Minireview

Inverted repeat domains in membrane proteins

Owen Pornillos, Geoffrey Chang*

The Scripps Research Institute, Department of Molecular Biology, 10550 North Torrey Pines Road, CB-105, La Jolla, CA 92037, United States

Received 8 December 2005; accepted 19 December 2005
Available online 28 December 2005

Edited by Sandro Sonnino

Abstract With the upsurge in known membrane protein structures, common structural themes have started to emerge. One of these is the inverted repeat, a tandem of \( \alpha \)-helical domains that have similar tertiary folds but opposite membrane orientations. In all previously known examples, both repeat units were encoded in a single continuous polypeptide. Recent structures of a bacterial multidrug transporter, EmrE, revealed an inverted repeat membrane protein wherein the two repeat units are assembled from two polypeptides with the same primary sequence. Here, we speculate on some of the implications of the EmrE structure with regards to our understanding of membrane protein evolution and topogenesis.

Keywords: Membrane protein structure; Evolution; Topology; Lipid; Membrane

1. Introduction

A surprisingly common feature in known structures of \( \alpha \)-helical membrane proteins is the existence of inverted repeats – tandem domains that have the same tertiary fold but opposite orientations in the lipid bilayer. These domains typically comprise three or more transmembrane (TM) segments, and can include extramembrane helices and coil regions as well. Examples of membrane protein structures with inverted repeats are shown in Fig. 1 [1–7]. They appear quite prevalent in channels and transporters.

Internal repeats in membrane proteins were first noted in the early 1990s from sequence analyses of aquaporins, the superfamily of channels that selectively conduct water, glycerol, and related molecules across cellular membranes [8,9]. The N- and C-terminal halves of these channels were shown to display clear sequence similarity, with up to \( \sim 25\% \) identity [8,10]. Even though it was appreciated then that the repeats were likely to adopt the same tertiary fold, the prevalent view was that they would also be inserted in the same orientation in the lipid bilayer, an assumption in conflict with the observation that each repeat unit contained an odd number of putative TM helices [9]. With the determination of high-resolution structures of human Aqp1 by electron crystallography [3,4] and the glycerol channel GlpF by X-ray crystallography [11] 10 years later, it was firmly established that the two aquaporin repeats are indeed inserted in opposite orientations in the membrane. The structures revealed that each repeat unit is composed of three helices that span the membrane and a fourth that only extends halfway through the bilayer (Fig. 1, upper left). As expected from the sequence, the tertiary folds of the two repeat units are highly similar; the two Aqp1 repeats superimpose with a root mean square deviation of 1.24 \( \text{Å} \) over equivalent \( \alpha \)-helical carbon atom positions [3]. Over the past several years, inverted repeats have now been observed in a growing number of unrelated \( \alpha \)-helical membrane protein structures [1–7,12–15]. In some cases, the presence of an inverted repeat was unanticipated from the protein sequence, and conserved sequence elements were identified only after subsequent alignments were made based on the crystal structure [1,6,7].

The topology of a membrane protein is fundamental to its function, because the two leaflets of cellular lipid bilayers face dramatically different environments and have distinct compositions [16]. Moreover, membrane proteins are synthesized at the cytoplasm, and with few exceptions, insert into cell membranes in a highly regulated manner [17,18]. The existence of inverted repeats, therefore, leads to interesting questions regarding membrane protein topogenesis and the evolution of membrane protein structures.

2. Small multidrug resistance (SMR) transporters as “inverted repeat” membrane proteins

In almost all known examples of inverted repeat membrane proteins, the two repeat units are encoded by tandem sequences in a single, continuous polypeptide. A first question is, how did this domain arrangement evolve? It is generally believed that membrane proteins with inverted repeats arose in two steps: (i) gene duplication of ancestral genes, followed by (ii) inversion of one domain’s topology. This evolutionary view is supported by studies of the Small Multidrug Resistance (SMR) family of transporters, proteins that couple the transport of small molecules to proton/cationic electrochemical gradients across bacterial cell membranes [19–21].

