FEBS Letters 581 (2007) 1335–1341

Membrane protein assembly patterns reflect selection for non-proliferative structures

Arianna Rath[a,b], Charles M. Deber[a,b,*]

[a] Division of Molecular Structure and Function, Research Institute, Hospital for Sick Children, 555 University Avenue, Toronto, Ont., Canada M5G 1X8
[b] Department of Biochemistry, University of Toronto, Toronto, Ont., Canada M5S 1A8

Received 29 January 2007; revised 14 February 2007; accepted 20 February 2007

Available online 1 March 2007

Edited by Maurice Montal

Abstract **Membrane proteins that regulate solute movement are often built from multiple copies of an identical polypeptide chain. These complexes represent striking examples of self-assembling systems that recruit monomers only until a prescribed level for function is reached. Here we report that three modes of assembly – distinguished by sequence and stoichiometry – describe all helical membrane protein complexes currently solved to high resolution. Using the 13 presently available non-redundant homo-oligomeric structures, we show that two of these types segregate with protein function: one produces energy-dependent transporters, while the other builds channels for passive diffusion. Given such limited routes to functional complexes, membrane proteins that self-assemble exist on the edge of aggregation, susceptible to mutations that may underlie human diseases.**

Keywords: Membrane protein; Protein folding; Quaternary protein structure; Classification

1. Introduction

Proteins embedded in the cytoplasmic membrane have been estimated to constitute ~20–30% of all proteins in sequenced genomes [1], and regulate the trafficking water, ions, and other molecules into and out of the cell. Many of these proteins do not function as individual polypeptides but as complexes of more than one copy of a polypeptide chain [2]. The essential function of these oligomers in maintaining homeostasis means that their assembly must be carefully controlled: structures of the correct size and shape must be built, and the probability of protein aggregation minimized. Disease-causing mutations that promote non-native associations are common in membrane-spanning domains [3], implying that the disruption or failure of assembly constraints may not only lead to a loss of protein function but also promote uncontrolled and potentially pathogenic polypeptide self-assembly.

Recognition of the principles that underlie complexation in the membrane is facilitated for α-helical membrane proteins by the conceptual division of folding into two major steps [4]. The first – insertion of transmembrane (TM) segments into the membrane – specifies topology and secondary structure. The second – adoption of tertiary and/or quaternary structure within the membrane – relies mostly on lateral interactions between helices [5–7] and is facilitated by the reduced entropic penalty of helix–helix association in the bilayer [2]. Final complex assembly therefore strongly relies on a specific network of helix–helix partnerships mediated by TM helical faces – defined as individual sets of residues on distinct surfaces of each α-helical TM segment – that form the required inter-chain contacts.

What mechanisms operate to constrain the assembly process to the production of the different structures necessary for homeostasis while evading aggregation? Here, we devise a classification system to describe membrane protein oligomers and apply it to a non-redundant set of three-dimensional structures of membrane protein complexes. We find that two distinct architectures can be distinguished by sequence and stoichiometry. Subdividing proteins into these categories reveals that one produces energy-dependent transporters, while the other builds channels for passive diffusion. Each assembly pattern reflects selection for non-proliferating structures and provides useful constraints for structure and function predictions.

2. Materials and methods

2.1. Data set construction

PDB files were obtained from a database of membrane proteins of known structure (http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html); the glycophorin A structure was obtained from the Protein Data Bank. As of October 2006, this database contained ~30 entries corresponding to α-helical homo-oligomeric membrane proteins (where proteins of the same type from different species are considered unique). Structures were selected for analysis from this pool using two criteria: (i) the structure of the complete complex was represented in the PDB file; and (ii) the oligomeric state present in the PDB file was supported by biochemical evidence. Because we wished to consider only helix–helix interactions in this work, proteins where large cofactors (such as heme, chlorophyll, and/or lipid molecules) were either coordinated within the monomer fold and/or between subunit molecules were excluded. We note that each structure that met the above criteria could be classified into Types I, II, and/or III. However, to avoid bias in residue composition analysis, we filtered the transmembrane (TM) domain sequences (including loops) of these structures for redundancy by sequence alignment and percent identity calculations using ClustalX [8]. Amino acid sequences with $\geqslant 30\%$ identity were considered homologous. For sequences with $\geqslant 30\%$ sequence

*Corresponding author. Address: Division of Molecular Structure and Function, Research Institute, Hospital for Sick Children, 555 University Avenue, Toronto, Ont., Canada M5G 1X8. Fax: +1 416 813 5005.
E-mail address: deber@sickkids.ca (C.M. Deber).

Abbreviations: TM, transmembrane; LASA, lipid accessible surface area; ASA, accessible surface area; CA, contact area

0014-5793/$32.00
doi:10.1016/j.febslet.2007.02.050

identity, a single PDB file corresponding to the highest-resolution structure available was included in the data set as a representative of the group. KcsA (1BL8, 3.2 Å) was selected as representative of potassium channel tetramers rather than its Kv1.2 homolog (2A79, 2.9 Å) based on the clarity of electron density data in the TM regions (mean B factors 90–110 Å^2 [9] vs. 159–162 Å^2 [10]). This filtering process resulted in a data set with 13 unique (<30% identical) polypeptide sequences (see Table 1).

