Published in final edited form as:

Proteins. 2018 January ; 86(1): 57–74. doi:10.1002/prot.25402.

Comparison of NMR and crystal structures of membrane proteins and computational refinement to improve model quality

Julia Koehler Leman^{1,2,*}, Andrew R. D'Avino^{1,3}, Yash Bhatnagar¹, and Jeffrey J. Gray¹

¹Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD 21218

²Center for Computational Biology, Flatiron Institute, Simons Foundation, 162 Fifth Avenue, New York, 10010

³Department of Biology, Johns Hopkins University, Baltimore, MD 21218

Abstract

Membrane proteins are challenging to study and restraints for structure determination are typically sparse or of low resolution because the membrane environment that surrounds them leads to a variety of experimental challenges. When membrane protein structures are determined by different techniques in different environments, a natural question is “which structure is most biologically relevant?” Towards answering this question, we compiled a dataset of membrane proteins with known structures determined by both solution NMR and X-ray crystallography. By investigating differences between the structures, we found that RMSDs between crystal and NMR structures are below 5 Å in the membrane region, NMR ensembles have a higher convergence in the membrane region, crystal structures typically have a straighter transmembrane region, have higher stereochemical correctness, and are more tightly packed. After quantifying these differences, we used high-resolution refinement of the NMR structures to mitigate them, which paves the way for identifying and improving the structural quality of membrane proteins.

Keywords

membrane proteins; protein structure; protein modeling; structure refinement; structural quality; Rosetta software

*corresponding author: Julia Koehler Leman (julia.koehler.leman@gmail.com).

Current address: Center for Computational Biology, Flatiron Institute, Simons Foundation, 162 Fifth Avenue, New York, NY 10010, Jeffrey J. Gray (jgray@jhu.edu), Department of Chemical and Biomolecular Engineering, Johns Hopkins University, 208 Maryland Hall, 3400 North Charles Street, Baltimore, MD, 21218; Andrew R. D'Avino (drewdavino@gmail.com), Department of Chemical and Biomolecular Engineering, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218; Yash Bhatnagar (yashdbhatnagar@gmail.com), Department of Chemical and Biomolecular Engineering, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218

Author contributions

YB worked on the database, ARD analyzed the data, JKL conceived the idea, worked on the database, analyzed the data and wrote the paper with help from JJG.

Conflicts of Interest

This arrangement has been reviewed and approved by the Johns Hopkins University in accordance with its conflict of interest policies.

References for table:

2KIH⁶³BKID⁶⁴2K73⁶⁵2ZUQ⁶⁶1R84⁶⁷1M0L⁶⁸2KSY⁶⁹1H2S⁷⁰2L6X⁷¹4KNF⁷²1LN6⁷³1U19⁷⁴2KDC³⁸4UXX⁷⁵2M07⁷⁶1QJ8⁷⁷2N2M⁷⁸3QRC⁷⁹1MM5⁸⁰3GP6⁸¹2GE4⁸²1QJP⁸³2MHL⁸⁴2F1V⁸⁵2JQY⁸⁶2X9K⁸⁷2K4T⁸⁸3EMN⁸⁹

Introduction

Membrane proteins (MPs) are important drug targets¹, yet remain extremely difficult to study because of the hydrophobic environment that surrounds them. This situation leads to a number of questions, some of which we attempt to investigate here: How well do membrane mimetics (see Supplement) used for structure determination mimic native membranes? How do MP structures differ from their soluble counterparts? How many of these differences can be attributed to (1) the hydrophobic environment of the membrane; (2) the chosen method for structure determination; (3) the different environments in a crystal vs. solution; (4) the influence of the membrane mimetics; and (5) the different environment in specific lipids or detergents? How can we identify and quantify ‘artifacts’ arising during structure determination? And how can we correct such features?

The membrane bilayer is a heterogeneous environment composed of chemically different lipids^{2,3} that are in constant movement around the protein. An asymmetric distribution between the inner and outer leaflet⁴, lateral diffusion, and formation of ordered microdomains (i.e. lipid rafts⁵) make the natural bilayer difficult to model experimentally or computationally. Further, to determine MP structures, the protein must be concentrated in a chosen membrane mimetic, that often depends on the method for structure determination, where the exact conditions are found through trial and error^{6,7}. Crystallographers often use micelles or lipidic cubic phases⁸, optimizing conditions based on crystal quality and diffracting resolution, typically sacrificing a native-like environment. Similarly, for NMR spectroscopy ‘optimal conditions’ are found by improving spectral quality to obtain the largest number of restraints, which, for MPs, are often sparse and of low resolution. These restraints are then used to compute structural models via MD simulations, which in turn work best with rich restraint datasets that are typically difficult to obtain for MPs⁹. In comparison to crystallography where protein flexibility often leads to missing density and hence to missing atom coordinates, a lack of restraints in NMR does not lead to missing atom coordinates, but rather to less reliable ones. The latter poses a problem when a single NMR model is used as a representative of an entire ensemble, which is unfortunately done too often.

One can therefore anticipate that some of the differences between NMR and crystal structures arise from the method for structure determination. For soluble proteins, Garbuzynskiy and co-workers compiled a set of 78 proteins¹⁰ and found that the RMSD between NMR and crystal structures are generally larger than the RMSD within the NMR ensemble, which indicates well-converged NMR ensembles. Montelione and co-workers came to similar conclusions on 148 structure pairs¹¹, where for 76% of the pairs this difference was larger than a factor of two. On average, the backbone RMSD between crystal and NMR structure was about 1.0 and 1.4 Å over core and all residues, respectively, and was independent of age or quality of the NMR models. Carugo et al. saw similar RMSDs (in the range of 1.5-2.5 Å) on 109 soluble protein structure pairs¹². They found that sidechain conformations of hydrophobic amino acids were more similar in crystal and NMR structures than of hydrophilic ones, which is expected because hydrophobic amino acids are well-packed and buried in the protein interior. Further, differences between NMR and crystal structures were on average smaller for β-strands than for helix or loop residues, differences

in loops were independent of crystal packing effects, and rotamer variations of buried sidechains were rare.

Here, we identify measures to quantify structural differences in MPs determined by crystallography and NMR. So far, there are no well-established metrics to measure the biological validity of a MP structure, and experimental verification remains difficult as structure determination in native-like membranes or vesicles remains virtually impossible to date. Functional assays can verify biological activity of the protein, yet, these are often carried out in vesicles or cells under very different conditions than those being used for structure determination¹³. Conversely, functional assays in membrane mimetics used for structure determination (micelles, bicelles or nanodiscs) are practically challenging. Determining structures in multiple mimetics with different techniques provides insights, but is extremely labor intensive and costly. Nevertheless, this has been accomplished for some proteins¹³, for instance the influenza virus A M2 protein¹⁴ and bacteriorhodopsin.

It was suggested that, once non-native features are identified in a MP structure, computational methods, such as restrained molecular dynamics^{15,16}, may be used for structural refinement¹³. Lower-cost alternatives are Monte-Carlo simulations, for instance in the Rosetta software suite. For soluble proteins, Mao et al. used restrained refinement and *de novo* structure prediction in Rosetta in conjunction with NMR data to refine NMR models¹⁷. Their refined models had smaller RMSDs to the crystal structures and were better suited for phasing them. Further, the refined models had on average fewer dihedral angle restraint violations and had improved stereochemistry.

Here we investigate the differences between NMR and crystal structures of MPs on a dataset of 14 structure pairs; to our knowledge, this is the largest comparison of MP structures so far. We use this dataset to establish measures to evaluate the structural quality, identify differences from structure determination, and test an efficient computational refinement approach in Rosetta to reduce them. While questions about the influence of distinct membrane mimetics and lipid specificity remain due to the small number of currently available structure pairs, we provide measures to identify and correct for non-native features in MP structures.

Materials and Methods

Dataset generation

We manually curated the PDB to find membrane proteins that had both crystal and NMR structures and found 14 proteins that match this criterion. All proteins had more NMR or crystal structures than the single pair. For crystal structures, we only added the highest resolution structures to our dataset. For NMR structures, we only considered solution state structures, since solid-state structures only exist for M2 (2L0J¹⁸, 2KQT¹⁹ and others) and one each for DsbB (2LTQ²⁰) and sensory rhodopsin (2M3G²¹). Further, for the few available solid-state NMR structures, the type of data used varies widely between *de novo* structure determination (without template) for sensory rhodopsin, ssNMR data in conjunction with data from Xray reflections and MD simulations for DsbB. And while the chemokine

receptor CXCR1 has a ssNMR structure (2LNL²²), no crystal or solution state NMR structures exist for it.

We then manually identified the ‘best’ representatives for a pair of structures (NMR and crystal structure), given the following considerations: (1) If possible, the structures should be from the same organism (not the expression system); (2) no structural intermediates, for instance for Bacteriorhodopsin ~100 structures exist in the PDB to date, some of which are from intermediates; (3) high convergence for NMR ensembles as visually identified; (4) large number of resolved residues or similar number of residues in both structures; (5) small number of mutations between the sequences in the pair. However, it cannot be precluded that thermostabilizing mutations exist that affect the structure, as was for instance shown for the DAGK crystal structure by solid-state NMR measurements²³. Details about structural and sequence alignments can be found in the Supplement. The final dataset for the pairs of structures is shown in Table 1 and Figure 1.

RMSD calculations

Two different RMSD calculation approaches were necessary since all models in the NMR ensemble contain the same number of residues, yet the NMR and Xray structures may not.

RMSDs within NMR ensemble—Structure files were ‘cleaned’, i.e. HETATMs removed, one occurrence of seleno-methionine was mutated back to its original isoleucine, and renumbered consecutively for input into Rosetta. We used PyMOL’s *align* routine with its 5 default cycles to find an optimal superposition of the models. This routine discards atoms in ill-defined regions of the protein in an iterative fashion to find an optimal superposition. Backbone RMSDs were computed with the *rms_cur* routine on *all* backbone atoms in the NMR structures, irrespective of whether the align routine had previously discarded them. Rosetta span files were computed from the first NMR model in the PDB using the *mp_span_from_pdb* application²⁴. For the TM region backbone RMSDs, only the backbone atoms of the residues in the span file were used.

RMSD between Xray and NMR—From the processed PDBs (see Supplemental Experimental Procedures: Structural and sequence alignments) we created trimmed models which only contain the residues that are present in both structures of the pair to allow for direct comparison between the structures and the quality measures computed for them. Structure files were then ‘cleaned’, seleno-methionine was mutated back to methionine, and the residues were renumbered. Rosetta ‘span files’ were computed from the trimmed PDBs²⁴ (note that these structures are already embedded into the membrane bilayer as described in the Supplement). Structures were superimposed in PyMOL²⁵ with the *align* routine, which iteratively discards atoms in ill-defined regions of the protein to find an optimal superposition. Then, backbone RMSDs were computed for *all* backbone atoms in the trimmed structures using PyMOL’s *rms_cur* routine which computes the RMSD without additional fitting or superposition. For the TM region backbone RMSDs were computed from backbone atoms of the residues in the span file.

Computing quality measures

TM bend angles—Bend angles were defined as displayed in Figure 5B and Supplementary Figure 12: From the span files, the lower, center, and upper residues of the TM spans ($z > 0$, $z \sim 0$, $z < 0$) were used to define three ‘layers’. For each layer, the xyz coordinates of the Ca atoms were used to compute the center-of-masses and the average radii. From the average radii and an assumed hydrophobic thickness of the membrane bilayer of 30 Å, the opening angle was computed (center panel in Supplementary Figure 12) and the bend angle was defined as $(180^\circ - \text{opening angle})/2$. The bend angle is therefore not necessarily symmetric across the membrane midplane.

MolProbity for structural quality—Molprobity²⁶ was downloaded from <https://github.com/rlabduke/MolProbity> and invoked by `/MolProbity/cmdline/multichart <PDB file>`. The output is written into html format which was analyzed by Python scripts. Details about how per-residue clash scores were computed in Rosetta, are given in the Supplement.

Naccess for relative surface areas—Relative per-residue surface areas were computed with the Naccess program²⁷ with a probe size of 2.0 Å that is representative of a methyl group from a lipid acyl chain. The command line used was `naccess protein.pdb -p probe_size -r vdw.radii -s standard.data`

High-resolution refinement

Rosetta high-resolution refinement was carried out on the trimmed structures with a recently developed MPRangeRelax algorithm (manuscript in preparation, see Supplement). This protocol carries out small dihedral angle perturbations and up to three cycles of rotamer refinement and minimization of the protein.

The algorithm is available in the Rosetta software suite, freely available for download for academic users at www.rosettacommons.org. Details and command lines can be found in the Supplement. From 1,000 built models for each NMR structure (only MODEL 1 was considered), the top-scoring models were used for evaluation. Backbone RMSDs to the trimmed crystal structures were automatically computed during the refinement and output into the score file. TM region backbone RMSDs were computed with the `rms_cur` routine in PyMOL as described above.

Results and Discussion

We curated a dataset of MP structure pairs determined by both crystallography and solution state NMR. Solid-state NMR structures were excluded, see Methods. We chose the pairs according to resolution, convergence in the NMR ensemble, and sequence similarity (with fewest mutations as possible) and removed all ligands and hetero-atoms (see Methods). The final dataset contained seven α -helical and seven β -barrel MPs with protein sizes from 100 to 363 residues (Table 1 and Figure 1). While this dataset is too small to derive information of statistical significance, it is an excellent starting point to identify differences and trends within those structure pairs that can be substantiated when more structures become available.

The biggest differences between NMR and crystal structures occur for (1) DAGK, which has a domain swap in the NMR structure that is not present in the crystal structure (more on this below); (2) the NMR structure of bovine rhodopsin (BovRh) is one of the earliest structures computed from overlapping fragments determined in DMSO – this structure was added for completeness purposes; (3) M2, which has different functional states with a different opening angle. All crystal structures have a large pore opening (the 3BKD structure has a detergent molecule wedged in the pore) and all NMR structures have a small pore opening, which impacts superposition. Both structures are also from different constructs (see Figure 1) and the NMR structure is a drug-resistant mutant while the crystal structure is a wild-type construct.

The TM regions of NMR ensembles converge within 5 Å RMSD

For NMR structures, convergence is typically used as a measure of structural quality. To estimate convergence of an NMR ensemble, we computed backbone RMSDs between the top NMR model to all other models in the ensemble. We use the top model as the reference because PDB deposition requirements state that a representative of the ensemble should be placed as the first model in the PDB entry (http://deposit.pdb.org/depoinfo/print_nmr.html). Figure 4 shows that the backbone RMSDs over all residues extends as high as ~14 Å due to regions in the protein that are not well-restrained. We see such less restrained regions for the amphipathic helices in both DAGK and PagP (Figures 1 A.7 and B.3) and for the long, unstructured C-termini of DsbB and SensRh (Figures 1 A.2 and A.4); unstructured termini were excluded for these calculations (see Methods). Further, long loops in OmpA, OmpW, and OmpG give rise to RMSDs up to ~10 Å (Figures 1 B.4–6). When considering only TM regions, the backbone RMSDs are below ~5 Å for all proteins with the largest ranges for VDAC, OmpW, and DAGK. The coordinates of BactRh have an extraordinarily high convergence, with all 12 NMR models being within 0.1 Å RMSD!

Convergence in the NMR ensemble is required but insufficient to satisfy model quality measures

So how does convergence of the NMR ensemble relate to ‘structural quality’? To answer this question, we distinguish here between (1) the convergence *within* the NMR ensemble and (2) the differences *between* NMR and crystal structures. Note that quality assessment of NMR structures has been discussed recently in several papers^{28,29}.

First, convergence is a measure that describes the NMR ensemble, i.e. the relation of different models with respect to each other (completely independent of the crystal structure). Convergence simply means that there is a sufficient number of restraints to create models with little variation between them. Hence there is higher convergence in the TM region due to the larger number of measured restraints. Further, there are no enforced standards on how many NMR models are reported in the PDB and how many models are computed. For instance, the top 10 structures from 1000 computed trajectories might have a better convergence, yet expose subtleties of the energy function used in the calculation, while the top 10 structures from 50 generated models might show more structural diversity.

Secondly, the NMR and crystal structures are almost always determined under different environmental conditions, which can give rise to small structural variations. That means that a well-converged NMR ensemble can still differ from the crystal structure³⁰, as we can see in the splaying of OmpA (Figure 3 B.4). Also, the crystal structure of *E. coli* OmpX shows a more elliptic barrel, while it is more circular in the well-converged NMR ensemble (data not shown). These differences might also be explained by the type of restraints used for NMR model calculations. For β -barrels, NMR restraints from backbone Nuclear Overhauser Enhancements (NOE's, i.e. distance restraints up to 6 Å between atoms more than 4 residues apart) used for structure calculations are long-range restraints in terms of sequence separation, while they are local restraints considering the measured distance. However, (in contrast to crystallographic electron density maps) their long-range character is insufficient to reach across the barrel to distinguish between a circular or elliptic barrel. Further, differences between NMR and crystal structures may also originate in the fact that MD simulation methods used for structure calculations are not designed for sparse datasets, and tools that are better suited are not widely used for structure calculations. For example, the RASREC protocol in Rosetta has been shown to produce higher-quality solution structures with sparse data^{9,31}.

Interestingly, we observe a correlation ($R = 0.84$) between the maximum backbone RMSD within the NMR ensembles and the backbone RMSDs between the crystal and NMR structures (Supplemental Figure 4), when DAGK is removed because of its domain swap, the correlation is as high as 0.90. This correlation likely arises because atoms, that have large deviations from the average model within the NMR ensemble, also deviate considerably from the crystal structure (i.e. these would be weakly restrained regions).

The type of NMR restraints influences structural quality

Similarly, as the structure determination method can influence the structural representation, the number and types of NMR restraints can influence the quality of the NMR structure. Additionally, which becomes especially important for large proteins like MPs, the type of the protein (α -helical bundle or β -barrel) influences the type of measurable restraints as well as the type of restraints required for protein fold determination or structure calculation. For β -barrels, the resonances in the NMR spectra have a larger spectral dispersion, leading to less signal overlap. This facilitates and therefore increases the number of possible resonance assignments and the acquisition of restraints. Further, protein fold determination for β -barrels is possible through backbone hydrogen bonds between neighboring strands, which requires 'as little as' backbone assignments and the measurement of long-range NOEs between neighboring strands. For α -helical proteins however, the spectral dispersion is reduced, which results in larger resonance overlap, complicating and therefore decreasing the number of resonance assignments and hence the number of measured restraints. It is much more difficult to obtain sidechain assignments and restraints involving sidechain atoms between neighboring α -helices, which are especially important for protein fold determination of α -helical proteins. Including long-range restraints from sidechain atoms in structure calculations can drastically improve structural quality, as has been shown for SensRh, for which over 1500 long-range sidechain restraints could be measured via ILV methyl labeling (Supplemental Table 1). If such NOEs are unavailable for helical bundles,

PREs or RDCs are commonly acquired to compensate for the lack of long-range restraints. In addition to the difference in measurable restraints and structure calculation, no standard format exists on how structural quality measures are reported, even if NOEs are reported. For instance, when gathering information about the published structures for Supplemental Table 1, NOEs for α -helical proteins were reported separately for intra-residue, sequential, medium-range and long-range NOEs. For β -barrels, even though most of the NOEs are long-range (see VDAC in Supplemental Table 1), no such distinction was made for most of the proteins.

For the α -helical structures in our database, SensRh has a well-defined ensemble with the largest number of long-range NOEs per residue. For the β -barrels, the structure can be determined by a comparably small number of long-range NOEs alone: For OmpX, the *Yersinia pestis* construct has about three times the number of long-range NOEs as the *E. coli* construct, yet the backbone RMSDs in the TM region are similarly small. OmpA is the only barrel for which RDCs were measured and the TM domain has the best convergence among β -barrels, as indicated by the smallest backbone RMSDs within the ensemble.

TM regions between crystal and NMR structures agree within 5 Å RMSD

To identify differences between NMR and crystal structures, we computed the backbone RMSDs between them for trimmed models that had the same number of residues (see Methods). The backbone RMSDs for the full-length structures can vary considerably (Figure 5A). DAGK contains a domain swap of the helices between the NMR and crystal structures^{23,32} (see below for discussion on this); the RMSD over the TM regions without the domain swap is 6.9 Å. Large differences are also seen for PagP (Figure 3 B.3), which contains an amphipathic N-terminal helix (which is either dynamic, poorly restrained, or both – see below), and OmpG (Figure 3 B.6), which has a long loop that protrudes straight in the crystal structure while wrapping around the micelle in the NMR model. OmpG therefore seems to have a thin hydrophobic dimension in the NMR structure while having an unusually thick one in the crystal structure (Fig. 3.B.6). Further, the NMR model of BovRh is distorted since it was computed from structures of overlapping fragments in DMSO solvent. For M2, the structural alignment is inferior due to opening angle differences between NMR and crystal structures that comes from different functional states.

Figure 5A further shows that while RMSDs in the TM region vary across α -helical MPs (1.0 – 5.6 Å without DAGK), they are relatively uniform for β -barrels (1.1 – 2.9 Å). The small range for β -barrels suggests little flexibility to satisfy a more rigid backbone hydrogen bond network, which is required to maintain protein stability in the membrane. In contrast, α -helical proteins may permit more flexibility in their TM region due to the irregular hydrogen bond network among sidechain atoms, in which bonds can be broken and reformed between different residues without compromising overall protein stability in the membrane.

For full-length constructs, backbone RMSDs between crystal and NMR structures are influenced by loop regions that are either dynamic, poorly restrained, or both. As seen from Figure 5A, backbone RMSDs between the full-length proteins and the TM regions differ substantially for β -barrel proteins (2.2 – 11.8 Å) due to their long loops. This is not the case for α -helical proteins (0.5 – 1.8 Å difference) in which long loops are absent, except for

DsbB whose long loop between TM3 and TM4 has a similar conformation in both the NMR and the crystal structure.

And what about dynamics?

One of the challenges in comparing NMR and crystal structures is the consideration of dynamics³³. While crystal structures are static snapshots, the B-factor can indicate thermal motion within an observable range³³. However, when certain regions are too dynamic, they don't give rise to well-defined electron density and therefore often have missing atom coordinates in the crystal structure. One of the major advantages of NMR spectroscopy is the ability to measure dynamics^{34–36}, yet the structures are often misinterpreted: variable regions in an NMR ensemble are often considered dynamic, which might or might not be true. In reality, variable regions in an NMR ensemble only point to fewer available restraints for these regions, while dynamics can only be measured by relaxation experiments such as T1, T2 or heteronuclear NOEs, which are not captured in the structural ensemble. Due to these inherent differences in methods and interpretations, it remains difficult to address whether 'flexible' regions are poorly restrained, truly dynamic, or both, without investigating each region individually.

Crystal structures more often have smaller bend angles indicating a straight, cylinder-like TM region

While most crystal structures have a straight TM region that resembles a cylinder, some of the NMR structures have TM regions that resemble stretched or compressed cylinders (concave like a one-sheet hyperboloid vs. convex like a wine barrel). Examples are DAGK, OmpA, and OmpG (Figures 1 A.7, B.4 and B.6) – for some of these examples, one can imagine that the loop regions wrap around the surface of the micelle in which the structure was determined. Because such deviations are not present in the crystal structure, we created a bend angle measure to quantify them (see Methods and Figure 5B): a smaller bend angle represents a straighter TM region. Note that the bend angle here defines the general straightness of the entire TM region (see inset in Figure 5B), but is independent of kinks or bends in individual helices or strands.

For many of the proteins in our dataset, the bend angles are 5–10° smaller for the crystal structures, indicating that the TM regions for those structures are straighter. For M2, both NMR and crystal structures splay (the crystal structure slightly more than the NMR structure) which seems an inherent characteristic of the structure. For ProtRh the structural differences are small: while both original PDBs contain the retinal ligand, the crystal structure is more tightly packed, while the NMR structure has a larger ligand cavity.

The largest variations in bend angles (>5°) can be seen for DAGK, BovRh, PagP, OmpA, OmpG, and VDAC. BovRh is an outlier due to its poor structural quality. DAGK has marked differences in bend angles (Figure 5B), which might be explained in different ways. Proteins crystallize more readily in the absence of dynamics and when they are in a low-energy state, which could be achieved in a dense packing arrangement in the crystal. For DAGK, the tight packing arrangement might lead to a straightening of the helices (see Supplemental Figure 3). Further, the absence of orientational NMR restraints (such as RDCs) that report on helix

bending might lead to distortions in the structure. For DAGK, the Q-value³⁷ between experimental and back-calculated RDCs is much smaller in the NMR structure (Q=0.03) than for the crystal structure (Q=0.55), which points to the micellar environment as a major source of differences.

DAGK also has a domain swap in the NMR structure, which is not present in the crystal structure. The domain swap could be explained by slight differences in refolding protocols^{38,39} and/or the ambiguity of NMR restraints used for structure calculation of the homo-oligomeric protein. In fact, Bruce Donald's group showed that there are nine conformational solutions for DAGK that satisfy the NMR restraints and both the NMR and the crystal structure are valid solutions³².

While it is interesting that many (especially β -barrel) structures have larger bend angles in the NMR structure, the origin of these differences remains somewhat elusive (for a summary, see Table 2 and discussion below). Our initial thought was that the smaller lateral pressure in micelles could lead to a less stable environment and therefore to bending. However, many of the crystal structures considered here are also determined in micelles. Another possibility is that loop regions reach out to interact with the curved micelle surface in the NMR structure while they are thermally restricted in the crystal, where they interact with residues of other subunits. Further, crystal structures are generally straighter, irrespective of the membrane mimetic. It is interesting to note that all three structures determined in LCP (BactRh, DAGK, SensRh) and both structures determined in bicelles (ProtRh and VDAC) show lamellar (i.e. stacked) bilayers, whereas all except one structure in micelles (M2) show sponge-like surfaces (Supplemental Figure 3). DsbB has large contacts between the antibody fragments that are inserted for crystallization purposes. Finally, we want to emphasize that while we quantify bend angles here, we don't want to imply native likeness of one or the other since the membrane mimetics used for structure determination by neither method are biochemically similar to native membranes that occur in the cell.

Crystal structures have generally better stereochemistry

To obtain more detailed measures of the structural quality, we used the MolProbity structural validation software²⁶. The overall MolProbity score combines measures about steric clashes, Ramachandran outliers and rotamer outliers into a single score; the lower the MolProbity score, the higher the structural quality. Compared to the other measures of bend angles and packing, stereochemistry for membrane proteins may be less important for identifying the most native-like structure because membrane protein structures are currently only determined to low resolution and may have larger structural distortions due to their environment.

Figure 5C shows that the MolProbity scores are typically worse for the NMR structures, except in cases for DsbB, *E. coli* OmpX, and OmpW, with no correlation between the year the structure was determined and its MolProbity score. Supplemental Table 2 contains the individual scores and shows that crystal structures usually have more favored rotamers and Ramachandran dihedrals, while having fewer rotamer and Ramachandran outliers. At the same time, deviations in C β positions, bond lengths, bond angles, and proline geometry are often smaller for NMR structures. This seems to indicate that MD simulation software with

which NMR structures are computed, is well able to produce good local geometry in bond lengths and angles etc. while it produces less optimal backbone dihedrals and rotamers than in crystal structures. This can occur when the NMR restraints are not turned off (or down-weighted) during the final refinement stage, where they compete with optimal rotamers and Ramachandran angles. Finally, both BovRh and DsbB have bad clash scores for both NMR and crystal structures.

The overall MolProbity clash score (Supplemental Table 2) is similar between NMR and crystal structures. However, the per-residue Rosetta full-atom repulsive score as an indicator of clashes, shows a distinct pattern (Figure 6 and Supplemental Figure 5); While NMR structures typically have more clashes of low severity, crystal structures have few severe clashes. For NMR structures, this may be because a typical criterion for structural quality is satisfaction of restraints from which the model is computed. Considering the large number of required restraints and their typically low quality for MPs, it is difficult to satisfy all restraints at the same time, especially given errors in the data and/or the interpretation of the data, for instance due to incorrect resonance assignments of overlapping peaks. In contrast, in crystallography, the protein structure is traced along the electron density and structure refinement may be carried out with a high degree of automation. It is conceivable that most residues fit the density well but that a few residues are difficult to place due to the low resolution of the electron density and that such errors may easily be overlooked during refinement.

Crystal structures generally pack more tightly than NMR structures

We used different measures to probe packing differences between NMR and crystal structures. Figure 5D shows the relative per-residue accessible surface areas (rASA²⁷) that are averaged over the number of residues. Except for M2 and BovRh which both have very similar values, the rASA is smaller for crystal structures, indicating that they are more tightly packed. The differences between NMR and crystal structure are usually larger than ~5% rASA.

Tighter packing was also confirmed through calculations of total ASA, protein volume, solvent volume in the protein interior, surface crevices with the 3V software⁴⁰ and packing scores from RosettaHoles⁴¹ (Supplemental Figure 6). All data reinforce that the NMR structures are less tightly packed, leading to a larger surface area and more rugged surfaces (Supplemental Figure 8) that are more accessible to solvent within the protein and on its surface. We also discuss Glycine accessibility in Supplemental Figure 7.

Tighter packing for crystal structures is consistent with the notion that proteins crystallize more readily in a low-energy state when motion is minimal. Thermal expansion could be considered as the root of packing differences as for all but one (VDAC at 298K) crystal structures data was collected at 100K while for the NMR structures, spectra were acquired between 283 and 323K. However, this is inconsistent with the fact that for (at least) four MP structures crystallized at different temperatures, the RMSDs between the high- and low-temperature structures were well below 1 Å:

- δ -opioid receptor with PDBIDs 4RWD/4RWA/4N6H at 294/100/100K have an RMSD of 0.8Å
- adenosine receptor: 5K2C/4EIY at 292/100K have an RMSD of 0.4Å
- photosystem II: 5WS6/3WU2 at 293/100K have an RMSD of 0.3Å
- photosynthetic reaction center: 5M7K/2WJN at 293/100K have an RMSD of 0.4Å.

Since thermal expansion is less likely as the source of packing differences, we argue that the different environment between a crystal and solution is a major source of packing differences.

Packing density could be influenced by the membrane environment

Packing density can further be influenced by the hydrophobic environment of the membrane bilayer. Several groups have examined packing differences between helical proteins in solution and the membrane. When considering structurally similar helix-helix interfaces, MPs pack more tightly because of their large number of small amino acids such as Gly, Ala, and Ser^{42,43}. Soluble helix-helix interfaces are often lined by larger hydrophobic residues that cannot make close contacts⁴⁴. In contrast, for whole proteins it was found that MPs pack more loosely than soluble proteins and have more functionally important sites in the cavities^{45–47}. These seemingly contrary conclusions may be explained by (1) differing datasets, i.e. structurally similar helix-helix interfaces vs. entire proteins. (2) The hydrophobic effect for soluble proteins leads to tight packing, while the smaller difference in hydrophobicity between the membrane and the hydrophobic interior of the MP (except for pores) may lead to weaker packing. (3) Further, MPs are often transporters and channels with cavities and pores that are functionally relevant and MPs require some structural flexibility to carry out their functions⁴⁸. To investigate the energetics of packing for soluble and MPs, Bowie and coworkers examined the decrease of packing stability with increasing cavity volume between T4-lysozyme and bacteriorhodopsin and found little difference between the two proteins, suggesting similar energetics⁴⁹. However, generalization of this statement requires investigation on a larger database.

To decipher the influence of membrane mimetics, a larger dataset is needed

Structural differences might also be explained by differences in membrane mimetics (micelles, bicelles, nanodiscs, lipidic cubic phases (LCP)) and lipids/detergents used. While some structures exhibiting marked differences were determined in micelles by both NMR and crystallography (OmpA), other structures exhibiting large differences were determined in different mimetics. For example, the structure of DAGK was determined in detergent micelles by NMR and in lipidic cubic phases by crystallography. Since micelles have increased water penetration in the TM region due to decreased lateral pressure, the structural differences in DAGK likely stem from different membrane mimetic environments²³. As shown above, back-calculated RDCs from the NMR structure match the NMR structure much better than the crystal structure, which points to the micelle environment as the source of differences. Further, it was shown that predicted resonances from the crystal structure matched data from oriented-sample solid-state NMR much better than for the solution NMR

structure²³, indicating that liquid crystalline bilayers used in ssNMR are likely similar to the LCP environment in the crystal. For the M2 channel from the influenza A virus a variety of structures is available that were determined in different environments with various techniques (Xray crystallography, ssNMR, solution NMR) for different functional states^{50,51}. It is therefore difficult to find a clear correlation between environment (pH, membrane mimetic, specific lipids or detergents). A detailed, statistical analysis of membrane mimetics and an investigation of their influence on structural differences is undoubtedly useful, however difficult to carry out on a large scale due to either lack of data or ambiguity in reporting.

Quality of NMR structures can be improved via high-resolution refinement

Since we now have metrics to identify differences between NMR and crystal structures and have identified non-native features in NMR structures (see Supplement), we tested whether we could computationally refine them to improve some of these features. Recently, two independent groups have shown that refinement can yield more biologically relevant models when the NMR structure contains distortions that are not native-like⁵²: either the refined NMR structure⁵³ (MD simulations without NMR restraints) or *ab initio* Rosetta models⁵⁴ of the single span MP KCNE1 (PDBID 2K21⁵⁵) were docked into a potassium channel KCNQ1 model. The generated models could explain activation characteristics that have been validated experimentally.

In another study¹⁷, the NMR structures of soluble proteins were refined in Rosetta with and without NMR restraints^{56,57}. The refined models, especially the ones refined with NMR restraints, had improved stereochemistry and smaller RMSDs compared to the crystal structure. The refined models were also better suited for phasing crystal structures. Here, we abstained from the use of NMR data because: (1) We wanted to remove artifacts that stem from an artificial micelle environment, but instead want to create protein models that are consistent with a planar membrane bilayer. For DAGK for instance, the Q-value between experimental and back-calculated RDCs is much smaller for the NMR structure (Q=0.03) than for the crystal structure (Q=0.55), indicating that the large bend angle might stem from the artificial micelle environment. Refining NMR structures with data from this artificial environment would not produce models compatible with a planar bilayer. For this reason, we also abstained from comparing the refined models with the experimental NMR data. (2) Further, especially helical MPs suffer from low spectral quality (line broadening, etc.) due to the addition and influence of detergent, which makes restraint collection more challenging^{58,59} with conceivably lower-quality restraints and higher error rates; we did not want to negatively bias refinement towards this data; and (3) we wanted to test the simplest protocol.

We refined the top NMR conformer and created 1,000 models (see Methods), the top-scoring model of which was evaluated with the metrics outlined above. Comparisons of the NMR model, the top-scoring refined NMR model, the refined NMR model closest to the crystal structure (by RMSD), and the crystal structure, are shown in Figures 5 and Supplemental Figure 9. Figure 8A compares the refined and unrefined NMR structures to the crystal structures. The refined NMR structures have smaller differences (RMSDs) to the

crystal structure than the unrefined ones, illustrating reduced splaying (Figure 8B). The refined models also have improved stereochemistry as seen from lower MolProbity scores (Figure 8C) and are more tightly packed than the unrefined ones (Figure 8D). These data show that even a simple refinement protocol can nudge NMR structures of MPs toward smaller deviations from the crystal structures and provides a simple means to improve models of MP NMR structures.

Currently, structures added to the PDB include both the experimental NMR data as well as models computed with NMR restraints. In light of our findings here, it may be time to also deposit structures created from the final NMR ensemble but computationally refined with a flat bilayer membrane potential and without NMR data. Such a flat bilayer refined ensemble will have additional deviations from the NMR data when the latter is acquired in membrane mimetics with a different shape. However, a flat bilayer refined ensemble may be more functionally relevant because the membrane model more closely resembles a native cell membrane.

To further improve and facilitate large-scale quality assessment of membrane proteins from structures in the PDB, we suggest a standardization of deposition criteria, especially for NMR structures of membrane proteins. These should (1) include the exact detergent/lipid composition in the final sample, (2) include the type of membrane mimetic (micelle, bicelle, nanodisc, LCP, etc.), which is sometimes implied but not explicitly stated, (3) settle on a number of parallel trajectories to compute (say 50) and the number of models to deposit in the PDB (say 20), and (4) deposit all NMR data including chemical shifts used for the modeled ensemble, since links between PDBIDs and BMRB records may be ambiguous. Establishing these standards would improve data analysis and studies of structural quality, allowing further examination of influences of crystal or solution environment, membrane mimetics and detergent/lipid composition.

Structure determination method influences structural quality

The differences between NMR and crystal structures of MPs may arise from a combination of differences in (1) structure determination methods, (2) environment of a packed crystal versus solution, (3) the hydrophobic environment of the membrane bilayer, (4) the membrane mimetic, and (5) specific lipid/detergent types used⁶⁰. Based on the data presented here and in the literature, we believe that the strongest influence stems from different environments in a crystal vs. in solution, the method of structure determination, and membrane mimetics (see Table 2 for a summary): large bend angles likely arise from the micellar environment in solution while protein stacking in a crystal might circumvent large bend angles (see Supplementary Figure 3) through crystal contacts in loop regions and the TM domain. Most of the proteins in our dataset exhibit some crystal contacts in loop regions (OmpG, PagP), the TM domain (OmpX-Yp, OmpW) or both (BovRh, SensRh, DAGK, DsbB). An extreme example of crystal contacts in our datasets is OmpX-Ec for which the loop regions form extended strands protruding into solution, which hydrogen-bond to the strands of another protein molecule to form a slightly curved sheet. Stereochemistry might be mostly influenced by the static environment in a crystal which could restrict rotamers and local geometry. In addition, for structure calculations using NMR restraints, the restraints

work orthogonal to the energy function in the MD structure calculation software. Careful balancing of these effects is required to avoid overfitting to either one or the other. Considering packing interactions, the crystal environment likely contributes to a higher packing density. However, the micellar environment in solution may play a role due to smaller lateral pressure than is present in bilayers.

Given the structural differences outlined above, a natural question is: Which structure is more native-like, the crystal or the NMR structure? Since measurement of a ‘true’ structure of a MP in a native membrane bilayer remains difficult, we cannot answer that question conclusively. The structural biology community has typically used crystal structures as the standard, which might be because of the larger number of crystal structures (only about 9% of the PDB are NMR structures), the notion that proteins crystallize at an energetically favorable state, and direct detection of the electron densities (in contrast to structure calculations from NMR restraints). Even though crystal structures do not report on dynamics and can have artifacts from crystal contacts, crystal structures are typically used for method development and benchmarking: methods such as MolProbity²⁶, the Dunbrack rotamer library⁶¹, and parts of the high-resolution Rosetta score function⁶² were derived from high-resolution crystal structures. It is therefore somewhat expected that MolProbity reports better stereochemistry for the crystal structures in our dataset and that Rosetta’s high-resolution refinement protocol creates models that are more similar to the crystal structure than to the starting NMR structure. Still, it is encouraging that a simple refinement method combined with a reliable score function can mitigate features in NMR structures that are inherent in the method, the environment, or stem from membrane mimetics. The protocol is applicable to refine NMR structures of MPs, which currently make up about 11% of the MP structures in the PDB.

Many questions remain unanswered: The impact of membrane mimetics remains somewhat elusive; the effect of different types of lipids, detergents and their mixtures on the structures are still unknown, as are the structural changes induced by other environmental variables such as pH, buffer conditions and salt concentration. Other open questions include the influence of the native environment and membrane remodeling such as curvature or lipid composition on MP structure.

Conclusion

In this work, we have examined the differences between MP structures determined by X-ray crystallography and NMR spectroscopy, discussed the origins of these differences and tested an efficient computational refinement method to minimize them.

In addition to convergence in the NMR ensemble, which measures a combination of restraint satisfaction and the number of available restraints, we have provided four orthogonal measures of structural quality: (A) the RMSD between crystal and NMR structure describes the differences between the two, (B) the bend angle measures visible, large-scale deformations, (C) the MolProbity score defines subtler measures such as clashes, dihedral angles and rotamers that aren’t necessarily immediately visible, and (D) the overall relative accessible surface area (rASA) defines packing.

We found that TM regions of NMR ensembles typically differ within 5 Å, which is about the same range as the differences in TM regions between NMR and crystal structures. These ranges are larger than what was found for soluble proteins^{10–12,30}. Crystal structures usually have straighter TM regions and are more tightly packed than NMR structures. Crystal structures also have enhanced stereochemistry, as shown by smaller MolProbity scores. However, while crystal structures have fewer clashes than NMR models, those that exist are typically more severe. We have discussed the origins of these differences considering the different techniques by which the structures were determined, the environment of a packed crystal vs. solution, and have examined the influence of the membrane mimetics onto the structure. With this quantitative description of structural differences, we then used Rosetta high-resolution refinement on the NMR structures to try to minimize them. Our results show that refinement of the NMR structures leads to models that are more similar to the crystal structures in terms of RMSD, straightness in the TM region, stereochemistry, and packing. The protocol we present therefore provides an excellent way to improve structural quality while reducing artificial features from structure determination.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Meera Valliath for work on the database, Dr. Georg Künze for the calculation of the Q-factors for DAGK, and RosettaCommons for helpful suggestions. We also thank the reviewers for their constructive comments and sharing detailed observations about specific structures in our database. Funding was provided by RosettaCommons to JKL and NIH R01 GM-078221 to JJG and JKL. JKL is funded by the Flatiron Institute at the Simons Foundation.

Funding for the study described in this article was provided by Rosetta Commons. Dr. Gray is also an unpaid member of the Executive Board of the Rosetta Commons. Under an institutional participation agreement between the University of Washington, acting on behalf of the Rosetta Commons, and the Johns Hopkins University (JHU), JHU may be entitled to a portion of revenue received on licensing of software described in this article.

Abbreviations

30-Fos	2-undecylphosphocholine
C8E4	tetraethylene glycol monooctyl ether
C8POE	octyl-poly-oxyethylenoxide
CHAPSO	3-([3-Cholamidopropyl]dimethylammonio)-2-hydroxy-1-propanesulfonate
DDM	n-dodecyl β-D-maltoside
DePC	decylphosphocholine
DHPC	dihexanoylphosphatidylcholine
diC7PC	1,2-diheptanoyl- <i>sn</i> -glycero-phosphocholine
DMPC	1,2-dimyristoyl- <i>sn</i> -glycero-phosphocholine

DMSO	dimethyl sulfoxide
DPC	<i>n</i> -dodecyl phosphatidylcholine
LCP	lipidic cubic phase
LDAO	lauryldimethylamine <i>N</i> -oxide
OG	octyl β -D-glucopyranoside
SDS	sodium dodecyl sulfate

References

1. Yildirim MA, Goh KI, Cusick ME, Barabási AL, Vidal M. Drug-target network. *Nat Biotechnol.* 2007; 25:1119–26. [PubMed: 17921997]
2. Van Meer G, Voelker DR, Feigenson GW, van Meer G, Voelker DR, Feigenson GW. Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol.* 2008; 9:112–24. [PubMed: 18216768]
3. van Meer G, de Kroon AIPM. Lipid map of the mammalian cell. *J Cell Sci.* 2011; 124:5–8. [PubMed: 21172818]
4. Holthuis JCM, Menon AK. Lipid landscapes and pipelines in membrane homeostasis. *Nature.* 2014; 510:48–57. [PubMed: 24899304]
5. Lorent JH, Levental I. Structural determinants of protein partitioning into ordered membrane domains and lipid rafts. *Chem Phys Lipids.* 2015; 192:23–32. [PubMed: 26241883]
6. Bolla JR, Su CC, Yu EW. Biomolecular membrane protein crystallization. *Philos Mag.* 2012; 92:2648–2661.
7. Wiener M. A pedestrian guide to membrane protein crystallization. *Methods.* 2004; 34:364–372. [PubMed: 15325654]
8. Johansson LC, Wöhri AB, Katona G, Engström S, Neutze R. Membrane protein crystallization from lipidic phases. *Curr Opin Struct Biol.* 2009; 19:372–8. [PubMed: 19581080]
9. Lange OF, Rossi P, Sgourakis NG, Song Y, Lee HW, Aramini JM, Ertekin A, Xiao R, Acton TB, Montelione GT, Baker D. Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples. *Proc Natl Acad Sci U S A.* 2012; 109:10873–8. [PubMed: 22733734]
10. Garbuzynskiy SO, Melnik BS, Lobanov MY, Finkelstein AV, Galzitskaya OV. Comparison of X-ray and NMR structures: is there a systematic difference in residue contacts between X-ray- and NMR-resolved protein structures? *Proteins.* 2005; 60:139–47. [PubMed: 15856480]
11. Andrec M, Snyder DA, Zhou Z, Young J, Montelione GT, Levy RM. A large data set comparison of protein structures determined by crystallography and NMR: statistical test for structural differences and the effect of crystal packing. *Proteins.* 2007; 69:449–65. [PubMed: 17623851]
12. Sikic K, Tomic S, Carugo O. Systematic comparison of crystal and NMR protein structures deposited in the protein data bank. *Open Biochem J.* 2010; 4:83–95. [PubMed: 21293729]
13. Zhou HX, Cross TA. Influences of membrane mimetic environments on membrane protein structures. *Annu Rev Biophys.* 2013; 42:361–92. [PubMed: 23451886]
14. Zhou HX, Cross TA. Modeling the membrane environment has implications for membrane protein structure and function: influenza A M2 protein. *Protein Sci.* 2013; 22:381–94. [PubMed: 23389890]
15. Cheng X, Im W. NMR observable-based structure refinement of DAP12-NKG2C activating immunoreceptor complex in explicit membranes. *Biophys J.* 2012; 102:L27–9. [PubMed: 22500771]
16. Sharma M, Yi M, Dong H, Qin H, Peterson E, Busath DD, Zhou HX, Cross TA. Insight into the mechanism of the influenza A proton channel from a structure in a lipid bilayer. *Science.* 2010; 330:509–12. [PubMed: 20966252]

17. Mao B, Tejero R, Baker D, