FOR THE RECORD

Architecture of β -barrel membrane proteins: Analysis of trimeric porins

K. SESHADRI, ROBERT GAREMYR, ERIK WALLIN, GUNNAR VON HEIJNE,
AND ARNE ELOFSSON

Department of Biochemistry, Stockholm University, S-106 91 Stockholm, Sweden

(RECEIVED February 25, 1998; ACCEPTED May 15, 1998)

Abstract: We have analyzed the known three-dimensional structures of trimeric porins from bacterial outer membranes. The distribution of surface-exposed residues in a direction perpendicular to the membrane is similar to that in helical membrane proteins, with aliphatic residues concentrated in the central 20 Å of the bilayer. Outside these residues is a layer of aromatic residues, followed by polar and charged residues. Residues in the trimer interface are more conserved than residues not in the interface. By comparing the interface and noninterface residues, an interface preference scale has been derived that may be used as a basis for predicting interface surfaces in monomer models.

Keywords: β -barrel; membrane protein; porin; three-dimensional structure

As far as is known from high-resolution structural data, integral membrane proteins belong either to the helix-bundle or the β -barrel class (von Heijne, 1997). In both cases, hydrophobic amino acids provide the interface to the surrounding lipid bilayer, but, beyond this, the two architectures are very different. Thus, while transmembrane α -helices are more or less uniformly hydrophobic (Bowie, 1997; Wallin et al., 1997), transmembrane β -strands tend to be amphiphilic, with polar and charged residues projecting into a central pore. Moreover, most transmembrane α -helices are believed to be individually stable in the bilayer (Hunt et al., 1997; Popot & Engelman, 1997) whereas individual transmembrane β -strands certainly are not. Their respective modes of membrane insertion are thus very different (von Heijne, 1997).

All β -barrel outer membrane proteins (porins), for which high-resolution structures are available form oligomers, and, thus, bury part of their outer surface in a protein-protein interface (Fig. 1). In contrast to water-soluble proteins, little is known about the factors that drive the association between protein subunits in a lipid environment. To provide a basis for understanding general aspects of membrane protein structure, we have analyzed the known high-

resolution porin structures in terms of which kinds of residues are exposed to the lipid environment, buried within monomers, exposed to the central pore and buried between subunits. In particular, we show that subunit interfaces can be roughly identified on the basis of amino acid composition and residue conservation, suggesting that it may be possible to extend current methods for predicting transmembrane β -strands and overall monomer fold (Schirmer & Cowan, 1993; Gromiha et al., 1997) to also include the oligomeric structure.

Results and discussions: Table 1 presents some basic statistics for the proteins used in this study. The porins of known structure can be divided into three groups according to the number of β -strands forming the pore. Porins from *Rhodopseudomonas blautica* (1prn) and from *Rhodobacter capsulatus* (2por and 3por, two different crystal forms of the same protein) both have 16 strands and are of similar size. Further they show a weak sequence identity (28%). The matrix and phosphoporins from *Escherichia coli* (2omf and 1pho) have 17 strands and 62% sequence identity. Finally the LamB porin from *E. coli* (1mal) has 18 strands and is significantly larger than the others.

Residue classes: To describe the porin structures in more detail, a number of different residue classes were defined (see Materials and methods). The basic distinction is between residues in the barrel β -strands and in the loops between the strands. β -Strand residues were subdivided into those facing away from the pore (outside residues) and those facing toward the pore (inside residues). The “outside” residues were further subdivided into residues that are in the trimer interface (interface residues) and those that are not (noninterface residues). Finally, all β -strand residues were classified as either exposed or buried on the basis of their solvent accessibility in the monomer structure. The so-called constriction loop that projects into the pore itself was included in the barrel β -strand class (and thus, also, in the “inside” class).

Alignment to the membrane-normal: Porins differ significantly in size and shape. Due to these structural differences, it is not entirely

Reprint requests to: Arne Elofsson, Department of Biochemistry, Stockholm University, S-106 91 Stockholm, Sweden; e-mail: arne@biokemi.su.se.

