W394–W399 *Nucleic Acids Research, 2004, Vol. 32, Web Server issue*
DOI: 10.1093/nar/gkh351

BOMP: a program to predict integral β-barrel outer membrane proteins encoded within genomes of Gram-negative bacteria

Frode S. Berven, Kristian Flikka[1,*], Harald B. Jensen and Ingvar Eidhammer[2]

Department of Molecular Biology, [1]Computational Biology Unit, Bergen Centre for Computational Science and [2]Department of Informatics, University of Bergen, 5020 Bergen, Norway

Received December 18, 2003; Revised and Accepted February 18, 2004

ABSTRACT

This work describes the development of a program that predicts whether or not a polypeptide sequence from a Gram-negative bacterium is an integral β-barrel outer membrane protein. The program, called the β-barrel Outer Membrane protein Predictor (BOMP), is based on two separate components to recognize integral β-barrel proteins. The first component is a C-terminal pattern typical of many integral β-barrel proteins. The second component calculates an integral β-barrel score of the sequence based on the extent to which the sequence contains stretches of amino acids typical of transmembrane β-strands. The precision of the predictions was found to be 80% with a recall of 88% when tested on the proteins with SwissProt annotated subcellular localization in *Escherichia coli* K 12 (788 sequences) and *Salmonella typhimurium* (366 sequences). When tested on the predicted proteome of *E.coli*, BOMP found 103 of a total of 4346 polypeptide sequences to be possible integral β-barrel proteins. Of these, 36 were found by BLAST to lack similarity (*E*-value score < 1e−10) to proteins with annotated subcellular localization in SwissProt. BOMP predicted the content of integral β-barrels per predicted proteome of 10 different bacteria to range from 1.8 to 3%. BOMP is available at http://www.bioinfo.no/tools/bomp.

INTRODUCTION

A Gram-negative cell envelope is typically comprised of two membranes, the cytoplasmic (CM) and the outer membrane (OM) (1). Both membranes contain integral membrane proteins that generally are involved in the transport of various molecular compounds across the membranes. Nevertheless, the integral membrane proteins of the CM and OM differ greatly in structure. Integral membrane proteins of the CM consist largely of α-structures, where the membrane spanning regions are hydrophobic α-helical stretches that typically consist of 15–25 mostly non-polar amino acids (2). Several computer programs are available for the prediction of this type of membrane proteins (3), with TMHMM (4) reported to have the best performance (5). The integral OM proteins (OMPs) generally consist of β-structures, and form monomeric, dimeric or trimeric transmembrane (TM) β-barrels containing between 8 and 22 TM β-strands (6). The function of these proteins in the cell is mainly passive nutrient intake and active ion transport, but they can also serve as membrane anchors and defence against attack proteins, and a few have been characterized as enzymes (7). The integral β-barrel proteins of the OM have proven more difficult to predict than the integral CM proteins, mainly due to much shorter TM stretches of amino acids with highly variable properties (8). In general, the amino acids in the TM β-strands alternate between being polar and non-polar, with non-polar residues facing the lipid bilayer and the protein interfaces, and the polar residues pointing into the interior of the barrel. Residues pointing inwards in the barrel can also be non-polar, obstructing the regular alternation between polar and non-polar residues, making this feature less suitable to use when predicting integral β-barrel proteins (7). Recent publications have described different approaches to recognize β-barrel OMPs from polypeptide sequences (9–11), but none of the programs has been made available for public use. The only program currently available for such a task is PSORT (www.psort.org) (12), which has a rather low accuracy, whereas the new and improved PSORT B does not discriminate between integral β-barrel proteins and OM lipoproteins (13). In this article, we describe the development

*To whom correspondence should be addressed. Tel: +47 55584032; Fax: +47 55584199; Email: flikka@ii.uib.no

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

of a program named 'The β-barrel Outer Membrane protein Predictor' (BOMP), which predicts whether or not a polypeptide sequence from a Gram-negative bacterium is an integral β-barrel OMP. BOMP can scan the entire collection of predicted polypeptide sequences encoded within a bacterial genome, and generates a list in which all the predicted β-barrel proteins are categorized according to reliability of the prediction. The performance of BOMP was tested on the polypeptides from *Escherichia coli* K 12 and *Salmonella typhimurium* with annotated subcellular localization in SwissProt release 42 (14), as both these organisms contain high numbers of well-annotated integral β-barrel proteins. The entire predicted proteome of *E.coli* was also analyzed by BOMP in order to identify possible new β-barrel OMPs not previously annotated as such and without homologues with known subcellular localization.

PROGRAM COMPONENTS

The program combines two independent methods for identifying the possible integral OMPs, and a filtering mechanism to remove false positives. In addition, we have added an optional BLAST function that is not part of the prediction. This function will find polypeptide sequences that have high similarity to proteins with annotated subcellular localization in SwissProt, and either support or contradict the prediction results.

C-terminal pattern

The first method to recognize integral β-barrel proteins relies on a pattern extracted from the last 10 amino acids in the C-terminal end of 12 integral OMPs with resolved crystal structure and less than 70% conserved residues (Table 1). The last TM β-strand of these sequences was at the far C-terminal end, with an aromatic amino acid, most often phenylalanine, in the last position (30). Therefore, the pattern extracted from these sequences had an aromatic residue as the last position. In the positions pointing inwards in the barrel, all amino acids except cysteine were allowed. In the positions pointing towards the membrane, the amino acids YFWKLHVITMAD were included in the pattern. In addition, the pattern had to match at the far C-terminal end of a sequence with a minimal length of 110 amino acids in order to give a valid hit. This is a relatively safe length criterion, as OmpX of *E.coli* with 171 amino acids are considered to be a very short integral β-barrel protein (8). The resulting pattern was .{100,} [^C] [YFWKLHVITMAD] [^C] [YFWKLHVITMAD] [^C] [YFWKLHVITMAD] [^C] [YFWKLHVITMAD] [^C] [FYW].

Integral β-barrel score

The second method is based on the data and algorithm given by W. C. Wimley (10) in which he identified the membrane interacting surfaces of 15 nonredundant integral β-barrel OMPs with resolved crystal structure (Table 1). From this information, he calculated the abundance of each amino acid in the external and internal positions of the membrane spanning segments, relative to the genomic abundance. We utilized this normalized amino acid distribution to score a 10-residue sliding window by taking the maximum of two scores: the scores obtained when summarizing the amino acids in the window starting with either an internal or external amino acid, as described by Wimley (10). In order to obtain a total sequence score (integral β-barrel score) from the score of each window, we used the average of the eight highest-scoring non-overlapping windows. We observed that the integral β-barrel proteins in general had less low-scoring windows than other protein types. The average of the 12 lowest scoring non-overlapping windows was therefore added to the integral β-barrel score. Several different numbers of high- and low-scoring windows to be included in the integral β-barrel score were tested, but the figures used above were found to give the best predictive power in this program component. Proteins with integral β-barrel score above an empirically found threshold were considered to be possible β-barrel OMPs. In addition, we defined a high integral β-barrel score limit which included very few false positives. This high scoring limit was used in the categorization of the predicted β-barrel OMPs, which indicates the likelihood of the prediction being correct.

Amino acid distribution filter

In order to limit the number of wrongly predicted integral OMPs (false positives), we developed a final filtering procedure. The proteins that have the C-terminal pattern or a significant integral β-barrel score are compared to a reference set

Table 1. Proteins used to generate the C-terminal amino acid pattern (the first 12), and the proteins used as a basis for the integral β-barrel score calculation (all)

SwissProt entry	Organism	C-terminal β-strand	β-strands	PDB code	Reference
PORI_RHOBL	*Rhodopseudomonas blastica*	VADVGVRFDF	16	1PRN	(15)
SCRY_SALTY	*S.typhimurium*	SFGVQMETWF	18	1AOT	(16)
OM32_COMAC	*Comamonas acidovorans*	GVQVGIRHAF	16	1E54	(17)
OMPX_ECOLI	*E.coli*	TWIAGVGYRF	8	1QJ9	(18)
PORI_RHOCA	*Rhodobacter capsulatus*	VADLGVKFKF	16	2POR	(19)
OMPF_ECOLI	*E.coli*	TVAVGIVYQF	16	1OPF	(20)
LAMB_SALTY	*Salmonella typhimurium*	TFGAQMEIWW	18	2MPR	(21)
FHUA_ECOLI	*E.coli*	QVVATATFRF	22	1BY5	(22)
FEPA_ECOLI	*E.coli*	TWYMSVNTHF	22	1FEP	(23)
OMPC_KLEPN	*Klebsiella pneumoniae*	VVALGLVYQF	16	1OSM	(24)
PA1_ECOLI	*E.coli*	GVGVMLNDLF	12	1QD6	(25)
PHOE_ECOLI	*E.coli*	IVAVGMTYQF	16	1PHO	(26)
OMPA_ECOLI	*E.coli*	—	8	1BXW	(27)
TOLC_ECOLI	*E.coli*	—	4	1EK9	(28)
HLA_STAAU	*Staphylococcus aureus*	—	2	7AHL	(29)

