA domain toward processing of the specific substrate. This would explain the sequence divergence for the TPSS POTRA2 from cyanobacteria and proteobacteria.

Some of the cyanobacterial TPSS proteins are encoded in operons with their putative substrates, in some cases with multiple copies of the putative substrate. An example for the latter case is the TPSS protein All5116 of *Nostoc* sp. PCC 7120, which forms an operon with the proteins All5110 to All5115, all featuring a hemagglutinin-like domain. A special case are TPSS proteins of *Trichodesmium erythraeum*, e.g. Tery_3487, which is a gene fusion with the transported substrate, resulting in a protein with a large N-terminal hemagglutinin-like domain followed by the TPSS part consisting of the two POTRA domains and the β -barrel.

Omp85 of Chlamydiaceae—The POTRA domains 2–5 of the Chlamydiaceae form separate clusters in close proximity to the respective Omp85 clusters (data not shown). However, POTRA1 of the Chlamydiaceae clusters apart from the rest. This is in strong contrast to the high similarity of POTRA1–5 seen for all other Omp85 proteins in this study. The strong sequence divergence of the chlamydial POTRA1 domain implies a specific role for this domain not present in other Omp85s.

Omp85-related Proteins with Unusual Numbers of POTRA Domains—The vast majority of bacterial Omp85 proteins in the data base include five POTRA domains. However, there are a few exceptions to the rule. The Omp85 of Fusobacteria (*fuOmp85*) have four POTRA domains, with POTRA1, -3, and -4 corresponding best to POTRA1–3 of *cyOmp85* (data not shown). The POTRA domain 2 of *fuOmp85* also clusters closest to POTRA3 of *cyOmp85*, indicating a domain duplication event, probably of a unit of two POTRA domains. Fusobacteria are tooth plaque bacteria that have been reported to be prone to horizontal gene transfer, their genome consisting of genes from clostridiales and Gram-negative bacteria, and a rather recent acquisition of the outer membrane from proteobacteria is discussed (47). However, the sequence analysis of the *fuOmp85* domains rather indicates a relationship to *cyOmp85*. This is further supported by the finding that the POTRA domains of Omp85 of other tooth plaque bacteria, for example *Selenomonas flueggii*, a member of the clostridiales, perfectly align to those of *cyOmp85* (see [supplemental material S4](#)).

In *Thermus* sp., *Deinococcus* sp., and some δ -Proteobacteria, the Omp85 homologues include six POTRA domains. Compared with other proteobacterial Omp85 proteins, there is one additional POTRA domain at the N terminus, which forms a separate cluster, whereas the consecutive potras (POTRA2–6) cluster with the respective POTRA1–5 from Omp85 (data not shown). The largest number of POTRA domains, namely seven, can be found in Omp85-related proteins in Myxococcales (which are δ -Proteobacteria) and Acidobacteria. These proteins are found in addition to the regular Omp85s with five POTRA domains within these species. Five of the POTRA domains show sequence similarity with the five POTRA domains of the respective Omp85 of each strain, but the positions of the additional two POTRA domains are different between strains. This provides evidence for recent and independent evolutionary events in these proteins. Again, the C-terminal POTRA7 clusters with the C-terminal POTRA domains of Omp85.

Cluster Analysis of the β -Barrel Domains of Omp85 and TPSS Homologues— β -Barrel outer membrane proteins are a large protein family, which evolved in a modular fashion by β -hairpin duplication (48, 49). Accordingly, the β -barrel domains of Omp85 and related proteins show a much higher degree in conservation than their POTRA domains, illustrated by the difference in *p*-value cutoffs between both cluster analyses. The clustering of the β -barrel domains yields in principle the same pattern as seen in the clustering of the POTRA domains (compare Fig. 2, A and B), albeit less complex as only one domain per protein is considered. *cyOmp85*, *Toc75-III*, and *Toc75-V* (*Oep80*) cluster strongly but form discrete subclusters (see Fig. 2B), similar to what is observed in the POTRA domain cluster map, and emphasizing the endosymbiotic origin of plant chloroplasts. These clusters are not directly connected to the mitochondrial (*Sam50*) clusters. The proteobacterial Omp85 β -barrels are found in a big cluster, with the sequences of *Bacteroides* sp., *Grammella* sp., *Cytophaga* sp., and *Chlorobium* forming satellite clusters. The cluster of *Fusobacterium* sp., *Veillonella*, and *Selenomonas* Omp85 β -barrels is located halfway between the *cyOmp85* and the Omp85, most likely the result of the evolution of these proteins from the recombination of horizontally acquired genes of cyanobacterial and proteobacterial ancestry as discussed above. The sequences of the proteobacterial TPSS β -barrels are forming a rather disperse cluster, which is distinct from the cyanobacterial TPSS. Note that the proteobacterial TPSS are only connected to the cyanobacterial TPSS cluster and not to the proteobacterial Omp85 cluster at the given cutoff. This provides further evidence for a common ancestry of all TPSS proteins as already seen in the POTRA domain cluster analysis. The *Sam50* clusters comprising fungi, protozoa, plants, and animals remain distinguishable. In contrast to the sequences of their POTRA domains, the β -barrels of yeast *Sam50* are found to cluster together with the fungal sequences.

Other Omp85-related Proteins—In the PSI-BLAST results, we found a number of proteobacterial YtfM-like proteins with three POTRA domains. YtfM has recently been reported to be nonessential, but the knock-out of the gene resulted in impaired growth (50). The clustering of the POTRA domains of these proteins resulted in defined clusters only for the C-terminal POTRA domain, although the others were quite dispersed in the cluster map, illustrating their strong sequence divergence. The same poor conservation is observed for the β -barrel domain (data not shown).

