Article

In Silico Structure and Sequence Analysis of Bacterial Porins and Specific Diffusion Channels for Hydrophilic Molecules: Conservation, Multimericity and Multifunctionality

Hilde S. Vollan [1,2], Tone Tannæs [1], Gert Vriend [3] and Geir Bukholm [2,4,*

[1] Department of Clinical Molecular Biology (EpiGen), Division of Medicine, Akershus University Hospital and University of Oslo, PO box 28, N-1478 Lørenskog, Norway; h.s.vollan@studmed.uio.no (H.S.V.); t.m.tannas@medisin.uio.no (T.T.)

[2] Norwegian Institute of Public Health, Box 4404 Nydalen, N-0403 Oslo, Norway

[3] Centre for Molecular and Biomolecular Informatics, Radboud University Medical Center, 6525 GA Nijmegen, The Netherlands; vriend@cmbi.ru.nl

[4] Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Pb 5003, N-1430 Ås, Norway

* Correspondence: geir.bukholm@fhi.no; Tel.: +47-99-291-018

Abstract: Diffusion channels are involved in the selective uptake of nutrients and form the largest outer membrane protein (OMP) family in Gram-negative bacteria. Differences in pore size and amino acid composition contribute to the specificity. Structure-based multiple sequence alignments shed light on the structure-function relations for all eight subclasses. Entropy-variability analysis results are correlated to known structural and functional aspects, such as structural integrity, multimericity, specificity and biological niche adaptation. The high mutation rate in their surface-exposed loops is likely an important mechanism for host immune system evasion. Multiple sequence alignments for each subclass revealed conserved residue positions that are involved in substrate recognition and specificity. An analysis of monomeric protein channels revealed particular sequence patterns of amino acids that were observed in other classes at multimeric interfaces. This adds to the emerging evidence that all members of the family exist in a multimeric state. Our findings are important for understanding the role of members of this family in a wide range of bacterial processes, including bacterial food uptake, survival and adaptation mechanisms.

Keywords: β-barrel membrane proteins; bacteria; porins; specific diffusion channels; structure analysis; multiple sequence analysis; entropy-variability analysis

1. Introduction

Bacteria thrive in an extremely wide range of environments. They can evolve rapidly to cope with continuously-changing environments. The cell envelope provides their main line of defense against their surroundings. The cell envelope must carefully discriminate between external molecules that are useful, e.g., nutrients, and those that are harmful, e.g., antibiotics. Gram-positive bacteria have a cell wall that consists of a single membrane covered with a thick peptidoglycan layer, while Gram-negative bacteria have two membranes that are separated by the periplasm [1,2]. Their outer membrane contains several outer membrane proteins (OMPs), which typically consist of a transmembrane barrel comprising of β-strands that are connected by short periplasmic loops (also called turns) and longer extracellular loops (see Figure 1) [3]. These OMPs regulate the passage of molecules through the outer membrane.
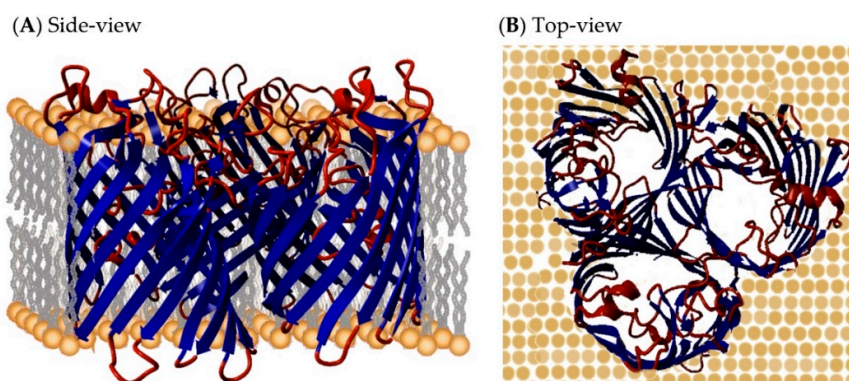
Figure 1. Outer membrane protein (OMP) structure. OMP structures are composed of a transmembrane motif that consists of β-strands that form a transmembrane β-barrel. These strands are connected by short loops (also called turns) in the periplasm and longer extracellular loops. The number of β-strands, the length of the β-strands and the loop lengths vary widely among OMPs. Here, the trimeric general diffusion porin (GDP) from *E. coli* (PDB (Protein Data Bank) ID: 2J1N, [4]) is shown as an example from (**A**) the side-view with the ß-barrel core embedded in the membrane and (**B**) the top-view, where the pores are visible viewed from the extracellular side. The β-barrel with loops and turns removed (blue) is referred to as the core.

Non-specific and specific diffusion channels form the largest OMP family, and they are involved in the selective uptake of molecules. The number of β-strands in a protein barrel ranges from eight to 26. The pore size and amino acid composition in the channel are the main determinants of porin specificity and function [3]. These proteins are often multifunctional, and they may also function as a bacteriocin receptor [5–7] or in host-cell interactions [8,9]. Non-specific channels, or porins, and specific diffusion channels are upregulated in the presence of nutrients and downregulated in the presence of toxins or other harmful molecules (e.g., antibiotics, heavy metals, detergents or bile salts) [10,11]. The gating mechanism of the larger channels is achieved through a constriction loop. This is usually the largest extracellular loop (loop L3) that is located inside the barrel (see Figure 2). Unlike other extracellular loops, this loop never faces the environment. The amino acid composition of this loop determines the substrate specificity of the pore [5].
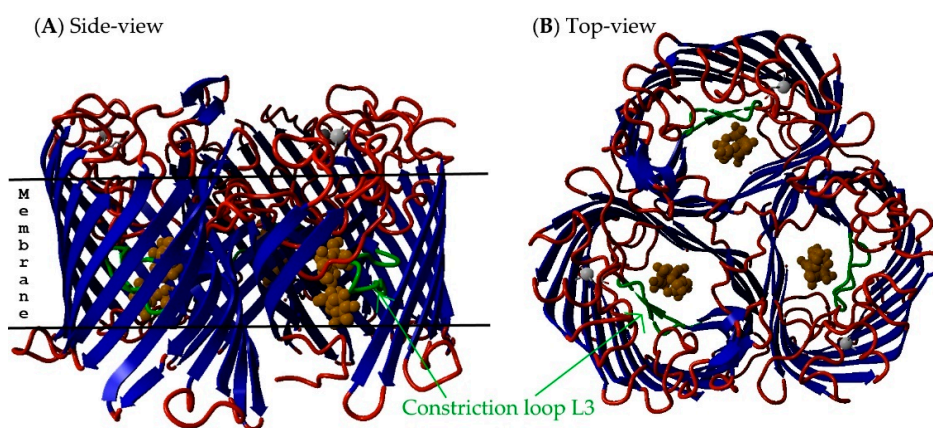

Figure 2. Porin structure. The constriction loop L3 of sugar-specific *Salmonella* Typhimurium ScrY porin is highlighted in green (PDB ID: 1OH2, [12]). The barrel core is colored blue; periplasmic turns and extracellular loops are red; the substrate is orange; and calcium ions are gray. (**A**) Side-view of the porin where strands 14, 15 and 16 of molecule P were removed to better visualize the constriction loop L3; (**B**) top-view of the porin.

Outer membrane protein channels involved in the diffusion of hydrophilic molecules have been divided into two groups in two different ways: specific diffusion channels *versus* non-specific porins [11,13–17] and monomeric *versus* trimeric channels [3,8,18]. Although these divisions are not universally accepted [3,8,11,17,19], there appears to be little doubt that some OM channels are monomeric and others are trimeric [5,11,20]. The subdivisions of the four main porin and specific diffusion channel groups, based on their multimeric state and specificity, are [3,8,11,13,15,21–24]:

1. Non-specific monomeric porins (NMPs): Outer membrane protein A (OmpA) and OmpG are NMPs. OmpA is an eight-stranded multifunctional protein where current research reveals porin activity [22,25–27], while OmpG is a 14-stranded, pH-dependent porin [5].
2. Non-specific trimeric porins (NTPs): NTPs, also known as general diffusion porins (GDPs), have 16-stranded β-barrels that allow the diffusion of hydrophilic substances smaller than 600 kDa [21].
3. Specific monomeric diffusion channels (SMDCs): The oligogalacturonate-specific KdgM channel, outer membrane porin B (OprB) and outer membrane carboxylate channel (Occ) proteins belong to the SMDC group. KdgM is an acidic, sugar-specific channel with a 12-stranded β-barrel [24]. OprB is a carbohydrate-specific channel with a 16-stranded β-barrel [23]. Finally, Occ porins are water-soluble, specific channel for small substrates with a carboxyl group with 18-stranded β-barrels [28,29].
4. Specific trimeric diffusion channels (STDCs): STDCs include OprP and maltoporins. OprP is a phosphate-specific porin that has a 16-stranded β-barrel [30]. Sugar-specific channels (maltoporins/ScrY) are slightly larger, as they consist of 18-stranded β-barrels [12,31].

The aim of this study was to examine the sequence- and structure-derived interaction patterns for non-specific porins and specific diffusion channels. To shed light on the structure-function relations of amino acids in these subclasses, MSAs (multiple sequence alignments) were constructed based on 3D structures of non-specific porins and specific diffusion channels. EVAs (Entropy-variability analyses) [32] were used to study the evolutionary footprints in the alignments.

2. Results

2.1. Data Collection

Eighty-nine unique OMP structures were extracted from the PDB, of which 34 belong to the porin and specific diffusion channel family. The 34 protein structures are listed in Table 1, together with biological and structural information.

Table 1. Protein structure information of bacterial porins and specific diffusion channels. This table lists protein names and biological and structural information of the structures used in our analyses. The size column lists the number of β-strands.

Groups	Name	Bacteriae	Size	PDB ID	Resolution (Å)
Monomeric proteins					
NMP	EcOmpA	*Escherichia coli*	8	1QJP	1.7
NMP	KpOmpA	*Klebsiella pneumonia*	8	2K0L	–
NMP	OmpG	*Escherichia coli*	14	2IWV	2.3
SMDC	NanC	*Escherichia coli*	12	2WJR	1.8
SMDC	KdgM	*Dickeya dadantii*	12	4FQE	1.9
SMDC	OprB	*Pseudomonas putida*	16	4GEY	2.7
SMDC	OccD	*Pseudomonas putida/ Pseudomonas fluorescens*	18	3SYS, 3SZD, 3SZV, 3T0S, 3T20, 3T24, 4FRT, 4FRX, 4FT6, 4FSO, 4FSP, 3JTY, 3SY7, 3SY9, 3SYB	Average: 2.4 (1.5–3.2)

Table 1. *Cont.*

Groups	Name	Bacteriae	Size	PDB ID	Resolution (Å)
Trimeric proteins					
NTP	EcOmpC	*Escherichia coli*	16	2J1N	2.0
NTP	StOmpC	*Salmonella* Typhimurium	16	3UPG	3.2
NTP	OmpK36	*Klebsiella pneumonia*	16	1OSM	3.2
NTP	StOmpF	*Salmonella* Typhimurium	16	3NSG	2.8
NTP	EcOmpF	*Escherichia coli*	16	4GCS	1.9
NTP	PhoE	*Escherichia coli*	16	1PHO	3.0
NTP	RcGDP	*Rhodobacter capsulatus*	16	2POR	1.8
NTP	RbGDP	*Rhodopseudomonas blastica*	16	1PRN	2.0
NTP	Omp32	*Delftia acidovorans*	16	2FGQ	1.5
STDC	EcMaltoporin	*Escherichia coli*	18	1AF6	2.4
STDC	StMaltoporin	*Salmonella* Typhimurium	18	2MPR	2.4
STDC	ScrY	*Salmonella* Typhimurium	18	1OH2	2.4
STDC	OprP	*Pseudomonas aeruginosa*	16	2O4V	1.9

EcMaltoporin (*Escherichia coli* maltoporin); EcOmpA (*E. coli* outer membrane protein A); EcOmpC (*E. coli* outer membrane protein C); EcOmpF (*E. coli* outer membrane protein F); KdgM (oligogalacturonate-specific channel); KpOmpA (*K. pneumonia* outer membrane protein A); NanC (*N*-acetylneuraminic acid-inducible outer-membrane channel); NMP (non-specific monomeric porins); NTP (non-specific trimeric porin); OccD (outer membrane carboxylate channel); Omp32 (outer membrane protein 32); OmpG (outer membrane protein G); OmpK36 (outer membrane porins of *K. pneumoniae*); OprB (outer membrane porin B); OprP (outer membrane porin P); PDB ID (Protein Data Bank Identifier); PhoE (phosphoporin); RbGDP (*R. blastica* general diffusion porin); RcGDP (*R. capsulatus* general diffusion porin); ScrY (Sugar specific porin); SMDC (specific monomeric diffusion channel); STDC (specific trimeric diffusion channel); StMaltoporin (*S.* Typhium maltoporin); StOmpC (*S.* Typhimurium outer membrane protein C); StOmpF (*S.* Typhimurium outer membrane protein F); Å (Angstrom).

The PDB contains structures of proteins, isolated from a wide variety of species, which engage in homotrimeric interactions (see Table 1). The NTP/GDP group contains structures isolated from γ-Proteobacteria (*Escherichia coli*, *Klebsiella pneumoniae* and *Salmonella enterica* serovar Typhimurium (*Salmonella* Typhimurium)), α-Proteobacteria (*Rhodobacter capsulatus* and *Rhodopseudomonas blastica*) and β-Proteobacteria (*Delftia acidovorans*). Despite the species variability among this group, they are all non-specific porins, and they have the same barrel size. The STDCs isolated from γ-Proteobacteria were subdivided into two groups based on their β-barrel size and specificity: the 18-stranded β-barrel, sugar-specific diffusion channels and the 16-stranded β-barrel, phosphate-specific channels.

2.2. Non-Specific Porin and Specific Diffusion Channel Secondary Structure Composition

The secondary structure composition of the four protein groups is summarized in Table 2. In total, 56.3% of the amino acids were found in β-strands; 30.8% were found in extracellular loops; and 12.5% were found in periplasmic turns. Their sizes varied from 171 (NMP; PDB ID 1QJP) to 422 (SMDC; PDB ID 3SY9) amino acids. The proportions of secondary structures remained roughly the same for all of the protein structures, with an average ratio of 6:1:3 for amino acids in the β-strand:periplasmic loop:extracellular loop.

