HHomp - Prediction and classification of outer membrane proteins
Supplementary material

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

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

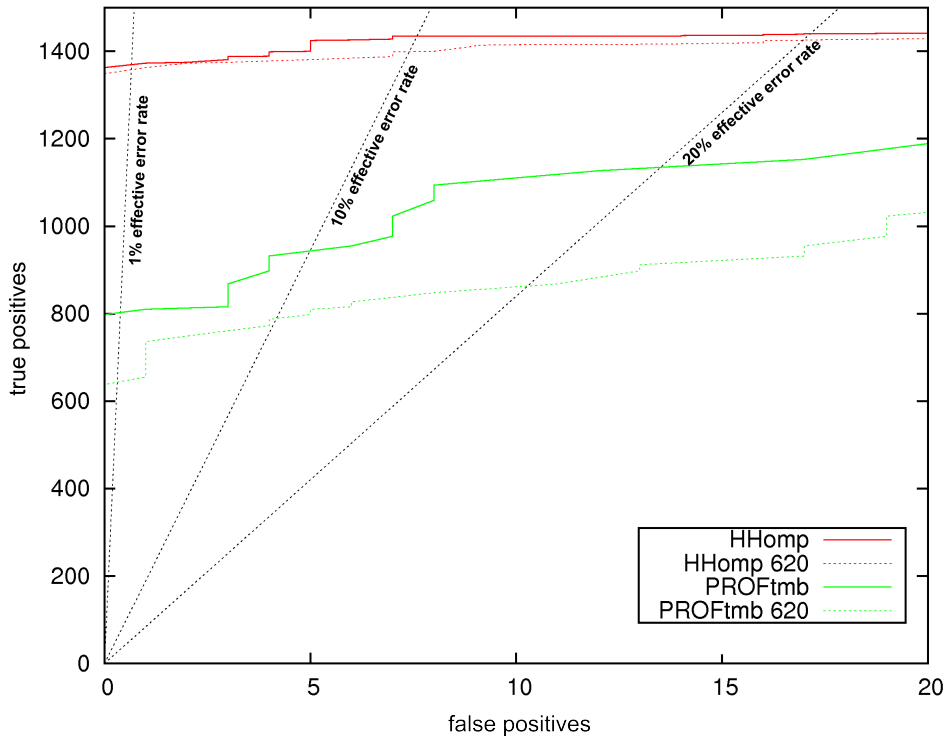

Fig. 1. ROC-plot showing the length dependence of PROFtmb. This benchmark uses as positive set 2164 OMPs from the TransportDB and two negative sets with different mean length, constructed by concatenating random sequences from the SCOP database. One negative set contains sequences with a mean length of 350 amino acids, which is equal to the mean sequence length for the SwissProt database, and the other one has a mean length of 620, which is equal to the mean length in the positive set. The sensitivity of PROFtmb decreases by $\sim 8\%$ to 25% when using the negative set with the higher mean length, whereas HHomp is not affected. The effective error rate is the number of effective false positives (FPs) divided by the sum of true positives (TPs) and effective FPs, where effective FPs are defined as FPs multiplied with 21 to obtain the same fraction of OMPs in the benchmark set as in Gram-negative genomes (2%).