Fig. 1. Molscript (Kraulis, 1991) view of the 2por trimer along the three-fold axis.

trivial to place all porins on a common z -axis normal to the membrane. As detailed in Materials and methods, we chose to define the $z = 0$ plane on the basis of the respective hydrophobicity profiles calculated for the “outside” β -strand residues that belong to the “exposed” class (i.e., β -strand residues facing away from the pore and exposed on the monomer surface).

In Figure 2, the average hydrophobicity of these residues is plotted against the z -axis after alignment. There is a central hydrophobic region of the residues facing the outside of the pore in the porins of approximately the same width, 30 Å, as the outer

Fig. 2. The average hydrophobicity of exposed β -strand residues facing the outside of the monomer is shown for the five different porins. The structures are overlapped so that the centers of the porin membrane region are aligned at $Z = 0$. The membrane region is defined to be between the two outermost peaks with an average hydrophobicity larger than 0.5. The negative side corresponds to the periplasmic interface. The average hydrophobicity was calculated in 1 Å slices cut out in the plane of the membrane with smoothing by point, hat-shaped running window averaging.

Table 1. Comparison of porins for which three-dimensional structures are available

Protein	2omf	1pho	1prm	2por	3por	1mal
Source	<i>E. coli</i> OMP matrix porin	<i>E. coli</i> phosphoporphin	<i>R. blastic</i>	<i>R. capsulatus</i> FormB	<i>R. capsulatus</i> FormA	<i>E. coli</i> Lamb
Resolution (Å)	2.4	3.0	2.0	1.8	2.5	3.1
Secondary structure information	DSSP	DSSP	PDB	PDB	PDB	PDB
Residues per monomer	240	328	289	301	301	421
Number of β -strands	17	17	16	16	16	18
Residues in β -strands	197 (58%)	181 (55%)	169 (58%)	178 (59%)	178 (59%)	220 (52%)
% of total accessible surface area in trimer interface	18	18	18	17	18	19
Sequence similarities (% identity)	1pho (62)	2omf (62)	2por, 3por (28)	3por (100), 1prm (28)	2por (100), 1prm (28)	None
Misc					Not used in this study since identical to 2por	

membrane in a bacteria. As the β -strands do not project very far into the periplasmic space, the periplasmic end of the hydrophobic region more or less coincides with the physical end of the β -strands approximately 10–15 Å from the central $z = 0$ plane, whereas in the extracellular domain where the β -strands project further, the hydrophobic region is followed by a hydrophilic region 15 to 20 Å away from the center of the membrane.

Residue distributions along the membrane normal: The distribution of different kinds of residues in the “outside” class along the z -axis is shown in Figure 3. In the central ± 10 Å region, there is a high concentration of aliphatic residues. The well-known “aromatic belts” (Weiss et al., 1991a) are clearly evident approximately ± 10 Å from the membrane center. The parts protruding outside the membrane are rich in charged and hydrophilic residues.

The amino acid distributions for the “outside” residues are similar to the amino acid distributions along the membrane normal in helical membrane proteins (Wallin et al., 1997), but with some notable differences. Thus, the central hydrophobic region is generally somewhat narrower and richer in aliphatic residues in the porins, and the aromatic belts are much more pronounced and also closer to each other.

As expected, the “inside” residues have a different overall amino acid composition (Fig. 3), with more charged and fewer hydrophobic residues. Acidic and basic residues have different distribution profiles, with peaks of acidic residues at the ends of the pore, i.e., ± 15 Å away from the membrane center.

Residues in the trimer interface are more conserved than residues not in the interface: Multiple sequence alignments and residue variability calculations can be used to identify residues that are conserved across a family of sequences. In helical membrane-proteins and in globular proteins, residues exposed to the surrounding are less conserved than residues buried in the protein interior (Wallin et al., 1997). All sequences with more than 35% sequence identity to one of the porins analyzed here were extracted from the TREMBL and SWISS-PROT databases (Bairoch & Apweiler, 1996), and multiple alignments were constructed using Clustal W (Thompson et al., 1994). The information content in each position in each alignment was calculated as described in Materials and methods, and averaged over the different residue classes in the five proteins. The overall results are shown in Figure 4. Residues in the β -strands are more conserved than residues in the loops, as has been reported before (Agterberg et al., 1989). The “inside” residues facing into the pore are slightly more conserved than those facing the outside, while, as expected, buried β -strand residues are more conserved than exposed ones.

Among the residues facing the outside, those in the trimer interface are more conserved than those not in the interface. As a further illustration of this, Figure 5A displays the “effective 120° sum of average” information content (see Materials and methods) around the barrel axis, for all five porins. It can be seen that the information content (i.e., the degree of conservation) is always highest in the trimer interface region.

The trimer interface region has distinct residue preferences: We also analyzed the differences in amino acid composition between the “outside” β -strand residues located in the trimer interface and those not in the interface. The number of residues in these two

classes and the calculated overall preference scale for all five porins are given in Table 2. Considering only residues with a reasonably large total count, we find that Phe is more abundant in the interface, while Tyr and Asp are more abundant in the non-interface region.

To test if these preferences can be used to identify interface surfaces in porins, we considered 2omf, 1prn, and 1mal that have no detectable sequence identity. Preference scales derived from any two of these proteins were used to analyze the third structure (jack-knife test). The “effective 120° sum of average” was calculated for the interface preference and is shown in Figure 5B. The interface preference peaks in the interface region in all three profiles. Together with the finding that the residues in the interface are the most conserved, this suggests that it may be possible to predict not only the transmembrane β -strands and the approximate monomeric structure of porins from their amino acid sequence (Schirmer & Cowan, 1993), but also to roughly identify the most likely interface surface.

In summary, while the secondary and tertiary structure architectures of the helix-bundle and β -barrel membrane proteins are very different, the amino acid distributions along the membrane normal are quite similar. In particular, both kinds of proteins have a very distinct core region rich in aliphatic residues facing the lipid tails, flanked by belts of aromatic residues (Tyr and Trp in particular) roughly located in the lipid headgroup region. The aromatic belts are particularly prominent in the porins. While apolar, the porin trimer interfaces have a different amino acid composition from the β -strand residues facing the outside of the trimer, and also have a higher degree of sequence conservation, suggesting that it may be possible to predict an approximate oligomeric structure from a porin monomer model.

Materials and methods: PDB files: The PDB entries used for the structural analysis were 2omf, matrix porin from *E. coli*; 1pho, phosphoporin from *E. coli* (Cowan et al., 1992); 1prn, porin from *R. blastic* (Kreusch et al., 1994); 2por from *R. capsulatus* (Weiss et al., 1990, 1991b; Weiss & Schulz, 1992); 1mal, maltoporin from *E. coli* (Schirmer et al., 1995). In all but 2omf, the PDB entries contained the coordinates of only one monomer. The entire trimer in these cases were generated using the symmetry transformations provided with the PDB files. Table 1 provides a comparison of the porins of known structure.

Definitions of different structural classes: All residues in each porin structure were subdivided into different structural classes. Similar classifications have been reported earlier for individual porins, but were all done manually. To ensure an objective classification, the process of assigning residues to different structural classes was automated. The basic distinction was made between residues in the barrel β -strands and in the loops between the strands. β -Strand residues were subdivided into those facing away from the pore (outside residues) and those facing toward the pore (inside residues). The “outside” residues were further subdivided into residues that are in the trimer interface (interface residues) and those that are not (noninterface residues). Finally, all barrel β -strand residues were classified as either exposed or buried on the basis of their solvent accessibility in the monomer structure. The so-called constriction loop (Schirmer et al., 1995; Przybylski et al., 1996) that projects into the pore itself was included in the β -strand class.

Fig. 3. Frequency distribution for different kinds of residues along the trimer axis in 1 Å slices smoothed as in Figure 2. The negative side corresponds to the side facing the periplasm. The left panels show residue frequencies for β -strand residues facing the outside of the porin monomers (thick) and for all residues in three transmembrane helical proteins (thin; PDB codes 1prc, 2brd, 1occ) (Wallin et al., 1997), while the right panel shows results for β -strand residues facing the inside of the porin barrel. From top to bottom the following residue classes are shown: hydrophobic residues (AILMV), aromatic residues (YFW), basic residues (KR), and acidic residues (DE). For the helical proteins, Phe is included in the hydrophobic class and not in the aromatic.

Fig. 4. Information content for different residue classes. The error bars show the significance of each value.

Table 2. Residue distribution and preference scales

Residue	Number in interface	Number not-in-interface	Preference
Ser	7	2	0.81
Met	10	6	0.59
Phe	24	18	0.50
Arg	5	5	0.36
His	2	2	0.36
Gly	17	20	0.28
Thr	12	15	0.25
Trp	7	10	0.17
Ala	20	32	0.10
Ile	13	22	0.07
Cys	0	0	—
Leu	17	34	-0.04
Val	17	36	-0.08
Gln	2	5	-0.20
Pro	1	3	-0.33
Tyr	9	49	-0.81
Asp	4	24	-0.89
Lys	2	14	-1.02
Asn	2	15	-1.08
Glu	0	9	-Inf
Total	171	321

Fig. 5. **A:** Information content for β -strand residues facing the outside of the monomer plotted versus the angle with respect to the line joining the barrel and the trimer axes. The average information content is plotted using the "effective 120° sum of average." The shaded area represents the number of residues in the trimer interface as a function of the angle relative to the trimer axis. **B:** Jack-knife interface prediction profiles for 20mf, 1prn, and 1mal (see Materials and methods). The x-axis is the same as in Figure 5. The preference is calculated using the "effective 120° sum of average." The shaded area represents the number of residues in the trimer interface as a function of the angle relative to the trimer axis.

The PDB entries except 20mf and 1pho contained explicit secondary structure definitions that were used for the β -strand/loop classification. For 20mf and 1pho, the elements of secondary structure were defined using the DSSP program (Kabsch & Sander, 1983).

Residues in the β -strand class with more than 40% of the accessible surface area exposed to the surrounding media were defined as exposed and the rest of the residues in the β -sheet class were considered buried. An improved algorithm for Connolly solvent accessible surface area (Connolly, 1983; Le Grand & Merz, 1993) was used to calculate the accessible surface area. The fraction of buried accessible surface area was calculated as the accessible surface area of a residue of type X divided by the accessible surface area of a residue of type X in a tripeptide Gly-X-Gly (Bowie et al., 1997).

To identify the residues that face the inside or outside of the β -barrel in each monomer, a "barrel axis" was defined by taking the geometric average of the $C\alpha$ coordinates of lower and upper ends of the β -strands and passing a line through these points. The residues of the β -barrel, whose $C\alpha$ - $C\beta$ vector points toward the

barrel axis, were defined as facing inside and the rest as facing outside. On comparison with the manual classifications of inside/outside available for 2omf and 1prn (Cowan et al., 1992; Kreusch et al., 1994), this algorithm classifies more than 95% of the residues in the same way as was done manually.

The β -strand residues in the "outside" category either face the surrounding media (phospholipids or solvent) or the trimer interface. The interface residues were identified by comparing the accessible surface area of the monomer when in isolation and when in association with the other two monomers. Residues that showed a difference in the accessible surface area greater than $0.01 \text{ \AA}^2$ were classified to belong to the interface, while all other residues were classified as noninterface residues.

Residue frequency distribution: The membrane normal was defined from the symmetry axis of the trimer. Residue distributions were calculated along the membrane normal as well as in the membrane plane.

Distribution along the membrane normal: The "exposed" β -strand residues facing the outside are, in general, hydrophobic. Thus, analysis of the average hydrophobicity along the membrane normal can serve as a criterion to align the proteins along the z -axis for comparison. For this purpose, the hydrophobicity along the membrane normal was plotted for the "exposed, outside β -strand residues" (Fig. 2), using the GES hydrophobicity scale (Engelman et al., 1986), with a $5 \text{ \AA}$ wide, hat-shaped window to smooth the plots (Claverie & Daulmiere, 1991). The average of the z -coordinates of the outermost positive peaks with a hydrophobicity above 0.5 was taken to represent the location of the center (i.e., the $z = 0$ plane) of the membrane for each protein.

All residues were assigned to the z -coordinate of their $C\alpha$ atom. The number of the 20 different residues in $1 \text{ \AA}$ slices along the z -coordinate for each structural class was calculated, and the corresponding frequencies were obtained by dividing this number by the total number of residues in the concerned structural class in the slice. A $5 \text{ \AA}$ wide, hat-shaped window was used to smooth the plots.

Distribution perpendicular to the membrane normal: To calculate the residue distributions around the β -barrel projected onto the membrane plane, the barrel axis was defined as described above. The $C\alpha$ positions of the barrel residues were projected onto the a plane perpendicular to the barrel axis. The positive x -axis of the plane was defined as the line joining the barrel axis and the trimer axis. In the resulting polar description, 36 sectors of 10° width were drawn in the plane. The number of the 20 different amino acids in each sector was calculated, and the corresponding frequencies were obtained by dividing this number by the total number of residues in the sector. An "effective 120° sum of average" for a given sector was computed by adding contributions from 12 sectors (120°) centered around the sector in question and subtracting the contributions from the remaining sectors (240°), and sliding over all the sectors.

Multiple alignments and residue variability: To assess the residue variability in different parts of the porin structures, related sequences were obtained by searching the cumulative update (November 19, 1997) of the TREMBL and SWISS-PROT databases (Bairoch & Apweiler, 1996) with FASTA (Pearson & Lipman,

1988). The sequences were aligned using Clustal W (Thompson et al., 1994). Default settings for search and alignment procedures were used. Only sequences with more than 35% identity with the query sequence were considered; for 2omf, 1pho, 2por, 1prn, and 1mal, the final alignments included 31, 31, 3, 2, and 6 sequences, respectively. For each position in the sequence, the information content (Schneider et al., 1986), a parameter that reflects residue variability, was calculated:

$$I_n = I_o + \sum_{\text{all 20 amino acids}} f(a, n) * \log_2 f(a, n) \quad (1)$$

where $f(a, n)$ is the frequency of amino acid a in position n and I_o is a constant that would be

$$I_o = \log_2(20) = 4.32 \text{ bits} \quad (2)$$

if all amino acids were equally common.

Residue preference scale: Residue preference parameters for the trimer interface were derived by the following relationship:

Preference = $\ln(p)$, where

$$p = \frac{p_{ij}}{p_i p_j} \quad (3)$$

where p_{ij} is the fraction of residues i in the interface region, p_i is the fraction of residues of type i , and p_j is the fraction of residues in the interface region. Only "outside" β -strand residues were included in the calculation.

Jack-knife prediction test: To assess the predictive ability of the interface-preference scale derived above, three test proteins (2omf, 1prn, 1mal) that have no significant sequence identity were analyzed. Interface preference scales were derived from two of the proteins in this set and was applied to the remaining protein for all the three possible combinations, and "effective 120° sum of average" profiles were calculated as above.

Acknowledgments: This work was supported by grants from the Swedish Natural Sciences Research Council and the Swedish Research Council for Engineering Sciences to AE and GvH.

References

- Agterberg M, Adriaanse H, Tjhaar E, Resink A, Tommassen J. 1989. Role of the cell surface-exposed regions of outer membrane protein PhoE of *Escherichia coli* K12 in the biogenesis of the protein. *Eur J Biochem* 185:365–370.
- Bairoch A, Apweiler R. 1996. The SWISS-PROT protein sequence data bank and its new supplement TREMBL. *Nucleic Acids Res* 24:17–21.
- Bowie JU. 1997. Helices packing in membrane proteins. *J Mol Biol* 272(5):780–789.
- Bowie JU, Lüthy R, Eisenberg D. 1997. A method to identify protein sequence that fold into a known three-dimensional structure. *Science* 253:164–170.
- Claverie J-M, Daulmiere C. 1991. Smoothing profiles with sliding windows: Better to wear a hat! *CABIOS* 7:113–115.
- Conolly MJ. 1983. Analytical molecular surface calculation. *J Appl Crystallogr* 17:548–558.
- Cowan SW, Schirmer T, Rummel G, Steiert M, Ghosh R, Pauptit RA, Jansonius JN, Rosenbusch JP. 1992. Crystal structures explain functional properties of two *E. coli* porins. *Nature* 358:727–733.

- Engelman DM, Steitz TA, Goldman A. 1986. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu Rev Biophys Chem* 15:321–353.
- Gromiha MM, Majumdar R, Ponnuswamy PK. 1997. Identification of membrane spanning β -strands in bacterial porins. *Protein Eng* 10(5):497–500.
- Hunt JF, Earnest TN, Bousche O, Kalghatgi K, Reilly K, Horvath C, Rothschild KJ, Engelman DM. 1997. A biophysical study of integral membrane protein folding. *Biochemistry* 36:15156–15176.
- Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637.
- Kraulis P. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Cryst* 24:946–950.
- Kreusch A, Neubuser A, Schlitz E, Weckesser J, Schulz GE. 1994. Structure of the membrane channel porin from *Rhodopseudomonas blastica* at 2.0 Å resolution. *Protein Sci* 3:58–63.
- Le Grand S, Merz KJ. 1993. Rapid approximation to molecular surface area via the use of boolean look-up tables. *J Global Optimiz* 3:49–66.
- Pearson WR, Lipman DJ. 1988. Improved tools for biological sequence analysis. *Proc Natl Acad Sci USA* 85:2444–2448.
- Popot JL, Engelman DM. 1997. Membrane protein folding and oligomerization: The 2-stage model. *Biochemistry* 29:4031–4037.
- Przybylski M, Glocker MO, Nestel U, Schnaible V, Bluggel M, Diederichs K, Weckesser J, Schad M, Schmid A, Welte W, Benz R. 1996. X-ray crystallographic and mass spectrometric structure determination and functional characterization of succinylated porin from *Rhodobacter capsulatus*: Implications for ion selectivity and single-channel conductance. *Protein Sci* 5:1477–1489.
- Schirmer T, Cowan SW. 1993. Prediction of membrane-spanning β -strands and its application to maltoporin. *Protein Sci* 2:1361–1363.
- Schirmer T, Keller TA, Wang Y-F, Rosenbusch JP. 1995. Structural basis for sugar translocation through maltoporin channels at 3.1 Å resolution. *Science* 267:512–514.
- Schneider TD, Stormo GD, Gold L, Ehrenfeucht A. 1986. Information content of binding sites on nucleotide sequences. *J Mol Biol* 188(3):415–431.
- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties weight matrix choice. *Nucleic Acids Res* 22:4673–4680.
- von Heijne G. 1997. Principles of membrane protein assembly and structure. *Progr Biophys Mol Biol* 66:113–139.
- Wallin E, Tsukihara T, Yoshikawa S, von Heijne G, Elofsson A. 1997. Architecture of helix bundle membrane proteins. An analysis of cytochrome *c* oxidase from bovine mitochondria. *Protein Sci* 6:808–815.
- Weiss MS, Abele U, Weckesser J, Welte W, Schlitz E, Schulz GE. 1991a. Molecular architecture and electrostatic properties of a bacterial porin. *Science* 254:1627–1630.
- Weiss MS, Kreusch A, Schiltz E, Nestel U, Welte W, Weckesser J, Schulz GE. 1991b. The structure of porin from *Rhodobacter capsulatus* at 1.8 Å resolution. *FEBS Lett* 280:379–382.
- Weiss MS, Schulz GE. 1992. Structure of porin refined at 1.8 Å resolution. *J Mol Biol* 227(2):493–509.
- Weiss MS, Wacker T, Weckesser J, Welte W, Schulz GE. 1990. The three-dimensional structure of porin *Rhodobacter capsulatus* at 3 Å resolution. *FEBS Lett* 267(2):268–272.