Table 2. Secondary structure composition. Secondary structure composition of the average non-specific porin and specific diffusion channel composition, which lists the average length (number of amino acids), percentages of amino acids located in β-strands, periplasmic loops and extracellular loops.

Porin Group	Average Length	Barrel (%)	Periplasmic Loop (%)	Extracellular Loop (%)
NMP	219	57	12	30
SMDC	385	56	14	28
NTP	325	57	11	30
STDC	417	54	11	34

NMP (non-specific monomeric porins); NTP (non-specific trimeric porin); SMDC (specific monomeric diffusion channel); STDC (specific trimeric diffusion channel).

2.3. Non-Specific Porin and Specific Diffusion Channel Phylogenetic Tree

A phylogenetic tree, which was constructed from the structure-based alignment of the 34 non-specific porin and specific diffusion channel sequences, is depicted in Figure 3. The sequences were colored according to the four groups discussed in the Introduction.
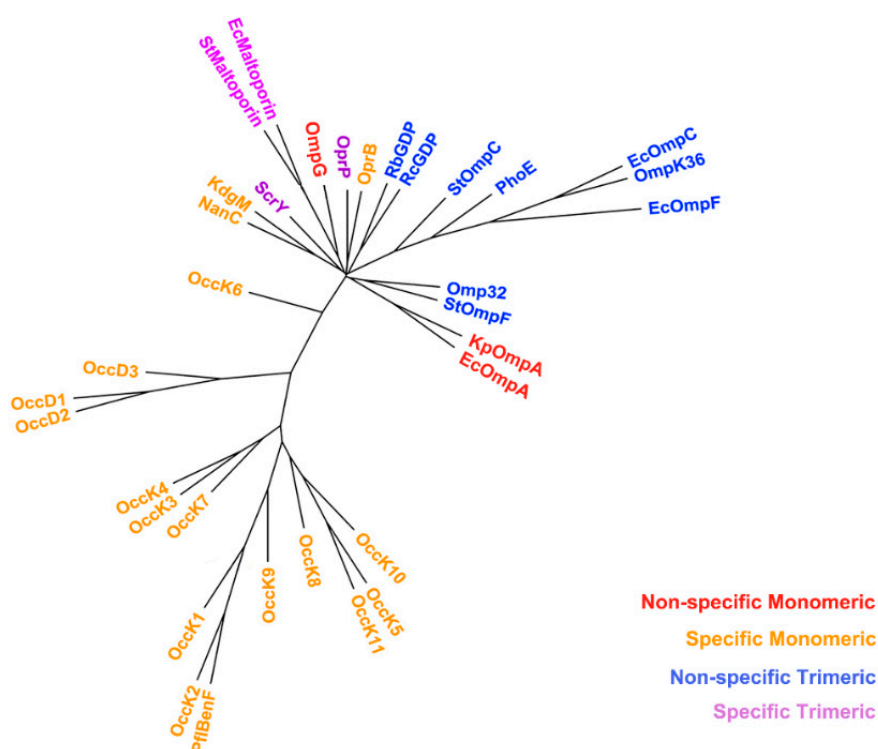

Figure 3. Phylogenetic tree of bacterial porins. Colors represent the four groups and the multimeric state of the porin based on the old classification system.

2.4. Non-Specific Porin and Specific Diffusion Channel Classification System

Several different OMP classification systems are based on protein structures. Many databases group non-specific porin and specific diffusion channel with transporters, surface proteins or mitochondrial proteins. For example, the "Class, Architecture, Topology, Homology" (CATH) database [33], "The Transporter Classification Database" (TCDB) [34] and the "Outer Membrane Protein Database" (OMPdb) [35] do not distinguish non-specific porin and specific diffusion channel from other OMPs. However, the OMPdb has 12 non-specific porin and specific diffusion channel classes that are relevant to our analysis. The remaining 79 OMP classes were excluded because they lacked a solved structure or because they were not non-specific porin nor the specific diffusion channel (see the Discussion and Table S1 for how the relevant non-specific porin and specific diffusion channel families correspond to the eight subclasses used in our analysis). Finally, only three porins (Omp32, OmpF and PhoE) were included in the "Structural Classification of Proteins-extended" (SCOPe) superfamily [36].

The non-specific porin and specific diffusion channel family are often divided into four groups, depending on their specificity and multimericity (see the Introduction). These four groups were used in Tables 1 and 2 and Figure 3. However, for the remaining analyses, we divided non-specific porins and specific diffusion channels into six classes with eight subclasses, based on their specificity and size. Table 3 lists class number and names, corresponding protein names (of structures solved to date) and PDB IDs.

Table 3. The non-specific porin and specific diffusion channel family distributed among six classes and further divided into eight subclasses. Corresponding OMPdb [35] classifications are found in Table S1. The empty classes (and subclasses) are reserved for future structures. "Size" refers to the number of strands.

Class		Subclass		Protein Structures	
Number	Size	Number	Name	Protein Name	PDB ID Template (Other)
1	8	1A	Non-specific, petite porin	OmpA	2K0L (1QJP)
2	10	2A	Non-specific, mini porin	–	–
3	12	3A	Non-specific, small porin	–	–
		3B	Oligogalacturonate-specific, small channel	KdgM and NanC	4FQE (2WJR)
4	14	4A	Non-specific, intermediate porin	OmpG	2IWV
5	16	5A	Non-specific, medium porin	OmpC, OmpK36, OmpF, PhoE, Omp32, RcGDP and RbGDP	2J1N (1PRN, 3UPG, 1OSM, 3NSG, 4GCS, 2POR, 2FGQ, and 1PHO)
		5B	Sugar-specific, medium channel	OprB	4GEY
		5C	Phosphate-specific, medium channel	OprP	2O4V
6	18	6A	Non-specific, large porin	–	–
		6B	Sugar-specific, large channel	Maltoporin and ScrY	2MPR (1AF6 and 1OH2)
		6C	Carboxyl-specific, large channel	Occ Channels	3SZV (3SYS, 3SZD, 3T0S, 3T20, 3T24, 4FRT, 4FRX, 4FT6, 4FSO, 4FSP. 3JTY, 3SY7, 3SY9, and 3SYB)

EcMaltoporin (*Escherichia coli* maltoporin); EcOmpA (*E. coli* outer membrane protein A); EcOmpC (*E. coli* outer membrane protein C); EcOmpF (*E. coli* outer membrane protein F); KdgM (oligogalacturonate-specific channel); KpOmpA (*K. pneumonia* outer membrane protein A); NanC (N-acetylneuraminic acid-inducible outer-membrane channel); OccD (outer membrane carboxylate channel); Omp32 (outer membrane protein 32); OmpG (outer membrane protein G); OmpK36 (outer membrane porins of *K. pneumoniae*); OprB (outer membrane porin B); OprP (outer membrane porin P); PDB ID (Protein Data Bank Identifier); PhoE (phosphoporin); RbGDP (*R. blastica* general diffusion porin); RcGDP (*R. capsulatus* general diffusion porin); ScrY (Sugar specific porin); StMaltoporin (*S.* Typhium maltoporin); StOmpC (*S.* Typhimurium outer membrane protein C); StOmpF (*S.* Typhimurium outer membrane protein F).

2.5. Non-Specific Porin and Specific Diffusion Channel Structure Analysis

A superposition of non-specific porin and specific diffusion channel core structures resulted in an average root-mean-square deviation (RMSD) of 1.2 Å over 95.2% of the aligned residues, with a 51.7% sequence identity per class (ranging from RMSDs of 0.9 to 1.6 Å over 90.7 to 98.7% of the aligned residues, with sequence identities between 25.6% and 96.1%). These results are listed in Table 4, while an analysis of the entire non-specific porin and specific diffusion channel structure is found in Table S2 and Table S2raw (these tables compare the core structure analysis with whole structures and the sequence analysis). The structure alignment of specific, small channels (Class 3B; KdgM and NanC) resulted in the lowest sequence identity score, while OmpA structures (Class 1A) resulted in the highest average sequence identity. Three of the subclasses could not be analyzed because they currently hold only one structure (Classes 4A, 5B and 5C).

Table 4. Structure alignment statistics of non-specific porins and specific diffusion channels. Core structures are analyzed after the loops were removed (so only the barrel was used). The percentages of aligned residues are averages for each pair-wise alignment possible in the subclass. Empty cells represent classes with only one structure. Minimum pore radii are determined using HOLE [37] for the whole monomeric protein and its β-barrel core, respectively.

Subclasses		Resolution	% Residues Superposed	% Sequence Identity		Pore Size	
Number	Name	RMSD (Å)	Mustang	Mustang	Clustal Ω	Core	Whole
1A	Non-specific, petite porin	1.6	98.7	96.1	93.6	~ 0	0.1
3B	Oligogalacturonate-specific, small channel	1.4	95.0	25.6	28.1	2.9	2.9
4A	Non-specific, intermediate porin	–	–	–	–	3.7	4.0
5A	Non-specific, medium porin	1.4	90.7	40.2	41.3	2.8	6.9
5B	Sugar-specific, medium channel	–	–	–	–	2.2	7.0
5C	Phosphate-specific, medium channel	–	–	–	–	1.6	6.4
6B	Sugar-specific, large channel	0.9	95.6	50.3	49.5	2.1	6.9
6C	Carboxyl-specific, large channel	0.9	96.2	46.4	45.7	1.8	7.8
	Average	1.2	95.2	52.0	51.6	2.4	7.5

Minimum pore size estimates are listed in Table 4 and visualized in Figure 4 (see the associated website for more information).
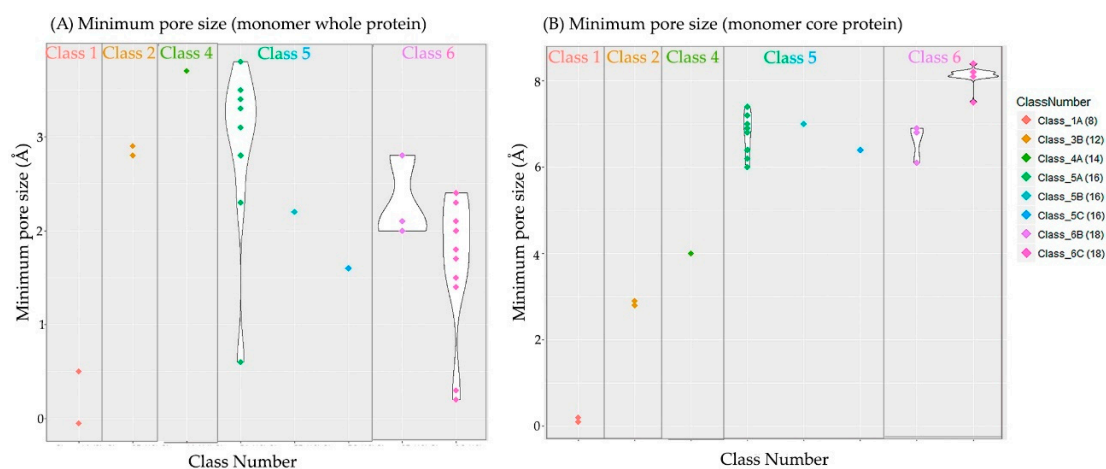

Figure 4. Estimated minimum Angstrom pore size radius for each subclass using the HOLE software [37]. Each colored diamonds represent one monomer structure. (**A**) Whole structure; (**B**) core structure where loops and turns were removed. The barrel size number of strands in each β-barrel) is listed in brackets next to the class name labels.

2.6. Multiple Sequence Alignments

A profile-based sequence alignment of the core proteins, the barrel structure, was used to generate an MSA; see Table 3 for the templates that were used to generate the initial profile. Based on the alignment results shown in Table 4 and Text S2, the core structure was defined to be the transmembrane β-barrel domain. Those classes with multiple structures were used to update the profile before including the sequences for the eight subclasses of the non-specific porin and specific diffusion channel protein family (see the Methods Section and Table S2). The number of sequences used in the MSAs of the eight different subclasses (listed in Table 5) ranged from 50 to 1384, with an average of approximately 490 sequences.

Table 5. Number of sequences used in MSA (multiple sequence alignment analyses). The number of sequences used in the final MSA that was generated for each class is listed together with the class number and name. The lacking classes (and subclasses) are reserved for future structures and highlighted with a question mark.

Subclass Number	Subclass Name	Number of Sequences Used in the MSA
1A	Non-specific, petite porin	389
3B	Oligogalacturonate-specific, small channel	246
4A	Non-specific, intermediate porin	50
5A	Non-specific, medium porin	725
5B	Sugar-specific, medium channel	319
5C	Phosphate-specific, medium channel	180
6B	Sugar-specific, large channel	663
6C	Carboxyl-specific, large channel	1394

2.7. Entropy-Variability Analysis

Entropy-variability analysis (EVA) was then used to obtain a better understanding of these alignments (see the Discussion for more information regarding EVA). Figure 5 shows the entropy-variability (EV) plots generated for the eight different subclasses. These EV plots show that the distribution of points over the plots varied from each subclass, but were scattered throughout the plot, thereby ensuring that there were not too many or too few sequence variations in the alignments. The plots also show that the number of conserved residues varied from each subclass.
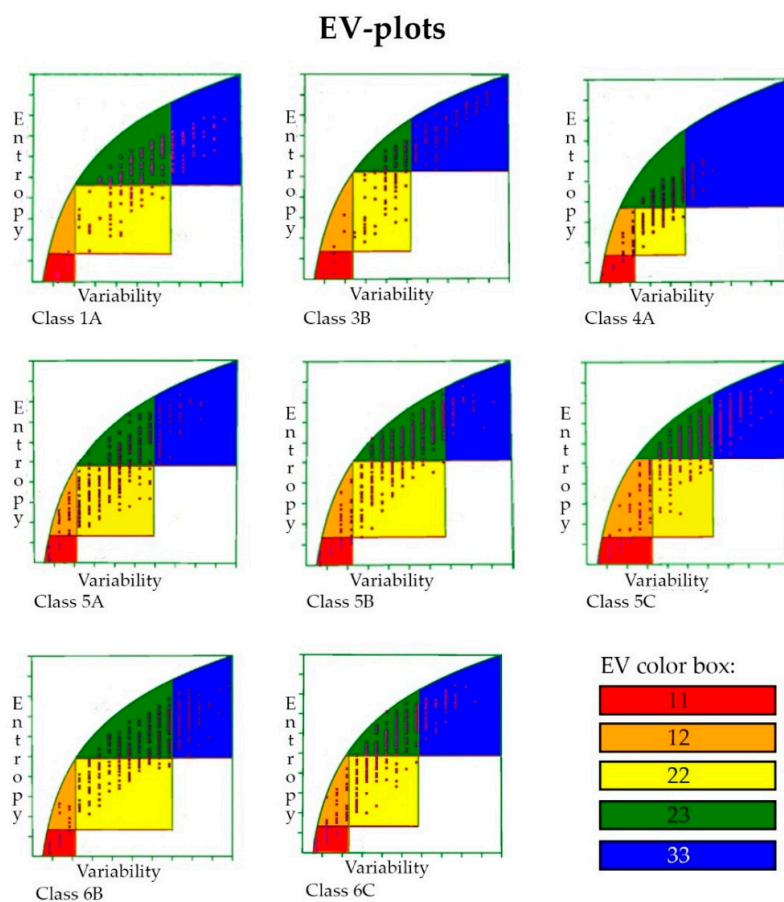

Figure 5. Entropy-variability (EV) plots of the eight subclasses. Red is for those residues likely to be in the main active site; orange is the main active site; green is the regulatory site; yellow communicates; blue is those with unknown function (see the Methods Section for more details).