GI	protein name	HHomp			PROFtmb
		Top-ranked cluster	prob.	β-strands	β-strands
16128048	organic solvent tolerance protein precursor	OMP.nn.14.2 org. solv. tol.	97.4	?	(30)
16128132	predicted outer membrane usher protein	OMP.nn.2.2 usher protein	100	?	(22)
16128143	ferrichrome outer membrane transporter	OMP.22.1.4 FhuA-like	100	22	24
16128170	hypothetical protein b0177	OMP.nn.3.1 Omp85-like	100	?	-
16128227	outer membrane phosphoporin protein E	OMP.16.1.1 PhoE-like	100	16	16
16128276	predicted aromatic compound dioxygenase	OMP.nn.2.3 usher protein	100	?	(18)
16128396	nucleoside channel	OMP.12.5.1 Tsx	100	12	12
16128516	predicted outer membrane usher protein	OMP.nn.2.2 usher protein	100	?	(24)
16128536	outer membrane porin	OMP.16.1.1. PhoE-like	100	16	16
16128548	outer membrane protease VII (OmpT)	OMP.10.1.1 OmpT	100	10	(10)
16128555	copper/silver efflux system	NodT.1 OprM	100	3×4	-
16128567	iron-enterobactin OMP transporter	OMP.22.4.2 FecA-like	100	22	22
16128605	palmitoyl transferase for Lipid A (PagP)	OMP.8.5.1 PagP	100	8	-
16128657	outer membrane porin	OMP.nn.9.1 OprD	98.9	?	(20)
16128773	predicted iron outer membrane transporter	OMP.22.1.3 FhuA-like	100	22	24
16128782	outer membrane protein X	OMP.8.3.1 OmpX	100	8	8
16128896	outer membrane porin 1a	OMP.16.1.1 PhoE-like	100	16	16
16128907	predicted outer membrane usher protein	OMP.nn.2.2 usher protein	100	?	(26)
16128924	outer membrane protein A	OMP.8.1.1 OmpA-like	100	8	(8)
16128949	predicted exopolysaccharide export protein	Wza.1	100	α-helical	-
16129065	ferric-rhodotorulic acid OM transporter	OMP.22.1.6 FhuA-like	100	22	24
16129165	predicted adhesin	OMP.12.1.4 Autotransp.	100	12	18
16129217	outer membrane protein W	OMP.8.2.1 OmpW	100	8	(8)
16129280	outer membrane porin	OMP.14.2.1 OmpG-like	100	14	14
16129338	outer membrane pore protein N, non-specific	OMP.16.1.1 PhoE	100	16	16
16129410	iron outer membrane transporter	OMP.22.4.2 FecA-like	100	22	22
16129454	predicted porin protein	OMP.22.10.3	100	22	26
16129464	hypothetical protein b1505	OMP.nn.2.2 usher protein	100	?	-
19129468	conserved protein, predicted pseudogene	OMP.12.1.4 Autotransp.	100	12	(14)
16130002	protein for capsular polysaccharide export	Wza.1	100	α-helical	-
16130047	predicted outer membrane protein	OMP.nn.2.2 usher protein	100	?	(26)
16130093	ferric iron-catecholate OM transporter	OMP.22.4.4 FecA-like	100	22	(24)
16130127	predicted autotransporter OMP	OMP.12.1.6 Autotransp.	100	12	16
16130152	outer membrane porin protein C	OMP.16.1.1 PhoE-like	100	16	16
16130168	adhesin	OMP.12.1.4 Autotransp.	100	12	28
16130277	long-chain fatty acid OMP	OMP.14.1.1 FadL-like	100	14	(16)
16131036	predicted outer membrane protein	OMP.nn.2.2 usher protein	100	?	(26)
16131106	predicted outer membrane protein	OMP.nn.2.2 usher protein	100	?	(24)
16131268	predicted fimbrial transporter	-	-	-	-
16131588	carbohydrate-specific outer membrane porin	OMP.18.1.1 LamB-like	100	18	18
16131671	outer membrane phospholipase A	OMP.12.6.1 OmpLA-like	100	12	(12)
16131715	predicted outer membrane porin L	OMP.nn.36.1	98.4	?	12
16131804	vitamin B12/cobalamin outer membrane transporter	OMP.22.4.5 FecA-like	100	22	(22)
16131862	maltoporin precursor	OMP.18.1.1 LamB-like	100	18	18
16131906	outer membrane factor of efflux pump	NodT.1 OprM	100	3×4	-
16132042	predicted OMP and surface antigen	OMP.nn.3.3 Omp85-like	100	?	-
16132112	ferric citrate outer membrane transporter	OMP.22.2.4 FhuA/FecA-like	100	22	22
16132138	outer membrane usher protein	OMP.nn.2.2 usher protein	100	?	(24)
49176177	phase-variable biofilm formation autotrans.	OMP.12.1.4 Autotransp.	100	12	20
49176211	predicted export usher protein	OMP.nn.2.2 usher protein	100	?	-
49176212	predicted export usher protein	OMP.nn.2.2 usher protein	100	?	-
49176305	predicted outer membrane usher protein	OMP.nn.2.2 usher protein	100	?	(24)
90111165	predicted outer membrane protein	OMP.nn.2.2 usher protein	100	?	(22)
90111306	predicted outer membrane porin protein	OMP.nn.37.1	94.0	?	-
90111396	predicted outer membrane protein, yohG	NodT.5	100	3×4	-
90111474	adhesin-like autotransporter	OMP.12.1.4 Autotransp.	100	12	-
90111528	outer membrane channel precursor protein	TolC.1	100	3×4	-
90111569	general secretory pathway component, cryptic	-	-	-	-
90111728	N-acetylnuraminic acid outer membrane channel	OMP.nn.36.1	98.4	?	12

Table 1
Prediction of all known outer membrane proteins in the *Escherichia coli K12* genome (see explanation below). HHomp detects 57 of these 59 proteins with $> 94\%$ probability, 52 of them have a probability of 100%. PROFtmb identifies only 23 (42) with a score > 10 (> 7).