of proteins by considering the relative abundances of two amino acids. The reference set was collected from SwissProt release 42, where all Gram-negative bacterial proteins with subcellular localization 'Integral membrane protein. outer membrane' or 'Outer membrane.' were considered to be true integral OMPs. Polypeptide sequences annotated with a subcellular localization not related to the OM were also added, to constitute a potential false positive group. This set was reduced so that all proteins were ensured to have less than 40% sequence identity to each other, leading to a reduced reference set containing 1231 sequences, of which 110 have OM as localization (available at http://www.bioinfo.no/tools/bomp). This sequence collection was further reduced to include only the proteins that were predicted to be integral OMPs by the C-terminal pattern or the integral β-barrel score. This is the final reference set used in the filtering process, containing both real integral OMPs and some proteins with other localization (available at http://www.bioinfo.no/tools/bomp). When searching for a good discriminator between the true positives and the false positives, we considered the relative abundances of all single amino acids, all pairs of amino acids (di-peptides) and a number of different groups of amino acids (hydrophobic, hydrophilic, charged etc.). By using principal component analysis (31) we were able to visualize the discriminative power of different subgroups of features, and then identify candidate subgroups of features to use as discriminators. We chose to use the relative abundance of Asparagine and Isoleucine as they gave the best separation between true and false positives in the reduced reference set containing 1231 sequences. In Figure 1, the proteins of the final reference set are shown in a plot with relative abundance values on the coordinate axis. When a protein is run through the filter, we compare it to the final reference set by using a k-nearest-neighbour method with $k = 5$ (33) to determine if the candidate is a true integral OMP. This is depicted in Figure 1, where the circle contains a candidate integral OMP and the five closest neighbours in the final reference set.

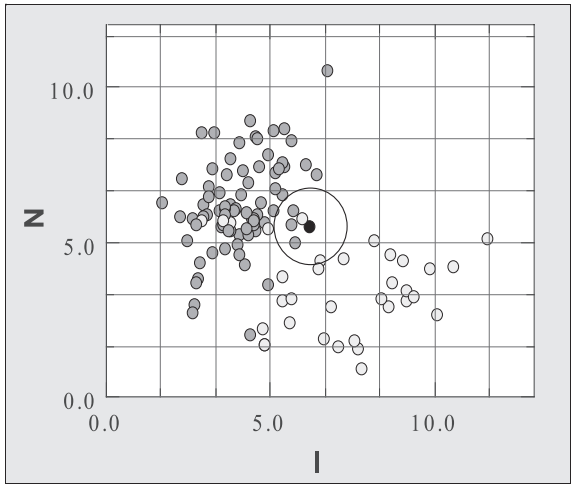

Figure 1. Relative amino acid abundance values for the reference set. Dark spots are integral OMPs, and light spots are other proteins. The black spot inside the circle represents a protein whose five closest neighbours are inside the circle. Three are integral OMPs and two are other proteins, thus the unknown candidate is predicted to be an integral OMP. Figure created with J-Express (32).

BLAST

As a supplement to the prediction methods outlined above, we added the possibility to include an automated BLAST search to be performed on the input sequences. We made a database containing the 10 618 Gram-negative polypeptide sequences in SwissProt release 42 with given subcellular localization, or those annotated with similarity to sequences with known localization (available at http://www.bioinfo.no/tools/bomp). Sequences with probable, possible or putative localization were excluded from the database. The input sequence is used to search against this database using BLAST in order to find the highest-scoring alignment with an *E*-value above 1e−10 (this value can be modified by the user, but 1e−10 is set as default) and a length of between 80% and 120% of the input sequence (13). The localization of the best database hit will either support or contradict the result from the prediction part of BOMP, and provide additional information about the input sequence to the user.