The EVA results were mapped onto structures representing the eight analyses (see Table 3, Figures 6–9 and the Methods Section): 5.6% of the residues were contained in Box 11 (highly conserved residues colored red); 7.0% of the residues were contained in Box 12 (very conserved residues colored orange); 26% of the residues were contained in Box 22 (quite conserved residues colored yellow); 40.9% of the residues were contained in Box 23 (moderately conserved residues colored green); and 19.8% of the residues were contained in Box 33 (highly variable residues with unknown function colored blue). The distributions of extracellular loops found in these were 0.5% (Box 11), 3.0% (Box 12), 21.2% (Box 22), 38.9% (Box 23) and 36.4% (Box 33). Of the highly variable (Box 33) extracellular loop residues, 62.9% were located in large loops (longer than 15% of the extracellular loop length for each subclass, except for Class 1A, in which three of the four loops were termed long loops). A constriction loop is essential for diffusion, and it contains a high number of conserved residues (and only 8% of such residues were found in Box 33; see Table S6). These residues are situated inside the barrel and, unlike the other loops, are not in contact with the environment. In summary, the constriction loops contained more conserved residues than the longer extracellular loops (more statistics are found in Tables S3–S6), and a trimeric pattern was observed in the extracellular loops (see next section; 2.8. Non-Specific Porin and Specific Diffusion Channel Multimericity).
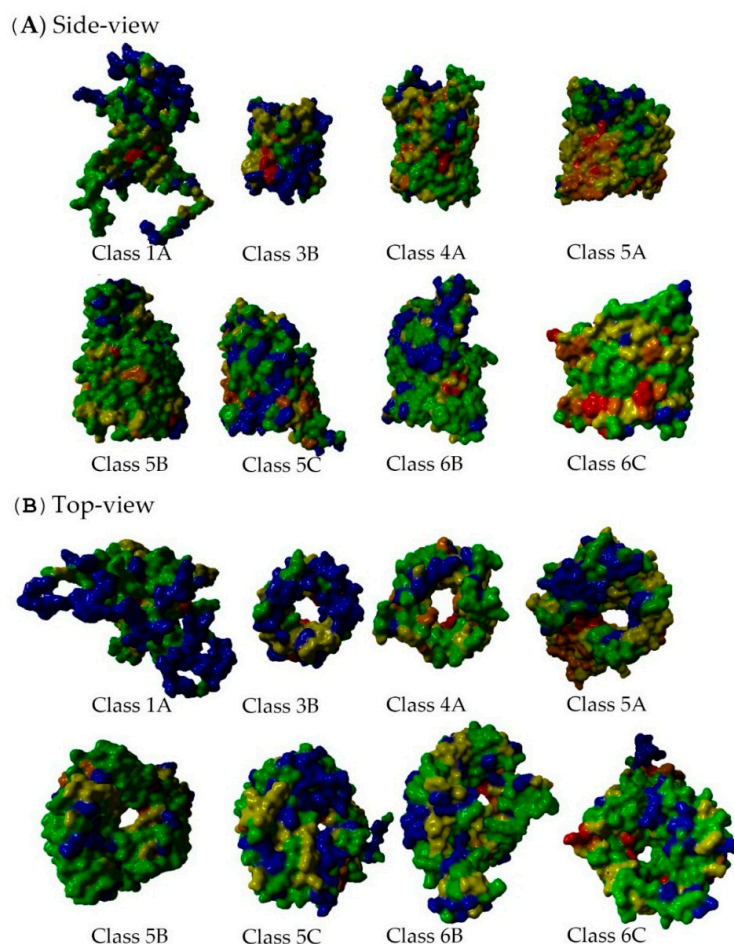

Figure 6. Entropy-variability (EV) results of the eight subclass from (**A**) side-view and (**B**) top-view. The structures are visualized as molecular surfaces and are colored according to residue conservation as described in the Methods Section. This figure illustrates differences in barrel size, pore size and loop variability. All structures are visualized in the monomeric form. These results can be downloaded and viewed in 3D from the associated websites (YASARA (Yet Another Scientific Artificial Reality Application) scene files; see the Supplementary Materials).
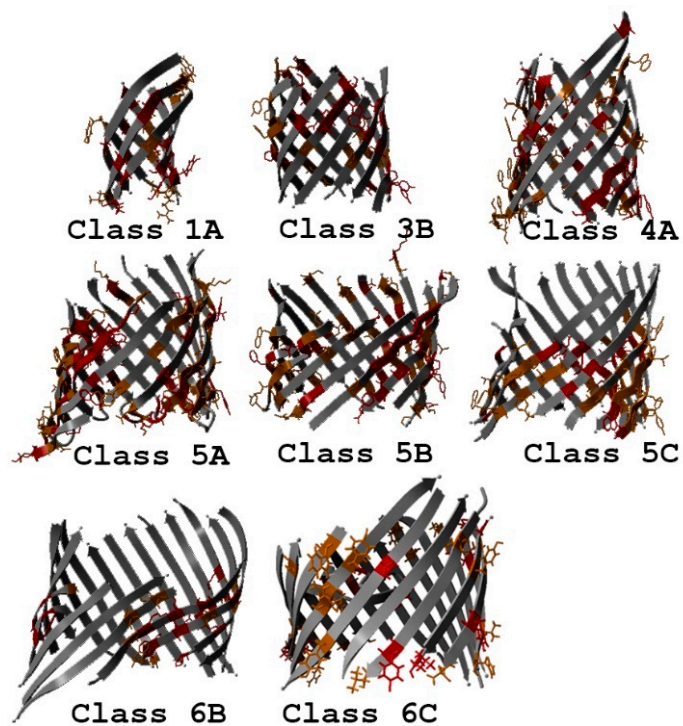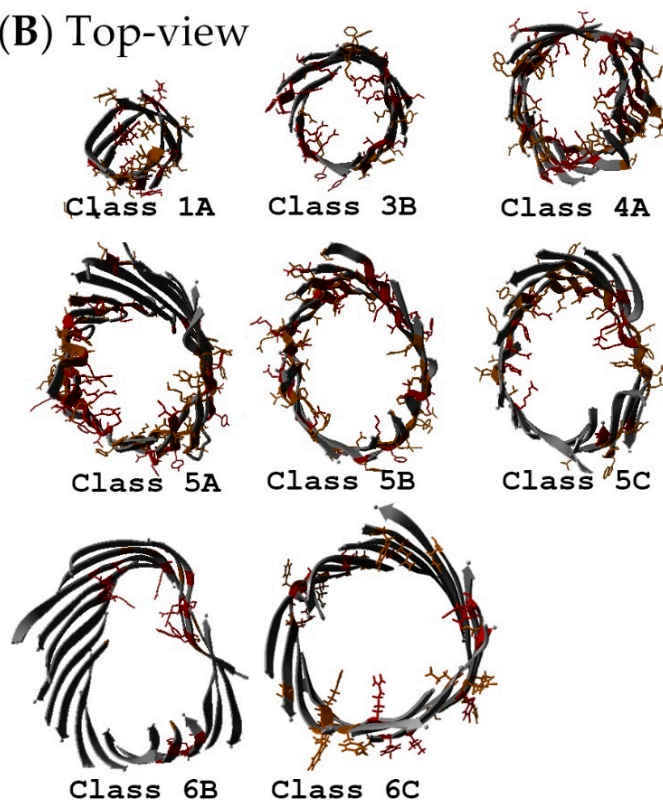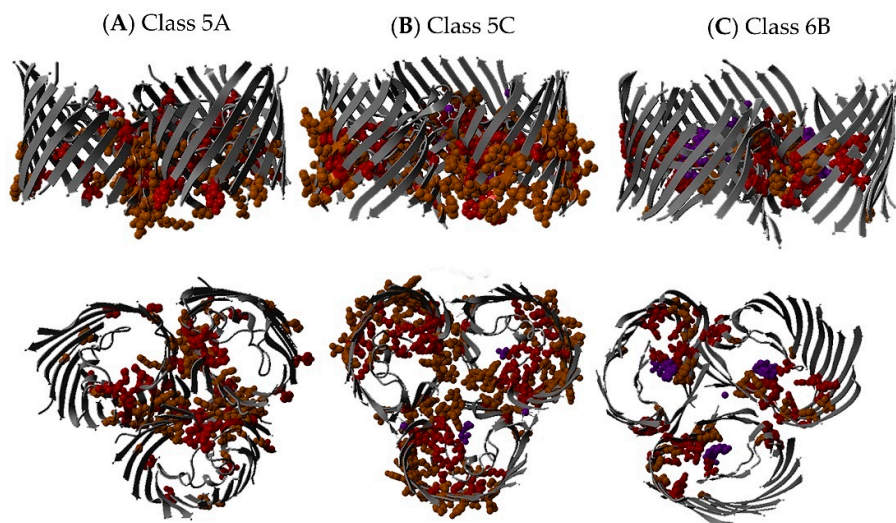
Figure 7. Conserved non-specific porin and specific diffusion channel residues. (**A**) Side view and (**B**) top view of conserved residues (Boxes 11, 12, and 22). Only the monomeric state is shown for all subclasses. Loops and turns are deleted to visualize the protein core. These results can be downloaded and viewed in 3D from the associated websites (YASARA scene files; see the Supplementary Materials).

Trimeric pattern (Class 5A, 5C, 6B)

(**A**) Class 5A　　　　　　　(**B**) Class 5C　　　　　　　(**C**) Class 6B

Trimeric models

(**D**) Class 1A　　　　　(**E**) Class 3B　　　　　(**F**) Class 4A

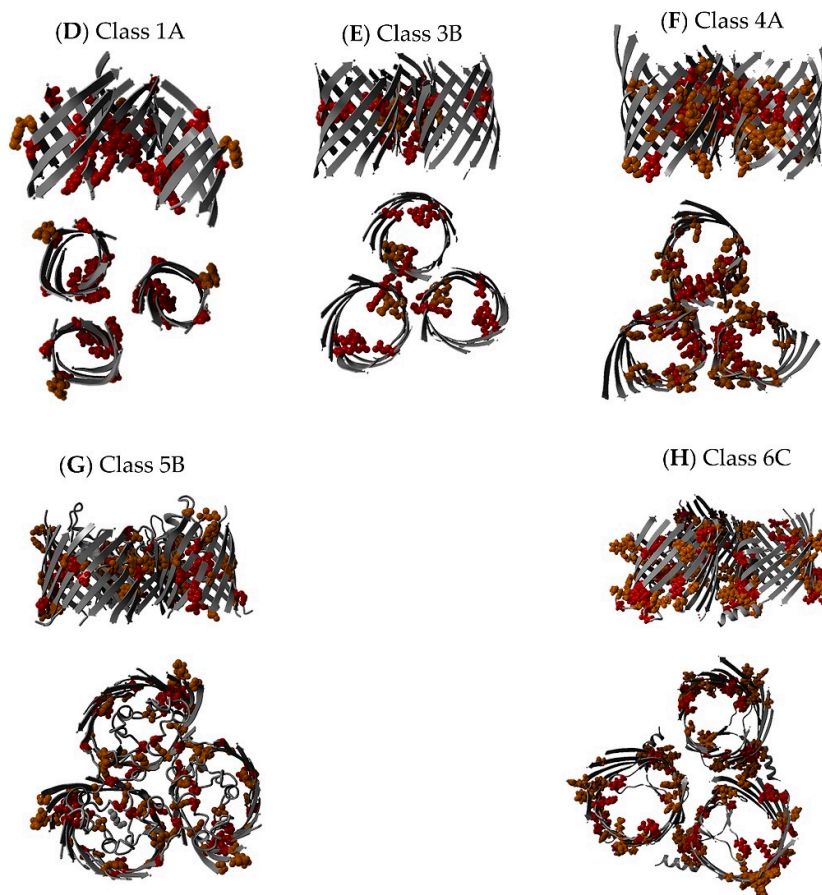(**G**) Class 5B　　　　　　　　　　　　(**H**) Class 6C

Figure 8. Non-specific porin and specific diffusion channel conservation. This figure illustrates differences in barrel size, pore size and loop variability. All structures are visualized in the trimeric form. Only the structures of Classes 5A (**A**); 5C (**B**); and 6B (**C**) proteins have been solved in an oligomeric conformation (all of them being homotrimeric). Class 1A (**D**); Class 3B (**E**); Class 4A (**F**); Class 5B (**G**) and Class 6C (**H**) are monomeric structures modeled in a trimeric conformation by hand. Same remark as the previous figures.

Trimeric structures

Class 5A

Class 5C

Class 6B

Trimeric Models

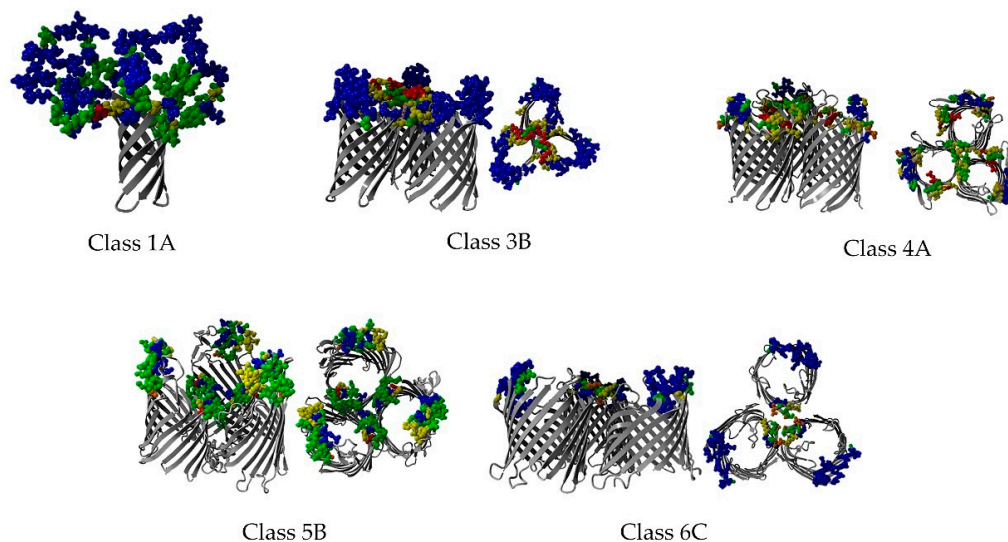Class 1A

Class 3B

Class 4A

Class 5B

Class 6C
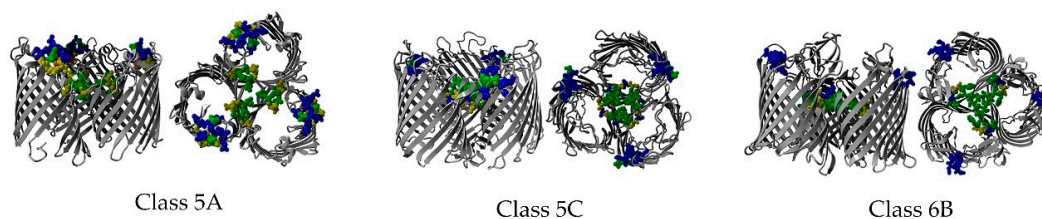
Figure 9. Porin variability. Side and top views displaying the loop variability of all protein subclasses analyzed (including the modeled trimeric interactions shown in Figure 8). All structures are visualized in the predicted trimeric form for Class 1A (2, 3, 5) and eight porins. Class 1A porins are represented in the monomeric state, and only conserved residues are highlighted (see the Discussion for more details). Red, orange and yellow represent conserved residues; green shows partially conserved residues involved in regulation; blue represents highly variable sites with unknown functions. Same as previous figures.

2.8. Non-Specific Porin and Specific Diffusion Channel Multimericity

Trimeric structures exist for Class 5A (GDP), Class 5C (OprP) and Class 6B (ScrY and maltoporins) (see Figure 8A–C). The trimeric models (Classes 1A, 3B, 4A, 5B and 6C; see Figure 8D–H) were handmade models using the YASARA-WHAT IF twinset that were used for illustrations [38,39]. Trimeric non-specific porins and specific diffusion channels were used as templates when possible (Classes 5C and 6B were used as templates to superpose three monomeric Class 5B and 6C structures). However, the barrels were rotated so that the conserved residues were facing each other, thereby illustrating the most likely orientation with regard to a trimeric interaction. The patterns of protein-protein interaction and multimericity in the different subclasses are illustrated in Figure 8D–H. No trimeric template is currently available for the trimeric motif alignment in Classes 1A, 3B and 4A; thus, the models of possible trimeric interactions among the monomeric non-specific porins and specific diffusion channels were manually generated (Figure 8D–F). The models were generated by manually aligning the most conserved residues (from the EVA) from each monomer to a forced, handmade, homotrimeric model.

The majority of the variable residues were found in the long extracellular loops that are located away from the oligomeric interaction interface. More variable residues were observed in the central loops in the larger barrels than in those in the smaller oligomers (Classes 1A to 4A and 5A). Presumably,

the larger barrels do not need these central loops for oligomerization as much as the smaller barrels. In some classes (*i.e.*, Classes 1A to 4A, and 5B), we even observed highly conserved (Boxes 11 and 12) residues in the central loops, indicating the substantial functional importance of these central loops. The central loops were usually shorter and less variable, while the distal loops were longer and more variable. This effect was more pronounced in non-specific porin and specific diffusion channel with smaller barrels (Classes 1A to 4A). We suggest that the short central loop L2 could be important for trimer stabilization, as has been shown for OmpF [40,41].

3. Discussion

3.1. Data Collection and Secondary Structure Composition

There has been an increase in the number of solved OMP structures; yet, the structures of many OMPs remain unsolved, and very few structures of OMP complexes have been determined [42]. Currently, only 0.1% of the protein structures deposited in the PDB belong to the OMP superfamily. Nearly 40% of the solved OMP structures are protein channels (non-specific porin or specific diffusion channel), indicating their importance in the bacterial protein field.

Despite the variations observed in protein size (the average length varied from 171 to 459 residues in KpOmpA and OccD3, respectively), the secondary structure composition reflected their transmembrane β-barrel motif (see the Introduction). The average 6:1:3 β-strand:periplasmic loop:extracellular loop residue ratio, shown in Table 2, confirmed that non-specific porins and specific diffusion channels are composed of a large number of β-strands that are connected by shorter periplasmic loops and longer extracellular loops.

3.2. Non-Specific Porin and Specific Diffusion Channel Phylogenetic Tree

The grouping of porin clades shown in Figure 3 reflects non-specific porin and specific diffusion channel function and size variations, rather than the current ideas about a particular porin's multimeric state. Non-specific monomeric trimeric porins (NTPs) formed three separate clades; α-Proteobacteria (*Rhodopseudomonas blastica* general diffusion porin (RbGDP) and *Rhodobacter capsulatus* general diffusion porin (RcGDP)); γ-Proteobacteria (*Salmonella* Typhimurium outer membrane protein C (StOmpC), PhoE, *Escherichia coli* outer membrane protein C (EcOmpC), OmpK36 and EcOmpF); and a clade with *Salmonella* Typhimurium outer membrane protein F (StOmpF) (γ-Proteobacteria) and Omp32 (β-Proteobacteria) that clustered with the non-specific smallest monomeric porins (OmpA). Two of the three specific trimeric channels (maltoporins) clustered together, while ScrY clustered with specific monomeric channels. OmpG (a non-specific monomeric rescue porin) clustered with maltoporins, which is not surprising, as it can mimic maltoporin uptake if needed [43]. Specific monomeric and specific trimeric *Pseudomonas* channels (OprB and OprP) clustered in one clade. Although function is the most likely explanation for these clades, size could also have an impact, as the largest porins (Occ channels) form a separate clade from the other eight- to 16-stranded β-barrels. Furthermore, the two OmpA sequences cluster together (eight-stranded β-barrel) as does the KdgM subclass (12-stranded β-barrel).

3.3. Non-Specific Porin and Specific Diffusion Channel Classification System

A naming scheme for protein families should be simple, robust and reproducible. It should also give space for future sequences to be added. We believe this has been accomplished for those proteins involved in the diffusion of hydrophilic molecules, which is based on available structure, sequence and literature information. This enabled us to analyze the six classes (with same β-barrel sizes) with eight different subclasses, looking at structure-function relationship. This classification system yields one ontology for those working with porins and specific diffusion channels. New information (new structures, new function studies) might require an updated schema with a more

detailed nomenclature than presented in this article. However, this classification system represents the groups of an OMP family that were required for the analyses performed.

3.4. Non-Specific Porin and Specific Diffusion Channel Structure Analysis

Non-specific porin and specific diffusion channel structures were used as a template to guide the MSA of all of the collected sequences (see the Methods Section for more details). The most unambiguous alignments tended to be obtained when only the barrels (with all loop residues removed) were aligned. Structure comparisons were possible for those subclasses containing more than one structure. Three of the eight subclasses have only one structure. Some of these alignments (e.g., Class 4A) resulted in fewer sequences than the optimum number required to harvest the full potential of the EV method used to analyze each protein subclasses. However, these alignments did yield relevant information, which will be discussed in light of other studies and published experiments. Table 1 lists the details of the structures found in each subclass (PDB ID, protein names, *etc.*).

Pore size may vary with environmental factors, e.g., pH-sensitivity of OmpF and OmpG channels [44,45]. Some of the structures analyzed have been proven to be in a closed state (e.g., OmpA, OccD1 and D2). Our pore size estimates complement literature findings (a complete list of literature estimates is found on the associated website). However, pore sizes analyzed using the whole protein revealed more variability within each subclass than the core structures (for those subclasses with more than one protein structure). Figure 3 shows that the pore size estimates correlate best with our classification system (based on barrel size, specificity and function).

The OmpF pore illustrates how estimations of the minimum core value can be useful. Experiments indicate that *E. coli* OmpF has a pore size of 10 to 12 Å. This would allow raffinose to pass through the pore [46]. Although this is about twice the structure-based pore size estimation using the entire protein, the minimum OmpF pore core radius is 6.2 Å. Removing the loops actually results in being closer to the estimates derived from laboratory experiments. This supports the theory of flexible loops affecting pore sizes [29,47–51]. The pore size calculations of the core protein are not occluded by any flexible loops. The constriction loop (usually loop L3) and extracellular loops govern the pore and limit pore size. These loops will shrink or increase the pore size in response to the continuously-changing environment. Having a minimum and maximum radius of a pore channel yields a better understanding of what might get through the pore.

3.5. Multiple Sequence Alignments

Each multiple sequence alignment (MSA) generated was based on available structure and sequence information (see Tables 1–4 for more details). The number of solved structures in each subclass varied from one to 15 (see Table 1 for a complete list of PDB IDs), and the number of sequences also varied greatly for each subclass. The generated MSAs were used to construct the EV plot and figures discussed below. Some of the subclasses held a less than optimal number of sequences for the EVA, although they contained enough variability to create sensible plots and figures.

3.6. Entropy-Variability Analysis

Each residue position in an aligned EV plot correlates with the residue′s function and structural characteristics. This sequence analysis technique is based on a combination of two commonly-used sequence variability measures. The first is variability, defined as the number of different amino acid types observed at each position. The second is Shannon entropy. Each residue position in the alignment is plotted on the EV plot. Boxes in this plot appear to represent groups of residues that share a common structural or functional characteristic. Conserved amino acids within a subclasses indicate that these residues are functionally important. The analysis is based on the collection of a large number of sequences that was used to filter the variability patterns. A profile alignment is used to identify conserved features in the structures. Finally, a plot is created based on where the residues are placed according to their structure-function characteristics. This plot is divided into five boxes

(Boxes 11, 12, 22, 23 and 33); each box contains the residues involved in the same functional category (see the Methods Section for more details) [32].

The most conserved residues that are important for protein functions are found in Box 11. Previous EV studies showed that highly-conserved residues (Box 11) are located in the active sites of proteins (e.g., G-protein-coupled receptors, globins, Ras-like proteins and proteases). Those residues supporting the active site were also quite conserved (Box 12). Box 22 contains the signal transducing residues between the modulator and main functions, while the residues found in Box 23 modulate the main function. The remaining residues that do not have any specific function are found in Box 33 [32,52,53].

According to the literature, the exterior loops of the non-specific porins and specific diffusion channels are continuously changing to avoid detection by the host immune system, phage invasion and as a response to ecological pressure. The exterior loops are the most variable regions in non-specific porins and specific diffusion channels, which reflects the adaptive traits accomplished through mutation or DNA rearrangement [10,54,55]. Both non-specific porins and specific diffusion channels had variable residues that were located mainly in extracellular loops. These proteins have long, highly variable, protruding loops, as well as more conserved, shorter loops that face the extracellular environment, as shown in Figure 8. This could be a mechanism in which the long, variable loops help to evade the host immune systems, while the shorter, conserved loops bind substrates [5,56]. For example, Class 1A is a multifunctional protein targeted by the immune system and is a bacteriophage receptor. These features are mainly conferred by the exterior loops [22]. Class 1A OmpA has probably adapted loop mutations to evade the immune system; e.g., loop L2 mutation may be the difference of an invasive and less invasive *E. coli* strain [57]. The fraction of variable, long loop residues is slightly lower in the specific diffusion channels (30%) than in the non-specific porins (42%) (see Table S3).

The conserved constriction loop residues observed in the EVA (for the larger barrels of Classes 5A–6C) support their importance in determining the substrate specificity. The presence of charged residues in the constriction loop (loop L3) of non-specific porins and specific diffusion channels creates an electrostatic field, which largely determines the permeability and ion selectivity of the pores [47,58,59].

This variability is mainly determined by differences in the number of sequences in the underlying MSA and by the average sequence identity between the sequences in the MSA. Nevertheless, a series of trends was clearly observable, but all classes have both conserved residues facing the core and the lipid membrane. These are important residues for protein function, as they are involved in either substrate or protein–protein interactions. It was not possible to directly discriminate between the monomeric and trimeric structures from these EV analyses, but mapping the residues in Boxes 11, 12 and 22 onto the protein structures revealed that all non-specific porins and specific diffusion channels have important residues pointing away from the core of the barrel. The only function imaginable for these residues is in protein–protein interactions. This observation strongly suggests that neither monomeric non-specific porins, nor specific diffusion channels exist. Figures 6 and 7 summarize the EV results for the barrel- and lipid-facing residues. These figures also show the observed differences in core size between the eight classes analyzed.

3.7. Non-Specific Porin and Specific Diffusion Channel Multimericity

The conserved residues facing the barrel core are likely to be important for pore activity. The conserved, lipid-facing residues are likely to be involved in protein-protein interactions, e.g., trimerization. Recent publications with laboratory data have verified these results (see Table 6). Loops involved in trimerization have a high proportion of residues in Boxes 22 and 23 and fewer residues in Box 33. Their interaction pattern is shown with the conserved residues in Figure 6A.

Table 6. Non-specific porin and specific diffusion channel oligomerization state. This table lists articles discussing the multimeric state of non-specific porins and specific diffusion channels, the protein name and information regarding the presumed multimeric state (monomeric, dimeric or oligomerization).

