ELSEVIER

Journal of Molecular Graphics and Modelling 20 (2001) 235–244

Journal of Molecular Graphics and Modelling

www.elsevier.com/locate/JMGM

Prediction of membrane protein orientation in lipid bilayers: a theoretical approach

Frederic Basyn [a], Benoit Charloteaux [a], Annick Thomas [b], Robert Brasseur [a,*]

[a] *Centre de Biophysique Moléculaire Numérique, Faculté Agronomique, Passage des déportés, FSAGX, 5030 Gembloux, Belgium*
[b] *INSERM Unité 410, Hôpital X. Bichat, 75018 Paris, France*

Received 30 January 2001; received in revised form 10 May 2001; accepted 17 May 2001

Abstract

Over the past few years, several three-dimensional (3-D) structures of membrane proteins have been described with increasing accuracy, but their relationship with membranes are still not well understood. Recently, we have developed an empirical method, Integral Membrane Protein and Lipid Association (IMPALA), to predict the insertion of molecules (lipids, drugs) into lipid bilayers (Proteins 30 (1998) 357). The IMPALA uses a Monte Carlo minimisation procedure to calculate the depth and the angle of insertion of membrane-interacting molecules taking into account the restraints dictated by a lipid bilayer. In this paper, we use IMPALA to test the insertion of 23 integral membranous proteins (IMPs) and 2 soluble proteins into membranes. Four IMP are studied in detail: OmpA, maltoporin, MsCl channel and bacteriorhodopsin. The 3-D structures of the proteins are kept constant and the insertion into membrane is monitored by minimising the value of the restraint representing the sum of two terms, one for lipid perturbation and the other for hydrophobicity. The two soluble proteins are rejected from the membrane whereas, under the same conditions, all the membrane proteins remain inside, if the solvent accessible surface of the amino acids located inside the pore of porins is ignored. The results give the tilt angle of the IMP helices or strands with respect to the membrane surface and the depth of the protein mass centre insertion. We conclude that the restraint terms of IMPALA could be used to study the insertion of model structures or complexes of proteins within membranes.

Keywords: Membrane protein; MHP; Molecular modelling; Hydrophobicity; Structure; IMP

1. Introduction

Integral membrane proteins (IMPs) are essential for cell life because they catalyse interactions between the cytoplasm and the extracellular environment or the intravesicular media. They are present in all cells and it is estimated that as many as 30–50% of the higher eukaryote genes encode membrane proteins. However, as few as 25 three-dimensional (3-D) structures of IMP have been determined [2]. Therefore, homology modelling, a major tool for determination of a 3-D model of globular proteins from sequence, is not appropriate for membrane proteins due to the lack of known template structures. Moreover, a prerequisite to explain the insertion of IMP in membranes should be a detailed understanding of their 3-D structures in that membrane [3].

Membrane proteins whose 3-D structures are well described are α-helical bundles or β-barrels. The accuracy of these structures is generally low but a recent improvement resulted in a structure at 1.8 Å resolution [4]. For many years, the reference protein for α-helical IMP has been bacteriorhodopsin because it was the first IMP described [5]. For β-proteins, the first IMP described was an outer membrane protein from Gram-negative bacteria, a porin. Similarly-structured proteins were suspected in the outer membranes of mitochondria and chloroplasts [6]. Porins are between 28 and 48 kDa and contain 8–22 antiparallel strands organised as β-sheet folded into a barrel that creates a channel. They enable the transfer of small hydrophilic molecules but not of macromolecules and are associated with peptidoglycan and lipopolysacharides [7].

In some instances (mainly the α-helix bundles), the hydrophobicity profile of a protein sequence can be related to its secondary structure in the membrane [8]. However, this relationship is less obvious for β IMP. Moreover, it is still very difficult to predict how α-helices and β-strands associate in the membrane [3].

In molecular modelling approaches, the simulation of all the components of a membrane requires very intense calculations. To shortcut the problem, we have developed a method, named Integral Membrane Protein and Lipid Association (IMPALA), which studies the association of

Abbreviations: MHP, molecular hydrophobicity potential; IMP, integral membrane protein; IMPALA, IMP and lipid association

* Corresponding author. Tel.: +32-816-22521; fax: +32-816-22522.
E-mail address: brasseur.r@fsagx.ac.be (R. Brasseur).

1093-3263/01/$ – see front matter
PII: S1093-3263(01)00114-0

Table 1
Description of the proteins used in this study[a]

Protein membrane name	PDB code	Source	Residues	Res (Å)	Polymer chains	Authors
α-protein						
Photosynthetic reaction centre	1AIG	*Rhodobacter sphaeroides*	1696	2.6	6	[30]
Cytochrome bc1	1BGY	*Bos taurus*	4332	3	22	[31]
Potassium channel protein (KcSA)	1BL8	*Streptomyces lividans*	388	3.2	4	[32]
Bacteriorhodopsin	1BRR	*Halobacterium salinarium*	741	2.9	3 (homotrimer)	[33]
Bacteriorhodopsin	1BRX	*Halobacterium salinarium*	247	2.3	1	[25]
Prostaglandine H2 synthase	1CQE	*Ovis aries*	1160	3.1	2 (dimer)	[34]
Fumarate reductase	1FUM	*Escherichia coli*	2184	3.3	8 (2 heterotetramer)	[35]
Light harvesting complex II	1LGH	*Rhodospirillum molischianum*	404	2.4	8 (octamer)	[36]
MsCl (mechanosensitive ion channel)	1MSL	*Mycobacterium tuberculosis*	545	3.5	5 (homopentamer)	[24]
Cytochrome C oxidase	1OCR	*Bos taurus*	3612	2.35	26 (homodimer)	[37]
Squalene hopene cyclase	3SQC	*Alicyclobacillus acidocaldarius*	1893	2.8	3 (homodimer)	[38]
Rhodopsin	1F88	Not available	696	2.8	2 (homodimer)	[39]
β-protein						
OmpA (outer membrane protein A)	1BXW	*Escherichia coli*	171	2.5	1	[21]
FepA (ferric enterobactin receptor)	1FEP	*Escherichia coli*	724	2.4	1	[40]
OmpX	1QJ9	*Escherichia coli*	148	2.1	1	[14]
Ompf	2OMF	*Escherichia coli*	340	2.4	1	[41]
Sucrose specific porin	1A0S	*Salmonella typhimurium*	1239	4.2	3 (homotrimer)	[42]
Fhua (ferric hydroxamate uptake receptor)	2FCP	*Escherichia coli*	723	2.5	1	[43]
Ompla (outer membrane phospholipase)	1QD6	*Escherichia coli*	506	2.1	4	[44]
OMPK36 (osmoporin)	1OSM	*Klebsiella pneumoniae*	1026	3.2	3 (homotrimer)	[45]
Porin	1PRN	*Rhodopseudomonas blastica*	289	1.96	1	[46]
Porin (crystal form B)	2POR	*Rhodobacter capsulatus*	301	1.8	1	[4]
Maltoporin	1AF6	*Escherichia coli*	1263	2.4	3 (homotrimer)	[22]