Dividing the predicted sequences into categories

The sequences predicted to be integral β-barrel proteins are divided into five categories in order to give the user additional information about the reliability of the predictions (Table 2). The categories are named 1–5 in the output, and the probability of being a β-barrel protein increases with higher category numbers. If a sequence is not found by the predictions, but BLAST detects a possible homologue with localization in the OM, this sequence will appear in the output as category 0. If a polypeptide sequence is predicted to be an integral β-barrel protein in category 1–5, and the best BLAST hit has localization other than (integral) OM, a conflict will be reported in the output as a star beside the category number.

EVALUATION OF PERFORMANCE

In order to evaluate the performance of BOMP, we extracted all the polypeptide sequences of *E.coli* and *S.typhimurium* with known subcellular localization annotated in SwissProt release 42, excluding those with possible, probable or putative localization. All proteins with subcellular localization 'Integral membrane protein. outer membrane' or 'Outer membrane.', including those annotated with similarity to such proteins, were considered as integral β-barrel OMPs. The two protein collections obtained from *E.coli* and *S.typhimurium* were separately run through BOMP after removing all the sequences with more than 40% sequence identity to any sequence in the test set under consideration from the reference set used in the filtering process. The accuracy was measured to an average of 88% recall [true hits/(true hits + false

Table 2. Dividing the predicted integral β-barrel proteins into categories based on aprobability of correct prediction

Category number	Pattern match	Integral β-barrel score
5	Yes	High
4	No	High
3	Yes	Above limit
2	Yes	Below limit
1	No	Above limit

Table 3. Evaluation of the performance of BOMP

Organism	Sequences	Integral β-barrel proteins	TP	FP	FN	Recall (%)	Precision (%)	MCC
E.coli	788	40	34	12	6	85	73.9	0.781
S.typhimurium	366	19	18	1	1	94.7	94.7	0.944
Total	1154	59	52	13	7	88	80	0.831

True positives (TP), false positives (FP) and false negatives (FN).

Table 4. Measuring the performance of the individual components of BOMP

	Recall (%)	Precision (%)	MCC
E.coli			
Pattern	52.5	63.6	0.558
Integral β-barrel score	80.0	72.7	0.749
Filter[a]	100	73.9	0.718
S.typhimurium			
Pattern	57.9	64.7	0.592
Integral β-barrel score	89.5	77.3	0.822
Filter[a]	100	94.7	0.918
Total			
Pattern	54.2	64.0	0.569
Integral β-barrel score	83.1	74.2	0.773
Filter[a]	100	80.0	0.696

[a]The performance of the filter is measured on the sequences predicted as integral OMPs by the Pattern and/or the Integral β-barrel score.

negatives)], and 80% precision [true hits/(true hits + false positives)]. Matthews correlation coefficient (MCC) is another measure of performance, which accounts for both over- and under-prediction. MCC is represented by the formula $MMC = (pn-ou)/\{\sqrt{[(p+o)(p+u)(n+o)(n+u)]}\}$, where p and n represent true positives and negatives, respectively, while o and u are the number of false positives and negatives, respectively. Average performance calculated by this formula was 0.831 (Table 3). The performance of the separate prediction features in BOMP is given in Table 4. The categories assigned to the BOMP-predicted β-barrel proteins help to determine the reliability of each prediction. During the testing outlined above, all the falsely predicted β-barrel proteins were assigned as category 1 or 2 proteins. After analysing the predicted proteomes of 10 different bacteria, none of the sequences predicted in category 4 and 5 was found to have homologues with localization other than the OM (besides 'cell wall' and 'secreted' autotransporters), indicating very high reliability of the predicted proteins assigned to these two categories. During the testing of the program, we observed that polypeptide sequences with less than eight transmembrane β-strands, e.g. sequences in the TolC family, were a group generally not predicted by BOMP. In general, this group of sequences obtained low sequence score and did not end with a TM β-strand. The BLAST function is not a prediction component in BOMP, and was not used when measuring the accuracy of the program.