Class Number	Class Name	Function	Monomer	Dimer	Oligomer
1A	Non-specific, petite porin	Abundant multifunctional porin; host evasion [22]	EcOmpA [3,41,59–65], KpOmpA [56]	EcOmpA [26,27,66,67]	FopA [68], Pgm6/7 [69], EcOmpA [70]
3B	Oligogalacturonate-specific, small channel	Oligogalacturonate-specific channel [71]	KdgM [71–73], NanC [24]	NanC [74]	–
4A	Non-specific, intermediate porin	pH-dependent rare rescue porin [75]	EcOmpG [44,75–78]	EcOmpG [44]	–
5A	Non-specific, medium porin	Classical porins or general diffusion porin (GDP) [11,79]	–	–	GDP [3,4,16,80–86]
5B	Sugar-specific, medium channel	Oligosaccharide specific channel [23]	OprB [23,87].	–	OprB [88]
5C	Phosphate-specific, medium channel	Phosphate specific channel [89]	–	–	OprP [30,90], OprO/P heterotrimer [91]
6B	Sugar-specific, large channel	Sugar specific channel [31]	–	–	EcMaltoporin [31], StMaltoporin [92], ScrY [12]
6C	Carboxyl-specific, large channel	Small water-soluble specific channel (Occ channels) [29]	Occ [28,29], OccD1 (OprD) [93].	–	OccK1 (OpdK) [94], AbOprD [95]

AbOprB (*Acinetobacter baumannii* outer membrane porin B); EcMaltoporin (*Escherichia coli* maltoporin); EcOmpA (*E. coli* outer membrane protein A); EcOmpC (*E. coli* outer membrane protein C); EcOmpF (*E. coli* outer membrane protein F); KdgM (oligogalacturonate-specific channel); KpOmpA (*K. pneumonia* outer membrane protein A); NanC (*N*-acetylneuraminic acid-inducible outer-membrane channel); OccD (outer membrane carboxylate channel); Omp32 (outer membrane protein 32); OmpG (outer membrane protein G); OmpK36 (outer membrane porins of *K. pneumoniae*); OprB (outer membrane porin B); OprP (outer membrane porin P); PDB ID (Protein Data Bank Identifier); PhoE (phosphoporin); RbGDP (*R. blastica* general diffusion porin); RcGDP (*R. capsulatus* general diffusion porin); ScrY (Sugar specific porin); StMaltoporin (*S.* Typhium maltoporin); StOmpC (*S.* Typhimurium outer membrane protein C); StOmpF (*S.* Typhimurium outer membrane protein F).

Table 6 describes the eight subclasses used in our analyses (including MSAs and EVAs), and it provides information regarding the presumed multimeric state (monomeric, dimeric or oligomeric protein channels). Classical porins (belonging to Class 5A) have long been described as trimeric porins. Trimeric structures of Class 5C (OprP) and Class 6B (ScrY and maltoporins) porins have also been solved. Several publications have shown that the classic monomeric non-specific porins and specific diffusion channels actually oligomerized in the right conditions (e.g., lower temperatures or less detergents) [44,70,74,88,94–97]. Thus, unconventional methods might be necessary to detect protein oligomerization of non-specific porins and/or specific diffusion channels. These proteins can form less stable trimers compared to those in Classes 5A, 6 and 7. This is likely due to the environmental changes, such as high temperature and detergents, which has been shown to break protein subunit interactions. The weaker non-specific porin or specific diffusion channel oligomerization does not need to imply a non-essential function. The OM efflux pump TolC can be used as an example of this. One TolC subunit form 1/3 of a β-barrel; hence, a TolC trimer forms one β-barrel structure (a trimer is essential for a functional protein) when not boiled using SDS–PAGE [98].

Only one NMR structure is available from the PDB for Class 1A porins. The structures of the loops in this porin were not determined experimentally, which makes it nearly impossible to construct a trimeric model (and consequently, it is missing from Figure 7). However, we know that the structure contains four long loops ranging from 18 to 25 residues in length. The shortest loop (18 residues) contains many conserved residues, including one in Box 11 (see Table S3). This loop sequence variability analysis corroborates the idea that Class 1A porins are trimeric. Figure 8D–H highlight the possibility

that monomeric porins can form homotrimers. This model only shows three monomeric barrels, with the most conserved sites turned toward each other. The most conserved outward-facing residues are aligned to interact with each other to simulate the trimeric pattern observed in the EVA of the trimeric structures. The interaction would be better in a trimeric crystal (like those of the homotrimeric Class 5A structures), where the binding sites are optimized compared to those of the monomeric structures. Monomeric, dimeric and trimeric barrels may show different conformations in monomeric and oligomeric states, e.g., the outer membrane phospholipase A (OMPLA) and OprP structures [5,99].

There is much debate about Class 5B OprB multimericity. Based on the results from two-dimensional gel electrophoresis and liquid chromatography-tandem mass spectrometry analyses, Shrivastava *et al.* [88] concluded that *Pseudomonas* OprB-specific diffusion channels form homotrimers. This conclusion has been debated in the literature, and arguments for a monomeric protein have been made. The results of Shrivastava *et al.* did not take large micelles into account when calculating protein size, which according to van den Berg [23], would result in OprB forming monomers, rather than homotrimers. In addition, the OprB channel lacks the amino-terminal region found in the *Pseudomonas* OprP trimeric structure [23]. Stabilization of trimers by amino-terminal strand exchange has thus far only been identified in OprP trimerization [5,30]. This does not necessarily have to be a common trait only for *Pseudomonas* channels; it could be a common trait for all phosphate-specific channels. However, the motif aligner suggested that the amino-terminal OprB helices take part in protein–protein interactions. These preliminary results would also explain the conserved residues found in this region According to our analyses, Class 5B channels are also likely to be trimeric (see Figure 6G). OprB is similar to trimeric structures (e.g., those of Class 5A), and it has an asymmetrical extracellular structure. Class 5B channels exhibit a conservation pattern similar to that observed in other trimeric structures. This supports the findings by Shrivastava *et al.* [88]. The conserved residues found on the opposite side of the proposed trimeric interaction site reveal protein interactions that are important to this subclass. It is also possible that this is the site of trimerization, and the amino-terminal loops are involved in protein binding in the periplasmic space. Further studies are needed to better understand the protein interactions of non-specific porins and specific diffusion channels.

The periplasmic amino-terminal residues found in the Class 5C OprP channel structure have a stabilizing effect on the trimer [30]. This is likely to be a *Pseudomonas* sp.-specific feature that is not found throughout the Class 5C MSA (see Figure 8B). This indicates that trimerization in Class 5C channels is likely to be caused by other residues in the trimer interface. The conserved trimeric pattern of Class 6B channels is shown in Figure 8C. The conserved residues are involved in binding and loop specificity.

Regarding the possibility of Class 6C oligomerization, our results support the findings of Biswas *et al.* [93], in which polyacrylamide gel electrophoresis of OccD1 (OprD) revealed the presence of oligomers. They concluded that OccD1 might exist as a trimer (although the trimer was less stable than those of other homotrimeric channels). Figure 8H illustrates a homotrimeric channel based on the conformation of another protein structure, although it requires experimental validation (a manually-generated model would probably rotate the monomers into a potentially more stable conformation). Class 6C channels displayed a trimeric pattern in the EVA (see Text S1 for more information). Conserved residues were concentrated on the largest side of the asymmetric barrel facing out toward the lipid membrane. In addition, Class 6C channels had conserved residues facing the pore, thereby providing substrate specificity. High loop variability was observed on the amino-terminal end (facing the periplasm). The loops on the side of the barrel pointing away from the trimer interface may be involved in interactions with host proteins.

We generally found that the patterns observed for channels were highly conserved, indicating a specific protein-protein interaction. Protein-lipid interactions are seldom highly conserved, as there are many amino acids with similar properties [100]. The assembly machinery would require conserved residues for all classes; however, the pattern we observed was not identical between the eight subclasses

analyzed. The number of conserved residues and side-chain features varied, indicating that each subclass has its own interaction pattern, which supports the homotrimeric hypothesis.

Oligomeric proteins generate stable non-specific porins and specific diffusion channels that can withstand harsh environments. The oligomerization of OmpF occurs in a step-wise manner. Naveed *et al.* [101] demonstrated that these trimeric porins may also have stable monomeric and dimeric states. An oligomeric analysis of Class 1A OmpA porins showed that dimers and monomers were found in the same population, but the physiological role of the dimers remains unknown [27]. These results suggest that some porins may oligomerize only under certain physiological conditions. Further studies are required to fully understand when non-specific porin and specific diffusion channel oligomerize, but our analyses indicate that they all trimerize at some point.

4. Materials and Methods

4.1. Non-Specific Porin and Specific Diffusion Channel Multimericity

This is an *in silico* study based on the extraction of bacterial β-barrel OMP structures from the Protein Data Bank (PDB). The extraction was performed on February 2014. From this collection, non-specific porins and specific diffusion channels were manually selected and divided into four groups (see the numbered list in the Introduction). Two NMP structures are available from the PDB: KpOmpA (2K0L holds 20 NMR structures of this *K. pneumoniae* protein) and EcOmpA (1QJP holds an X-ray structure of this *E. coli* protein). The latter structure lacks 34 extracellular loop residues. Although both structures were used in the analyses, only KpOmpA (Model 2) was used in the structure superposition that guided the MSAs.

Structure and literature databases were searched for relevant information. We compared the current classification system with structural and functional information before a new schema was constructed. Search criteria included porins (specific and non-specific channels) and OMPs. All non-bacterial hits were filtered out. Literature searches included searching for each porin's name, structure and/or function to retrieve more information for each class. Relevant information discussing the function of porin structures was included.

4.2. Structure Analysis

All structure analyses were performed using the YASARA-WHAT IF twinset [38,39]. Water molecules, co-factors, substrates and lipids were removed, and YASARA's "clean" function was used on all structures. For trimeric molecules, only the first structure (labeled "A") in the PDB file was used for analyses that were not explicitly related to multimeric interactions. This molecule was renumbered before starting the analyses. Structure alignments were performed on the monomeric structures, using the MUSTANG (A multiple structural alignment algorithm) pairwise motif aligner [102] as implemented in YASARA. These structure analyses were compared to the pairwise sequence identity gathered using EBI Clustal Ω [103,104] (see Tables S1–S5 for an overview of the core and whole protein structure analyses and Table S2raw for the complete list of pairwise identities used to calculate the average values found in Table 2).

Pore sizes were estimated using the HOLE software (Version 2.2.004, SmartSci, Cambridge, UK) [37]. The minimum pore size was calculated for both the whole protein and core protein (where all loops and turns were removed) using default settings. Our associated web page also contains pore size estimation of poly-alanine mutated β-barrel structures (where loops and turns were removed).

4.3. Multiple Sequence Analysis

The workflow of the MSA is described by Kuipers *et al.* [105] and depicted in Figure 10. The first step in this workflow is data collection. This was accomplished through DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST algorithm) searches to collect the top 1000 sequences with an *e*-value <0.1 [103,104]. OMPdb [35] sequences were also included in the analyses. The second

step is an iterative profile alignment step using the WHAT IF module in the YASARA/WHAT IF twinset [32,38,39,106] (see Appendix A for more information). The collected sequences were aligned to the profile generated for each subclass.
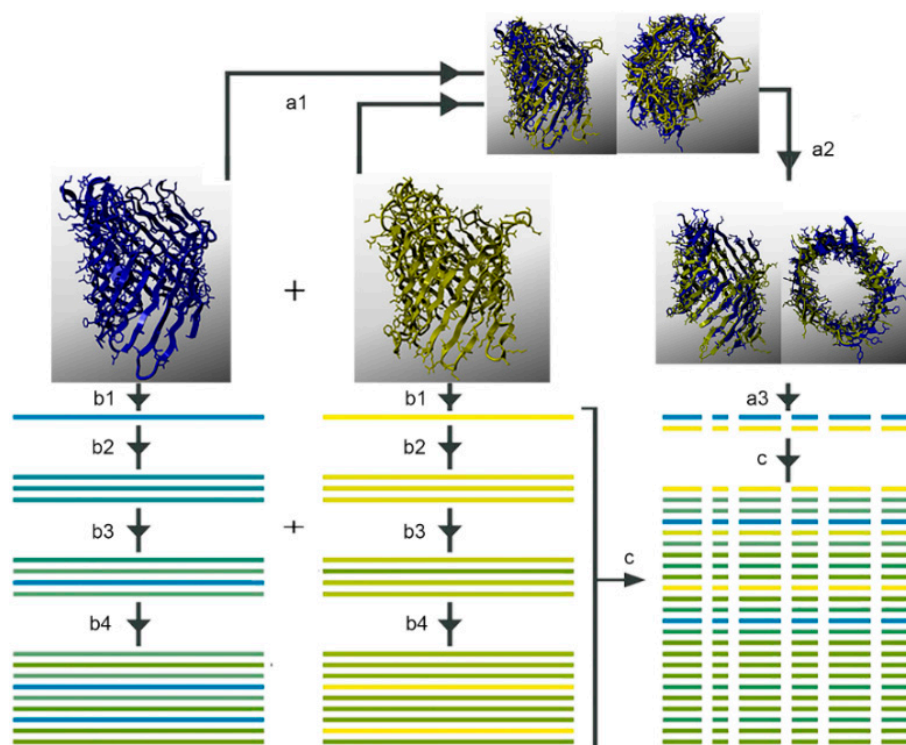

Figure 10. The sequence analysis workflow [102] (copied with permission from Bio-Prodict) shows the process of structure superposition guided by a profile alignment of the sequences. Step a1 shows the superposition of two structures (Omp32, colored blue, and PhoE in yellow; PDB IDs: 2FGQ and 1PHO, respectively); Step a2 shows the core alignment of the barrel structures in which loops have been removed; Step a3 is core alignment of similar structure; Step b is a sequence alignment against the nearest structure. All sequences and structures from one subclass are aligned together in step c. Both structures used in this MSA (multiple sequence alignment) belong to the general diffusion protein subclass. The WHAT IF module of the YASARA/WHAT IF twinset was used for these analyses [39].

4.4. Sequence Filtering

Sequence identities >90% were filtered out in WHAT IF using the DMATCH (asks for sequence identity cutoff value) option. This is an iterative algorithm that uses pairwise sequence comparisons (after sequence alignments) to identify and remove sequences with >90% sequence identities. Sequences with low identities (<20% to 30%) were also filtered out. This step removes little information, but throws out a lot of noise. It ensures an unbiased alignment and aids the entropy-variability calculations.

4.5. Entropy-Variability Analysis

EVAs were used to develop an evolutionary model [32]. The entropy and variability terms in the EVAs provide two different descriptions of the variability patterns for individual residue positions in the MSAs. The Shannon entropy (E_i) at each position is defined as $\Sigma P \log(P)$, where P represents the frequency of occurrence of a given amino acid in the MSA at position i. The variability V_i is the number of amino acid types observed (for more than 0.5%) at position i.

By plotting the entropy, E_i, *versus* the variability, V_i, for all residue positions i in the MSA, the EV plot can be divided into five sectors, each of which tends to contain residues mainly involved in

one broad functional category [32]. The low entropy and low variability box (Box 11), colored red, contains residues in the main active site; the intermediate entropy and low variability box (Box 12), colored orange, contains residues that support the structure of the main active site (often situated next to the red residues); the intermediate entropy and intermediate variability box (Box 22), colored yellow, contains residues involved in communication between the main active site and regulatory sites (further away from the core than the red and orange residues); the high entropy and intermediate variability box (Box 23), colored green, contains residues involved in the regulation of protein activity (modulators may be located on the surface or in the core of proteins); and the high entropy and high variability box (Box 33), colored blue, contains residues for which no function is known. EVA is based on well-established experimental methods from multiple, large protein families: globin chains, G-protein-coupled receptors, Ras-like proteins and serine-proteases [33,53]. Signal transduction residues have also been identified in other protein families, including the nuclear receptor family [52]. Figures 6 and 7 can be downloaded as YASARA scene files from the associated website. This allows users to see in 3D what we study. See the Supplementary Materials for more information.

4.6. Phylogenetic Analysis

An MSA of non-specific porins and specific diffusion channels with solved structures was generated (see Step b in the workflow found in Figure 10). The OccK3 channel sequence was used to guide the MSA. A maximum-likelihood phylogenetic tree was built with PhyML Version 3.0 (French National Institute of Bioinformatics, Montpellier, France) [107] using 1000 bootstrap replicates for branch support tests [108]. The Whelan and Goldman (WAG) model of amino acid substitution, with nine γ-distributed rate categories (WAG + G), and nearest-neighbor interchange methods of a tree topology search were used [107,109]. Molecular Evolutionary Genetics Analysis software Version 6.0 (MEGA 6.0, Biodesign Institute, Tempe, AZ, USA) [110] selected WAG + G to be the best-fit model [109]. The results were combined by "majority rule" consensus using Consense [111]. The final phylogenetic tree was visualized as unrooted radial trees in FigTree v 1.4.2 (Prof. Rambaut, Molecular Evolution, Phylogenetics and Epidemiology, University of Edinburgh, Edinburgh, Scotland).

5. Conclusions

EVA results were correlated to known structural and functional aspects of non-specific porins and specific diffusion channels, e.g., structural integrity, multimericity, specificity or biological niche adaptations. All non-specific porins have highly variable loops facing the environment, which strengthens the hypothesis that they have been subject to positive selection. Loop variability among many of the porins (especially Class 5A and Class 1A porins) supports the high mutation rate needed for events, such as host evasion. Loop variability is a pathogenic mechanism. However, conserved protein binding sites can also reveal cooperative interactions. Cooperative porin binding to bacterial pili and the human CR3 receptor has been observed during host invasion [112]. Our results indicate that all non-specific porins and specific diffusion channels are multimers, and the residues involved in protein-protein interactions can be determined in all eight classes analyzed. Trimeric interaction patterns were identified in non-specific porins and specific diffusion channels that are currently classed as monomeric, with conserved residues found at the interface of the trimeric interactions. EVA showed that conserved residues are concentrated in the trimeric interface facing inward and outward from the barrel. The loops of sugar- and phosphate-specific channels were more highly conserved than those of OmpA and other non-specific porins. This indicates that loop specificity might be needed in some porins (or specific diffusion channels) to "capture" the right molecules in the extracellular milieu. Not enough data are available for an extensive study of non-porin OMPs. Our preliminary work on OMPLA indicates that this protein is dimeric, as commonly stated in the literature [99,113,114].

Most classification systems that exist today (including CATH [33], SCOPe [36], TCDB [34] and OMPdb [35] classifications) rely on a division between monomeric and trimeric non-specific porins

and specific diffusion channels. We suggest that it is time to discuss a new classification scheme that discriminates between structure and function, rather than relying upon a classification system based on multimeric states. The three trimeric non-specific porin and specific diffusion channel families that exist today are the sugar-specific channels, the phosphate-specific channels and the GDPs (porins). Evidence that many of these proteins exist in multimeric states is emerging in the literature (see Table 6 for references). Our data support these findings, which suggest that most non-specific porins and specific diffusion channels are multimeric and are likely to form homotrimeric interactions.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/17/4/599/s1. Our associated website (www.hildevollan.no/publications/porins) will be updated with relevant analyses, an abstract video, and 3D structures (that can be viewed by downloading YASARA View free from www.yasara.org).

Author Contributions: Hilde S. Vollan planned the study design and performed all the bioinformatic analyses; Tone Tannæs planned the study design; Gert Vriend supervised the bioinformatic procedures and participated in planning the study design; Geir Bukholm supervised the study. All authors read and approved the final article.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

AbOprD	*Acinetobacter baumannii* OprD
EcMaltoporin	*E. coli* maltoporin
EcOmp	*E. coli* outer membrane protein
EV	entropy-variability
EVA	entropy-variability analysis
GDP	general diffusion porin
KpOmpA	*K. pneumoniae* outer membrane protein A
MSA	multiple sequence alignment
NanC	*N*-acetylneuraminic acid-inducible outer-membrane channel
NMP	non-specific monomeric porin
NTP	non-specific trimeric porin
Occ	outer membrane carboxylate channel
OMP	outer membrane protein
OmpK36	outer membrane porin of *K. pneumoniae*
OMPLA	outer membrane phospholipase A
Opr	outer membrane porin
PhoE	Phosphoporin
PDB	Protein Data Bank
RbGDP	*R. blastica* general diffusion porin
RcGDP	*R. capsulatus* general diffusion porin
RMSD	root-mean-square deviation
SMDC	specific monomeric diffusion channel
Salmonella Typhimurium	*Salmonella enterica* serovar Typhimurium
StMaltoporin	*Salmonella* Typhimurium maltoporin
StOmp	*Salmonella* Typhimurium outer membrane protein
STDC	specific trimeric diffusion channel

Appendix A

Iterative Profile Alignment Method

The commands used to perform an iterative profile alignment in WHAT IF is shown in Figure A1.

	Walign	// More info about this menu: http://swift.cmbi.ru.nl/whatif/html/chap46.html
	Getpir	// Import the template porin sequence.
Initial profile Based on the template structure	Makprf	// Generate a profile based on template sequence. Edit profile manually.
	Getprf	// Retrieve the manually updated profile.
	Getseq	// Import sequences form other porin structure (of same class).
	Aliprf	// Aligns sequences to a profile.
	Srtpct	// Sorts sequences by similarity to a profile.
Second profile Based on all available structures	Makprf	// Make an updated profile with the added sequences from known structures.
	Getseq	// Get the collected sequences (for the porin class to be analyzed).
	Kildbl	// Deletes identical sequences.
	Aliprf	// Aligns all sequences to a profile.
Iterative step Repeated until the final profile is generated	Srtpct	// Sorts sequences by similarity to a profile.
	Updprf	// Asks for a sequence range. Updates the profile.
	Orgbck	// Recovers the original sequences .
Other useful options	Dmatch	// Asks for sequence identity cutoff (e.g. 90%).
	Mview	// Shows a MSA of all sequences aligned to a profile.
	Makhsp	// This generates alignments modelled after hssp procedures [http://swift.cmbi.ru.nl/gv/hssp].
	Wavsho	// EVA results (table output).
	Walgra	// Activates the graphical sequence menu.
	Wavgra	// Generates EVA plot.

Figure A1. This shows the commands used to generate a profile in WHAT IF. This method was used to generate the multiple sequence alignment (MSA) and to run the entropy-variability analysis (EVA) discussed in the article.

References

1. Gupta, R.S. Origin of diderm (Gram-negative) bacteria: Antibiotic selection pressure rather than endosymbiosis likely led to the evolution of bacterial cells with two membranes. *Antonie Van Leeuwenhoek* **2011**, *100*, 171–182. [CrossRef] [PubMed]
2. Bos, M.P.; Robert, V.; Tommassen, J. Biogenesis of the gram-negative bacterial outer membrane. *Annu. Rev. Microbiol.* **2007**, *61*, 191–214. [CrossRef] [PubMed]
3. Koebnik, R.; Locher, K.P.; van Gelder, P. Structure and function of bacterial outer membrane proteins: Barrels in a nutshell. *Mol. Microbiol.* **2000**, *37*, 239–253. [CrossRef] [PubMed]
4. Baslé, A.; Rummel, G.; Storici, P.; Rosenbusch, J.P.; Schirmer, T. Crystal structure of osmoporin OmpC from *E. coli* at 2.0 a. *J. Mol. Biol.* **2006**, *362*, 933–942. [CrossRef] [PubMed]
5. Galdiero, S.; Falanga, A.; Cantisani, M.; Tarallo, R.; Della Pepa, M.E.; D'Oriano, V.; Galdiero, M. Microbe-host interactions: Structure and role of gram-negative bacterial porins. *Curr. Protein Pept. Sci.* **2012**, *13*, 843–854. [CrossRef] [PubMed]
6. Zakharov, S.D.; Eroukova, V.Y.; Rokitskaya, T.I.; Zhalnina, M.V.; Sharma, O.; Loll, P.J.; Zgurskaya, H.I.; Antonenko, Y.N.; Cramer, W.A. Colicin occlusion of OmpF and TolC channels: Outer membrane translocons for colicin import. *Biophys. J.* **2004**, *87*, 3901–3911. [CrossRef] [PubMed]
7. Housden, N.G.; Wojdyla, J.A.; Korczynska, J.; Grishkovskaya, I.; Kirkpatrick, N.; Brzozowski, A.M.; Kleanthous, C. Directed epitope delivery across the *Escherichia coli* outer membrane through the porin OmpF. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 21412–21417. [CrossRef] [PubMed]
8. Nikaido, H. Molecular basis of bacterial outer membrane permeability revisited. *Microbiol. Mol. Biol. Rev.* **2003**, *67*, 593–656. [CrossRef] [PubMed]

9. Lin, J.; Huang, S.; Zhang, Q. Outer membrane proteins: Key players for bacterial adaptation in host niches. *Microb. Infect.* **2002**, *4*, 325–331. [CrossRef]

10. Achouak, W.; Heulin, T.; Pages, J.M. Multiple facets of bacterial porins. *FEMS Microbiol. Lett.* **2001**, *199*, 1–7. [CrossRef] [PubMed]

11. Pages, J.M.; James, C.E.; Winterhalter, M. The porin and the permeating antibiotic: A selective diffusion barrier in gram-negative bacteria. *Nat. Rev. Microbiol.* **2008**, *6*, 893–903. [CrossRef] [PubMed]

12. Forst, D.; Welte, W.; Wacker, T.; Diederichs, K. Structure of the sucrose-specific porin ScrY from *Salmonella typhimurium* and its complex with sucrose. *Nat. Struct. Biol.* **1998**, *5*, 37–46. [CrossRef] [PubMed]

13. Schirmer, T. General and specific porins from bacterial outer membranes. *J. Struct. Biol.* **1998**, *121*, 101–109. [CrossRef] [PubMed]

14. Faraldo-Gómez, J.D.; Sansom, M.S. Acquisition of siderophores in gram-negative bacteria. *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 105–116. [CrossRef] [PubMed]

15. Nikaido, H. Porins and specific channels of bacterial outer membranes. *Mol. Microbiol.* **1992**, *6*, 435–442. [CrossRef] [PubMed]

16. Welte, W.; Nestel, U.; Wacker, T.; Diederichs, K. Structure and function of the porin channel. *Kidney Int.* **1995**, *48*, 930–940. [CrossRef] [PubMed]

17. Aguilella, V.M.; Queralt-Martín, M.; Alcaraz, A. Bacterial porins. In *Electrophysiology of Unconventional Channels and Pores*; Delcour, H.A., Ed.; Springer International Publishing: Cham, Albania, 2015; pp. 101–121.

18. Jap, B.K.; Walian, P.J. Structure and functional mechanism of porins. *Physiol. Rev.* **1996**, *76*, 1073–1088. [PubMed]

19. Cox, G.; Wright, G.D. Intrinsic antibiotic resistance: Mechanisms, origins, challenges and solutions. *Int. J. Med. Microbiol.* **2013**, *303*, 287–292. [CrossRef] [PubMed]

20. Freixeiro, P.; Dieguez-Casal, E.; Costoya, L.; Marzoa, J.; Ferreiros, C.M.; Criado, M.T.; Sanchez, S. High resolution clear native electrophoresis (hrCNE) allows a detailed analysis of the heterotrimeric structure of recombinant neisseria meningitidis porins inserted into liposomes. *J. Proteome Res.* **2013**, *12*, 777–784. [CrossRef] [PubMed]

21. Schulz, G.E. The structure of bacterial outer membrane proteins. *Biochim. Biophys. Acta* **2002**, *1565*, 308–317. [CrossRef]

22. Smith, S.G.; Mahon, V.; Lambert, M.A.; Fagan, R.P. A molecular swiss army knife: OmpA structure, function and expression. *FEMS Microbiol. Lett.* **2007**, *273*, 1–11. [CrossRef] [PubMed]

23. Van den Berg, B. Structural basis for outer membrane sugar uptake in pseudomonads. *J. Biol. Chem.* **2012**, *287*, 41044–41052. [CrossRef] [PubMed]

24. Wirth, C.; Condemine, G.; Boiteux, C.; Berneche, S.; Schirmer, T.; Peneff, C.M. NanC crystal structure, a model for outer-membrane channels of the acidic sugar-specific KdgM porin family. *J. Mol. Biol.* **2009**, *394*, 718–731. [CrossRef] [PubMed]

25. Sharma, M. *Facilitating Folding of Outer Membrane Proteins, Roles of the Periplasmic Chaperone Skp and the Outer Membrane Lipoprotein Bamd*; Universität Kassel: Kassel, Germany, 2014.

26. Whitelegge, J. Gas-phase structure of the *E. coli* OmpA dimer. *Structure* **2014**, *22*, 666–667. [CrossRef] [PubMed]

27. Marcoux, J.; Politis, A.; Rinehart, D.; Marshall, D.P.; Wallace, M.I.; Tamm, L.K.; Robinson, C.V. Mass spectrometry defines the C-terminal dimerization domain and enables modeling of the structure of full-length OmpA. *Structure* **2014**, *22*, 781–790. [CrossRef] [PubMed]

28. Eren, E.; Vijayaraghavan, J.; Liu, J.; Cheneke, B.R.; Touw, D.S.; Lepore, B.W.; Indic, M.; Movileanu, L.; van den Berg, B. Substrate specificity within a family of outer membrane carboxylate channels. *PLoS Biol.* **2012**, *10*, e1001242. [CrossRef] [PubMed]

29. Liu, J.; Eren, E.; Vijayaraghavan, J.; Cheneke, B.R.; Indic, M.; van den Berg, B.; Movileanu, L. OccK channels from *Pseudomonas aeruginosa* exhibit diverse single-channel electrical signatures but conserved anion selectivity. *Biochemistry* **2012**, *51*, 2319–2330. [CrossRef] [PubMed]

30. Moraes, T.F.; Bains, M.; Hancock, R.E.; Strynadka, N.C. An arginine ladder in OprP mediates phosphate-specific transfer across the outer membrane. *Nat. Struct. Mol. Biol.* **2007**, *14*, 85–87. [CrossRef] [PubMed]

31. Wang, Y.F.; Dutzler, R.; Rizkallah, P.J.; Rosenbusch, J.P.; Schirmer, T. Channel specificity: Structural basis for sugar discrimination and differential flux rates in maltoporin. *J. Mol. Biol.* **1997**, *272*, 56–63. [CrossRef] [PubMed]

32. Oliveira, L.; Paiva, P.B.; Paiva, A.C.; Vriend, G. Identification of functionally conserved residues with the use of entropy-variability plots. *Proteins* **2003**, *52*, 544–552. [CrossRef] [PubMed]

33. Sillitoe, I.; Lewis, T.E.; Cuff, A.; Das, S.; Ashford, P.; Dawson, N.L.; Furnham, N.; Laskowski, R.A.; Lee, D.; Lees, J.G.; *et al.* CATH: Comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res.* **2015**, *43*, D376–D381. [CrossRef] [PubMed]

34. Saier, M.H., Jr.; Reddy, V.S.; Tamang, D.G.; Vastermark, A. The transporter classification database. *Nucleic Acids Res.* **2014**, *42*, D251–D258. [CrossRef] [PubMed]

35. Tsirigos, K.D.; Bagos, P.G.; Hamodrakas, S.J. OMPdb: A database of β-barrel outer membrane proteins from gram-negative bacteria. *Nucleic Acids Res.* **2011**, *39*, D324–D331. [CrossRef] [PubMed]

36. Fox, N.K.; Brenner, S.E.; Chandonia, J.M. Scope: Structural classification of proteins-extended, integrating scop and astral data and classification of new structures. *Nucleic Acids Res.* **2014**, *42*, D304–D309. [CrossRef] [PubMed]

37. Smart, O.S.; Neduvelil, J.G.; Wang, X.; Wallace, B.A.; Sansom, M.S. HOLE: A program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* **1996**, *14*, 354–360. [CrossRef]

38. Krieger, E.; Koraimann, G.; Vriend, G. Increasing the precision of comparative models with yasara nova—A self-parameterizing force field. *Proteins* **2002**, *47*, 393–402. [CrossRef] [PubMed]

39. Vriend, G. What if: A molecular modeling and drug design program. *J. Mol. Graph.* **1990**, *8*, 52–56. [CrossRef]

40. Phale, P.S.; Philippsen, A.; Kiefhaber, T.; Koebnik, R.; Phale, V.P.; Schirmer, T.; Rosenbusch, J.P. Stability of trimeric OmpF porin: The contributions of the latching loop l2. *Biochemistry* **1998**, *37*, 15663–15670. [CrossRef] [PubMed]

41. Tamm, L.K.; Hong, H.; Liang, B. Folding and assembly of β-barrel membrane proteins. *Biochim. Biophys. Acta* **2004**, *1666*, 250–263. [CrossRef] [PubMed]

42. Fairman, J.W.; Noinaj, N.; Buchanan, S.K. The structural biology of β-barrel membrane proteins: A summary of recent reports. *Curr. Opin. Struct. Biol.* **2011**, *21*, 523–531. [CrossRef] [PubMed]

43. Subbarao, G.V.; van den Berg, B. Crystal structure of the monomeric porin OmpG. *J. Mol. Biol.* **2006**, *360*, 750–759. [CrossRef] [PubMed]

44. Mari, S.A.; Köster, S.; Bippes, C.A.; Yildiz, O.; Kühlbrandt, W.; Muller, D.J. Ph-induced conformational change of the β-barrel-forming protein OmpG reconstituted into native *E. coli* lipids. *J. Mol. Biol.* **2010**, *396*, 610–616. [CrossRef] [PubMed]

45. Rokitskaya, T.I.; Kotova, E.A.; Naberezhnykh, G.A.; Khomenko, V.A.; Gorbach, V.I.; Firsov, A.M.; Zelepuga, E.A.; Antonenko, Y.N.; Novikova, O.D. Single channel activity of OmpF-like porin from yersinia pseudotuberculosis. *Biochim. Biophys. Acta* **2016**, *1858*, 883–891. [CrossRef] [PubMed]

46. Rostovtseva, T.K.; Nestorovich, E.M.; Bezrukov, S.M. Partitioning of differently sized poly(ethylene glycol)s into OmpF porin. *Biophys. J.* **2002**, *82*, 160–169. [CrossRef]

47. Besya, A.B.; Mobasheri, H.; Ejtehadi, M.R. Gating and conduction of nano-channel forming proteins: A computational approach. *J. Biomol. Struct. Dyn.* **2013**, *31*, 818–828. [CrossRef] [PubMed]

48. Tamber, S.; Ochs, M.M.; Hancock, R.E. Role of the novel OprD family of porins in nutrient uptake in *Pseudomonas aeruginosa*. *J. Bacteriol.* **2006**, *188*, 45–54. [CrossRef] [PubMed]

49. Niramitranon, J.; Sansom, M.S.; Pongprayoon, P. Why do the outer membrane proteins OmpF from *E. coli* and OprP from *P. aeruginosa* prefer trimers? Simulation studies. *J. Mol. Graph. Model.* **2016**, *65*, 1–7. [CrossRef] [PubMed]

50. Patel, D.S.; Re, S.; Wu, E.L.; Qi, Y.; Klebba, P.E.; Widmalm, G.; Yeom, M.S.; Sugita, Y.; Im, W. Dynamics and interactions of OmpF and LPS: Influence on pore accessibility and ion permeability. *Biophys. J.* **2016**, *110*, 930–938. [CrossRef] [PubMed]

51. Huang, H.; Hancock, R.E. The role of specific surface loop regions in determining the function of the imipenem-specific pore protein oprd of *Pseudomonas aeruginosa*. *J. Bacteriol.* **1996**, *178*, 3085–3090. [PubMed]

52. Folkertsma, S.; van Noort, P.; Van Durme, J.; Joosten, H.J.; Bettler, E.; Fleuren, W.; Oliveira, L.; Horn, F.; de Vlieg, J.; Vriend, G. A family-based approach reveals the function of residues in the nuclear receptor ligand-binding domain. *J. Mol. Biol.* **2004**, *341*, 321–335. [CrossRef] [PubMed]

53. Oliveira, L.; Paiva, P.B.; Paiva, A.C.; Vriend, G. Sequence analysis reveals how g protein-coupled receptors transduce the signal to the g protein. *Proteins* **2003**, *52*, 553–560. [CrossRef] [PubMed]

54. LeClerc, J.E.; Li, B.; Payne, W.L.; Cebula, T.A. High mutation frequencies among *Escherichia coli* and salmonella pathogens. *Science* **1996**, *274*, 1208–1211. [CrossRef] [PubMed]

55. Easton, D. Bacterial outer membrane porins—Are they decoy agents. *Aust. J. Med. Sci.* **2005**, *26*, 1038–1643.

56. Iordanov, I.; Renault, M.; Reat, V.; Bosshart, P.D.; Engel, A.; Saurel, O.; Milon, A. Dynamics of *Klebsiella pneumoniae* OmpA transmembrane domain: The four extracellular loops display restricted motion behavior in micelles and in lipid bilayers. *Biochim. Biophys. Acta* **2012**, *1818*, 2344–2353. [CrossRef] [PubMed]

57. Power, M.L.; Ferrari, B.C.; Littlefield-Wyer, J.; Gordon, D.M.; Slade, M.B.; Veal, D.A. A naturally occurring novel allele of *Escherichia coli* outer membrane protein A reduces sensitivity to bacteriophage. *Appl. Environ. Microbiol.* **2006**, *72*, 7930–7932. [CrossRef] [PubMed]

58. Pezeshki, S.; Chimerel, C.; Bessonov, A.N.; Winterhalter, M.; Kleinekathofer, U. Understanding ion conductance on a molecular level: An all-atom modeling of the bacterial porin OmpF. *Biophys. J.* **2009**, *97*, 1898–1906. [CrossRef] [PubMed]

59. Eppens, E.F.; Saint, N.; van Gelder, P.; van Boxtel, R.; Tommassen, J. Role of the constriction loop in the gating of outer membrane porin phoe of *Escherichia coli*. *FEBS Lett.* **1997**, *415*, 317–320. [CrossRef]

60. Sugawara, E.; Nikaido, H. Pore-forming activity of OmpA protein of *Escherichia coli*. *J. Biol. Chem.* **1992**, *267*, 2507–2511. [PubMed]

61. Pautsch, A.; Schulz, G.E. Structure of the outer membrane protein A transmembrane domain. *Nat. Struct. Biol.* **1998**, *5*, 1013–1017. [CrossRef] [PubMed]

62. Pautsch, A.; Schulz, G.E. High-resolution structure of the OmpA membrane domain. *J. Mol. Biol.* **2000**, *298*, 273–282. [CrossRef] [PubMed]

63. Kleinschmidt, J.H. Membrane protein folding on the example of outer membrane protein A of *Escherichia coli*. *Cell. Mol. Life Sci.* **2003**, *60*, 1547–1558. [CrossRef] [PubMed]

64. Dieguez-Casal, E.; Freixeiro, P.; Costoya, L.; Criado, M.T.; Ferreiros, C.; Sanchez, S. High resolution clear native electrophoresis is a good alternative to blue native electrophoresis for the characterization of the *Escherichia coli* membrane complexes. *J. Microbiol. Methods* **2014**, *102*, 45–54. [CrossRef] [PubMed]

65. Stenberg, F.; Chovanec, P.; Maslen, S.L.; Robinson, C.V.; Ilag, L.L.; von Heijne, G.; Daley, D.O. Protein complexes of the escherichia coli cell envelope. *J. Biol. Chem.* **2005**, *280*, 34409–34419. [CrossRef] [PubMed]

66. Dovling Kaspersen, J.; Moestrup Jessen, C.; Stougaard Vad, B.; Skipper Sorensen, E.; Kleiner Andersen, K.; Glasius, M.; Pinto Oliveira, C.L.; Otzen, D.E.; Pedersen, J.S. Low-resolution structures of OmpA*DDM protein-detergent complexes. *Chembiochem* **2014**, *15*, 2113–2124. [CrossRef] [PubMed]

67. Patel, G.J.; Kleinschmidt, J.H. The lipid bilayer-inserted membrane protein BamA of *Escherichia coli* facilitates insertion and folding of outer membrane protein A from its complex with Skp. *Biochemistry* **2013**, *52*, 3974–3986. [CrossRef] [PubMed]

68. Manchur, M.A.; Kikumoto, M.; Kanao, T.; Takada, J.; Kamimura, K. Characterization of an OmpA-like outer membrane protein of the acidophilic iron-oxidizing bacterium, acidithiobacillus ferrooxidans. *Extremophiles* **2011**, *15*, 403–410. [CrossRef] [PubMed]

69. Nagano, K.; Read, E.K.; Murakami, Y.; Masuda, T.; Noguchi, T.; Yoshimura, F. Trimeric structure of major outer membrane proteins homologous to OmpA in *Porphyromonas gingivalis*. *J. Bacteriol.* **2005**, *187*, 902–911. [CrossRef] [PubMed]

70. Wang, H.; Andersen, K.K.; Vad, B.S.; Otzen, D.E. OmpA can form folded and unfolded oligomers. *Biochim. Biophys. Acta* **2013**, *1834*, 127–136. [CrossRef] [PubMed]

71. Hutter, C.A.; Lehner, R.; Wirth, C.; Condemine, G.; Peneff, C.; Schirmer, T. Structure of the oligogalacturonate-specific KdgM porin. *Acta Crystallogr.* **2014**, *70*, 1770–1778. [CrossRef] [PubMed]

72. Pellinen, T.; Ahlfors, H.; Blot, N.; Condemine, G. Topology of the erwinia chrysanthemi oligogalacturonate porin KdgM. *Biochem. J.* **2003**, *372*, 329–334. [CrossRef] [PubMed]

73. Blot, N.; Berrier, C.; Hugouvieux-Cotte-Pattat, N.; Ghazi, A.; Condemine, G. The oligogalacturonate-specific porin KdgM of erwinia chrysanthemi belongs to a new porin family. *J. Biol. Chem.* **2002**, *277*, 7936–7944. [CrossRef] [PubMed]

74. Dunton, T.A.; Goose, J.E.; Gavaghan, D.J.; Sansom, M.S.; Osborne, J.M. The free energy landscape of dimerization of a membrane protein, NanC. *PLoS Comput. Biol.* **2014**, *10*, e1003417. [CrossRef] [PubMed]

75. Yildiz, O.; Vinothkumar, K.R.; Goswami, P.; Kühlbrandt, W. Structure of the monomeric outer-membrane porin OmpG in the open and closed conformation. *EMBO J.* **2006**, *25*, 3702–3713. [CrossRef] [PubMed]

76. Zhuang, T.; Chisholm, C.; Chen, M.; Tamm, L.K. NMR-based conformational ensembles explain pH-gated opening and closing of OmpG channel. *J. Am. Chem. Soc.* **2013**, *135*, 15101–15113. [CrossRef] [PubMed]

77. Liang, B.; Tamm, L.K. Structure of outer membrane protein G by solution nmr spectroscopy. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 16140–16145. [CrossRef] [PubMed]

78. Fajardo, D.A.; Cheung, J.; Ito, C.; Sugawara, E.; Nikaido, H.; Misra, R. Biochemistry and regulation of a novel *Escherichia coli* K-12 porin protein, OmpG, which produces unusually large channels. *J. Bacteriol.* **1998**, *180*, 4452–4459. [PubMed]

79. Schiltz, E.; Kreusch, A.; Nestel, U.; Schulz, G.E. Primary structure of porin from *Rhodobacter capsulatus*. *Eur. J. Biochem. FEBS* **1991**, *199*, 587–594. [CrossRef]

80. Balasubramaniam, D.; Arockiasamy, A.; Kumar, P.D.; Sharma, A.; Krishnaswamy, S. Asymmetric pore occupancy in crystal structure of OmpF porin from *Salmonella typhi*. *J. Struct. Biol.* **2012**, *178*, 233–244. [CrossRef] [PubMed]

81. Dutzler, R.; Rummel, G.; Alberti, S.; Hernandez-Alles, S.; Phale, P.; Rosenbusch, J.; Benedi, V.; Schirmer, T. Crystal structure and functional characterization of OmpK36, the osmoporin of *Klebsiella pneumoniae*. *Structure* **1999**, *7*, 425–434. [CrossRef]

82. Weiss, M.S.; Schulz, G.E. Structure of porin refined at 1.8 Å resolution. *J. Mol. Biol.* **1992**, *227*, 493–509. [CrossRef]

83. Efremov, R.G.; Sazanov, L.A. Structure of *Escherichia coli* OmpF porin from lipidic mesophase. *J. Struct. Biol.* **2012**, *178*, 311–318. [CrossRef] [PubMed]

84. Zachariae, U.; Kluhspies, T.; De, S.; Engelhardt, H.; Zeth, K. High resolution crystal structures and molecular dynamics studies reveal substrate binding in the porin Omp32. *J. Biol. Chem.* **2006**, *281*, 7413–7420. [CrossRef] [PubMed]

85. Cowan, S.W.; Schirmer, T.; Rummel, G.; Steiert, M.; Ghosh, R.; Pauptit, R.A.; Jansonius, J.N.; Rosenbusch, J.P. Crystal structures explain functional properties of two *E. coli* porins. *Nature* **1992**, *358*, 727–733. [CrossRef] [PubMed]

86. Kreusch, A.; Neubuser, A.; Schiltz, E.; Weckesser, J.; Schulz, G.E. Structure of the membrane channel porin from *Rhodopseudomonas blastica* at 2.0 Å resolution. *Protein Sci.* **1994**, *3*, 58–63. [CrossRef] [PubMed]

87. Adewoye, L.O.; Tschetter, L.; O'Neil, J.; Worobec, E.A. Channel specificity and secondary structure of the glucose-inducible porins of *Pseudomonas* spp. *J. Bioenerg. Biomembr.* **1998**, *30*, 257–267. [CrossRef] [PubMed]

88. Shrivastava, R.; Basu, B.; Godbole, A.; Mathew, M.K.; Apte, S.K.; Phale, P.S. Repression of the glucose-inducible outer-membrane protein OprB during utilization of aromatic compounds and organic acids in *Pseudomonas putida* CSV86. *Microbiology* **2011**, *157*, 1531–1540. [CrossRef] [PubMed]

89. Pongprayoon, P.; Beckstein, O.; Wee, C.L.; Sansom, M.S. Simulations of anion transport through OprP reveal the molecular basis for high affinity and selectivity for phosphate. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 21614–21618. [CrossRef] [PubMed]

90. Angus, B.L.; Hancock, R.E. Outer membrane porin proteins F, P, and D1 of *Pseudomonas aeruginosa* and PhoE of *Escherichia coli*: Chemical cross-linking to reveal native oligomers. *J. Bacteriol.* **1983**, *155*, 1042–1051. [PubMed]

91. Hancock, R.E.; Egli, C.; Benz, R.; Siehnel, R.J. Overexpression in *Escherichia coli* and functional analysis of a novel PPi-selective porin, oprO, from *Pseudomonas aeruginosa*. *J. Bacteriol.* **1992**, *174*, 471–476. [PubMed]

92. Meyer, J.E.; Hofnung, M.; Schulz, G.E. Structure of maltoporin from salmonella typhimurium ligated with a nitrophenyl-maltotrioside. *J. Mol. Biol.* **1997**, *266*, 761–775. [CrossRef] [PubMed]

93. Biswas, S.; Mohammad, M.M.; Patel, D.R.; Movileanu, L.; van den Berg, B. Structural insight into OprD substrate specificity. *Nat. Struct. Mol. Biol.* **2007**, *14*, 1108–1109. [CrossRef] [PubMed]

94. Biswas, S.; Mohammad, M.M.; Movileanu, L.; van den Berg, B. Crystal structure of the outer membrane protein OpdK from *Pseudomonas aeruginosa*. *Structure* **2008**, *16*, 1027–1035. [CrossRef] [PubMed]

95. Catel-Ferreira, M.; Nehme, R.; Molle, V.; Aranda, J.; Bouffartigues, E.; Chevalier, S.; Bou, G.; Jouenne, T.; Dé, E. Deciphering the function of the outer membrane protein OprD homologue of *Acinetobacter baumannii*. *Antimicrob. Agent Chemother.* **2012**, *56*, 3826–3832. [CrossRef] [PubMed]

96. Ishida, H.; Garcia-Herrero, A.; Vogel, H.J. The periplasmic domain of *Escherichia coli* outer membrane protein A can undergo a localized temperature dependent structural transition. *Biochim. Biophys. Acta* **2014**, *1838*, 3014–3024. [CrossRef] [PubMed]

97. Anand, A.; LeDoyt, M.; Karanian, C.; Luthra, A.; Koszelak-Rosenblum, M.; Malkowski, M.G.; Puthenveetil, R.; Vinogradova, O.; Radolf, J.D. Bipartite topology of treponema pallidum repeat proteins C/D and I: Outer membrane insertion, trimerization, and porin function require a C-terminal β-barrel domain. *J. Biol. Chem.* **2015**, *290*, 12313–12331. [CrossRef] [PubMed]

98. Agarwal, R.; Zakharov, S.; Hasan, S.S.; Ryan, C.M.; Whitelegge, J.P.; Cramer, W.A. Structure-function of cyanobacterial outer-membrane protein, slr1270: Homolog of *Escherichia coli* drug export/colicin import protein, TolC. *FEBS Lett.* **2014**, *588*, 3793–3801. [CrossRef] [PubMed]

99. Baaden, M.; Meier, C.; Sansom, M.S. A molecular dynamics investigation of mono and dimeric states of the outer membrane enzyme OMPLA. *J. Mol. Biol.* **2003**, *331*, 177–189. [CrossRef]

100. Adamian, L.; Naveed, H.; Liang, J. Lipid-binding surfaces of membrane proteins: Evidence from evolutionary and structural analysis. *Biochim. Biophys. Acta* **2011**, *1808*, 1092–1102. [CrossRef] [PubMed]

101. Naveed, H.; Jimenez-Morales, D.; Tian, J.; Pasupuleti, V.; Kenney, L.J.; Liang, J. Engineered oligomerization state of OmpF protein through computational design decouples oligomer dissociation from unfolding. *J. Mol. Biol.* **2012**, *419*, 89–101. [CrossRef] [PubMed]

102. Konagurthu, A.S.; Whisstock, J.C.; Stuckey, P.J.; Lesk, A.M. Mustang: A multiple structural alignment algorithm. *Proteins* **2006**, *64*, 559–574. [CrossRef] [PubMed]

103. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; *et al.* Clustal W and clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [CrossRef] [PubMed]

104. Goujon, M.; McWilliam, H.; Li, W.; Valentin, F.; Squizzato, S.; Paern, J.; Lopez, R. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* **2010**, *38*, W695–W699. [CrossRef] [PubMed]

105. Kuipers, R.K.; Joosten, H.J.; Van Berkel, W.J.; Leferink, N.G.; Rooijen, E.; Ittmann, E.; van Zimmeren, F.; Jochens, H.; Bornscheuer, U.; Vriend, G.; *et al.* 3DM: Systematic analysis of heterogeneous superfamily data to discover protein functionalities. *Proteins* **2010**, *78*, 2101–2113. [CrossRef] [PubMed]

106. Jones, D.T.; Taylor, W.R.; Thornton, J.M. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **1992**, *8*, 275–282. [CrossRef] [PubMed]

107. Guindon, S.; Dufayard, J.F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **2010**, *59*, 307–321. [CrossRef] [PubMed]

108. Felsenstein, J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **1985**, *39*, 783–791. [CrossRef]

109. Whelan, S.; Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **2001**, *18*, 691–699. [CrossRef] [PubMed]

110. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **2013**, *30*, 2725–2729. [CrossRef] [PubMed]

111. Hordijk, W.; Gascuel, O. Improving the efficiency of SPR moves in phylogenetic tree search methods based on maximum likelihood. *Bioinformatics* **2005**, *21*, 4338–4347. [CrossRef] [PubMed]

112. Criss, A.K.; Seifert, H.S. A bacterial siren song: Intimate interactions between neisseria and neutrophils. *Nat. Rev. Microbiol.* **2012**, *10*, 178–190. [CrossRef] [PubMed]

113. Stanley, A.M.; Chuawong, P.; Hendrickson, T.L.; Fleming, K.G. Energetics of outer membrane phospholipase A (OMPLA) dimerization. *J. Mol. Biol.* **2006**, *358*, 120–131. [CrossRef] [PubMed]

114. Bishop, R.E. Structural biology of membrane-intrinsic β-barrel enzymes: Sentinels of the bacterial outer membrane. *Biochim. Biophys. Acta* **2008**, *1778*, 1881–1896. [CrossRef] [PubMed]

115. Voorintholt, R.; Kosters, M.T.; Vegter, G.; Vriend, G.; Hol, W.G. A very fast program for visualizing protein surfaces, channels and cavities. *J. Mol. Graph.* **1989**, *7*, 243–245. [CrossRef]

116. Rasmussen, A.A.; Eriksen, M.; Gilany, K.; Udesen, C.; Franch, T.; Petersen, C.; Valentin-Hansen, P. Regulation of ompa mrna stability: The role of a small regulatory RNA in growth phase-dependent control. *Mol. Microbiol.* **2005**, *58*, 1421–1429. [CrossRef] [PubMed]

117. Confer, A.W.; Ayalew, S. The OmpA family of proteins: Roles in bacterial pathogenesis and immunity. *Vet. Microbiol.* **2013**, *163*, 207–222. [CrossRef] [PubMed]

118. Datta, D.; Vaidehi, N.; Floriano, W.B.; Kim, K.S.; Prasadarao, N.V.; Goddard, W.A., 3rd. Interaction of *E. coli* outer-membrane protein a with sugars on the receptors of the brain microvascular endothelial cells. *Proteins* **2003**, *50*, 213–221. [CrossRef] [PubMed]

119. Maruvada, R.; Kim, K.S. Extracellular loops of the *Eschericia coli* outer membrane protein A contribute to the pathogenesis of meningitis. *J. Infect. Dis.* **2011**, *203*, 131–140. [CrossRef] [PubMed]

120. Mittal, R.; Krishnan, S.; Gonzalez-Gomez, I.; Prasadarao, N.V. Deciphering the roles of outer membrane protein A extracellular loops in the pathogenesis of *Escherichia coli* k1 meningitis. *J. Biol. Chem.* **2011**, *286*, 2183–2193. [CrossRef] [PubMed]

121. Koebnik, R. Structural and functional roles of the surface-exposed loops of the β-barrel membrane protein OmpA from *Escherichia coli*. *J. Bacteriol.* **1999**, *181*, 3688–3694. [PubMed]

122. Andersen, C.; Bachmeyer, C.; Täuber, H.; Benz, R.; Wang, J.; Michel, V.; Newton, S.M.; Hofnung, M.; Charbit, A. *In vivo* and *in vitro* studies of major surface loop deletion mutants of the *Escherichia coli* k-12 maltoporin: Contribution to maltose and maltooligosaccharide transport and binding. *Mol. Microbiol.* **1999**, *32*, 851–867. [CrossRef] [PubMed]

123. Gouaux, E. Roll out the barrel. *Nat. Struct. Biol.* **1998**, *5*, 931–932. [CrossRef] [PubMed]

124. Sidorova, O.V.; Khomenko, V.A.; Portnyagina, O.Y.; Likhatskaya, G.N.; Vakorina, T.I.; Kim, N.Y.; Chistyulin, D.K.; Solov'eva, T.F.; Novikova, O.D. Mutant OmpF porins of yersinia pseudotuberculosis with deletions of external loops: Structure-functional and immunochemical properties. *Biochem. Biophys. Res. Commun.* **2014**, *445*, 428–432. [CrossRef] [PubMed]

125. Stenkova, A.M.; Isaeva, M.P.; Shubin, F.N.; Rasskazov, V.A.; Rakin, A.V. Trends of the major porin gene (OmpF) evolution: Insight from the genus yersinia. *PLoS ONE* **2011**, *6*, e20546. [CrossRef] [PubMed]

126. Modi, N.; Bárcena-Uribarri, I.; Bains, M.; Benz, R.; Hancock, R.E.; Kleinekathöfer, U. Role of the central arginine r133 toward the ion selectivity of the phosphate specific channel OprP: Effects of charge and solvation. *Biochemistry* **2013**, *52*, 5522–5532. [CrossRef] [PubMed]

127. Siehnel, R.J.; Egli, C.; Hancock, R.E. Polyphosphate-selective porin OprO of *Pseudomonas aeruginosa*: Expression, purification and sequence. *Mol. Microbiol.* **1992**, *6*, 2319–2326. [CrossRef] [PubMed]

128. Dumas, F.; Koebnik, R.; Winterhalter, M.; Van Gelder, P. Sugar transport through maltoporin of *Escherichia coli*. Role of polar tracks. *J. Biol. Chem.* **2000**, *275*, 19747–19751. [CrossRef] [PubMed]

129. Chatterjee, S.; Rothenberg, E. Interaction of bacteriophage λ with its *E. coli* receptor, LamB. *Viruses* **2012**, *4*, z3162–3178. [CrossRef] [PubMed]