As described above, the PSI-BLAST search for Omp85 homologues yielded three uncharacterized proteins of green sulfur bacteria harboring an additional N-terminal domain (see Fig. 1, *Omp-Patatin*). The sequence of CT1556 of *Chlorobium tepidum* was analyzed using HHPred (36), searching proteins of known structure in the Protein Data Bank. The N-terminal domain is homologous to patatin of *Solanum cardiophyllum*, a nightshade plant (Protein Data Bank code 1oxw (51)). The second half of the protein yields the TPSS protein FhaC as best ranking hit (Protein Data Bank code 2qdz; *p* value 2.8×10^{-31} for residues 345–752 (17)). 43 protein sequences were found in the follow-up search, including sequences from pathogenic bacteria like *Ralstonia* sp., *Burkholderia* sp., *Vibrio* sp., and *Pseudomonas* sp. Among the results, three proteins, all from

Chlorobium sp., have two POTRA domains, and all others only one. The cluster analysis of the POTRA domains shows a rather disperse distribution of the sequences. The C-terminal β -barrel domain clusters apart from the TPSS and Omp85s (data not shown). The N-terminal patatin-like domain shows the highest degree of conservation of the three domains. We assume that because of a selection for the interaction with the patatin domain, the sequence of the transporting domains diverged from those of TPSS or Omp85 proteins.

None of the proteins found in the query has been characterized so far, but patatin from *S. cardiophyllum* has been shown to be an unspecific lipid acyl hydrolase with insecticidal activity. The catalytic Ser-Asp dyad is conserved in the bacterial proteins found in the query. The architecture resembles that of the TPSS protein Tery_3487, which is a fusion of the transported substrate and the secretion proteins (see Fig. 1). The patatin domain might be secreted through the pore and either released after cleavage or presented to the outside medium in an autotransporter fashion. Similar to TPSS proteins, the POTRA domains could assist in the secretion process. The function and the topology of this protein family need to be addressed experimentally.

FtsQ and DivIB—FtsQ is a component of the essential Fts complex of proteins that mediate the septum formation initiating the cell division (52). Bioinformatic analysis has revealed that this protein also harbors a POTRA domain and the structure reported recently confirmed this finding (20, 53). In the cluster analysis of FtsQ/DivIB together with Omp85 POTRA domains, the FtsQ/DivIB sequences form a compact cluster with very remote connections to all POTRA clusters of Omp85 (data not shown). As for the POTRAs of Omp85-like patatins, the sequences probably diverged because of the optimization for different interaction partners, which are in this case components of the Fts-Div complex. Yet the hydrophobic residues that provide the POTRA signature are conserved also in FtsQ/DivIB. The POTRA domain is located at the cytoplasmic membrane surface, preceded by the N-terminal transmembrane segment (see Fig. 1). The structural similarity (53) and the proposed similar function, which is protein-protein interaction for both Omp85 and FtsQ POTRAs, provide evidence that the domains share a common origin.

Electrophysiological Characterization of TeOmp85 and TeOmp85-C

The C-terminal domain of *TeOmp85* forms a β -barrel in the outer membrane. The existence of a periplasmic domain at the N terminus, formed by the POTRA domains and an additional disordered proline-rich region N-terminal to the POTRA domains, prompted us to compare the electrophysiological properties of the pore in the presence and absence of the soluble portion of the protein. We conducted the electrophysiological characterization in a black lipid bilayer setup using *TeOmp85* and a truncated mutant, *TeOmp85-C*. The latter includes only the C-terminal β -barrel pore (residues 281–635).

Recombinantly produced *TeOmp85* was added to the *cis*-side of the membrane and inserted not immediately but very reliably after some time at voltages around +100 mV. The measurements were recorded at a constant potential of +100 mV. Several different conductance states were observed, one