RESULTS

The whole predicted proteome of *E.coli* (http://us.expasy.org/sprot/hamap/) was analysed using BOMP, predicting 91 integral β-barrel proteins. In addition, 12 sequences not predicted to be integral β-barrel proteins were found to have similarity to proteins localized in the outer membrane when the additional BLAST function was used. Of these 103 possible *E.coli* integral β-barrel proteins found by BOMP, 67 were either previously annotated with (integral) OM as the subcellular localization in SwissProt or found to have similarity to such proteins by BLAST. Seven of the predicted proteins were found to possibly have localization other than OM, but might not be false positives since the best BLAST hits of five of them were annotated with the localizations membrane-associated, secreted or cell wall in SwissProt. The remaining 36 predicted β-barrel proteins were not found to have similar sequences in our BLAST database, and could be previously undiscovered integral β-barrel OMPs in *E.coli*. Eleven of the 36 proteins were also predicted to be OMPs by Casadio *et al*. (9) using their Hunter program (not publicly available). However, none of these sequences was predicted as an OMP by PSORT-B, as all the *E.coli* sequences predicted as OMPs by this program had similarity to sequences with OM localization in their BLAST database. BOMP found that the predicted proteomes of 10 different Gram-negative bacteria contained between 1.8% and 3% β-integral barrel proteins (http://www.bioinfo.no/tools/bomp). These results are in line with the 1.5–2.4% OMPs per genome predicted from nine different bacteria by Hunter (9).

DISCUSSION

In this article, we describe the development of the first publicly available program that predicts with good accuracy the integral β-barrel OM proteins from a collection of polypeptide sequences from Gram-negative bacteria. The development of a reliable program to perform this task has previously proven to be a bottleneck in the area of TM protein prediction (3). The most common way to identify integral β-barrel proteins from predicted proteomes has so far been the use of annotation information in addition to PSORT I (34,35). PSORT I, with a precision of 65.3% and recall of 54.5% in the prediction of all types of OMPs, was recently replaced by a new and improved version, PSORT B, with a reported recall of 90.3% and precision of 98.8%. PSORT B does not, however, separate the integral β-barrel proteins from the lipoproteins. When examining all the PSORT B-predicted OMPs from *E.coli* and six other precomputed genomes (*Helicobacter pylori J99*, *S.typhimurium*, *Haemophilus influenzae*, *Fusobacterium nucleatum*, *E.coli O157:H7 Sakai* and *Xanthomonas campestris*), we found that all the predicted OMPs were recognized by the PSORT B BLAST module. No additional sequences without known homologues were predicted by the other program modules. This indicates that PSORT B will probably have little chance of identifying novel OMPs without already-known homologues. At least three other

programs for integral β-barrel prediction have been developed over the last couple of years, Hunter (9), the β-barrel finder (11) and a simple algorithm developed by Wimley (10). Hunter is mainly based on signal sequence prediction, and a predictor of topography to recognize all-β-membrane proteins, whereas the β-barrel finder is based on secondary structure predictions together with hydropathy and amphipathicity information. Wimley developed a simple algorithm to calculate the β-barrel score of sequences based on the relative abundance of amino acids in the TM β-strands of 15 different integral β-barrel proteins with known crystal structures (10). Unfortunately, none of these programs has been available for performance testing, and Hunter is the only one to report its accuracy, with a recall of 82.4% and a precision of 90.3% for the prediction of well-annotated integral β-barrel proteins in *E.coli*. This is slightly poorer recall, but higher precision, than BOMP. Unlike Hunter, BOMP is not based on signal sequence prediction, giving BOMP an advantage when it comes to predicting integral β-barrel proteins from translated open reading frames since in some cases they can have been given the wrong start site, which might lead to difficulties in signal sequence prediction.

From the discussion outlined above, it is obvious that BOMP will close a gap in the collection of currently available prediction tools for TM proteins. This program will provide fast and reliable information for the experimental analysis of β-barrel OMPs. When analysing a predicted proteome with BOMP, the resulting overview of the predicted integral β-barrel OM subproteome will provide important information on how to approach the experimental proteomic work, and will speed up the experimental analysis of integral β-barrel proteins in the laboratory. Due to the good prediction accuracy, several previously hypothetical annotated polypeptide sequences can now be given a likely localization, which will narrow down possible function(s) of these proteins. An overview of the predicted integral β-barrel subproteome will also narrow down the number of proteins to be selected for experimental investigation with respect to identifying proteins that might serve as vaccine candidates in pathogenic bacteria. BOMP also opens up the possibility of comparing the predicted integral β-barrel subproteome of two different strains of the same bacterium, in order to find differences that might explain pathogenesis of one of the strains.