2.2. Lipid accessibility calculations

Side chain accessibility was estimated using the program NACCESS [11] by submitting each PDB file to the program and determining the lipid accessible surface area (LASA) of each side chain in Å^2 as described [12]. Briefly, a probe of radius 1.88 Å was used to approximate the radius of a methylene group on a lipid acyl chain. Side chain relative LASA values were obtained by dividing the absolute LASA in Å^2 of each side chain in the protein by its LASA in Å^2 calculated in an α-helical Gly-X-Gly tripeptide. Substrate molecules within the lumen of channels or pores were omitted from PDB files before analysis. The percentage of side chain accessibility of each subunit in isolation was calculated using PDB files where all atoms excepting those in a single monomer were deleted. This method considers each subunit as a rigid body and does not address any conformational changes that might occur upon oligomerization. The percent burial for each residue was averaged for all monomers in each structure with the exception of the aquaporin (2ABM) structure, where the conformation of subunit A differs from that of subunits B–D [13]. As such, the percent burial for each residue was averaged for subunits B–D.

2.3. Identification of interacting helix surfaces and classification

The relative accessible surface area (ASA) thresholds used the literature for defining buried vs. solvent-exposed positions in protein structures vary, with some groups utilizing 20% thresholds in membrane protein structure analysis [14], and others defining thresholds in soluble protein structures of 7% using statistical methods [15] or 5% via site-directed mutagenesis [16]. In this work, residues were divided into buried and surface positions using a side chain relative LASA threshold of 20%. The change in side chain LASA upon oligomerization (ΔLASA) was then calculated for each residue in each subunit using absolute side chain LASA values in Å^2, as follows:

$$\Delta\text{LASA} = \frac{(\text{side chain LASA}_{\text{monomer}} - \text{side chain LASA}_{\text{complex}}) \times 100\%}{\text{side chain LASA}_{\text{monomer}}}$$

Residues on the surface of the subunit (side chain relative LASA > 20%) with a >50% ΔLASA value were considered as inter-chain interaction sites. The location(s) of these residues was mapped onto each structure and used in conjunction with sequence for classification.

2.4. Contact area calculations

The contact areas (CAs) on the TM domain of each monomer were calculated as the difference between the sum of the individual side chain absolute LASA values in Å^2 in the TM regions of each polypeptide in isolation and as part of its complex, as follows:

$$\text{CA}(\text{Å}^2) = \sum \text{side chain LASA}_{\text{monomer}}\ (\text{TM segments}) - \sum \text{side chain LASA}_{\text{complex}}\ (\text{TM segments})$$

CAs were averaged for each subunit within the structure, with the exception of aquaporin, where chain A was excluded as above. TM segment boundaries were selected with reference to the literature where available, or with reference to the position of aromatic, basic and/or acidic residues in each structure.

2.5. Residue and motif composition calculations

The 20 amino acid residues were grouped into 11 categories by physicochemical character (Ala/Gly, Cys/Ser, Asp/Glu/Asn/Gln, Phe/Trp/Tyr, His/Lys/Arg, Ile, Leu, Met, Pro, Thr, Val) in order to ensure large enough counts of each residue for statistical testing. The percentage occurrence in the TM domains of each of the 11 resulting amino acid categories was then evaluated for each polypeptide within each complex for each of the total, buried, surface, and interfacial positions, as follows:

Percentage occurrence of amino acid *i*:

$$\%\ \text{Occurrence}_i = \frac{\text{No. of } i \text{ residues} \times 100\% \text{ at total, buried, surface, or interfacial position}}{\text{total No. residues at total, buried, surface, or interfacial position}}$$

Mean values for % Occurrence$_i$ were calculated for the Type II and Type III groups. Surface/buried residues and interfacial residues were defined as described as above. Because it is a single-pass membrane protein, phospholamban (1ZLL) does not have buried residues and was omitted from calculations of percent occurrence at buried positions in the Type III polypeptide set. We noted that the exact assignment of the TM region of the helices did not significantly influence the residue composition analysis, in agreement with the results of other groups [14].

Small-xxx-small motifs were defined using any combination of Ala, Gly, and/or Ser residues separated by three variable residues. The number of motifs was counted for each polypeptide in each of the total, buried, surface, and interfacial categories and averaged within the Type II and Type III groups. Both small residues of the motif were required to meet the above mentioned criteria for buried, surface, and/or interfacial residues in order for the motif to be counted in the category.

2.6. Statistical analysis and Type II/Type III comparisons

Sample sizes of the Type II and Type III groups ($n = 4$ and $n = 7$, respectively) reflect the limited availability of non-redundant, high-resolution homo-oligomeric structures where complete oligomer coordinates are available. Statistical testing on small sample sizes is complicated by the fact that they contain little information concerning their underlying distributions. With small sample sizes, normality testing has little power to discriminate between Gaussian and non-Gaussian distributions, and non-parametric tests such as the Mann–Whitney (or rank-sum) test – while they do not require normality – assume that

