Membrane Protein Structure Determination using Paramagnetic Tags

Soumya Ganguly,[1] Brian E. Weiner,[1] and Jens Meiler[1,*]
[1]Departments of Chemistry, Pharmacology, and Biomedical Informatics, Center for Structural Biology, Institute for Chemical Biology, Vanderbilt University, 7330 Stevenson Center, Station B 351822, Nashville, TN 37235, USA
*Correspondence: jens.meiler@vanderbilt.edu
DOI 10.1016/j.str.2011.03.008

The combination of paramagnetic tagging strategies with NMR or EPR spectroscopic techniques can revolutionize de novo structure determination of helical membrane proteins. Leveraging the full potential of this approach requires optimal labeling strategies and prediction of membrane protein topology from sparse and low-resolution distance restraints, as addressed by Chen et al. (2011).

Based on their secondary structure composition, membrane proteins can be classified into α-helical bundles, examples of which are voltage-gated ion channels and G protein coupled receptors (GPCRs), and β-barrels such as outer membrane proteins or porins. About 30% of the sequenced genome represents integral membrane proteins, many of which are important drug targets. Determining the three-dimensional structure of membrane proteins is the starting point toward an atomic-detail understanding of the mechanisms that define their biological function.

Despite their biological significance, membrane proteins comprise only about 1%–2% of experimentally determined protein structures available today, mostly due to the difficulties of producing suitable amounts of protein for structural studies. Additional complication arises from the fact that membrane proteins need to be incorporated into a membrane mimetic, such as lipid bilayers, micelles, or bicelles, to retain their native fold. This leads to the third challenge of adapting techniques for protein structure determination to work with a protein in the membrane mimetic.

The use of nuclear magnetic resonance (NMR) spectroscopy is limited by the size of the protein/membrane-mimetic complex and resulting difficulties in unambiguous signal assignment, in particular for amino acid side chains. In turn, distance restraints from nuclear overhauser effects (NOEs) used for structure calculations are often limited to the protein backbone. In particular for helical membrane proteins, this links residues that are close in sequence but excludes distance restraints between remote amino acids. However, these high "information content" restraints define the protein fold (Alexander et al., 2008). Topology determination through residual dipolar couplings (RDCs), another NMR technique that gives information on the protein fold, is hampered, as many of the traditional alignment media do not apply to membrane proteins. Nevertheless, substantial progress has been made in the development of specialized NMR methods such as transverse-relaxation optimized spectroscopy (TROSY), alternative labeling strategies, and perdeuteration in conjunction with the increase in magnetic field strength (Sanders and Sonnichsen, 2006).

The introduction of paramagnetic labels is an alternative strategy to obtain distance restraints for helical membrane proteins. Typically, a paramagnetic probe is attached to the protein of interest through a cysteine residue via a disulfide linkage. The cysteine is strategically introduced using site-directed mutagenesis of a native residue—"site directed spin labeling (SDSL)." The probe selectively broadens NMR signals in its proximity in a distance-dependent manner. These paramagnetic relaxation enhancements (PREs) result primarily from a dipole-dipole interaction between an unpaired electron and a nucleus. Distances of up to 25 Å can be measured through monitoring the changes in the intensities of the peaks in NMR spectra (Battiste and Wagner 2000). This technology proved critical for the structure determination of helical membrane proteins such as DsbB (Zhou et. al. 2008) and DAGK (Van Horn et al., 2009). Similarly, the introduction of two spin labels allows distance determi-

nation in the range from 5Å–80Å by measuring the dipole-dipole interaction between the unpaired electrons using electron paramagnetic resonance (EPR) spectroscopy (Borbat et al., 2002).

Since a dedicated mutant protein needs to be prepared for every measurement, along with verifying the structural and functional integrity of the mutant, SDSL-NMR and SDSL-EPR experiments are resource intensive. Effective labeling strategies are needed to minimize the number of experiments for unambiguous topology determination. Additionally, distance restraints obtained from these experiments are not only sparse but intrinsically low in resolution. The unpaired electron resides on the tip of the tag up to 8.5 Å from the Cβ-atom of the cysteine and is connected to the backbone through a flexible linker arm. As a result, the distance restraints alone are insufficient to define the protein backbone at atomic detail accuracy.

This setting poses three formidable challenges for computational structural biology (Figure 1): (1) to determine optimal labeling strategies that minimizes the number of experiments needed to determine the membrane protein topology unambiguously; (2) to define the topology of the transmembrane segments from the low-resolution distance restraints and assign a confidence measure; and (3) to complete and refine these initial models to atomic detail that is invisible in the experimental data. Recently, a number of isolated computational techniques have been introduced that, when combined, have the potential to provide an integrated approach to tackle all three challenges.
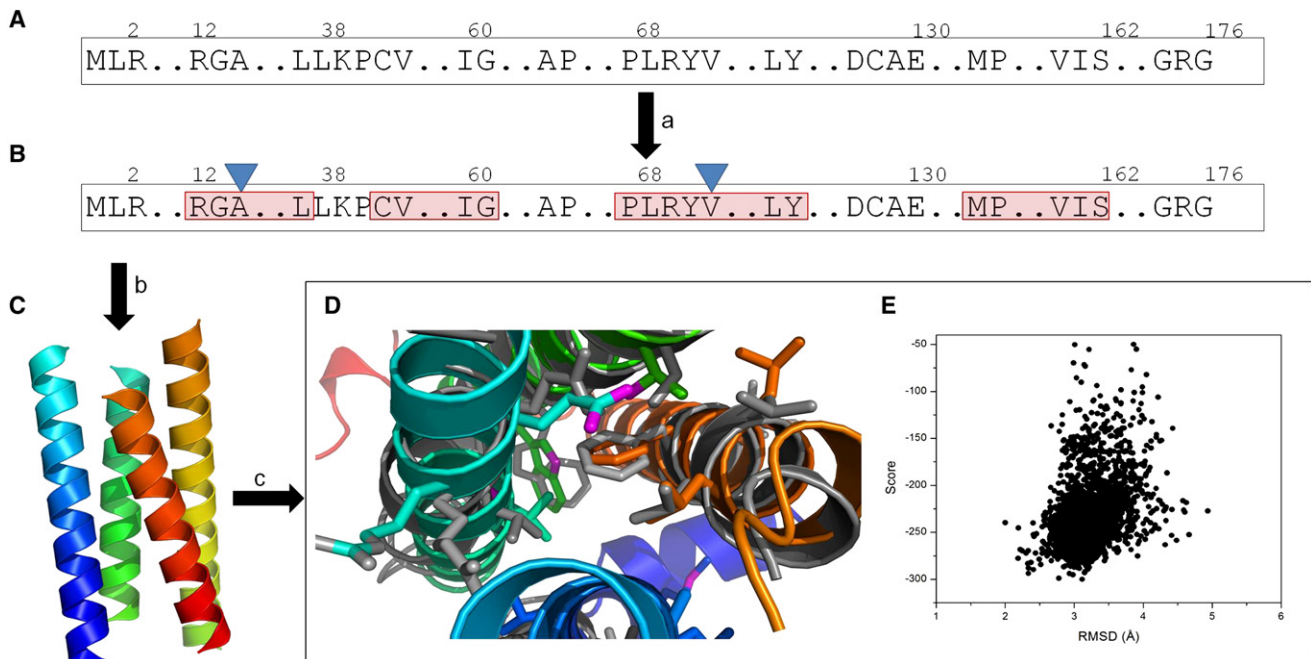
Figure 1. Protocol for Structure Prediction from Paramagnetic Restraints
(A and B) Primary sequence (A) is used to predict secondary structure elements and transmembrane span region (specified by red frames) (B).
(C) PRE distance restraints from two different sites (blue triangles, B) drive determination of the topology of the four transmembrane helices with minimum backbone rmsd of 3.2 Å compared with the crystal structure.
(D) The Rosetta refined model (rainbow) has a backbone rmsd of 2.7 Å compared with the crystal structure (gray), with similar side-chain conformation in the protein core.
(E) A Rosetta score versus rmsd plot of the models generated during the refinement process.

