Evolution of outer membrane β-barrels from an ancestral $\beta\beta$ hairpin - Supplementary Information

Michael Remmert[1,2], Andreas Biegert[1,2], Dirk Linke[1], Andrei N. Lupas[1] and Johannes Söding[1,2]

[1]Gene Center Munich and Center for Integrated Protein Science (CIPSM), Department of Biochemistry, Ludwig-Maximilians-Universtät München, Feodor-Lynen-Str. 25, 81377 Munich, Germany; [2]Department of Protein Evolution, Max-Planck-Institute for Developmental Biology, Spemannstr. 35, 72076 Tübingen, Germany

PDB-ID	protein name	Organism	β-strands
1AF6	Maltoporin Sucrose Complex (LamB)	*Escherichia coli*	18
1FEP	Ferric Enterobactin Receptor (FepA)	*Escherichia coli*	22
1HXX	Porin OmpF	*Escherichia coli*	16
1I78	Outer Membrane Protease (Ompt)	*Escherichia coli*	10
1K24	Outer Membrane Adhesin/Invasin (OpcA)	*Neisseria meningitidis*	10
1KMO	Outer Membrane Transporter (FecA)	*Escherichia coli*	22
1OH2	Sucrose-specific Porin	*Salmonella typhimurium*	18
1P4T	Neisserial surface protein (NspA)	*Neisseria Meningitidis*	8
1QD6	Outer Membrane Phospholipase A (OmpLA)	*Escherichia coli*	12
1QJ8	Outer Membrane Protein X (OmpX)	*Escherichia coli*	8
1QJP	Outer Membrane Protein A (OmpA)	*Escherichia coli*	8
1T16	Fatty Acid Transporter (FadL)	*Escherichia coli*	14
1THQ	Outer Membrane Enzyme (PagP)	*Escherichia coli*	8
1TLY	Bacterial nucleoside transporter (Tsx)	*Escherichia coli*	12
1UYN	Translocator Domain Of Autotransporter (Nalp)	*Neisseria Meningitidis*	12
2FCP	Ferric Hydroxamate Uptake Receptor (FhuA)	*Escherichia coli*	22
2FGQ	Porin Omp32	*Comamonas acidovorans*	16
2GUF	Cobalamin Transporter (BtuB)	*Escherichia coli*	22
2POR	Porin	*Rhodobacter capsulatus*	16
2QDZ	Omp85/TPSB transporter family protein (FhaC)	*Bordetella pertussis*	16
2VQI	P Pilus Usher Translocation Pore (PapC)	*Escherichia col*i	24
3EMN	VDAC1	*Mus musculus*	19
3PRN	Porin E1M	*Rhodopseudomonas blastica*	16

Table S1: 23 representative single-chain OMBBs used for the direct HMM-HMM comparisons in Fig. 2A. We obtained the representative sequences from the SCOP database (version 1.73, filtered for 25% sequence identity), and added the sequences of VDAC1, FhaC and PapC that were not yet contained in SCOP v1.73. The last column gives the number of β-strands forming the β-barrel.

PDB-ID	Protein name	Organism	β-strands
1AF6	Maltoporin Sucrose Complex (LamB)	*Escherichia coli*	18
1E54	Anion-selective porin (Omp32)	*Comamonas acidovorans*	16
1FW3	Outer Membrane Phospholipase A (OmpLA)	*Escherichia coli*	12
1I78	Outer Membrane Protease (Ompt)	*Escherichia coli*	10
1K24	Outer Membrane Adhesin/Invasin (OpcA)	*Neisseria meningitidis*	10
1KMO	Outer Membrane Transporter (FecA)	*Escherichia coli*	22
1NQE	Outer Membrane Cobalamin Transporter (Btub)	*Escherichia coli*	22
1P4T	Neisserial surface protein (NspA)	*Neisseria Meningitidis*	8
1PHO	Phosphoporin (Phoe)	*Escherichia coli*	16
1Q9F	Outer membrane protein (OmpX)	*Escherichia coli*	8
1T1L	Long-Chain Fatty Acid Transporter (Fadl)	*Escherichia coli*	14
1THQ	Outer Membrane Enzyme (PagP)	*Escherichia coli*	8
1TLY	Bacterial nucleoside transporter (Tsx)	*Escherichia coli*	12
1UYN	Translocator Domain Of Autotransporter (Nalp)	*Neisseria Meningitidis*	12
2ERV	Outer Membrane Enzyme (Pagl)	*Pseudomonas aeruginosa*	8
2F1V	Outer membrane protein (OmpW)	*Escherichia coli*	8
2GE4	Outer Membrane Protein A Transmembrane Domain (OmpA)	*Escherichia coli*	8
2O4V	Porin P (OprP)	*Pseudomonas aeruginosa*	16
2POR	Porin	*Rhodobacter capsulatus*	16

Table S2: 19 representative OMBBs used as starting point for transitive profile searches with HHsenser. We obtained the sequences by filtering the sequences of all bacterial OMBBs in the SCOP database (version 1.71) for a maximum of 25% pairwise sequence identity. The last column gives the number of β-strands forming the β-barrel.

GI	Protein name	Organism
121533976	TonB-dependent receptor	*Thermosinus carboxydivorans*
121534322	TonB-dependent receptor	*Thermosinus carboxydivorans*
121534812	TonB-dependent receptor	*Thermosinus carboxydivorans*
121534955	surface antigin (D15)	*Thermosinus carboxydivorans*
121534975	hypothetical protein TcarDRAFT_1330	*Thermosinus carboxydivorans*
89210169	S-layer-like region	*Halothermothrix orenii H 168*
89210191	surface antigen (D15):Surface antigen variable number	*Halothermothrix orenii H 168*
89211242	hypothetical protein HoreDRAFT_0748	*Halothermothrix orenii H 168*
13940157	tetrachloroethylene dehalogenase	*Clostridium bifermentans*

Table S3: Proteins from Gram-positive bacteria in the OmpT clustermap. All proteins belong to species in the genus *Clostridiae*, the first two of which are known to have an outer membrane (Sokolova et al. 2004; Cayol et al. 1994).