SMR proteins are typically 105–121 amino acids long, have only 4 TM helices, and function as oligomers [19]. Recent studies indicate that pairs of homologous SMR proteins, such as EbrA/EbrB and YkkC/YkkD of Bacillus subtilis, and YdgE/YdgF of Escherichia coli, combine to form active transporters, possibly as heterodimers [21–23]. Most interesting, topology prediction algorithms indicate that these protein pairs are inserted in opposite orientations in bacterial cell membranes. Experimental data has now confirmed these predictions for the YdgE/YdgF pair, and show that YdgE is inserted with a N<sub>in</sub>/C<sub>out</sub> topology, while YdgF is N<sub>out</sub>/C<sub>in</sub> [24]. Thus, the...
The putative YdgE/YdgF heterodimer is antiparallel, and given the high degree of sequence homology between these two proteins (32% identity, 65% similarity), their ternary complex is very likely to have the characteristics of an inverted repeat structure. Additionally, the \textit{ydgEF} genes are arranged as an operon in the \textit{E. coli} genome and expressed under the control of one promoter \cite{19,21}, reinforcing the idea that these two genes arose from a duplication event. It is easy to imagine how, over time, mutations in the intervening regions between the two genes may eventually result in a single polypeptide with an inverted repeat.

3. \textit{EmrE} is a homodimeric SMR transporter composed of a dual topology polypeptide

Not all SMR proteins function in pairs, and indeed, the best-characterized SMR transporter, \textit{E. coli} EmrE, is well documented to function as a homooligomer \cite{20,25–27}. EmrE expression in bacterial cells confers a multidrug resistant phenotype \cite{20}, and proteoliposomes reconstituted from pure EmrE protein and synthetic lipids display robust proton-dependent drug transport activity \cite{20,28}. Recent structures of EmrE in complex with a transport substrate, tetraphenylphosphonium (TPP), revealed that its basic structural and functional unit is indeed a homodimer, and that the bound drug and putative transport pathway are located at the dimerization interface \cite{5,29}. Remarkably, the two subunits are antiparallel, in agreement with the predicted topological arrangement of paired SMR transporters \cite{5}. In each EmrE subunit, the first three TM helices form a similar tertiary fold, a left-handed three-helix bundle – in the dimer, these six helices form an “inverted repeat” structure (colored yellow and green in Fig. 1, upper right). The EmrE-TPP structures therefore indicate that in the cell membrane, EmrE polypeptides are inserted in both possible orientations, that is, with dual topology. Although a previous study has suggested that EmrE has a unique topology \cite{30}, more recent analyses of the \textit{E. coli} inner membrane proteome by von Heijne and colleagues indicate that EmrE and other homomeric SMR proteins are likely to have mixed orientations, in agreement with the structural data \cite{24,31}. Taken together, we believe the above studies of EmrE, YdgE/YdgF, and other SMR proteins provide a convincing argument for a unifying structural model of SMR transporters as being composed of antiparallel “inverted repeat” membrane protein dimers.

4. Possible mechanism for dual topology of EmrE

Membrane protein insertion into cellular lipid bilayers is generally a co-translational process that begins when the signal sequence or first TM helix emerges from the ribosome and is targeted to the Sec protein translocation machinery (recently reviewed in Refs. \cite{17,18}). Once there, the signal/TM sequence is inserted into the heterotrimeric SecYEG protein translocation channel (Sec61 in eukaryotes), which incidentally, also contains an inverted repeat \cite{6} (Fig. 1, lower right). The TM is subsequently released into the bilayer, and successive TM segments follow until translation is complete. The current dogma is that the orientation of the first TM helix defines the topology of the entire protein \cite{17}, although there is evidence that topological signals in subsequent TM helices may also play an active role \cite{32}.

---

**Fig. 1.** Examples of known membrane protein structures containing inverted repeats. The structural repeats are colored green (N-terminal) and yellow (C-terminal), while unpaired segments are colored gray. With the exception of EmrE, all repeat units in these examples are a single polypeptide. Structural representations were created using PyMOL (Delano Scientific).
How can a membrane protein acquire dual topology? The membrane orientation of a TM helix is defined by several factors [33]. These include the disposition of charged residues flanking the hydrophobic core (positively-charged regions are generally cytosolic, a phenomenon known as the “positive-inside rule”) [34], the hydrophobicity of the TM sequence itself (which can “override” the charge bias) [35,36], and folding of hydrophilic domains N-terminal to the helix [37]. Intriguingly, recent studies by Spiess and co-workers suggest that signal/TM sequences initially insert head-on, in a N\text{out}/C\text{in} orientation [38]. The TM may then reorient until protein synthesis is completed or until further reorientation is terminated by an unknown mechanism after 40–50 s [33,38]. In principle, therefore, signal and TM sequences could be modulated in a manner that would produce equal populations in N\text{out}/C\text{out} and N\text{out}/C\text{in} topologies, and indeed, this has been done in experimental model systems [35,38]. We believe that, to a first approximation, the first TM helix of EmrE contains a naturally evolved sequence to produce such a result, although this requires direct experimental confirmation. Other variables may also come into play in determining dual topology, and further research into this area is certainly warranted.