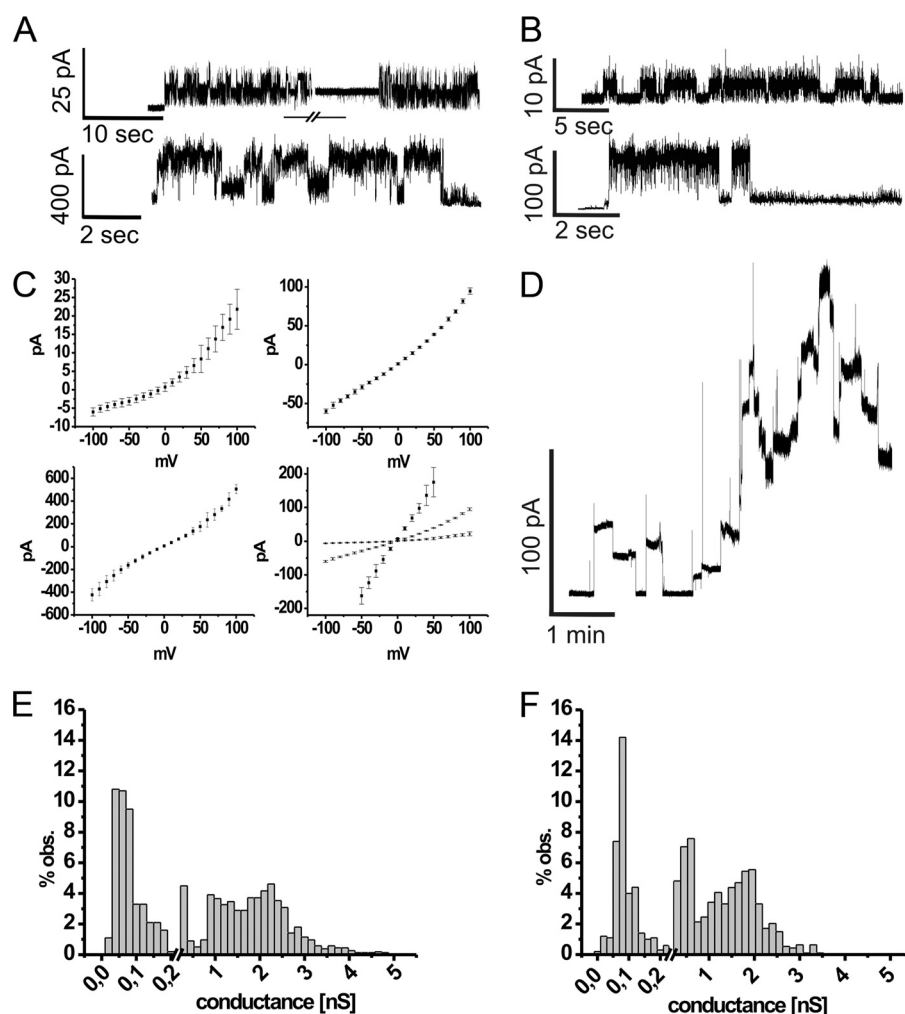


FIGURE 3. Electrophysiological characterization of *TeOmp85* and *TeOmp85-C* (which represents only the β -barrel domain of *TeOmp85*). *A*, current traces of *TeOmp85* showing the low conductance state (*upper panel*) and the high conductance state (*lower panel*), at 100 mV. *B*, current traces of *TeOmp85-C* showing the low conductance state (*upper panel*) and the high conductance state (*lower panel*) at 50 mV. *C*, *I/V* plots of *TeOmp85*, showing the low (*upper left*), medium (*upper right*), and high (*lower left*) conductance state. *Lower right*, all three conductance states. *D*, *TeOmp85-C*, multiple channels inserting into the membrane. Stepwise increase of the current flow, usually leading to membrane rupture within minutes. *E*, conductance histogram of *TeOmp85*, measured at 100 mV, 1559 events. *F*, conductance histogram of *TeOmp85-C*, measured at 50 mV, 936 events.

abundant state at ~ 80 pS, a medium conductance state at ~ 600 pS, and a high conductance state at ranging from 1.5 to 4 nS.

After the insertion, *TeOmp85* usually flickers in an erratic fashion to eventually establish a channel at a conductance level of ~ 80 pS (see Fig. 3*A*). Sometimes the channel immediately opened up to form higher conductance level from 500 pS to 4 nS. After lowering the membrane potential for some minutes to 50 mV, the channel reduced its activity, and one could only observe the low conductance state again or the channel closed completely. This phenomenon can also be observed in the *I/V* plots of *TeOmp85* (Fig. 3*C*, average of three voltage ramps started from negative membrane potentials), where the slope of the current becomes steeper with increasing voltage, showing the opening of additional channels and an increased altitude of flickering. In contrast, when negative voltage is applied, the channel tends to flicker to the closed state.

The medium conductance state of ~ 600 pS at -100 mV and 1 nS at $+100$ mV showed almost no conductance fluctua-

tions, which is also illustrated by the small error bars in the plot (see Fig. 3*C*, *upper right*). The high conductance state of ~ 4 nS showed an almost point symmetric *I/V* curve with the channel sometimes flickering to a higher conductance state with increased voltage, resulting in bigger standard errors (see Fig. 3*C*, *lower left*).

In contrast to *TeOmp85*, the truncated protein *TeOmp85-C* refolded from inclusion bodies inserted readily and more quickly into the membrane, often in a way that several proteins followed the insertion of the first one, leading to a stepwise increase of the current (see Fig. 3, *B* and *D*) and finally to membrane rupture within a matter of minutes. However, we also obtained measurements where single channels were established. The channel seemed to become very unstable with increasing voltage, so that all long time measurements had to be conducted at $+50$ mV to allow recording of discrete conductance steps. Measured at 50 mV, the channel showed a similar gating behavior as the full-length protein (measured at $+100$ mV), and also the same three conductance states could be observed, with the medium conductance state of ~ 600 pS stronger represented as in the distribution for *TeOmp85* (compare Fig. 3, *E*, *TeOmp85*; *F*, *TeOmp85-C*).

The soluble N terminus might fulfill a stabilizing role for the chan-

nel, accomplished by interactions of the C-terminal POTRA domain with the periplasmic turns connecting the β -strands of the pore. This can also be observed in the crystal structure of the related TPSS protein FhaC from *B. pertussis*, where the N terminus forms a helix inside the pore and the second POTRA domain in its close proximity (17). Sequence and secondary structure alignments with HHPred indicate a pore of at least equal size as the one of FhaC ([supplemental material S5](#)). The two α -helices and the adjacent loops from and to β -strands 1 and 2 of POTRA2 of *B. pertussis* FhaC wrap around the β -barrel turns 1 and 2 forming several salt bridges. A similar scenario can be expected for *TeOmp85*. The alignment shows an even more extended loop from β -strand 1 to α -helix 1 of POTRA3 than in FhaC POTRA2 (each representing the most C-terminal POTRA domain). Moreover, the sequence of this region features a rather high portion of charged residues as follows: 20 of 62 (6 arginines, 3 lysines, 6 aspartate, and 5 glutamates) compared with 10 of 44 in POTRA2 (3 arginines, 2 lysines, 1 aspar-

tate, and 4 glutamates). The unstructured, proline-rich N terminus might fold back into the channel just as the N-terminal helix does in FhaC, but this should lead to a strongly decreased conductance of the full-length protein, which is not observed.

The conductance of *TeOmp85* is in the range of other cyanobacterial Omp85s, plant Toc75 and TPSS proteins reported before (17, 22, 54–56) and higher than that of proteobacterial Omp85 (2, 54, 57, 58) and Sam50 (6, 54). For *TeOmp85*, the deletion of the POTRA domains leads to a destabilization of the pore at higher voltages. The deletion of the POTRA domains has been shown to have diverse effects on the channel conductance of Omp85 and TPSS proteins. For Omp85 of *Nostoc* or Toc75 of *Pisum sativum*, the removal of the POTRA domains resulted in a reduction of the conductance fluctuations. This is in contrast to the results reported for HMWIB of *Haemophilus influenzae* (55) or *E. coli* BamA (58), where the N-terminally truncated proteins show stronger conductance fluctuations. Albeit being more closely related to plant Toc75 and *cyOmp85* from *Nostoc*, *TeOmp85-C* behaves rather like HMWIB-C or BamA-C in our conductance measurements. The frequently observed multiple insertions of *TeOmp85-C* often seem cooperative, implying either the autocatalysis of membrane insertion or a propensity to oligomerize for the *TeOmp85* β -barrel. This effect is not observed in the presence of the POTRA domains.