Table 1
Database of solved membrane protein structures used for analysis

PDBID	Type	N[a]	Protein name	Resolution (Å)	Reference
1AFO	I	2	Glycophorin A	NMR	[37]
1L7V	II	2	BtuCD vitamin B12 transporter	3.2	[38]
1OTS	II	2	H^+/Cl^- exchange transporter	2.5	[39]
2HYD	II	2	Sav1866 multidrug transporter	3.0	[40]
2ABM	II	4	AQPZ aquaporin water channel	3.2	[13]
1BL8	III	4	KcsA potassium channel	3.2	[9]
1MSL	III	5	MscL mechanosensitive ion channel	3.5	[41]
1MXM	III	7	MscS voltage-modulated mechanosensitive ion channel	3.9	[42]
1YCE	III	11	F-type Na^+-ATPase rotor	2.4	[36]
1ZLL	III	5	Phospholamban	NMR	[27]
2BBJ	III	5	CorA Mg^{2+} transporter	3.9	[43]
2BL2	III	10	V-type Na^+-ATPase rotor	2.1	[35]
1XFH[b]	II + III	3	Glutamate transporter homologue Glt_{Ph}	3.5	[21]

[a]Stoichiometry of complex.
[b]This complex is built from Type II and Type III contacts. See text for details.

the shape and spread of the two distributions is identical. We therefore have generally used the ratio of the mean values of a given parameter in Type II vs. Type III groups (or vice-versa) as a basis of comparison. In order to illustrate the potential relationship between these ratios and statistical significance, for distributions of contact areas, stoichiometries, numbers of TM segments, and numbers of small-xxx-small motifs, we assumed normal distributions and compared the mean values of these parameters for the Type II and Type III groups with unpaired two-tailed *t* tests. Means derived from normalized values – residue percent occurrences and number of small-xxx-small motifs/number of TM segments – were also compared using unpaired two-tailed *t*-tests. *P* values of 0.10 or less were deemed marginally statistically significant, and values of 0.01 or less as highly statistically significant, with $P \leqslant 0.05$ the conventional standard of statistical significance.

3. Results and discussion

We envisaged the possible ways that membrane protein complexes could be assembled in a controlled manner. TM helices within each polypeptide were first divided into equivalent (homologous or identical in sequence) or non-equivalent (neither homologous nor identical) groups. We found that the face-to-face TM helix associations that connect polypeptides within a complex could then be organized into three possible categories (symmetric (Type I), non-symmetric (Type II), and asymmetric (Type III) interactions, see Fig. 1). This categorization showed that the number, arrangement, and Type of inter-chain contacts between TM helices describes the controlled assembly of three corresponding membrane protein complex Types with distinct architectures (Fig. 1).

Whether actual membrane protein complexes exploit these theoretical assembly modes was evaluated with reference to the structural database. We constructed a non-redundant data set of 13 α-helical homo-oligomers where no two sequences were more than 30% identical (Table 1). Analysis of the lipid-accessible surface area on the structure of the assembled complex vs. its component polypeptide chains was used to define and map buried positions, surface positions, and inter-chain interaction sites. The number and location of inter-chain contact sites was then used in conjunction with TM helix sequences for classification. We found that every structure could be described as one or more of Types I, II, and/or III (see Fig. 2 for examples; Table 1 for a complete classification).

While inter-chain contact areas (CAs) could not discriminate between Type II and Type III groups, these architectures could be readily delimited by stoichiometry, and the number of TM helices/polypeptide (Table 2). The number of polypeptide chain copies present in Type II structures was significantly lower than Type III (<2-fold, see Table 2). Why are fewer chains required for function of Type II structures? It has been previously noted that 7–10 TM helices are needed to move molecules across the membrane [17]; we observe that the average number of helices present in Type II chains (~10 TM heli-

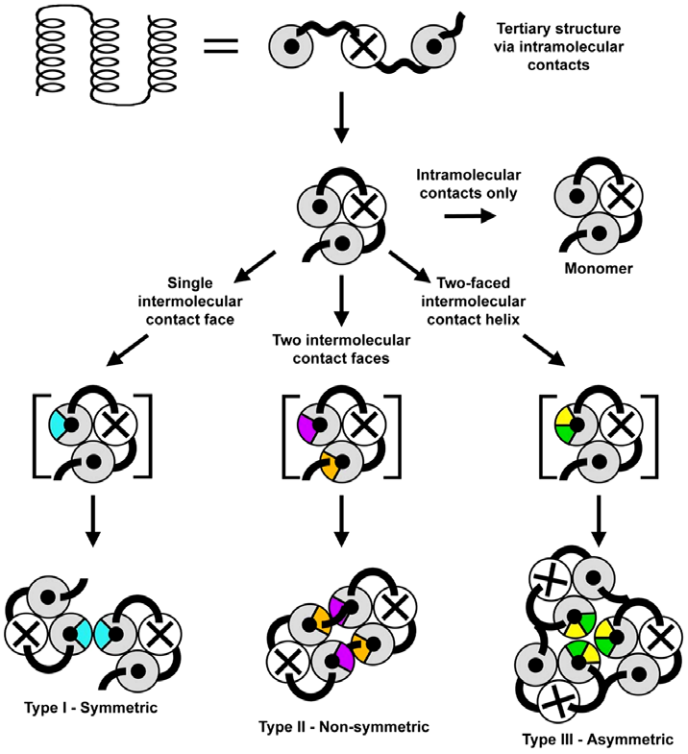

Fig. 1. Three categories of helix–helix contacts describe membrane protein complexes. Assembly is diagrammed with a model three-helix bundle monomer where circles represent end-on views of TM helices, connecting loops are shown as solid lines, dots (●) indicate an N-to-C terminal helix orientation with the C-terminus pointing out of the page, and crosses (X) indicate N-to- C-terminal helix orientation with the C-terminus oriented into the page. Grey shading indicates helices with their C-termini out of the page. Helix faces bridging chains are colored. In *Type I* or *symmetric complexes*, a single helix face (blue) contacts its counterpart on an equivalent/homologous helix. *Type II* or *non-symmetric complexes* assemble via contacts between faces (orange and purple) of two or more non-equivalent/non-identical helices. Complexes built from non-identical/non-equivalent chains therefore represent a special case of Type II assembly where each subunit is a unique polypeptide with its own set of interactive helix faces. In *Type III* or *asymmetric complexes*, two distinct interactive surfaces (green and gold) found on the same equivalent helix make "two-faced" contacts such that each retains an interaction-competent face when matched with its counterpart in another monomer.