ACKNOWLEDGEMENTS

We wish to thank Johan R. Lillehaug for critically reading through the manuscript, Bjarte Dysvik for help with graphics and Trond Hellem Bø for useful discussions on statistical topics. Lillehaug is at the Department of Molecular Biology, Dysvik and Bø are at the Department of Informatics, University of Bergen. This work was supported in part by grants from the Norwegian Research Council [SUP 140785/420 (GABI); FUGE/CBU–151899/ISO], and the Meltzer Foundation, University of Bergen.

REFERENCES

1. Lugtenberg,B. and Van Alphen,L. (1983) Molecular architecture and functioning of the outer membrane of *Escherichia coli* and other gram-negative bacteria. *Biochim. Biophys. Acta*, **737**, 51–115.
2. Santoni,V., Molloy,M. and Rabilloud,T. (2000) Membrane proteins and proteomics: un amour impossible? *Electrophoresis*, **21**, 1054–1070.
3. Chen,C.P. and Rost,B. (2002) State-of-the-art in membrane protein prediction. *Appl. Bioinf.*, **1**, 21–35.
4. Sonnhammer,E.L.L., von Heijne,G. and Krogh,A. (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. In Glasgow,J., Littlejohn,T., Major,F., Lathrop,R., Sankoff,D. and Sensen,C. (eds), *Proceedings Sixth International Conference Intelligence Systems Molecular Biology* AAAI Press, Menlo Park, CA, pp. 176–182.
5. Moller,S., Croning,M.D.R. and Apweiler,R. (2001) Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*, **17**, 646–653.
6. Tamm,K. L., Arora,A. and Kleinschmidt,H. J. (2001) Structure and assemly of beta-barrel membrane proteins. *J. Biol. Chem.*, **276**, 32399–32402.
7. Schulz,G.E. (2000) Beta-barrel membrane proteins. *Curr. Opin. Struct. Biol.*, **10**, 443–447.
8. Koebnik,R., Locher,K.P. and Van Gelder,P. (2000) Structure and function of bacterial outer membrane proteins: Barrels in a nutshell. *Mol. Microbiol.*, **37**, 239–253.
9. Casadio,R., Fariselli,P., Finocchiaro,G. and Martelli,P.L. (2003) Fishing new proteins in the twilight zone of genomes: the test case of outer membrane proteins in *Escherichia coli* K12, *Escherichia coli* O157:H7, and other Gram-negative bacteria. *Protein Sci.*, **12**, 1158–1168.
10. Wimley,C.W. (2002) Towards genomic identification of beta-barrel membrane proteins: composition and architecture of known structures. *Protein Sci.*, **11**, 301–312.
11. Zhai,Y. and Saier,H.M.J. (2002) The beta-barrel finder (BBF) program, allowing identification of outer membrane beta-barrel proteins encoded within prokaryotic genomes. *Protein Sci.*, **11**, 2196–2207.
12. Nakai,K. and Kanehisa,M. (1991) Expert system for predicting protein localization sites in gram-negative bacteria. *Proteins*, **11**, 95–110.
13. Gardy,L.J., Spencer,C., Wang,K., Ester,M., Tusnády,E.G., Simon,I., Hua,S., deFays,K., Lambert,C., Nakai,K. *et al.* (2003) PSORT-B: improving protein subcellular localization prediction for Gram-negative bacteria. *Nucleic Acids Res.*, **31**, 3613–3617.
14. Boeckmann,B., Bairoch,A., Apweiler,R., Blatter,M.-C., Estreicher,A., Gasteiger,E., Martin,M.J., Michoud,K., O'Donovan,C., Phan,I. *et al.* (2003) The Swiss-Prot protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.*, **31**, 365–370.
15. Kreusch,A. and Schulz,G.E. (1994) Refined structure of the porin from *Rhodopseudomonas blastica*. Comparison with the porin from *Rhodobacter capsulatus*. *J. Mol. Biol.*, **243**, 891–905.
16. Forst,D., Welte,W., Wacker,T. and Diederichs,K. (1998) Structure of the sucrose-specific porin ScrY from *Salmonella typhimurium* and its complex with sucrose. *Nat. Struct. Biol.*, **5**, 37–46.
17. Zeth,K., Diederichs,K., Welte,W. and Engelhardt,H. (2000) Crystal structure of Omp32, the anion-selective porin from *Comamonas acidovorans*, in complex with a periplasmic peptideat 2.1 A resolution. *Structure*, **8**, 981–992.
18. Vogt,J. and Schulz,G.E. (1999) The structure of the outer membrane protein Ompx from *Escherichia coli* reveals mechanisms of virulen. *Structure*, **7**, 1301–1309.
19. Weiss,M.S. and Schulz,G.E. (1992) Structure of porin refined at 1.8 A resolution. *J. Mol. Biol.*, **227**, 493–509.
20. Cowan,S.W., Garavito,R.M., Jansonius,J.N., Jenkins,J.A., Karlsson,R., Konig,N., Pai,E.F., Pauptit,R.A., Rizkallah,P.J. and Rosenbusch,J.P. (1995) The structure of OmpF porin in a tetragonal crystal form. *Structure*, **3**, 1041–1050.
21. Meyer,J.E., Hofnung,M. and Schulz,G.E. (1997) Structure of maltoporin from *Salmonella typhimurium* ligated with a nitrophenyl-maltotrioside. *J. Mol. Biol.*, **266**, 761–775.
22. Locher,K.P., Rees,B., Koebnik,R., Mitschler,A., Moulinier,L., Rosenbusch,J.P. and Moras,D. (1998) Transmembrane signaling across the ligand-gated FhuA receptor: crystal structures of free and ferrichrome-bound states reveal allosteric changes. *Cell*, **95**, 771–778.
23. Buchanan,S.K., Smith,B.S., Venkatramani,L., Xia,D., Esser,L., Palnitkar,M., Chakraborty,R., van der Helm,D. and Deisenhofer,J. (1999) Crystal structure of the outer membrane active transporter FepA from *Escherichia coli*. *Nat. Struct. Biol.*, **6**, 56–63.
24. Dutzler,R., Rummel,G., Alberti,S., Hernandez-Alles,S., Phale,P., Rosenbusch,J., Benedi,V. and Schirmer,T. (1999) Crystal structure and