Structure of *TeOmp85-N*

The N-terminal domain of *TeOmp85* was crystallized in the cubic space group *I432* with one monomer in the asymmetric unit. The structure solution by heavy atom derivatization using platinum salts was not successful; however, the preparation of protein with SeMet allowed the determination by single-wavelength anomalous diffraction methods on the basis of the single methionine residue Met²²⁷. We determined the structure to a resolution of 1.97 Å, without the N-terminal residues 1–45 that are missing due to disorder in the crystal packing. This residue range includes a large number of prolines and is predicted to be disordered in different secondary structure prediction tools. This proline-rich sequence is found in almost all cyanobacterial Omp85 proteins. It forms a sequence stretch of variable length and remarkably low complexity, with a disproportionate content of prolines (~10–40%), alanines, and glutamates, leading to a negative net charge. The topology of this part of the protein is not known. In *B. pertussis* FhaC, the N terminus of the protein forms an α -helix within the pore, but the insertion of a charged moiety like the glutamate-rich region of *cyOmp85* into the pore would severely affect its electrophysiological characteristics. This is not observed as *TeOmp85-C* and full-length *TeOmp85* (which includes the proline-rich region) show similar conductance states (see above). The function of this conserved region still needs to be revealed. This proline-rich region is in all cyanobacteria followed by three POTRA domains. In the structure, POTRA1 (residues Leu⁵¹–Leu¹¹⁰), POTRA2 (residues Gly¹¹¹–Asp¹⁸⁷), and POTRA3 (residues Gly¹⁸⁸–Glu²⁷⁷) were modeled. The domains align with each other by a small rotation and translation between POTRA1 and POTRA2 and a translation between POTRA2 and POTRA3, respectively. Together, they form an extended banana-shaped scaffold with a 10-nm extension (see Fig. 4A and Fig. 5A). Interactions between the individual POTRA domains are relatively weak. Interactions between

POTRA1 and POTRA2 are mediated by residues Asn¹⁵³ and Arg¹⁵⁵ and by main chain atoms of residues located on the loop between α 2 and β 2 (Ala⁹⁴–Gly⁹⁶). A similar arrangement is observed between POTRA2 and POTRA3, where Asn²³⁶ and Gln²³⁷ stabilize the equivalent loop of POTRA2 (residues Asn¹⁷¹ and Gly¹⁷²). In addition, one H-bond is formed by Gln¹⁹³ and Arg²⁶⁷ (see Fig. 4A). The labile arrangement of POTRA domains may have consequences for their plasticity, and indeed, several different conformations were reported for the *E. coli* homologue.

POTRA domains form a compact $\beta\alpha\alpha\beta\beta$ mixed α/β -fold with a small three-stranded β -sheet covered by two helices slightly twisted against each other (see Fig. 4B). This arrangement forms a conserved hydrophobic core, with the POTRA sequence fingerprint responsible for structure maintenance (17, 25). Mostly medium size hydrophobic residues (valines and leucines) are conserved between the individual domains.

When superimposed individually, POTRA1 and POTRA3 of *TeOmp85-N* align significantly better than POTRA1/POTRA2 or POTRA3/POTRA2, respectively, with a root mean square deviation (r.m.s.d.) of 1.6 Å and 17% identical residues for POTRA1/POTRA3, 2.4 Å r.m.s.d. and 18% identical residues for POTRA1/POTRA2, and 2.2 Å r.m.s.d. and 18% identical residues for POTRA2/POTRA3 (see Fig. 4C). The structural conservation confirms the findings of the cluster analysis of the POTRA domains, which showed that POTRA1 and -3 of cyanobacterial Omp85s cluster much closer together than either does with POTRA2.

In the crystal packing the extended structure forms an interface of 1100 Å² with a crystallographically related molecule, which at first raised the question whether this interface would represent a biologically relevant dimerization interface. A similar interface has been reported for Omp85 (BamA), here of 1900 Å² (25). However, in agreement with the findings for Omp85 using SAXS methods, neither size exclusion chromatography nor cross-linking gave any indication for an oligomeric form of this domain (24).

To get an idea how the POTRA architecture would connect to the β -barrel domain, we superimposed the structure of *TeOmp85-N* onto the FhaC protein, the integral membrane interaction partner of the TPSS. This member of the membrane transporter superfamily forms a 16-stranded β -barrel with two N-terminally localized POTRA domains. POTRA domains 2 and 3 of the cyanobacterial protein align well to the POTRA domains of FhaC with an overall r.m.s.d. of 3.4 Å (see Fig. 4E; for 140 C α atoms, 2.6 Å r.m.s.d. for POTRA2_{*TeOmp85-N*}/POTRA1_{FhaC} and 2.2 Å r.m.s.d. for POTRA3_{*TeOmp85-N*}/POTRA2_{FhaC}). The similarity of the POTRA2_{*TeOmp85-N*} and POTRA1_{FhaC} is in good agreement with the result of our cluster analysis, where these domains clustered close together, showing very low pairwise *p*-values (see Fig. 2A). In the FhaC protein, several residues were tested for their possible influence on substrate binding, and only residues on α -helix 2 of POTRA1 were found to suppress export of the FHA substrate protein (17). This helix 2 corresponds to helix α 4 on POTRA2 of *TeOmp85-N*. Residues Tyr¹⁰⁶, Asp¹⁰⁷, and Arg¹⁰⁸ of FhaC align well with residues Arg¹⁶⁹, Asp¹⁷⁰, and Asn¹⁷¹ in *TeOmp85-N*, all having the potential to form hydrogen bonds with potential substrate proteins; note especially the exposed conserved aspartate. It is tempting to spec-