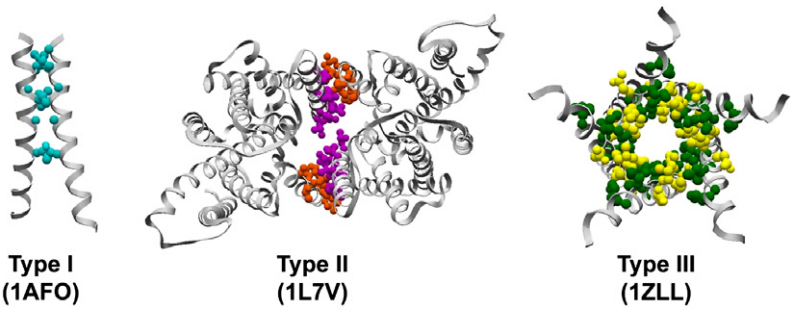

Fig. 2. Examples of each interaction Type. Ribbon diagrams of the glycophorin A (Type I), BtuC (Type II), and phospholamban (Type III) structures are shown with interfacial residues (those with >50% ΔLASA) in ball-and-stick representation. Interfacial residues are colored according to Type following the schemes given in Fig. 1. Figure produced with SwissPDB Viewer [34].

ces) falls within this range, while Type III chains are limited to ~2 TM helices each (Table 2). Once constructed, the numbers of helices in each complex are essentially identical. The lower stoichiometry of Type II vs. Type III structures therefore appears to be offset by the larger number of TM helices in Type II polypeptide chains.

Type II and Type III chains were strikingly similar in the composition of amino acids at their buried, surface, and importantly inter-chain interaction sites (Table 3). Nevertheless, certain statistically significant variations were observed: Type II chains had 16-fold more basic residues buried within their individual structures than Type III; basic residues were more abundant on Type II surfaces; and Met was more frequent at Type III inter-chain contact sites. We found the strong enrichment of basic residues in Type II polypeptide interiors to be intriguing, as the basic residue Arg is thought to play a role in permeation through aquaporins [13]. To discern any other sequence patterns that could underlie the construction of such different and distinct groups of membrane proteins, we investigated the well-characterized small-xxx-small/Gly-xxx-Gly-like interaction motif that mediates strong intra- and inter-chain helix–helix contacts (where Ala, Gly, and Ser are small residues, see [18] for review). Type II polypeptides had significantly more small-xxx-small motifs in total (~3-fold) and buried in their interiors (~5-fold) than Type III chains (Table 4). This distinction disappears when motif counts within each group are normalized to the total number of TM segments in each chain. On the other hand, the occurrence of motifs on the surfaces and at inter-chain contact sites in Type III subunits were markedly enriched when normalized, with >8-fold more motifs/TM helix at surface positions, and ~20-fold more motifs/TM helix at sites of inter-subunit contact than Type II structures (Table 4). The consistent pattern of enrichment of small-xxx-small motifs on the surfaces and at the inter-chain interaction sites of Type III vs. Type II structures led us to suspect that these motifs play a central role in the construction of Type III architectures in preference to Type II.

Once we considered the overall trends that distinguish the two groups of structures – Type II have (i) lower stoichiometry; (ii) component chains that exceed the number of TM helices needed for membrane permeation; (iii) more basic residues and more small-xxx-small motifs in their subunit interiors; and (iv) fewer of these motifs on their surfaces and inter-chain contact sites, than Type III – it seemed as though each architecture Type could have been recruited through evolution to function in characteristic ways. The cellular role(s) of each member of the Type II or Type III groups in our data set was thus examined (Fig. 3). Remarkably, we found that 3 of 4 Type II structures catalyze the uphill transport of ions and/or small molecules against their electrochemical gradient, while 0 of 7 Type III structures were capable of active transport. Instead, 6 of the 7 Type III structures promote the downhill movement of ions. We also noted that the Type II channel and channel-like structures – aquaporin and the H^+/Cl^- exchanger[1] each feature one solute pathway per chain, whereas 0 of the Type III channel subunits have this capability. For Type III structures, formation of a single conduit across the membrane requires assembly of all subunits. Interestingly, the Type II + Type III mixed-architecture homo-oligomer (1XFH, see Table 1) functions both as an active transporter and as a passive channel: (i) it couples the uphill movement of Glu to the energetically favorable movement of Na^+, K^+ and H^+ ions [19]; and (ii) catalyzes the downhill movement of Cl^- ions [20]. Like other Type II structures, each 1XFH protomer is suspected to have an independent Glu transport pathway, while the location of the chloride permeation pathway remains unknown [21]. The Type II + Type III architecture of 1XFH may therefore reflect energy-dependent transporter and -independent channel-like functionalities.