Other membrane proteins have also been documented to have dual topology (reviewed in Ref. [39]). These include the 4-TM eukaryotic protein ductin, which function as a component of gap junctions in one orientation and as a subunit of V-type ATPases in the other [40]. Mixed orientations have also been reported for epoxide hydrolase [41] and members of the cytochrome P450 family [42]. Although examples of dual topology membrane proteins are likely to remain small in number [24], they promise to further increase our understanding of membrane protein topogenesis. At a minimum, we envision that EmrE and these membrane proteins may provide a useful complement to existing experimental model systems.

5. Possible evolutionary model for inverted repeat membrane proteins

In our view, it is likely that at least some membrane proteins with inverted repeats arose from primordial proteins similar to EmrE – dimeric (or homooligomeric) proteins with dual topology whose genes were duplicated. This view takes into account the extensive “dimeric” (or higher-order) interactions among all pairs of repeat domains observed so far. The proposed lineage is nicely illustrated by the Drug/Metabolite Transporter (DMT) superfamily of transporters, of which SMR proteins form a subgroup [43]. In addition to the 4-TM SMR family, this superfamily also includes 5-TM proteins of the Bacterial/Archaeal Transporter (BAT) family, and 10-TM proteins of the Drug/Metabolite Exporter (DME) family [43,44]. SMR and BAT proteins are homologous in sequence except that BAT family members have an extra N-terminal TM not present in SMRs. DME family members are internally duplicated in sequence, with two antiparallel 5-TM segments equivalent to those of BAT proteins [43,45]. It has been proposed that the 4-TM proteins of the DMT Superfamily gave rise to the 5-TM proteins (or vice versa), and the 10-TM proteins arose from gene duplication of the 5-TM proteins [43] (illustrated in Fig. 2). In this scheme, EmrE can be thought of as the “evolutionary link” between transmembrane domains with unique orientations and internally duplicated membrane proteins with antiparallel tandem domains. In some cases, interdigitation of the repeat units also suggest that gene insertion events may have also occurred prior to duplication, as proposed for the major facilitator superfamily of transporters [13].

6. Implications on membrane protein assembly and oligomerization

The EmrE structure also presents additional puzzles with regards to membrane protein folding and oligomerization. The two subunits in the EmrE dimer adopt slightly different tertiary folds, despite having identical primary sequence [5]. As mentioned above, the first three TM helices in each EmrE subunit fold into two three-helix bundles that form the inverted repeat. However, the disposition of TM4 is different in the two subunits (colored gray in Fig. 1, upper right). In one subunit, TM4 is packed against TM2, while the other subunit’s TM4 is packed against its TM3. These helices form part of the dimer interface, suggesting the alternate EmrE folds may be required for dimerization. Such assembly-driven alternative tertiary folds are well documented in viral capsid proteins [46]. Alternative folding of EmrE also appears to have functional consequences, and we have proposed that it allows the two subunits to dimerize in a manner that creates an asymmetric substrate translocation pathway, thereby imposing unidirectionality to drug transport [5]. More generally, the occurrence of inverted repeats further suggests that like soluble proteins, membrane proteins may be assembled in a modular
manner, using basic building blocks [2]. It is conceivable that each repeat domain represents a “folding unit” in the lipid bilayer, allowing segmental folding of large membrane proteins. Given that folding and assembly of membrane proteins are thought to be tightly linked to the membrane insertion process, these observations seem to point toward added levels of complexity to the already intricate events that must occur at and around the translocon. Conversely, given the parallels between membrane proteins and soluble proteins, it may turn out that folding and assembly principles are largely common to both protein types. In any case, further investigation should begin to unravel these events.

Acknowledgments: We are grateful to M.H. Saier, Jr. (UCSD) for discussions on membrane protein evolution, and W.M. Clemons, Jr. (Caltech) for discussions on membrane protein translocation. Work in the Chang Lab is supported by grants from the NIH (GM 67644), NASA, and the Beckman Young Investigator Award. O.P. is an NIH postdoctoral fellow.

References