In this issue of *Structure*, Chen et al. (2011) discuss a computational method to determine the optimal labeling sites for collecting PRE data for helical membrane proteins from sequence information alone. The complex three-dimensional packing problem was reduced to a problem of determining the two dimensional geometry of the interacting helices by assuming ideal helix geometries in the transmembrane region, parallel to each other and perpendicular to the membrane surface. Using distance geometry, the correct topology of the four-helix membrane protein DsbB was successfully determined by considering PRE data from two tagging sites. The concept was extended to predict the optimal tagging sites of five, six, and seven helix bundle proteins. The results suggest that, to correctly predict the topology, the tags should be attached to helices that are furthest apart in the structure as estimated by predicted lipophilicity.

Kazmier et al. (2011) describe a computational algorithm for the selection of optimized labeling sites for de novo structure determination of helical proteins from SDSL-EPR distance restraints. The data suggest that one distance restraint between each pair of helices is needed for efficient determination of protein topology. In another study, the structure prediction program Rosetta was used in conjunction with a cone model that maps distance information from the flexible spin label back onto the protein backbone to determinate the protein topology (Alexander et al., 2008, Hirst et al., 2011). The same authors demonstrate refinement of initial topology models to atomic detail accuracy using RosettaEPR. It is expected that this approach can be extended to membrane proteins, as Barth et al., (2009) predicted membrane protein structures with complex topologies using limited constraints in conjunction with Rosetta and refined some of these models to high resolution.

In the coming years, substantial progress in membrane protein structure determination from spectroscopic techniques is expected through development and integration of these and similar computational approaches. Figure 1 illustrates a possible protocol for this combined approach by generating a high-resolution model of DsbB topology that has been determined using the PRE distance restraints from labeling sites A14C and V72C (Chen et al., 2011). Transmembrane helices have been assembled using BCL::Fold (Lindert et al., 2009). The initial topology model has been refined using Rosetta to an accuracy of 2.7 Å and can be identified by superior agreement with experimental data as well as Rosetta energy.

REFERENCES

Alexander, N., Al-Mestarihi, A., Bortolus, M., Mchaourab, H., and Meiler, J. (2008). Structure *16*, 181–195.

Barth, P., Wallner, B., and Baker, D. (2009). Proc. Natl. Acad. Sci. USA *106*, 1409–1414.

Battiste, J.L., and Wagner, G. (2000). Biochemistry *39*, 5355–5365.

Borbat, P.P., Mchaourab, H.S., and Freed, J.H. (2002). J. Am. Chem. Soc. *124*, 5304–5314.

Chen, H., Ji, F., Olman, V., Mobley, J.K., Liu, Y., Zhou, Y., Bushweller, J.H., Prestegard, J.H., and Xu, Y. (2011). Structure *19*, this issue, 484–495.

Hirst, S.J., Alexander, N., McHaourab, H.S., and Meiler, J. (2011). J. Struct. Biol. *173*, 506–514.

Kazmier, K., Alexander, N.S., Meiler, J., and McHaourab, H.S. (2011). J. Struct. Biol. *173*, 549–557.

Lindert, S., Staritzbichler, R., Wötzel, N., Karakas, M., Stewart, P.L., and Meiler, J. (2009). Structure *17*, 990–1003.

Sanders, C.R., and Sonnichsen, F. (2006). Magn. Reson. Chem. *44* (Spec No), S24–S40.

Van Horn, W.D., Kim, H.J., Ellis, C.D., Hadziseli-movic, A., Sulistijo, E.S., Karra, M.D., Tian, C., Son-nichsen, F.D., and Sanders, C.R. (2009). Science *324*, 1726–1729.

Zhou, Y., Cierpicki, T., Jimenez, R.H., Lukasik, S.M., Ellena, J.F., Cafiso, D.S., Kadokura, H., Beckwith, J., and Bushweller, J.H. (2008). Mol. Cell *31*, 896–908.

Deuterium Reveals the Dynamics of Notch Activation

Kopan Raphael[1],*
[1]Department of Developmental Biology and Division of Dermatology, Washington University School of Medicine, Box 8103, 660 South Euclid Avenue, St. Louis, MO 63110-1095, USA
*Correspondence: kopan@wustl.edu
DOI 10.1016/j.str.2011.03.006

Notch activation requires unfolding of a juxtamembrane negative regulatory domain (NRR). Tiyanont et al. (2011) analyzed the dynamics of NRR unfolding in the presence of EGTA. As predicted from the crystal structure and deletion analyses, the lin-Notch repeats unfold first, facilitating access by ADAM proteases. Surprisingly, the heterodimerization domain remains stable.

Notch receptors play important roles in metazoans. In vertebrates, they undergo cleavage by a subtilisin-like proprotein convertase within the secretory pathway at a site called S1 to produce two polypeptides that remain associated (Kopan et al., 1996; Logeat et al., 1998) (Figure 1A). Notch activation requires ligand-mediated unfolding of the negative regulatory region (NRR), which normally forms a protective fortress preventing ADAM metalloproteases from cleaving Notch at a site called S2 (Brou et al., 2000; Mumm et al., 2000). Following S2 cleavage, γ-secretase cleaves Notch within its transmembrane domain (TMD) at the S3 site to release the Notch intracellular domain (NICD) and thereby regulate gene expression (Kopan and Ilagan, 2009).

The NRR is composed of two distinct structural elements. The globular heterodimerization domain (HD) is further divided into HD-N (teal, Figure 1) and HD-C (yellow, Figure 1) by S1 cleavage, which occurs within an unstructured loop that does not contribute to HD stability (Gordon et al., 2007, 2009). The HD forms extensive contacts with a second domain comprised of three

calcium-binding LNR modules (LNR-A, B, and C) (Figure 1). S2 is located in an inaccessible pocket within HD-C that is buried under LNR-A (Figure 1B). The molecule shown in Figure 1B models the structure without LNR-A to expose S2; deletion studies indicate that this molecule retains sufficient structural integrity to prevent efficient ADAM-mediated proteolysis (Gordon et al., 2007). The auto-inhibitory NRR thus keeps Notch receptors in an "off" state by preventing access to S2; if S2 is protected, γ-secretase cannot cleave Notch, and the intracellular domain remains tethered to the membrane and thus unable to affect transcription.

Clearly, the HD must change its conformational state when the LNR domains are peeled off (or undergo allosteric changes) to allow the deep catalytic pocket of ADAM10 access to the scissile bond at site 2. However, the exact events involved in NRR unfolding and transition to the "on" state are shrouded in mystery. Does unfolding lead to HD dissociation, and only then to S2 cleavage (Nichols et al., 2007)? Or can the HD retain its integrity while assuming a conformational state permitting access to the S2-containing

β strand? In this issue of *Structure*, Tiyanont et al. (2011) used elegantly designed experiments to address this question.

Using hydrogen exchange mass spectrometry (HX-MS), the investigators reasoned that they could indirectly observe unfolding of purified NRR by monitoring the exchange of deuterium between the solvent and the backbone amides. While the exterior of intact globular NRR in the off state is rapidly deuterated, the internal surfaces within the NRR exchange hydrogen for deuterium slowly, or not at all. To mimic ligand-mediated activation, the investigators used Ca^{2+} ion chelation (Rand et al., 2000), which results in ADAM-dependent activation of Notch proteins (Bozkulak and Weinmaster, 2009). The reactions were quenched at different time points, followed by pepsin protease treatment. The differences in masses (i.e., deuteration level) of the pepsin-generated peptides can therefore be used to determine when a peptide that was protected from the solvent in the intact NRR became exposed to the solvent after EGTA addition. From this information, one can infer the conformational states of the NRR at different time points following EGTA addition.