The Type/function segregation that we observe as a consequence of our classification is readily connected to stoichiometry, subunit size, and residue composition patterns in the Type II vs. Type III groups. For example, Type III relies strongly on the small-xxx-small interaction motif to mediate inter-chain contacts than Type II. In terms of residue composition, Type III assembly sites thus more closely resemble the folds of membrane protein interiors – where large numbers of Ala, Gly, and Ser residues are buried [14,22] – than do Type II interfaces. This preference may have a structural origin in the small TM domain size of Type III subunits. With too few TM segments to form channels independently [17], their inter-chain contacts may require a high degree of stability in order to ensure maintenance of a functional pore. Small residue motifs may also additionally provide pivot points for structural rearrangements, helping to confer the flexibility required for gating [23]. Conversely, the larger size of Type II subunits – within the range of TM helices needed to form a

[1]The structure examined here functions as an energy-dependent exchanger; other ClC family members are Cl-channels.

Table 2
Properties of Type II and Type III polypeptides and complexes

Property	Average[a]		Ratio[b]	P[c]
	Type II	Type III		
CA/polypeptide (Å^2)	1725 ± 476	1801 ± 836	1.0 (1.0)	0.870
Stoichiometry	2.5 ± 1.0	6.7 ± 2.8	0.4 (2.7)	0.020
No. TM helices/polypeptide	9.8 ± 4.5	2.3 ± 1.0	4.3 (0.2)	0.043
No. TM helices/complex	23.0 ± 8.9	16.6 ± 12.2	1.4 (0.7)	0.372

[a]Means ± S.D. of parameter value in Type II or Type III groups.
[b]Ratio of average value in Type II/Type III. The inverse ratio is given in parentheses. Values are rounded to the nearest decimal place.
[c]*P*-value determined using unpaired two-tailed *t*-tests.

Table 3
Comparison of Type II and Type III amino acid compositions

Residue/group	Ratio[a]			
	Total	Buried[b]	Surface	Interface
L	1.1 (0.9)	1.6 (0.6)	1.2 (0.8)	1.5 (0.7)
I	0.8 (1.2)	0.6 (1.7)	1.1 (0.9)	1.5 (0.7)
F, W, Y	1.2 (0.8)	2.0 (0.5)	1.4 (0.7)	1.3 (0.8)
M	0.7 (1.4)	0.5 (2.0)	0.8 (1.3)	**0.0**[c] (—)
V	0.7 (1.4)	0.8 (1.3)	0.6 (1.7)	0.6 (1.7)
C, S	0.9 (1.1)	0.5 (2.0)	1.4 (0.7)	1.2 (0.8)
T	1.1 (0.9)	2.3 (0.4)	0.5 (2.0)	0.5 (2.0)
A, G	0.9 (1.1)	0.9 (1.1)	0.6 (1.7)	0.6 (1.7)
D, E, N, Q	0.9 (1.1)	0.6 (1.7)	1.0 (1.0)	1.1 (0.9)
H, K, R	1.9 (0.5)	**16.9 (0.1)**	**2.2 (0.5)**	1.6 (0.6)
P	2.2 (0.5)	1.6 (0.6)	2.5 (0.4)	2.1 (0.5)

[a]Ratio of residue/group percentage occurrences in Type II/Type III. The inverse ratio is given in parentheses. Values are rounded to the nearest decimal place. Means were compared using two-tailed unpaired *t*-tests with $n = 4$ (Type II) and $n = 7$ (Type III), with the exception of buried positions, where $n = 6$ for Type III (see Section 2 for details). Ratios representing marginally significant ($P \leqslant 0.10$), significant ($P \leqslant 0.05$), or highly significant ($P \leqslant 0.01$) differences between groups are shown underlined, in bold type, or underlined and in bold type, respectively. Residues are sorted in order of decreasing hydropathy (in the case of grouped residues, mean hydropathy) according to the Liu–Deber scale [44].
[b]Buried and surface residues have ⩽20% and >20% side chain relative LASA, respectively. Interfacial residues are those surface residues with a >50% ΔLASA value.
[c]Met residues were not present at interfacial positions in the group of Type II polypeptides and had a percent occurrence of 3.1% in Type III inter-chain interaction sites.

Table 4
Small-xxx-small motifs in Type II and Type III chains

Residue/group	Ratio[a]			
	Total	Buried	Surface	Interface
No. small-xxx-small	**3.0 (0.3)**	**5.3 (0.2)**	0.5 (2.1)	0.2 (4.9)
No. small-xxx-small/ No. TMs	0.7 (1.4)	1.2 (0.8)	0.1 (8.8)	0.0 (20.7)

[a]Ratio of average value in Type II/Type III. The inverse ratio is given in parentheses. Values are rounded to the nearest decimal place. Ratios representing marginally significant or significant differences between groups are shown underlined or in bold type, respectively (see Section 2 for details).

tunnel across the membrane [17] – is a structural parameter that may relate to the ability of certain Type II subunits to form individual permeation pathways. This property – and/or the requirement for large structural shifts needed for transport – may also preclude the utilization of small-xxx-small motifs for strong inter-subunit contacts in favor of other residues [24].