Thermosynechococcus Omp85 Structure and Function

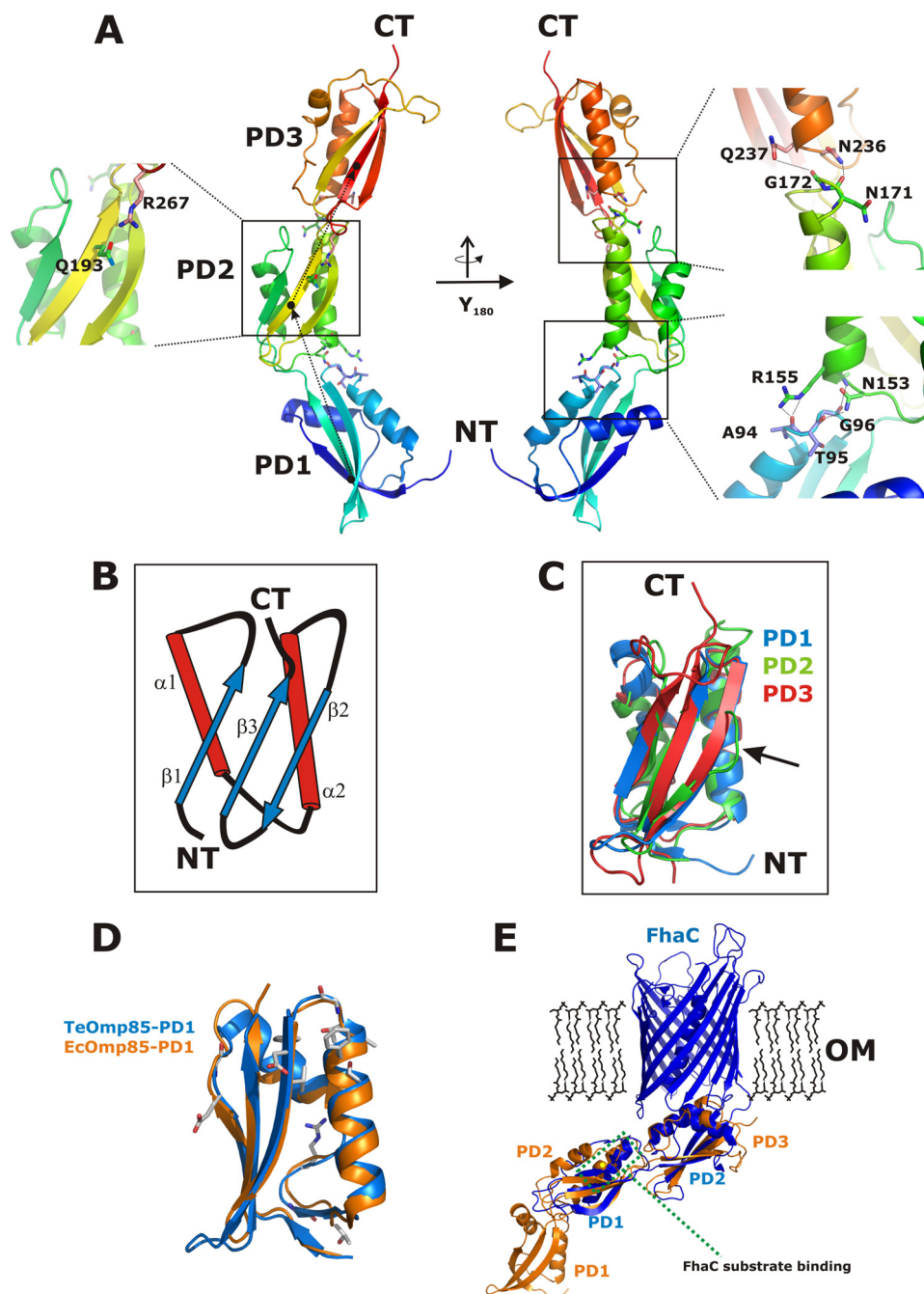


FIGURE 4. Structure of the N-terminal domain of *TeOmp85*. *A*, two ribbon plots of the *TeOmp85* domain architecture rotated by 180° show the three POTRA domains *PD1*, *PD2*, and *PD3*. The structure is color-coded from the N terminus (*NT*, blue) to the C terminus (*CT*, red). Interfaces connecting the POTRA domains are enlarged, and residues contributing to the interface contacts are labeled. *B*, schematic view of the POTRA domain fold. A β -sheet consisting of three β -strands is flanked by two α -helices, yielding a topology of β_1 - α_1 - α_2 - β_2 - β_3 . *C*, structure superposition of the three *TeOmp85* POTRA domains shown in blue (*PD1*), green (*PD2*), and red (*PD3*). An unusual protrusion present only in the β_2 strand of *PD2* is marked by an arrow. *D*, superposition of the two N-terminal POTRA domains of *TeOmp85* (*TeOmp85-PD1*) and *E. coli* *Omp85* (*BamA*, *EcOmp85-PD1*). Identical residues identified in the superposition are highlighted in stick representation. *E*, structure alignment of the entire *FhaC* Omp85-like export protein (in blue) with the N-terminal portion of *TeOmp85* (in orange). POTRA domains one and two of the integral transporter align well with POTRA domains *PD2* and *PD3* of the *TeOmp85* protein. The potential substrate-binding site located in the first POTRA domain of *FhaC* is marked. The overall representation is a good model for the full *cyOmp85* structure.

ulate that this represents a functionally important (and thus conserved) region, although the function of the two proteins is slightly different (protein export *versus* insertion). By contrast, the superposition of *TeOmp85-N* with the four N-terminal POTRA

domains of *Omp85/BamA* overall does not align as well, except for *POTRA1* (Fig. 4*D* shows the alignment of *POTRA1*, the rest of the structural alignment is not shown). In the *Omp85/BamA* protein, there is a significant kink between *POTRA2* and *POTRA3* that is comparable with the kink in the *TeOmp85-N* protein, here between *POTRA1* and *POTRA2*. Another feature in the *Omp85* structure was identified as β -strand augmentation between *POTRA3* and the β -strand 1 of the truncated *POTRA5* in the crystal (25). Although likely to be an artifact of crystallization, the authors mention the possible importance of this pairing *in vivo*.

We have also analyzed the conservation pattern of the C-terminal domain, which was predicted to contain 16 β -strands and displays a significant similarity to *FhaC*. The domain was modeled on the basis of *FhaC* structure (17). Three strongly conserved parts are observed between the cyanobacterial and the *FhaC* barrel domain (Fig. 5*B*). One area forms the interaction surface to *POTRA3* (*POTRA2* in *FhaC*), which is presumably important for the alignment of the POTRA domains relative to the transmembrane domain. A second patch of conserved residues is localized around the N- and C-terminal part of the barrel (strands β_1 , β_{16} , and β_{15}). These features may originate from two different constraints. First, the C termini of outer membrane proteins (OMPs) in bacteria and mitochondria are conserved in that they express so called β -signals, which are important for the recognition and insertion by the assembly machineries (which in this case is self-recognition). Second, the lateral release of a newly assembled OMP was postulated to occur as a stepwise process where part of the OMP is temporarily included into the vicinity between β_1 and β_{16} (6). A similar mechanism has been proposed for the

Sam50 protein from mitochondria. Finally, the long loop L6 folding back into the β -barrel is well conserved between *FhaC* and *cyOmp85*. This loop has been implied in *FHA* export and may be responsible for the assembly of OMPs by *cyOmp85* as well.

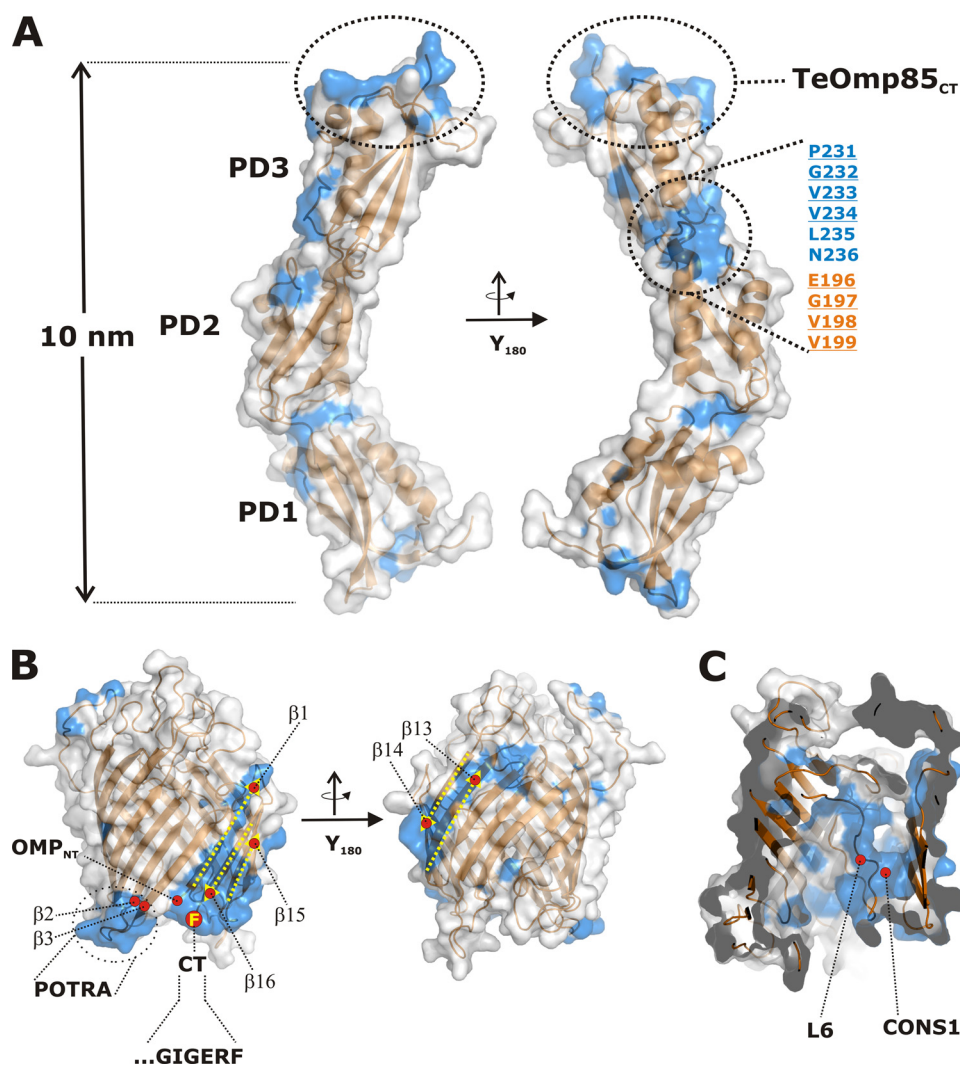


FIGURE 5. Analysis of conserved residues in the N-terminal domain and model of the C-terminal domain. *A*, surface representation of the N-terminal domain with two views of TeOmp85 rotated by 180°. Conserved residues are marked in blue on the surface. Patches of high conservation are mostly found at the C-terminal end of the structure, which presumably is involved in the interactions with the β -barrel part of the protein. Another region of high conservation is seen at the interface between PD2 and PD3. Individual conserved residues at this interface are marked in orange (for PD2) and blue (for PD3). *B*, surface representation of the C-terminal domain model of TeOmp85 (based on the structure of FhaC) in two different views. Note two areas of high conservation. One area is exposed in the loop between strands β 2 and β 3. Residues comprising this area are presumably involved in interactions with the C-terminal POTRA domain. Close by, a second region of high conservation is formed by the first β -strand together with the terminal three β -strands. As for most β -barrel proteins of the bacterial outer membrane, the C terminus is highly conserved (marked GIGERF) and carries a phenylalanine residue (marked *F*) similar to bacterial porins. *C*, representation of the surface exposed to the interior of the channel. The highly conserved loop L6 is shown aligned to another conserved patch (CONS1) at the inner surface of the barrel.

DISCUSSION

The endosymbiotic theory states that chloroplasts and mitochondria once were bacteria, which were taken up by other cells. Therefore, they still maintain a reduced genome and their own set of proteins (12, 13). Most of the proteins in these organelles, however, are produced in the endoplasmic reticulum of the cell and imported through the organelle outer membrane via Tom40 channels in mitochondria or Toc75-III channels in chloroplasts (9, 10). Mitochondrial outer membrane or chloroplast outer envelope proteins are then most probably inserted into the respective membrane with assistance of Sam50/Tob55 or Toc75-V (Oep80). The arrangements in our

cluster maps strongly support the endosymbiotic theory, showing the close relatedness of Sam50 and Omp85, Toc75, and cyOmp85 until today and also emphasizing their common functional role in protein transport and membrane protein insertion.

Some aspects of this overall similar function have changed, or rather diversified, however. In plant chloroplasts, Toc75-III has altered its function from the insertion of proteins into the outer membrane to translocation of proteins through the outer membrane. Plant Toc75-V probably then takes the role which cyOmp85 has in outer membrane assembly in cyanobacteria. Thus, we expect the structural (Fig. 4) and sequence data (supplemental material S4), which we provide here, should be valid for Toc75-V and, to a lesser extent, for Toc75-III.

The importance of the single POTRA domains and their overall number has been studied for *E. coli* BamA and *Neisseria meningitidis* Omp85, with partly conflicting results (21, 25). cyOmp85 only includes three POTRA domains, the first and last of which correspond best to the first and last (of five) POTRA domains in proteobacterial Omp85. We show this on the level of sequence (Fig. 2A) and, for POTRA1, structure (Fig. 4D) (note that there is no complete structural data for POTRA5 (25, 59)). This suggests that these two POTRA domains host the most central functions of the protein, which are substrate recognition and membrane insertion (in combination with the β -barrel domain, by a mechanism that is not yet understood).

The simplest architecture conceivable for Omp85 proteins is realized in Sam50, which has only a single POTRA domain. This domain is most closely related to the C-terminal POTRAS of all other Omp85 homologues. It is reasonable to assume that protein evolution typically proceeds from simple to more complex architectures and that a common ancestor of all Omp85 proteins included only one POTRA domain. More complex demands on the protein function then presumably lead to duplications of the POTRA domain. In the sequence space of today, we did not observe Omp85 homologues with only two POTRA domains that would correspond to POTRA1 and POTRA3 in cyOmp85 (POTRA1 and POTRA5 in proteobacte-

rial Omp85), albeit this would be a logical intermediate in the evolution of these proteins. The TPSS proteins do not represent this intermediate, as their C-terminal POTRA domain does not correspond to the C-terminal POTRA domain of all other proteins discussed here, and their N-terminal domain seems to correspond best to the much later acquired POTRA2 from both proteobacterial and cyanobacterial Omp85.

The different number of POTRA domains in Sam50 (one), cyOmp85 and Toc75 (three), and Omp85 (mostly five) has also to be seen in the context that the protein forms complexes with other proteins. To date, it remains unknown whether or not cyOmp85 forms complexes, but Sam50 forms a complex with Sam37, Sam35, and Mdm10; Toc75-III is found in a complex with Toc64, Toc34, and Toc159 (9) proteins, and several Omp85s have been reported to coordinate a number of lipoproteins (4), none of which are found in *T. elongatus*. The interaction with the complex partners is most likely also accomplished by the POTRA domains, so that not all POTRA domains are involved in interactions with the substrate proteins. Examples from Fusobacteria and Myxococcales show that until now the domains are recombined to form proteins with different numbers of POTRA domains, to optimize the still not understood mechanism of Omp85.

Acknowledgments—We thank Andrei Lupas for continuing support, and the beamline staff of the Swiss Light Source, Villigen, Switzerland, for providing an excellent technical facility. We thank Jan Kern and Athina Zouni for the *T. elongatus* cells used in this study.

REFERENCES

- Voulhoux, R., Bos, M. P., Geurtsen, J., Mols, M., and Tommassen, J. (2003) *Science* **299**, 262–265
- Robert, V., Volokhina, E. B., Senf, F., Bos, M. P., Van Gelder, P., and Tommassen, J. (2006) *PLoS Biol.* **4**, e377
- Volokhina, E. B., Beckers, F., Tommassen, J., and Bos, M. P. (2009) *J. Bacteriol.* **191**, 7074–7085
- Wu, T., Malinverni, J., Ruiz, N., Kim, S., Silhavy, T. J., and Kahne, D. (2005) *Cell* **121**, 235–245
- Habib, S. J., Waizenegger, T., Lech, M., Neupert, W., and Rapaport, D. (2005) *J. Biol. Chem.* **280**, 6434–6440
- Kutik, S., Stojanovski, D., Becker, L., Becker, T., Meinecke, M., Krüger, V., Prinz, C., Meisinger, C., Guiard, B., Wagner, R., Pfanner, N., and Wiedemann, N. (2008) *Cell* **132**, 1011–1024
- Paschen, S. A., Neupert, W., and Rapaport, D. (2005) *Trends Biochem. Sci.* **30**, 575–582
- Reumann, S., Inoue, K., and Keegstra, K. (2005) *Mol. Membr. Biol.* **22**, 73–86
- Schleiff, E., and Soll, J. (2005) *EMBO Rep.* **6**, 1023–1027
- Walthers, D. M., Rapaport, D., and Tommassen, J. (2009) *Cell. Mol. Life Sci.* **66**, 2789–2804
- Bölter, B., Soll, J., Schulz, A., Hinnah, S., and Wagner, R. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15831–15836
- Gould, S. B., Waller, R. F., and McFadden, G. I. (2008) *Annu. Rev. Plant Biol.* **59**, 491–517
- Gray, M. W., Burger, G., and Lang, B. F. (1999) *Science* **283**, 1476–1481
- Reumann, S., Davila-Aponte, J., and Keegstra, K. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 784–789
- Inoue, K., and Potter, D. (2004) *Plant J.* **39**, 354–365
- Patel, R., Hsu, S. C., Bédard, J., Inoue, K., and Jarvis, P. (2008) *Plant Physiol.* **148**, 235–245
- Clantin, B., Delattre, A. S., Rucktooa, P., Saint, N., Méli, A. C., Loch, C., Jacob-Dubuisson, F., and Villeret, V. (2007) *Science* **317**, 957–961
- Henderson, I. R., Navarro-Garcia, F., Desvaux, M., Fernandez, R. C., and Ala'Aldeen, D. (2004) *Microbiol. Mol. Biol. Rev.* **68**, 692–744
- Schiebel, E., Schwarz, H., and Braun, V. (1989) *J. Biol. Chem.* **264**, 16311–16320
- Sánchez-Pulido, L., Devos, D., Genevrois, S., Vicente, M., and Valencia, A. (2003) *Trends Biochem. Sci.* **28**, 523–526
- Bos, M. P., Robert, V., and Tommassen, J. (2007) *EMBO Rep.* **8**, 1149–1154
- Ertel, F., Mirus, O., Bredemeier, R., Moslavac, S., Becker, T., and Schleiff, E. (2005) *J. Biol. Chem.* **280**, 28281–28289
- Habib, S. J., Waizenegger, T., Niewianda, A., Paschen, S. A., Neupert, W., and Rapaport, D. (2007) *J. Cell Biol.* **176**, 77–88
- Knowles, T. J., Jeeves, M., Bobat, S., Dancea, F., McClelland, D., Palmer, T., Overduin, M., and Henderson, I. R. (2008) *Mol. Microbiol.* **68**, 1216–1227
- Kim, S., Malinverni, J. C., Sliz, P., Silhavy, T. J., Harrison, S. C., and Kahne, D. (2007) *Science* **317**, 961–964
- Prilipov, A., Phale, P. S., Van Gelder, P., Rosenbusch, J. P., and Koebnik, R. (1998) *FEMS Microbiol. Lett.* **163**, 65–72
- Miroux, B., and Walker, J. E. (1996) *J. Mol. Biol.* **260**, 289–298
- Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L., and Sweet, R. M. (1993) *Nature* **362**, 219–223
- Petty, K. J. (2001) *Curr. Protoc. Protein Sci.* Unit 9.4, 9.4.1–9.4.16
- Kabsch, W. (1993) *J. Appl. Crystallogr.* **26**, 795–800
- Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) *Methods Mol. Biol.* **364**, 215–230
- Perrakis, A., Morris, R., and Lamzin, V. S. (1999) *Nat. Struct. Biol.* **6**, 458–463
- Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
- Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255
- Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283–291
- Söding, J., Biegert, A., and Lupas, A. N. (2005) *Nucleic Acids Res.* **33**, W244–W248
- Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U., and Sali, A. (2007) *Curr. Protoc. Protein Sci.* Unit 2.9, 2.9.1–2.9.31
- Fischer, J. D., Mayer, C. E., and Söding, J. (2008) *Bioinformatics* **24**, 613–620
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402
- Biegert, A., Mayer, C., Remmert, M., Söding, J., and Lupas, A. N. (2006) *Nucleic Acids Res.* **34**, W335–W339
- Jones, D. T. (1999) *J. Mol. Biol.* **292**, 195–202
- Frickey, T., and Lupas, A. (2004) *Bioinformatics* **20**, 3702–3704
- Braun, V., Hobbie, S., and Ondraczek, R. (1992) *FEMS Microbiol. Lett.* **79**, 299–305
- Hertle, R. (2005) *Curr. Protein Pept. Sci.* **6**, 313–325
- Jacob-Dubuisson, F., Loch, C., and Antoine, R. (2001) *Mol. Microbiol.* **40**, 306–313
- Yang, F. L., and Braun, V. (2000) *Int. J. Med. Microbiol.* **290**, 529–538
- Mira, A., Pushker, R., Legault, B. A., Moreira, D., and Rodríguez-Valera, F. (2004) *BMC Evol. Biol.* **4**, 50
- Arnold, T., Poynor, M., Nussberger, S., Lupas, A. N., and Linke, D. (2007) *J. Mol. Biol.* **366**, 1174–1184
- Remmert, M., Biegert, A., Linke, D., Lupas, A. N., and Söding, J. (2010) *Mol. Biol. Evol.*, in press
- Stegmeier, J. F., Glück, A., Sukumaran, S., Mäntele, W., and Andersen, C. (2007) *Biol. Chem.* **388**, 37–46
- Rydell, T. J., Williams, J. M., Krieger, E., Moshiri, F., Stallings, W. C., Brown, S. M., Pershing, J. C., Purcell, J. P., and Alibhai, M. F. (2003) *Biochemistry* **42**, 6696–6708
- Errington, J., Daniel, R. A., and Scheffers, D. J. (2003) *Microbiol. Mol. Biol. Rev.* **67**, 52–65
- van den Ent, F., Vinkenvleugel, T. M., Ind, A., West, P., Veprintsev, D., Nanninga, N., den Blaauwen, T., and Löwe, J. (2008) *Mol. Microbiol.* **68**,

- 110–123
54. Bredemeier, R., Schlegel, T., Ertel, F., Vojta, A., Borissenko, L., Bohnsack, M. T., Groll, M., von Haeseler, A., and Schleiff, E. (2007) *J. Biol. Chem.* **282**, 1882–1890
55. Duret, G., Szymanski, M., Choi, K. J., Yeo, H. J., and Delcour, A. H. (2008) *J. Biol. Chem.* **283**, 15771–15778
56. Jacob-Dubuisson, F., El-Hamel, C., Saint, N., Guédin, S., Willery, E., Molle, G., and Loch, C. (1999) *J. Biol. Chem.* **274**, 37731–37735
57. Nesper, J., Brosig, A., Ringler, P., Patel, G. J., Müller, S. A., Kleinschmidt, J. H., Boos, W., Diederichs, K., and Welte, W. (2008) *J. Bacteriol.* **190**, 4568–4575
58. Stegmeier, J. F., and Andersen, C. (2006) *J. Biochem.* **140**, 275–283
59. Gatzeva-Topalova, P. Z., Walton, T. A., and Sousa, M. C. (2008) *Structure* **16**, 1873–1881