We identify all known OMPs in the *Escherichia coli K12* genome by performing BLAST searches against the TransportDB and all OMPs in the SCOP database. All *E. coli* proteins, that match a database protein with an E-value better than 10^{-10} over 80% of its length, were considered to be OMPs. We exclude four clear false positive hits to proteins in the TransportDB. These four proteins contain only the periplasmatic domain of OMPs (gi numbers: 16128716, 16129841, 16130526, 49176370). We obtain 59 OMPs in the *E. coli* genome of which HHomp was able to predict all but two, both of them from the Secretin family (Table S1).

PDB-ID	**protein name**	**Organism**
1AF6	Maltoporin Sucrose Complex (LamB)	Escherichia coli
1E54	Anion-selective porin (Omp32)	Comamonas acidovorans
1EK9	Outer membrane protein (TolC)	Escherichia coli
1FW3	Outer Membrane Phospholipase A (OmpLA)	Escherichia coli
1I78	Outer Membrane Protease (Ompt)	Escherichia coli
1K24	Outer Membrane Adhesin/Invasin (OpcA)	Neisseria meningitidis
1KMO	Outer Membrane Transporter (FecA)	Escherichia coli
1NQE	Outer Membrane Cobalamin Transporter (Btub)	Escherichia coli
1P4T	Neisserial surface protein (NspA)	Neisseria Meningitidis
1PHO	Phosphoporin (Phoe)	Escherichia coli
1Q9F	Outer membrane protein (OmpX)	Escherichia coli
1T1L	Long-Chain Fatty Acid Transporter (Fadl)	Escherichia coli
1THQ	Outer Membrane Enzyme (PagP)	Escherichia coli
1TLY	Bacterial nucleoside transporter (Tsx)	Escherichia coli
1UUN	Main Porin (Mspa)	Mycobacteria Smegmatis
1UYN	Translocator Domain Of Autotransporter (Nalp)	Neisseria Meningitidis
2ERV	Outer Membrane Enzyme (Pagl)	Pseudomonas aeruginosa
2F1C	Monomeric Porin (Ompg)	Escherichia coli
2F1V	Outer membrane protein (OmpW)	Escherichia coli
2GE4	Outer Membrane Protein A Transmembrane Domain (OmpA)	Escherichia coli
2J58	Outer membrane lipoprotein (Wza)	Escherichia coli
2O4V	Porin P (OprP)	Pseudomonas aeruginosa
2POR	Porin	Rhodobacter capsulatus

Table 2
Representative OMPs used as starting point for HHsenser searches. These 23 sequences were obtained by filtering the sequences of all bacterial OMPs in the PDB for a maximum of 25% pairwise sequence identity.

GI number	**protein name**	**HHomp probability**	**PROFtmb score**
6324302	Essential component of the Sorting and Assembly Machinery (SAM or Tob complex) of the mitochondrial outer membrane; Sam50p	100.00	-
6324273	Mitochondrial porin (voltage-dependent anion channel)	94.48	-
6322077	Putative mitochondrial porin (voltage-dependent anion channel)	92.15	-
6323859	Component of the TOM (translocase of outer membrane) complex; Tom40p	90.19	-

Table 3
Prediction of outer membrane proteins in the genome of *Saccharomyces cerevisiae* with an HHomp probability above 90%. None of these proteins have a PROFtmb score > 7.

GI number	protein name	OMP?	HHomp probability	PROFtmb score
31542301	sorting and assembly machinery component 50 homolog	yes	100.00	-
5174723	mitochondrial outer membrane protein TOM40	yes	94.55	-
4507879	voltage-dependent anion channel 1	yes	93.84	-
42476281	voltage-dependent anion channel 2	yes	93.65	-
25188179	voltage-dependent anion channel 3	yes	92.92	-
21362030	translocase of outer mitochondrial membrane 40 homolog-like	yes	92.42	-
113412878	PREDICTED: similar to voltage-dependent anion channel 2	likely	94.50	-
88955151	PREDICTED: similar to voltage-dependent anion channel 2	likely	94.32	-
20070205	ladinin 1	unlikely	97.01	-
113412549	PREDICTED: similar to ladinin 1	unlikely	97.00	-
113423316	PREDICTED: similar to mucin 19	unlikely	94.38	-
33359213	polycystin 1 isoform 2 precursor	unlikely	93.95	-
58331146	polycystin 1 isoform 1 precursor	unlikely	93.74	-
113423720	PREDICTED: similar to mucin 19	unlikely	93.04	-
32964830	collagen, type VIII, alpha 2	unlikely	92.13	8.2
17738302	alpha 1 type VIII collagen precursor	unlikely	91.28	8
18105032	collagen, type X, alpha 1 precursor	unlikely	91.28	7.5
32895368	alpha 1 type VIII collagen precursor	unlikely	91.28	8

Table 4
Prediction of outer membrane proteins in the genome of *Homo sapiens* with an HHomp probability above 90%. Shown PROFtmb hits have a score > 7.