Two structures, phospholamban (Type III) and aquaporin (Type II), appear to be outliers in the architecture/function relationship discerned in our analysis (Fig. 3). Phospholamban dissociates from its pentameric structure to associate with and regulate the activity of the sarcoplasmic reticulum calcium-ATPase pump in cardiomyocytes. Early ion conductance studies, however, suggest that the phospholamban homo-oligomer may also function as a Ca^{2+}-selective channel [25,26], a possibility revisited upon solution of its channel-like pentameric structure [27], and consistent with its Type III classification. Notably, phospholamban is also distinguished from the other Type III structures in terms of residue composition: it is the only protomer that contains zero small-xxx-small motifs. Aquaporin, on the other hand, differs in function from other Type II proteins in that it facilitates passive transport of water rather than active transport of ions or small molecules. Its structure is nevertheless distinct from Type III channels in that each its four component polypeptide chains has a single permeation pathway. In this, aquaporin subunits are similar to those of the homo-dimeric H^+/Cl^- exchanger, a Type II active transporter with membership in a protein family that includes pas-

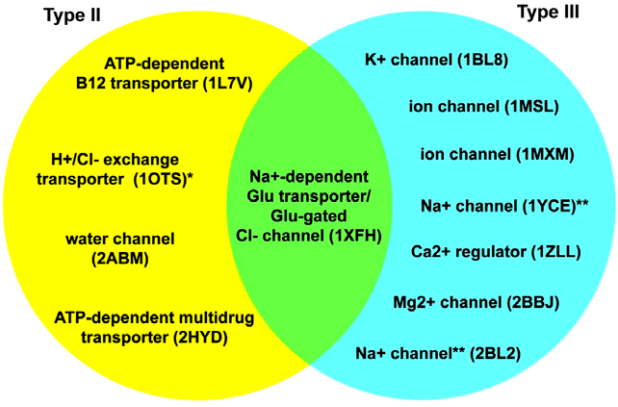

Fig. 3. Functional characteristics of Type II and Type III complexes. Type II and Type III structures are encircled in yellow and blue, respectively. The cellular role(s) of each protein is given along with its PDB identifier. Note that the 1XFH structure with mixed Type II + Type III architecture exhibits dual functionality. *Other members of the ClC family of proteins have channel rather than H^+/Cl^- exchange activity; the structure analyzed in this work is an exchanger. **These rotors catalyze the downhill movement of Na^+ ions in concert with a static membrane component of the ATPase machinery; the rotor must be assembled and in contact with additional ATPase subunits in order for sodium flux to occur [35,36].

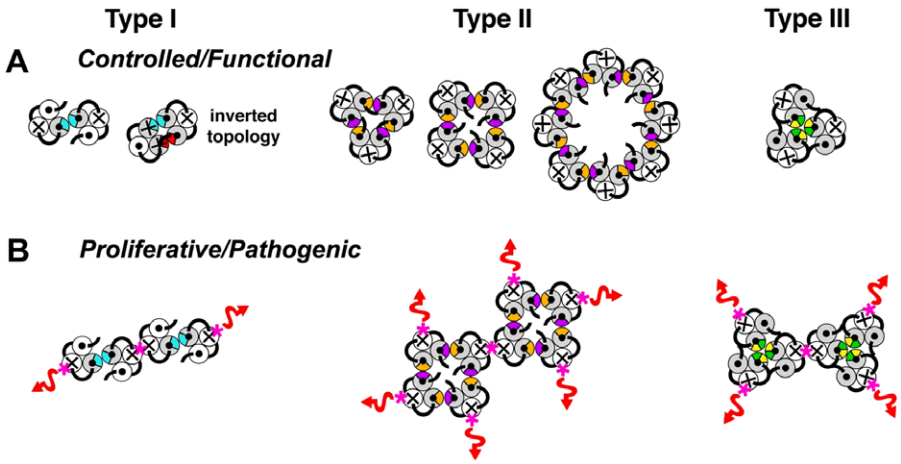

Fig. 4. Controlled (A) and proliferative (B) membrane protein aggregation. Self-assembly is diagrammed using the model three-helix-bundle described in Fig. 1. (A) Type I complexes are limited to dimer stoichiometry when a single helix face (blue) mediates inter-chain contacts, or by topology inversion when multiple symmetric interactive sites (blue or red) are present. In Type II complexes, the angle separating the interacting faces (purple and orange) on the monomer surface could potentially influence stoichiometry. Placement of inter-chain contact sites (green and yellow) on the same equivalent helix in Type III structures restricts their angle of separation, minimizing the possibility of propagation. (B) Creation of a single non-native symmetric interaction site (pink asterisk) can lead to loss of stoichiometric control and potentially pathogenic self-association in all three Types.

sive Cl^- channels. It is not yet clear in either case whether assembly is absolutely required for conduction. The potential role for complex formation in the proper intra-chain folding and/or cell surface targeting of these two proteins is also a fascinating possibility, especially given the complex "double-funnel" arrangement of their subunits [28–30].

Membrane protein complexes represent a striking example of self-assembling systems that must recruit identical polypeptides until a prescribed level for function is reached. The mechanisms by which these complexes are constructed should be distinct from proliferative structures such as amyloid. It is perhaps not surprising, then, that our scheme predicts Type-specific strategies that promote controlled association of identical chains (Fig. 4A). For example, Type I homo-oligomers are inevitably dimeric: if more than a single symmetric inter-chain interaction site participates in assembly, stoichiometry cannot be controlled without topology inversion. Utilization of Type II and/or Type III (also termed "two-faced" [31]) helix–helix contacts permit a wider sampling of sizes, and the angular requirements of placing both interactive surfaces on the same helix in Type III pairings may help to ensure eventual cyclization. Control of self-assembly is nevertheless readily disrupted; the presence of a single aberrant symmetric contact site supports out-of-control proliferation (Fig. 4B). Several circumstances could conceivably lead to this potentially pathogenic situation. Surfaces normally employed for intra-chain contacts could become exposed following improper folding of a wild-type monomer – or weakening of its fold by a residue substitution; atypical site(s) could also be created by mutation of surface residues. We note that such proliferative/pathogenic complexes are unlikely to be represented in the structural database, as these proteins should be identified as improperly folded by the cell and targeted for degradation.

Our classification system may prove useful in the prediction of protein architecture in the absence of high-resolution structures. Polypeptides with a large number of TM segments and low stoichiometry are likely to be Type II structures involved in active transport; conversely, chains with few TM segments known to specifically self-assemble are likely Type III complexes built by surfaces enriched in small residues that transport ions or small molecules down their concentration gradients. The number of TM helices/polypeptide chain can be estimated from amino acid sequence using a variety of programs (for reviews, see [32,33]), and biochemical experiments can be used to examine oligomeric state and/or determine functional properties. Given the limited routes to assembly of discrete complexes, membrane proteins that self-assemble for function exist on the edge of aggregation, highly susceptible to mutations that create non-native interaction sites.

Acknowledgements: The authors are indebted to Professor Raisa Deber for helpful discussions. We thank the anonymous reviewers for helpful suggestions and Fiona Cunningham for critical reading of an earlier manuscript version. This work was supported, in part, by a grant to C.M.D. from the Canadian Institutes of Health Research (CIHR). A.R. holds a post-doctoral award from the CIHR Strategic Training Program in Protein Folding: Principles and Diseases.

References

[1] Wallin, E. and von Heijne, G. (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. Protein Sci. 7, 1029–1038.
[2] Liu, Y., Gerstein, M. and Engelman, D.M. (2004) Transmembrane protein domains rarely use covalent domain recombination as an evolutionary mechanism. Proc. Natl. Acad. Sci. USA 101, 3495–3497.
[3] Partridge, A.W., Therien, A.G. and Deber, C.M. (2004) Missense mutations in transmembrane domains of proteins: phenotypic propensity of polar residues for human disease. Proteins 54, 648–656.
[4] Popot, J.L. and Engelman, D.M. (1990) Membrane protein folding and oligomerization: the two-stage model. Biochemistry 29, 4031–4037.
[5] DeGrado, W.F., Gratkowski, H. and Lear, J.D. (2003) How do helix-helix interactions help determine the folds of membrane

proteins? Perspectives from the study of homo-oligomeric helical bundles. Protein Sci. 12, 647–665.
[6] Engelman, D.M. et al. (2003) Membrane protein folding: beyond the two stage model. FEBS Lett. 555, 122–125.
[7] Bowie, J.U. (2005) Solving the membrane protein folding problem. Nature 438, 581–589.
[8] Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.
[9] Doyle, D.A., Morais Cabral, J., Pfuetzner, R.A., Kuo, A., Gulbis, J.M., Cohen, S.L., Chait, B.T. and MacKinnon, R. (1998) The structure of the potassium channel: molecular basis of K+ conduction and selectivity. Science 280, 69–77.
[10] Long, S.B., Campbell, E.B. and Mackinnon, R. (2005) Crystal structure of a mammalian voltage-dependent Shaker family K+ channel. Science 309, 897–903.
[11] Hubbard, S.J. and Thornton, J.M. (1993) 'NACCESS', Computer Program, Department of Biochemistry and Molecular Biology, University College London.
[12] Johnson, R.M., Rath, A., Melnyk, R.A. and Deber, C.M. (2006) Lipid solvation effects contribute to the affinity of Gly-xxx-Gly motif-mediated helix–helix interactions. Biochemistry 45, 8507–8515.
[13] Jiang, J., Daniels, B.V. and Fu, D. (2006) Crystal structure of AqpZ tetramer reveals two distinct Arg-189 conformations associated with water permeation through the narrowest constriction of the water-conducting channel. J. Biol. Chem. 281, 454–460.
[14] Javadpour, M.M., Eilers, M., Groesbeek, M. and Smith, S.O. (1999) Helix packing in polytopic membrane proteins: role of glycine in transmembrane helix association. Biophys. J. 77, 1609–1618.
[15] Hubbard, T.J. and Blundell, T.L. (1987) Comparison of solvent-inaccessible cores of homologous proteins: definitions useful for protein modelling. Protein Eng. 1, 159–171.
[16] Bajaj, K., Chakrabarti, P. and Varadarajan, R. (2005) Mutagenesis-based definitions and probes of residue burial in proteins. Proc. Natl. Acad. Sci. USA 102, 16221–16226.
[17] Liu, Y., Engelman, D.M. and Gerstein, M. (2002) Genomic analysis of membrane protein families: abundance and conserved motifs. Genome Biol. 3, (research0054).
[18] Senes, A., Engel, D.E. and DeGrado, W.F. (2004) Folding of helical membrane proteins: the role of polar, GxxxG-like and proline motifs. Curr. Opin. Struct. Biol. 14, 465–479.
[19] Zerangue, N. and Kavanaugh, M.P. (1996) Flux coupling in a neuronal glutamate transporter. Nature 383, 634–637.
[20] Fairman, W.A., Vandenberg, R.J., Arriza, J.L., Kavanaugh, M.P. and Amara, S.G. (1995) An excitatory amino-acid transporter with properties of a ligand-gated chloride channel. Nature 375, 599–603.
[21] Yernool, D., Boudker, O., Jin, Y. and Gouaux, E. (2004) Structure of a glutamate transporter homologue from *Pyrococcus horikoshii*. Nature 431, 811–818.
[22] Adamian, L. and Liang, J. (2001) Helix–helix packing and interfacial pairwise interactions of residues in membrane proteins. J. Mol. Biol. 311, 891–907.
[23] Curran, A.R. and Engelman, D.M. (2003) Sequence motifs, polar interactions and conformational changes in helical membrane proteins. Curr. Opin. Struct. Biol. 13, 412–417.
[24] Sansom, M.S. and Weinstein, H. (2000) Hinges, swivels and switches: the role of prolines in signalling via transmembrane alpha-helices. Trends Pharmacol. Sci. 21, 445–451.
[25] Kovacs, R.J., Nelson, M.T., Simmerman, H.K. and Jones, L.R. (1988) Phospholamban forms Ca^{2+}-selective channels in lipid bilayers. J. Biol. Chem. 263, 18364–18368.
[26] Arkin, I.T., Adams, P.D., Brunger, A.T., Smith, S.O. and Engelman, D.M. (1997) Structural perspectives of phospholamban, a helical transmembrane pentamer. Annu. Rev. Biophys. Biomol. Struct. 26, 157–179.
[27] Oxenoid, K. and Chou, J.J. (2005) The structure of phospholamban pentamer reveals a channel-like architecture in membranes. Proc. Natl. Acad. Sci. USA 102, 10870–10875.
[28] Fu, D., Libson, A., Miercke, L.J., Weitzman, C., Nollert, P., Krucinski, J. and Stroud, R.M. (2000) Structure of a glycerol-conducting channel and the basis for its selectivity. Science 290, 481–486.
[29] Dutzler, R., Campbell, E.B., Cadene, M., Chait, B.T. and MacKinnon, R. (2002) X-ray structure of a ClC chloride channel at 3.0 A reveals the molecular basis of anion selectivity. Nature 415, 287–294.
[30] Gouaux, E. and Mackinnon, R. (2005) Principles of selective ion transport in channels and pumps. Science 310, 1461–1465.
[31] Rath, A., Melnyk, R.A. and Deber, C.M. (2006) Evidence for assembly of small multidrug resistance proteins by a ''two-faced transmembrane helix. J. Biol. Chem. 281, 15546–15553.
[32] Lehnert, U. et al. (2004) Computational analysis of membrane proteins: genomic occurrence, structure prediction and helix interactions. Q. Rev. Biophys. 37, 121–146.
[33] Cuthbertson, J.M., Doyle, D.A. and Sansom, M.S. (2005) Transmembrane helix prediction: a comparative evaluation and analysis. Protein Eng. Des. Sel. 18, 295–308.
[34] Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18, 2714–2723.
[35] Murata, T., Yamato, I., Kakinuma, Y., Leslie, A.G. and Walker, J.E. (2005) Structure of the rotor of the V-Type Na+-ATPase from *Enterococcus hirae*. Science 308, 654–659.
[36] Meier, T., Polzer, P., Diederichs, K., Welte, W. and Dimroth, P. (2005) Structure of the rotor ring of F-Type Na+-ATPase from *Ilyobacter tartaricus*. Science 308, 659–662.
[37] MacKenzie, K.R., Prestegard, J.H. and Engelman, D.M. (1997) A transmembrane helix dimer: structure and implications. Science 276, 131–133.
[38] Locher, K.P., Lee, A.T. and Rees, D.C. (2002) The *E. coli* BtuCD structure: a framework for ABC transporter architecture and mechanism. Science 296, 1091–1098.
[39] Dutzler, R., Campbell, E.B. and MacKinnon, R. (2003) Gating the selectivity filter in ClC chloride channels. Science 300, 108–112.
[40] Dawson, R.J. and Locher, K.P. (2006) Structure of a bacterial multidrug ABC transporter. Nature 443, 180–185.
[41] Chang, G., Spencer, R.H., Lee, A.T., Barclay, M.T. and Rees, D.C. (1998) Structure of the MscL homolog from *Mycobacterium tuberculosis*: a gated mechanosensitive ion channel. Science 282, 2220–2226.
[42] Bass, R.B., Strop, P., Barclay, M. and Rees, D.C. (2002) Crystal structure of *Escherichia coli* MscS, a voltage-modulated and mechanosensitive channel. Science 298, 1582–1587.
[43] Lunin, V.V. et al. (2006) Crystal structure of the CorA Mg^{2+} transporter. Nature 440, 833–837.
[44] Liu, L.P. and Deber, C.M. (1998) Guidelines for membrane protein engineering derived from de novo designed model peptides. Biopolymers 47, 41–62.