[a] All structures used had been determined by X-ray crystallography. First column: name of the membrane protein (α- and β-proteins); second column: PDB code; third column: organism from which the protein was purified; fourth column: number of residues; fifth column: resolution of crystal structure; sixth column: the technique used to determine the structure of the protein; seventh column: the number of chains (monomer, dimer, multimer, etc) and eighth column: literature references.

peptides with membranes [1]. The IMPALA is an empirical approach that mimics the biological membrane effects by accounting for the water–lipid interfaces and describes the membrane as an apolar implicit layer whose water content varies from 1.0 (the water phase) to 0.0 (the acyl chain of lipids). The IMPALA has been useful to study the interaction of peptides with a membrane [9–11], their orientation and depth of penetration into the lipid bilayer. Transmembrane, amphipathic [1] and tilted peptides [12] were also studied and compared to experimental data when the latter were available. Predictions and experiments are in good agreement. For instance, the dimerisation of glycophorin A was studied and the restraint terms of IMPALA were shown to increase the efficiency of the Monte Carlo algorithm to determine a correct dimer structure [13].

The aim of this study was to analyse the insertion of large proteins in membranes by applying the IMPALA simulation to a series of 23 α- and β-proteins (Table 1).

2. Computational methods

2.1. IMPALA

The IMPALA is a programme developed by Ducarme et al. [1]. It describes the insertion of a molecule (protein or drug) into a bilayer by adding energy restraint functions to the usual energy description of molecules [13,14]. These additional describe the bilayer hydrophobicity and the molecule-bilayer interactions.

The function describing the bilayer properties is constant in the plane of the membrane (*x*- and *y*-axes). It varies along the bilayer thickness (*z*-axis) and more specifically at the lipid–water interface corresponding to the transition between the acyl chains of lipids (no water: hydrophobic core) and the outside medium (water: hydrophilic phase). The membrane properties are expressed by the empirical function, $C_{(z)}$:

$$C_{(z)} = 1 - \frac{1}{1 + e^{\alpha(z - z_0)}}$$

The *z*-axis has its origin at the centre of the bilayer (Fig. 1). $C_{(z)}$ varies from 0.0 (hydrophobic core) to 1.0 (hydrophilic phase). Values of z of $\pm\infty < z < \pm 18$ Å are designated the water phases, values of ± 18 Å $< z <$ ± 13.5 Å are within the region of the polar heads of lipids and values of ± 13.5 Å $< z < 0$ Å are in the region of the acyl chains of lipids. α is a constant equal to 1.99 and z_0 correspond to the middle of the polar head group region ($(13.5 + 18)/2 = 15.75$ Å).

The first term simulates the effect that pushes the hydrophobic segments of proteins into the membrane

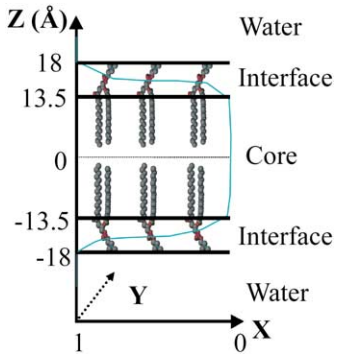

Fig. 1. Description of the model used for simulating the lipid–water interface with the IMPALA program. The water content of the membrane is described by a function $C_{(z)}$ which varies between 0 and 1 along the z-axis normal to the membrane surface (plane x–y). The origin of z is at the centre of the bilayer. $C_{(z)}$ is a symmetrical empirical function that is 1 from $-\infty$ to -18 Å (the water phase), varies from 1 to 0 from -18 to -13.5 Å and from 13.5 to 18 Å (lipid polar heads) and is null in the hydrocarbon core (-13.5 to $+13.5$ Å).

(hydrophobic restraint).

$$E_{\text{int}} = -\sum_{i=1}^{N} S_{(i)} E_{\text{tr}(i)} C_{(z_i)}$$

where N is the total number of atoms, $S_{(i)}$ the accessible surface of atom i to solvent, $E_{\text{tr}(i)}$ its transfer energy by unit of accessible surface area and $C_{(z_i)}$ the value of $C_{(z)}$ at the position z_i of the atom i.

The second restraint is a measure of the lipid–molecule interactions and is proportional to the accessible surface area of the molecule, which is the sum of $S_{(i)}$ of all atoms constituting the membrane protein [15]. This effect is simulated by the lipid perturbation restraint E_{lip}:

$$E_{\text{lip}} = a_{\text{lip}} \sum_{i=1}^{N} S_{(i)} (1 - C_{(z_i)})$$

where a_{lip} is an empirical factor fixed to 0.018 [1].

During the analysis, the protein structures remained constant; we analysed the insertion of the rigid body protein structures into a membrane. Therefore, the only energy changes occurring during the Monte Carlo procedure are due to the restraint term. Each simulation was as follows: temperature 25°C, 10^5 steps, random rotation ±5°, random translation ±4 Å. At the beginning, the mass centre of the molecule was set at the bilayer centre. The most stable position(s) and orientation(s) were finally those corresponding to the lower values of restraint. An average of three simulations has been performed per protein.

2.2. Molecular hydrophobicity potential (MHP)

The MHP is a 3-D plot of the hydrophobicity potential of a molecule in order to visualise its amphipathy. Molecules and MHP are drawn with WinMGM [16]. The hydrophobicity of a molecule is calculated using its partition coefficient between water and octanol. In order to visualise this amphipathy, a 3-D molecular plot of hydrophobicity was proposed by Fauchère et al. for small molecules [17]. The MHP approach was then extended to plot the surface of isopotential contours of larger molecules [18]. We postulate that the hydrophobicity induced by an atom i and measured at a point M of space decreases exponentially with the distance between this point M and the surface of atom i. The molecular hydrophobicity potential at a point M in space is, thus, estimated according to the equation: $\text{MHP}_{\text{M}} = \sum_{i=1}^{N} E_{\text{tr}(i)} \exp(r_i - d_i)$ where N are all the atoms of the molecule, $E_{\text{tr}(i)}$ the transfer energy of the atom i, r_i the radius of the atom i, and d_i the distance between the atom i and the point M. $E_{\text{tr}(i)}$ is the energy required to transfer an atom i from an hydrophobic to an hydrophilic medium. Atomic $E_{\text{tr}(i)}$ were calculated from the molecular E_{tr} compiled by Tanford [19] assuming that molecular E_{tr} are the sum of their atomic E_{tr}. Atomic E_{tr} values were derived for seven different types of atoms [1,18].

The isopotential surfaces of protein hydrophobicity were then calculated by a cross-sectional computational method. A 1 Å mesh-grid plane was set to sweep across the molecule by steps of 1 Å. At each step, the sum of the hydrophobicity and hydrophilicity values at all the grid nodes were calculated. The hydrophobic and hydrophilic MHP surfaces were then drawn by joining the isopotential values.