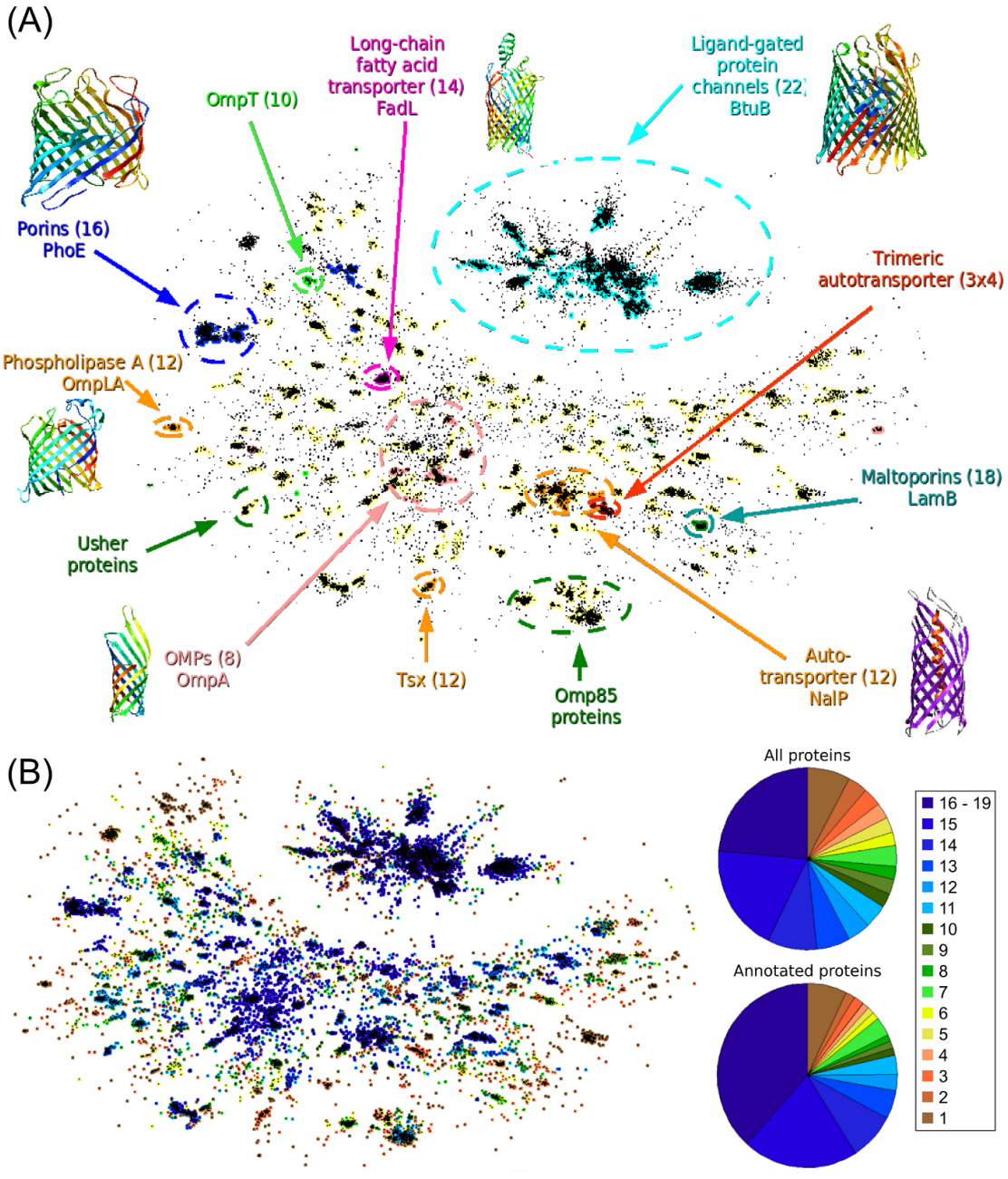

Figure S1: (A) Pooled cluster map of 21856 protein sequences found with the exhaustive transitive HHsenser search starting from 19 OMBBs with known structure (Table S2). The different colors indicate the different numbers of β-strands. (B) Most sequence clusters in our map of putative OMBBs are reliably linked to known OMBBs by multiple HHsenser searches. This map is colored according to the number of transitive searches detecting each protein. The pie charts on the right show the fractions of proteins that were detected by all 19 searches (dark blue), by some of the searches (blue to organge), down to only one search (brown). The upper chart refers to all 21856 proteins in the map, the lower chart only to the 12015 protein sequences annotated as OMPs. About two thirds of all proteins and more than 75% of all known OMPs on the map can be found in more than half of all transitive searches (blue to cyan).

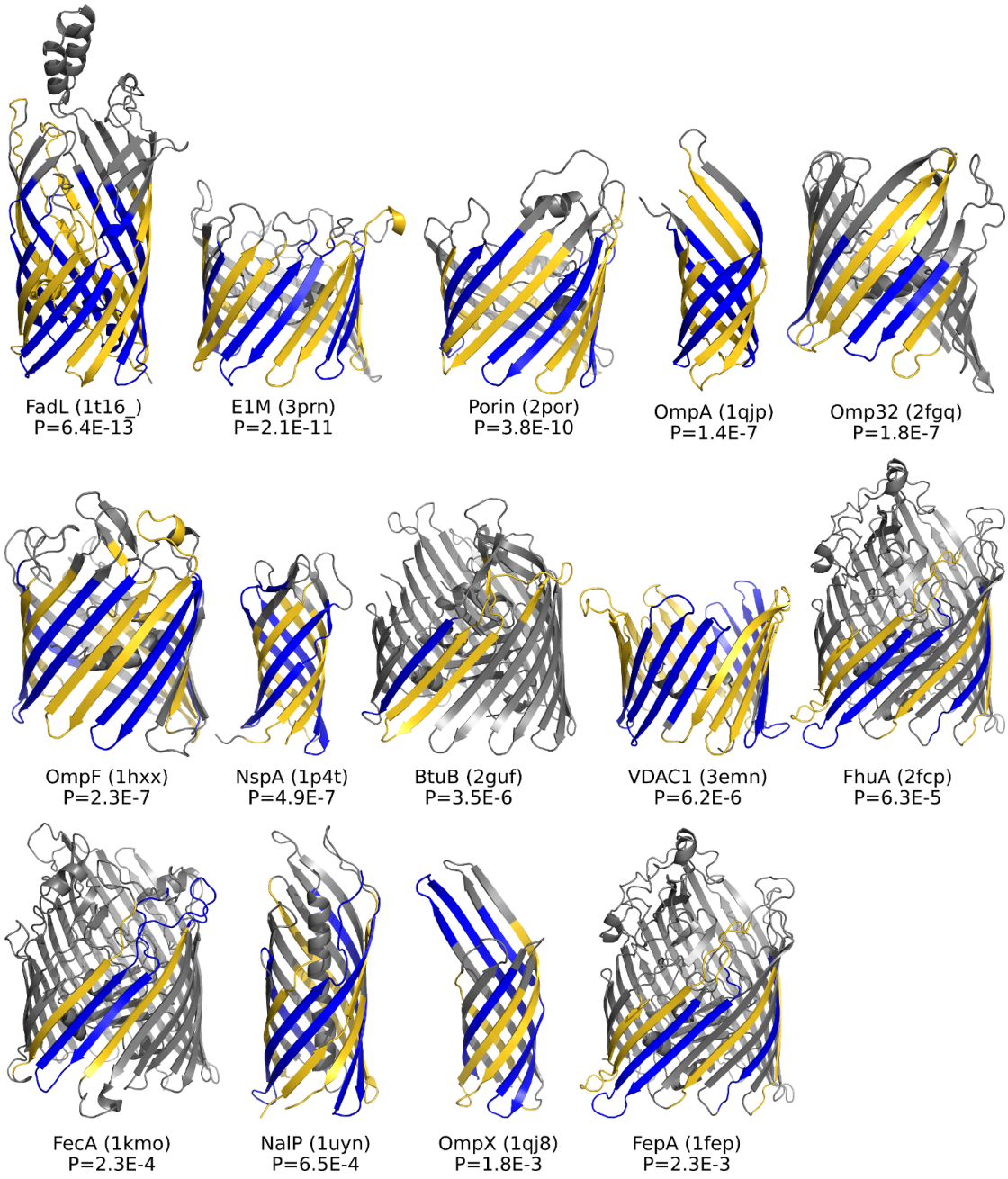

Figure S2: 3D structures of the 14 representative OMBBs for which HHrepID detects a significant off-diagonal alignment with P-value better than 10^{-2} (see Table 1 in main text). Predicted repeat units are highlighted in blue and yellow. Almost all predicted repeats coincide with a structural $\beta\beta$-hairpin. Note that the conservation of sequence repeats seems to be highest near the C-terminal, in accord with functional restraints related to membrane insertion (Robert et al. 2006).

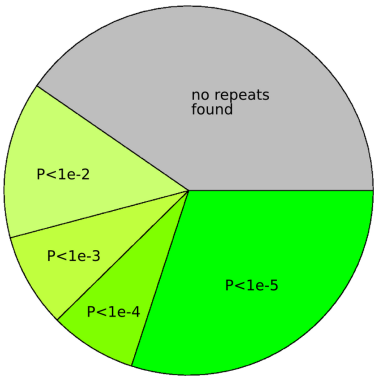

Figure S3: *De novo* repeat analysis using HHrepID on the 474 clusters of putative bacterial OMBBs from the pooled cluster map. We detect repeats with a P-value better than 10^{-2} in 281 (59%) clusters (green sections), showing that internal sequence symmetry is a general property of the majority of bacterial OMBBs.

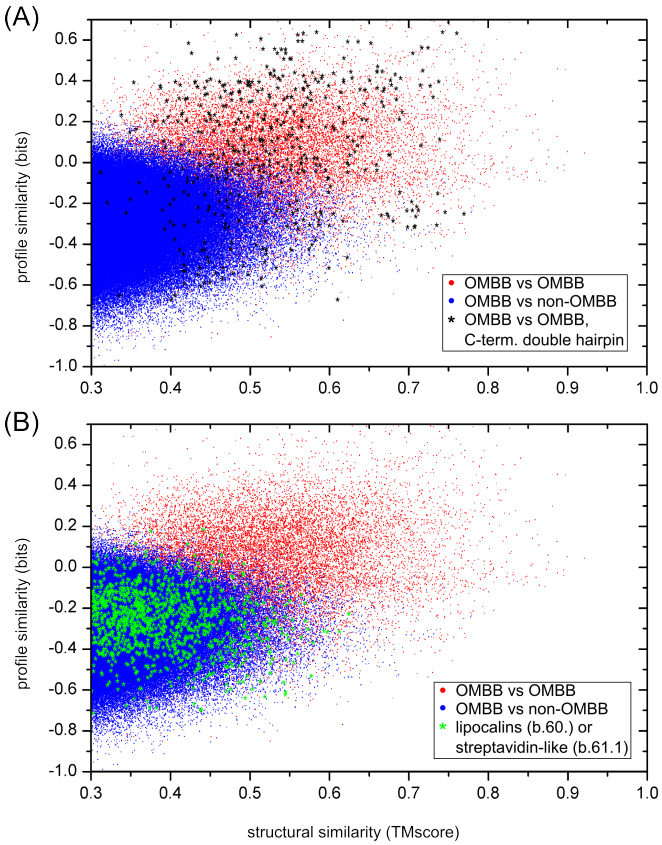

Figure S4: Profile-profile and structural similarity scores between all double hairpins from single-chain bacterial OMBBs and all other double hairpins from OMBBs (red), and between double hairpins and all proteins in the PDB minus OMBBs (blue). (A) Highlighted in black are all hits between double hairpins from OMBBs containing the last β-strand, which in most OMBBs carries a C-terminal signal sequence. Most of the hits are distributed similary as the red points, which indicates that there is only a weak influence of the functional constraints on the observed sequence similarities. (B) Same as A, but with all lipocalins and streptavidin-like proteins (SCOP IDs b.60. and b.61.1) highlighted in green. These two groups of β-barrel proteins are similar in structure to OMBBs. They lie well within the original blue distribution, implying that the differences in the red and blue distributions can not be explained through the similarities in global structural architecture among β-barrel proteins.

References

Cayol JL, Ollivier B, Patel BK, Prensier G, Guezennec J, Garcia JL. 1994. Isolation and characterization of Halothermothrix orenii gen. nov., sp. nov., a halophilic, thermophilic, fermentative, strictly anaerobic bacterium. Int J Syst Bacteriol. 44:534–540.

Robert V, Volokhina EB, Senf F, Bos MP, Van Gelder P, Tommassen J. 2006. Assembly factor Omp85 recognizes its outer membrane protein substrates by a species-specific C-terminal motif. PLoS Biol. 4.

Sokolova TG, González JM, Kostrikina NA, Chernyh NA, Slepova TV, Bonch-Osmolovskaya EA, Robb FT. 2004. Thermosinus carboxydivorans gen. nov., sp. nov., a new anaerobic, thermophilic, carbon-monoxide-oxidizing, hydrogenogenic bacterium from a hot pool of Yellowstone National Park. Int J Syst Evol Microbiol. 54:2353–2359.