functional characterization of Ompk36, the osmoporin of *Klebsiella pneumoniae*. *Structure*, **7**, 425–434.
25. Snijder,H.J., Ubarretxena-Belandia,I., Blaauw,M., Kalk,K.H., Verheij,H.M. and Egmond,M.R. (1999) Structural evidence for dimerization-regulated activation of an integral membrane phospholipase. *Nature*, **401**, 717–721.
26. Cowan,S.W., Schirmer,T., Rummel,G., Steiert,M., Ghosh,R., Pauptit,R.A., Jansonius,J.N. and Rosenbusch,J.P. (1992) Crystal structures explain functional properties of two *E. coli* porins. *Nature*, **358**, 727–733.
27. Pautsch,A. and Schulz,G.E. (1998) Structure of the outer membrane protein A transmembrane domain. *Nat. Struct. Biol.*, **5**, 1013–1017.
28. Koronakis,V., Sharff,A.J., Koronakis,E., Luisi,B. and Hughes,C. (2000) Structure of the bacterial membrane protein Tolc central to multidrug efflux and protein export. *Nature*, **405**, 914–919.
29. Song,L., Hobaugh,M.R., Shustak,C., Cheley,S., Bayley,H. and Gouaux,J.E. (1996) Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science*, **274**, 1859–1856.
30. Struyve,M., Moons,M. and Tommassen,J. (1991) Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein. *J. Mol. Biol.*, **218**, 141–148.
31. O'Connel,M.J. (1974) Search program for significant variables. *Comput. Phys. Commun.*, **8**, 49–55.
32. Dysvik,B. and Jonassen,I. (2001) J-Express: exploring gene expression data using Java. *Bioinformatics*, **17**, 369–370.
33. Ripley,B.D. (1996) Nearest neighbour methods. *Pattern Recognition and Neural Networks*. Cambridge University Press, pp. 191–201.
34. Chakravarti,D.N., Fiske,M.J., Fletcher,L.D. and Zagursky,R.J. (2001) Application of genomics and proteomics for the identification of bacterial gene products as potential vaccine candidates. *Vaccine*, **19**, 601–612.
35. Phadke,N.D., Molloy,M.P., Steinhoff,S.A., Ulintz,P.J., Andrews,P.C. and Maddock,J.R. (2001) Analysis of the outer membrane proteome of *Caulobacter crescentus* by two-dimensional electrophoresis and mass spectrometry. *Proteomics*, **1**, 705–720.