3. Results and discussion

A panel of 23 IMP (12 α and 11 β) was selected. These proteins are monomeric, dimeric or multimeric containing between 150 and 4300 residues (Table 1).

In IMPALA, the evolution of the restraint values with respect to the penetration depth and the insertion angle were tested (Table 2) (Figs. 2 and 3). Fig. 2 plots the nadir envelope of the restraint versus the mass centre penetration in the bilayer versus the molecule tilt with respect to the membrane surface. The nadir envelope plots the energy minima of the Monte Carlo minimisation procedure at each orientation angle and position depth. The minimal restraint corresponds to the most stable state, i.e. the most favourable position and orientation.

For the porins, we faced a problem due to the location of amino acids inside the pore. In native proteins, these amino acids are accessible to water: in the IMPALA procedure, they are accessible to lipids since they are within the membrane z co-ordinates. Under such conditions, the porins were unable to equilibrate in membranes because, inside the bilayer, the restraint terms E_{int} and E_{lip} are proportional to the solvent-accessible surface area of residues and E_{int} forces the polar residues out. To restore insertion into the membrane, we had to set all residues inside the pore with a null solvent accessible surface.

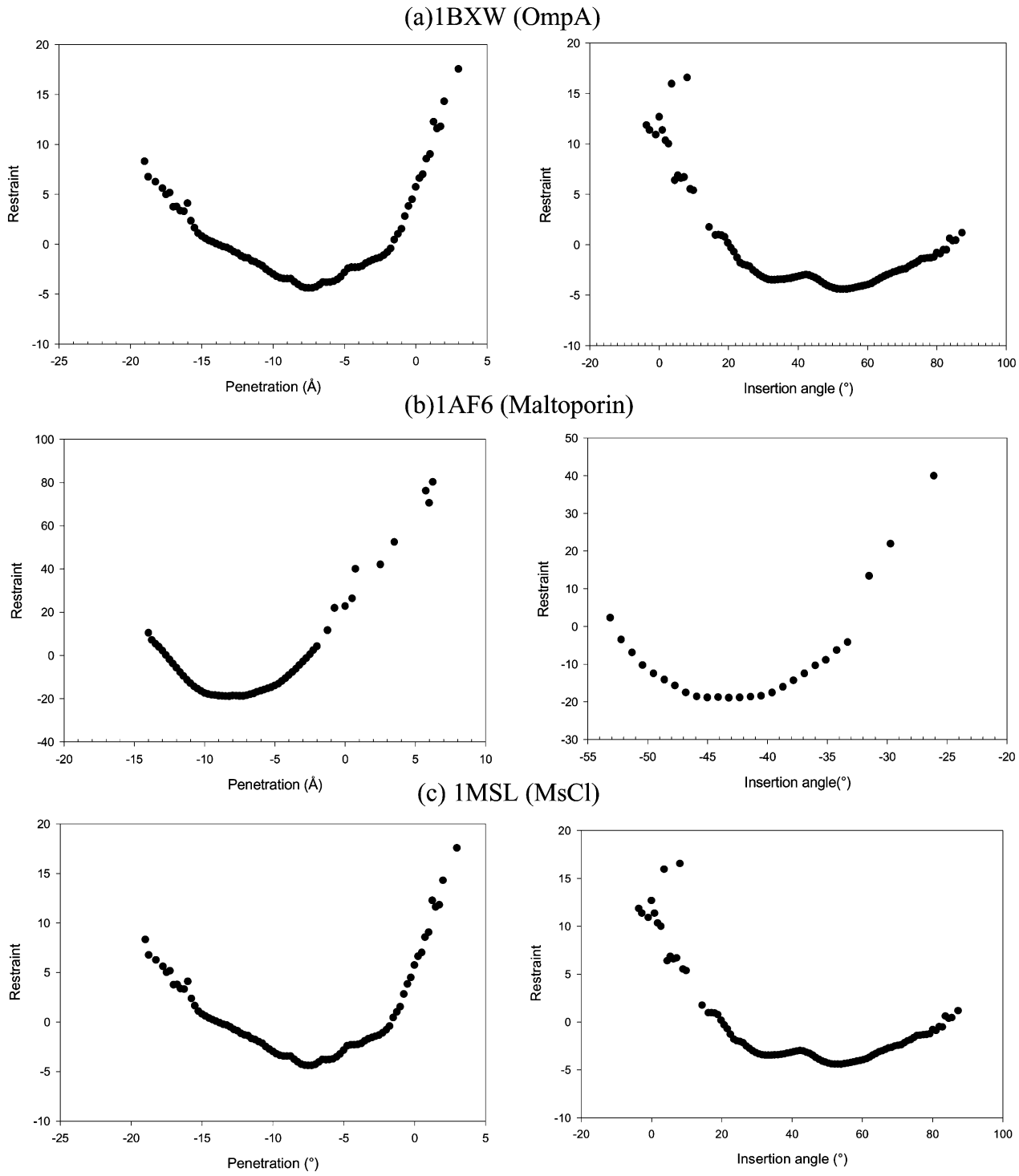

Fig. 2. Monte Carlo (MC) simulations of the insertion of proteins into membranes: (a) 1BXW: OmpA; (b) 1AF6: maltoporin; (c) 1MSL: MsCl; (d) 1BRX: bacteriorhodopsin. On each plot, the line corresponds to the nadir envelope of all Monte Carlo steps. On the left, the values of the restraints are plotted vs. the penetration (i.e. the *z* ordinate of the mass centre of the molecule). On the right, the restraint values are plotted vs. the insertion angle (i.e. the angle between the helix axis and the surface of the bilayer).

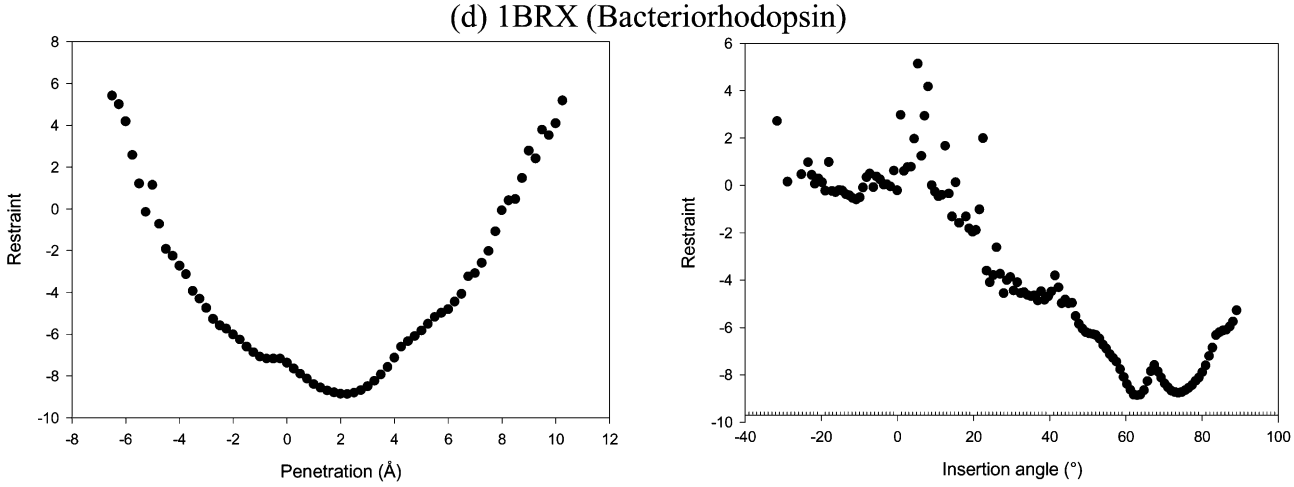

Fig. 2. (*Continued*).

The first protein tested was outer membrane protein A (OmpA) (PDB:1BXW), a β-barrel from the *Escherichia coli* outer membrane. It stabilises the outer membrane of the bacteria and is a receptor for bacteriophages. The OmpA has a transmembrane domain of 171 amino acids which consists of an eight-stranded β-barrel and was studied particularly for membrane protein folding [20]. The crystal structure of OmpA was determined at 2.5 Å resolution by Pautsch and Schulz [21]. The eight-strands start and finish at the following amino acids: T6-S16, K34-V45, V49-R60, Y72-P86, L91-T106, N114-A130, I135-W143 and M161-F170, respectively. The minimisation in IMPALA occurs when the

Table 2
Results of the IMPALA simulation for all the proteins tested[a]

	PDB code	Reference amino acids	Residues	Restraint value	Co-ordinate *z* (Å)	Insertion angle (°)
α-protein						
Photosynthetic reaction centre	1AIG	120E–131E	1696	40	9	86
Cytochrome bc1	1BGY	*	4332	16	−21	*
Potassium channel protein (KcSA)	1BL8	5B–22B	388	−7	1	−70
Bacteriorhodopsin	1BRR	203A–222A	741	−25	−2	74
Bacteriorhodopsin	1BRX	7A–22A	247	−9	2	63/75
Prostaglandine H_2 synthase	1CQE	309A–321A	1160	20	−35	−20
Fumarate reductase	1FUM	*	2184	3	40	*
Light harvesting complex II	1LGH	20A–34A	404	10	−1	−17
MsCl (mechanosensitive ion channel)	1MSL	77B–91B	545	10	3	−50
Cytochrome C oxidase	1OCR	*	3612	220	8	*
Squalene hopene cyclase	3SQC	161A–172A	1893	90	6	78
Rhodopsin	1F88	42B–57B	696	−15	−1	50
β-protein						
OmpA (outer membrane protein A)	1BXW	31B–36B	171	−5	−7	52
FepA (ferric enterobactin receptor)	1FEP	128D–135D	724	−6	11	−52
OmpX	1QJ9	43A–47A	148	−13	−6	57
Ompf	2OMF	254A–261A	340	−13	−9	−50
Sucrose specific porin	1A0S	282A–289A	1239	30	36	7
Fhua (ferric hydroxamate uptake receptor)	2FCP	465A–468A	723	−15	13	48
Ompla (outer membrane phospholipase)	1QD6	194C–200C	506	−20	3	43
OMPK36 (osmoporin)	1OSM	145A–150A	1026	−30	7	−52
Porin	1PRN	28A–35A	289	−1	−5	−31
Porin (crystal form B)	2POR	228A–238A	301	−1	14	50
Maltoporin	1AF6	40A–48A	1263	−20	−8	−43

[a] First column: the name of the protein; second column: PDB code; third column: the two atoms that define the reference vector used to calculate the tilt of the insertion angle during the IMPALA simulation. The number refers to the amino acid and the letter to the chain. In three cases (*), we were unable to choose reference atoms. Fourth column: number of residues; fifth column: minimal value of the restraint after the Monte Carlo minimisation; sixth column: the penetration of the mass centre corresponding at the optimal position and seventh column: the tilt of the reference vector of column third.

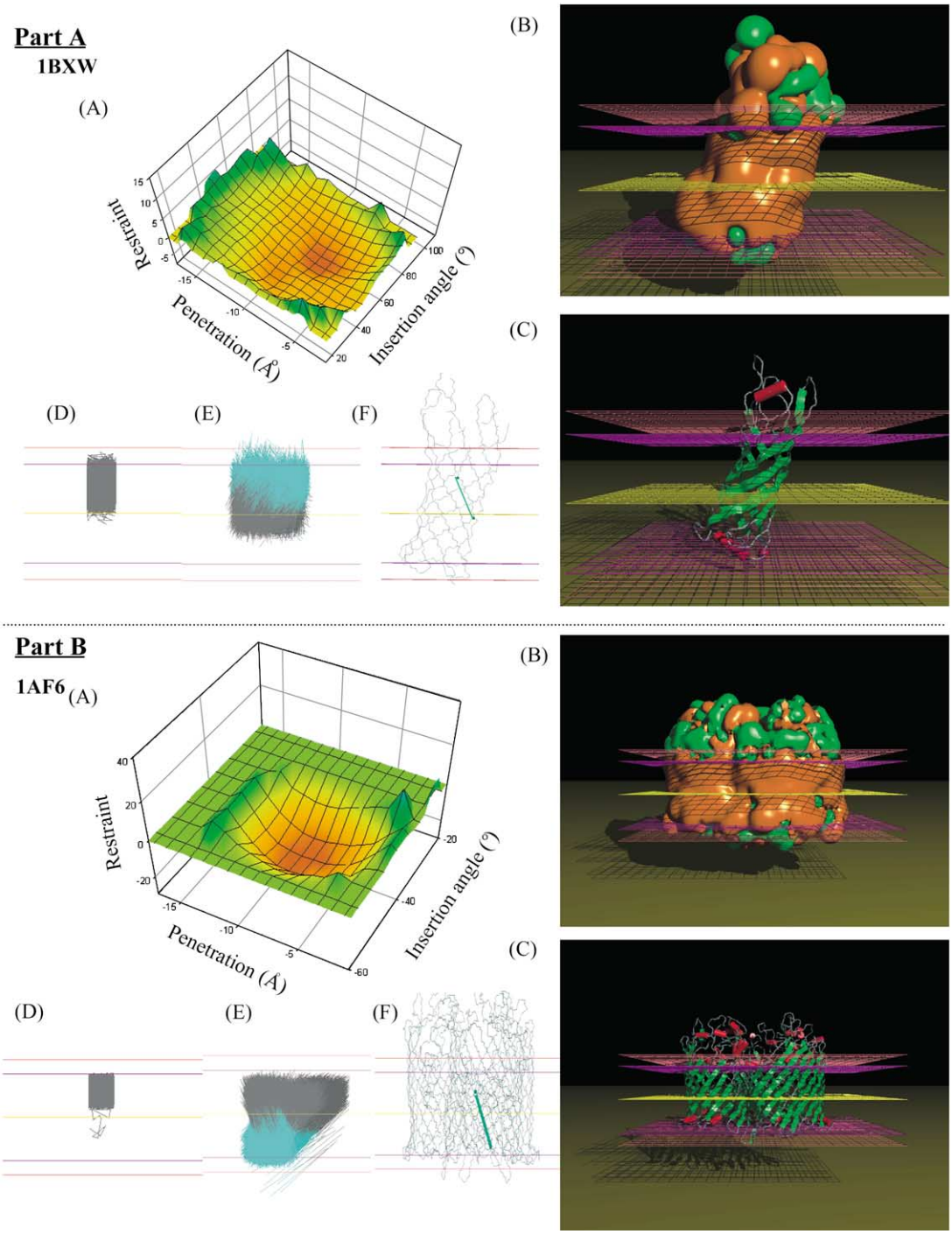

Fig. 3. The different plots correspond to the proteins OmpA (part A), maltoporin (part B), MsCl (part C) and bacteriorhodopsin (part D). IMPALA simulations are plotted (A) as 3-D plots of the restraint energy vs. the penetration protein mass centre and vs. the angle of insertion of a reference vector. For a reference vector, two atoms of a helix or of a strand are selected and the angle of this vector with the *xy* plane is an index of the orientation of the protein within the bilayer. In the *x*-axis is the protein mass centre penetration (0 is when the mass centre is at the bilayer centre); the *y*-axis is the angle of the reference vector. The restraint values are plotted along the *z*-axis. The energy minimum (orange patch) locates the most favourable positions of the protein. (B) and (C) are images of the most stable conformation of the protein in the IMPALA membrane. (B) is the MHP and (C) the ribbon plot of the protein. Green surfaces of MHP are hydrophilic, brown are hydrophobic. The yellow line is the centre of the bilayer ($z = 0$) and the purple line defines where the polar heads of lipids start ($z = \pm 13.5$ Å). The pink line is the water–lipid interface ($z = \pm 18$ Å). In D, the various positions of the protein mass centre during the simulation are given, one point corresponds to one step of Monte Carlo. Figure E plots the reference vector at each step of the Monte Carlo minimisation. In F, the reference vector of the best position is plotted on the skeleton drawing of the molecule.

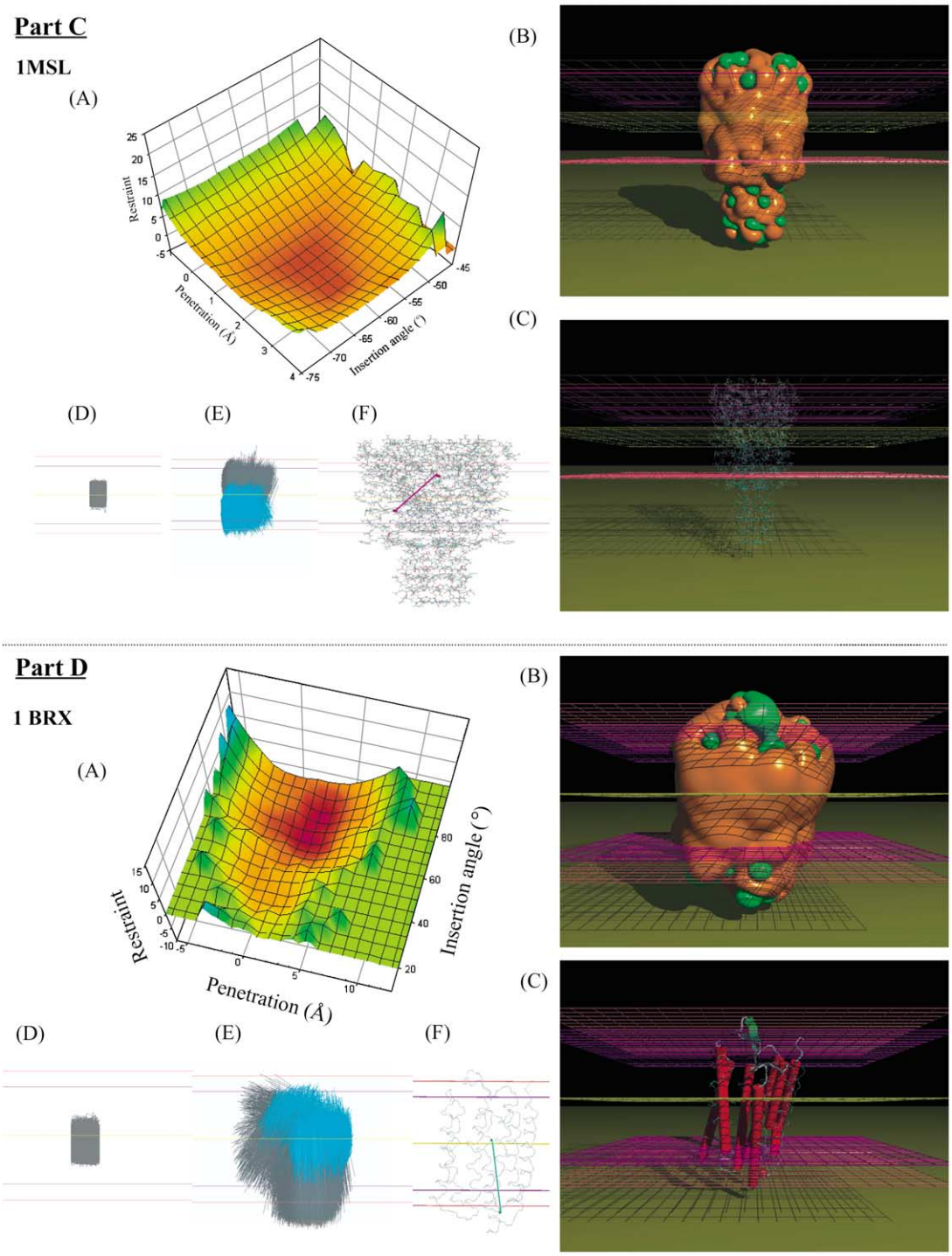

Fig. 3. (*Continued*).

mass centre of OmpA is moved by about −7 Å from the bilayer centre and when the angle of insertion of the reference strand (the second, K34-V45) (Fig. 3 part A(E)) is ∼50–55° with respect to the membrane surface. Thus, the final position is not symmetrical with respect to the bilayer centre. The energy minimisation resulted in partitioning part of the intracellular side out of the membrane. As shown on the MHP plot, this part has hydrophilic patches (green surfaces, Fig. 3 part A(B)). In the literature, all strands are suggested to be tilted by ∼45° with respect to the membrane surface [21] and all the amino acids of the eight-strand model [21] are inserted inside the membrane in our IMPALA simulation.

The second protein tested was maltoporin (PDB:1AF6), a β-barrel homotrimer from bacterial outer membranes

(*Escherichia coli*). The function of this protein is to facilitate the diffusion of maltodextrins across the outer membrane of *E. coli*. The crystal structure was determined at 2.4 Å resolution by Wang et al. [22]. The reference strand is the vector between T40 and Q48 of the A monomer. All strands are tilted by approximately ∼45° with respect to the barrel axis. The restraint nadir envelope shows a local minimum at approximately −(40–45°), thus, the barrel axis is normal on the membrane surface. Here, again the mass centre has moved down by −7 to −9 Å and the hydrophilic patches of the MHP are in the aqueous medium (Fig. 3 part B(B)).

The third protein tested was the mechanosensitive ion channel (Mscl, PDB:1MSL) from *Mycobacterium tuberculosis*. It is proposed to function in cellular homeostasis, probably by regulating osmotic pressure. This channel is a homopentamer and each monomer has two transmembrane a helices and a cytoplasmic domain. The transmembrane helices are suggested to be tilted at 62° with respect to the membrane surface [23]. The crystal structure was determined at 3.5 Å resolution by Chang et al. [24]. The reference atoms that were used for the tilt calculation are located in the second transmembrane helix (I77 and V91). In our calculations, the optimum insertion has a ∼60° tilt (Fig. 3 part C(A)) and, thus, is in agreement with the results from the literature.

Bacteriorhodopsin (PDB:1BRX) is the first IMP the structure of which was determined from a two-dimensional crystal [5] and then from 3-D crystals down to an atomic resolution of 1.8–2.3 Å [25]. It is the proton pump of the purple membrane of archaebacteria *Halobium salinarium*. Bacteriorhodopsin has seven transmembrane α-helices of 20–27 residues and binds a retinal chromophore which captures the light to supply energy for the proton pumping activity. Three helices are almost normal to the membrane surface while the other four are tilted by 70° [25–27]. The limits of the seven helices are; helix A (E9-V29), helix B (Y43-M60), helix C (A81-L94), helix D (I108-T128), helix E (R134-L152), helix F (T178-I191), helix G (T205-L221). The reference vector for the insertion angle is located in the first helix (W7-F22). In Fig. 3 part D(A), this helix has a tilt of ∼65° with respect to the 70° published in literature. A few residues, V148, A149, S150, T151, F152 which are described to belong to the transmembrane helix E in the literature are located in the extramembranous space in our data (>18 Å from the centre of the bilayer). The insertion of all the other amino acids in the bilayer is in agreement with the literature.

Table 2 summarises the data from the simulations. We can observe that large proteins have a high restraint even after minimisation. For instance, the large α-proteins, 1AIG, 1BGY, 1CQE, 1OCR, 3SQC, all have positive restraint values. Fumarate reductase is an exception. The molecule is large (2184 amino acids) but the restraint satisfaction value is low. This can be understood by considering the nature of the two components of the restraint term, the perturbation of the lipid bilayer by the insertion (E_{lip}) and the hydrophobic interactions of the inserted protein with its surrounding medium (E_{int}). Outside the membrane, E_{lip} is null and the restraint is given solely by E_{int}. Inside the hydrophobic core of the membrane ($-13.5 < z < 13.5$), E_{int} is null (C_z is 0). The value of E_{lip} is always positive and is proportional to the lipid-accessible surface of the molecule. The value of E_{int} is positive for hydrophobic and negative for hydrophilic interactions. For fumarate reductase, there are large polar cytoplasmic parts of the structure and so the positive E_{lip} is balanced by the negative value of E_{int}. Thus, the restraint tends to optimise the partition of the hydrophobic and hydrophilic surfaces between the bilayer and the aqueous phases. Indeed, for 1AF6, 1BRX, 1BXW and 1MSL, the

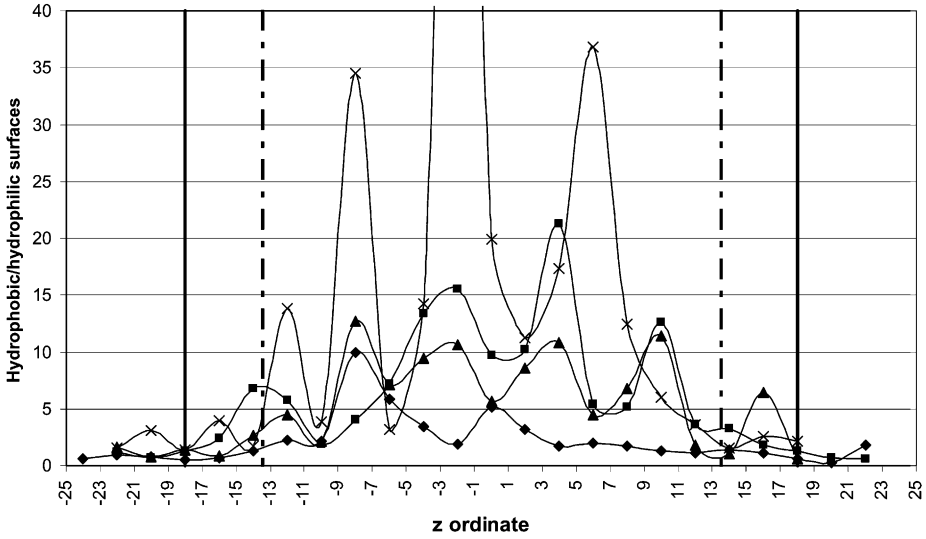

Fig. 4. Ratio of the hydrophobic vs. the hydrophilic accessible surfaces of residues in the membrane inserted proteins (1AF6 (circle), 1BRX (square), 1BXW (triangle) and 1MSL (cross). Dashed lines at ±13.5 Å is the interface between the lipid core and the polar heads of lipids. Plain lines at ±18 Å is the water–lipid interface.

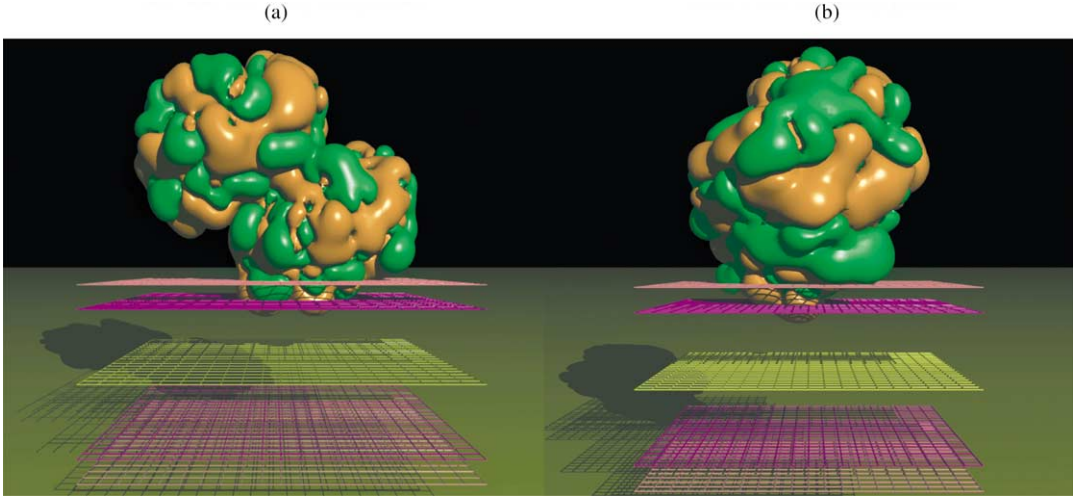

Fig. 5. MHP plot of lysozyme (PDB:1LYS) (a), and trypsin (PDB:1TPO) (b) after the IMPALA simulation. These proteins started at the centre of the bilayer and were ejected outside the bilayer during the simulation.

best conditions are obtained when the ratio of the hydrophobic to the hydrophilic solvent-accessible surfaces of residues are higher, between −13.5 and 13.5 Å (Fig. 4). Therefore, there are more hydrophobic residues between −13.5 and 13.5 Å and more hydrophilic residues outside ±13.5 Å.

Two soluble proteins, lysozyme (PDB:1LYS) [28,29] and trypsin (PDB:1TPO) [28] were tested in the same conditions (Fig. 5). For each assay, we started with the mass centre of the protein located at the bilayer centre. Both proteins were recovered outside the membrane after minimisation supporting the contention that the procedure is efficient and sufficient to drive large proteins out when required. The restraint is equal to −25 for lysozyme and trypsin after Monte Carlo minimisation.

For all the proteins and, especially for the four examples illustrated in Fig. 3, the hydrophobic domains, i.e. the brown surfaces of MHP are located inside the bilayer after energy minimisation whereas the green patches are in the aqueous phase and at the interface, i.e. the level of the polar heads of lipids.

Several criteria are used to judge the insertion: the MHP plot, the angle of a vector set between two atoms of the molecule, the mass centre *z*-value (penetration), and the hydrophobic/hydrophilic surface ratio. All can be used to compare the conformation to data from the literature. The IMPALA approach demonstrates that the 23 large membrane proteins prefer the bilayer as compared to water whereas, in the same conditions, the soluble proteins were ejected out from the membrane. Therefore, simple empirical terms, used here as the restraint consisting of two terms (lipid perturbation and hydrophobic restraint), are sufficient to obtain a rapid and relevant image of whether a structure will insert into a membrane or be soluble. In the future, such empirical restraints could be valuable to evaluate whether ab initio and homology models of large proteins will insert and stay in membranes or not.

4. Conclusions

We have applied a Monte Carlo simulation to study the orientation of large proteins in lipid bilayers. Assuming that the protein structure is constant, a simple empirical term is sufficient to predict the orientation and the depth of insertion. We, therefore, suggest that the procedure could be useful to test whether modelled structures are water- or membrane-soluble.

Acknowledgements

Robert Brasseur is Research Director of the National Funds For Scientific Research of Belgium (FNRS). This work was supported by the “Interuniversity Poles of Attraction Programme-Belgian state, Primer Minister’ office-Federal Office for Scientific, Technical and Cultural Affairs” PAI contract no. 4/03 and by the “Ministère de la Région Wallone” and the work of F. Basyn was supported by a grant FNRS-Télévie. The authors wish to thank Dr Philippa Talmud for her corrections of this manuscript.

References

[1] P. Ducarme, M. Rahman, R. Brasseur, IMPALA: a simple restraint field to simulate the biological membrane in molecular structure studies, Proteins 30 (1998) 357–371.
[2] R. Grisshammer, C.G. Tate, Overexpression of integral membrane proteins for structural studies, Q. Rev. Biophys. 28 (1995) 315–422.
[3] P.J. Booth, A.R. Curran, Membrane protein folding, Curr. Opin. Struct. Biol. 9 (1999) 115–121.
[4] M.S. Weiss, G.E. Schulz, Structure of porin refined at 1.8 Å resolution, J. Mol. Biol. 227 (1992) 493–509.
[5] R. Henderson, P.N. Unwin, Three-dimensional model of purple membrane obtained by electron microscopy, Nature 257 (1975) 28–32.
[6] W. Kühlbrandt, E. Gouaux, Membrane proteins (editorial overview), Curr. Opin. Struct. Biol. 9 (4) (1999) 445–447.

[7] S.K. Buchanan, Beta-barrel proteins from bacterial outer membranes: structure, function and refolding (see comments), Curr. Opin. Struct. Biol. 9 (1999) 455–461.
[8] S.H. White, Hydropathy plots and the prediction of membrane protein topology, in: S.H. White (Ed.), Membrane Protein Structure, New York, 1994, pp. 97–124.
[9] L. Lins, P. Ducarme, E. Breukink, R. Brasseur, Computational study of nisin interaction with model membrane, Biochim. Biophys. Acta 1420 (1999) 111–120.
[10] M. Lindgren, X. Gallet, U. Soomets, M. Hallbrink, E. Brakenhielm, M. Pooga, R. Brasseur, U. Langel, Translocation properties of novel cell penetrating transportan and penetratin analogues, Bioconjug. Chem. 11 (5) (2000) 619–626.
[11] G. Drin, H. Déméné, R. Brasseur, Translocation of pAntp peptide and its amphipathic analogue AP-2AL, Biochemistry 40 (2001) 1824–1834.
[12] L. Lins, B. Charloteaux, A. Thomas, R. Brasseur, Computational study of lipid-destabilising protein fragments: towards a comprehensive view of tilted peptides, Proteins 44 (2001) 435–447.
[13] P. Ducarme, A. Thomas, R. Brasseur, The optimisation of the helix/helix interaction of a transmembrane dimer is improved by the IMPALA restraint field, Biochim. Biophys. Acta 1509 (2000) 148–154.
[14] J. Vogt, G.E. Schulz, The structure of the outer membrane protein OmpX from *Escherichia coli* reveals possible mechanisms of virulence, Struct. Fold. Des. 7 (1999) 1301–1309.
[15] D.C. Rees, A.J. Chirino, K.-H. Kim, H. Komiya, Membrane protein structure and stability: implications of the first crystallographic analyses, in: S.H. White (Ed.), Membrane Protein Structure, New York, 1994, pp. 3–26.
[16] M. Rahman, R. Brasseur, WinMGM: a fast CPK molecular graphics program for analyzing molecular structure, J. Mol. Graph. 12 (3) (1994) 212–218, 198.
[17] J.L. Fauchère, P. Quarendon, L. Kaetterer, Estimating and representing hydrophobicity potential, J. Mol. Graph. 6 (1988) 203–206.
[18] R. Brasseur, Differentiation of lipid-associating helices by use of three-dimensional molecular hydrophobicity potential calculations, J. Biol. Chem. 266 (1991) 16120–16127.
[19] C. Tanford, The Hydrophobic Effect: Formation of Micelles and Biological Membranes, 1991, John Wiley and Sons, Malabar, Florida.
[20] J.H. Kleinschmidt, L.K. Tamm, Folding intermediates of a beta-barrel membrane protein. Kinetic evidence for a multi-step membrane insertion mechanism, Biochemistry 35 (1996) 12993–13000.
[21] A. Pautsch, G.E. Schulz, Structure of the outer membrane protein A transmembrane domain (see comments), Nat. Struct. Biol. 5 (1998) 1013–1017.
[22] Y.F. Wang, R. Dutzler, P.J. Rizkallah, J.P. Rosenbusch, T. Schirmer, Channel specificity: structural basis for sugar discrimination and differential flux rates in maltoporin, J. Mol. Biol. 272 (1997) 56–63.
[23] R.H. Spencer, G. Chang, D.C. Rees, Feeling the pressure: structural insights into a gated mechanosensitive channel, Curr. Opin. Struct. Biol. 9 (4) (1999) 448–454.
[24] G. Chang, R.H. Spencer, A.T. Lee, M.T. Barclay, D.C. Rees, Structure of the MscL homolog from *Mycobacterium tuberculosis*: a gated mechanosensitive ion channel, Science 282 (1998) 2220–2226.
[25] H. Luecke, H.T. Richter, J.K. Lanyi, Proton transfer pathways in bacteriorhodopsin at 2.3 Å resolution, Science 280 (1998) 1934–1937.
[26] J.-L. Popot, C. De Vitry, A. Atteia, in: S.H. White (Ed.), Folding and assembly of integral membrane proteins: an introduction, New York, 1994, pp. 41–96.
[27] S. Subramaniam, The structure of bacteriorhodopsin: an emerging consensus, Curr. Opin. Struct. Biol. 9 (1999) 462–468.
[28] K. Harata, X-ray structure of a monoclinic form of hen egg-white lysozyme crystallized at 313-K comparison of 2 independent molecules, Acta Crystallogr. D. Biol. Crystallogr. 50 (1994) 250.
[29] W. Bode, P. Schwager, The refined crystal structure of bovine beta-trypsin at 1.8 Å resolution. II. Crystallographic refinement, calcium binding site, benzamidine binding site and active site at pH 7.0, J. Mol. Biol. 98 (4) (1975) 693–717.
[30] M.H. Stowell, T.M. McPhillips, D.C. Rees, S.M. Soltis, E. Abresch, G. Feher, Light-induced structural changes in photosynthetic reaction centre: implications for mechanism of electron-proton transfer, Science 276 (1997) 812–816.
[31] S. Iwata, J.W. Lee, K. Okada, J.K. Lee, M. Iwata, B. Rasmussen, T.A. Link, S. Ramaswamy, B.K. Jap, Complete structure of the 11-subunit bovine mitochondrial cytochrome bc1 complex (see comments), Science 281 (1998) 64–71.
[32] D.A. Doyle, C.J. Morais, R.A. Pfuetzner, A. Kuo, J.M. Gulbis, S.L. Cohen, B.T. Chait, R. MacKinnon, The structure of the potassium channel: molecular basis of K^+ conduction and selectivity (see comments), Science 280 (1998) 69–77.
[33] L. Essen, R. Siegert, W.D. Lehmann, D. Oesterhelt, Lipid patches in membrane protein oligomers: crystal structure of the bacteriorhodopsin-lipid complex, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 11673–11678.
[34] D. Picot, P.J. Loll, R.M. Garavito, The X-ray crystal structure of the membrane protein prostaglandin H2 synthase-1 (see comments), Nature 367 (1994) 243–249.
[35] T.M. Iverson, C. Luna-Chavez, G. Cecchini, D.C. Rees, Structure of the *Escherichia coli* fumarate reductase respiratory complex (see comments), Science 284 (1999) 1961–1966.
[36] J. Koepke, X. Hu, C. Muenke, K. Schulten, H. Michel, The crystal structure of the light-harvesting complex II (B800-850) from *Rhodospirillum molischianum*, Structure 4 (1996) 581–597.
[37] S. Yoshikawa, K. Shinzawa-Itoh, R. Nakashima, R. Yaono, E. Yamashita, N. Inoue, M. Yao, M.J. Fei, C.P. Libeu, T. Mizushima, H. Yamaguchi, T. Tomizaki, T. Tsukihara, Redox-coupled crystal structural changes in bovine heart cytochrome c oxidase, Science 280 (1998) 1723–1729.
[38] K.U. Wendt, A. Lenhart, G.E. Schulz, The structure of the membrane protein squalene-hopene cyclase at 2.0 Å resolution, J. Mol. Biol. 286 (1999) 175–187.
[39] K. Palczewski, T. Kumasaka, T. Hori, C.A. Behnke, H. Motoshima, B.A. Fox, T.I. Le, D.C. Teller, T. Okada, R.E. Stenkamp, M. Yamamoto, M. Miyano, Crystal structure of rhodopsin: a G protein-coupled receptor (see comments), Science 289 (2000) 739–745.
[40] S.K. Buchanan, B.S. Smith, L. Venkatramani, D. Xia, L. Esser, M. Palnitkar, R. Chakraborty, H.D. van der, J. Deisenhofer, Crystal structure of the outer membrane active transporter FepA from *Escherichia coli* (see comments), Nat. Struct. Biol. 6 (1999) 56–63.
[41] S.W. Cowan, R.M. Garavito, J.N. Jansonius, J.A. Jenkins, R. Karlsson, N. Konig, E.F. Pai, R.A. Pauptit, P.J. Rizkallah, J.P. Rosenbusch, The structure of OmpF porin in a tetragonal crystal form, Structure 3 (1995) 1041–1050.
[42] D. Forst, W. Welte, T. Wacker, K. Diederichs, Structure of the sucrose-specific porin ScrY from *Salmonella typhimurium* and its complex with sucrose (see comments), Nat. Struct. Biol. 5 (1998) 37–46.
[43] A.D. Ferguson, E. Hofmann, J.W. Coulton, K. Diederichs, W. Welte, Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide (see comments), Science 282 (1998) 2215–2220.
[44] H.J. Snijder, I. Ubarretxena-Belandia, M. Blaauw, K.H. Kalk, H.M. Verheij, M.R. Egmond, N. Dekker, B.W. Dijkstra, Structural evidence for dimerization-regulated activation of an integral membrane phospholipase, Nature 401 (1999) 717–721.
[45] R. Dutzler, G. Rummel, S. Alberti, S. Hernandez-Alles, P. Phale, J. Rosenbusch, V. Benedi, T. Schirmer, Crystal structure and functional characterization of OmpK36, the osmoporin of *Klebsiella pneumoniae*, Struct. Fold. Des. 7 (1999) 425–434.
[46] A. Kreusch, G.E. Schulz, Refined structure of the porin from *Rhodopseudomonas blastica*. Comparison with the porin from *Rhodobacter capsulatus*, J. Mol. Biol. 243 (1994) 891–905.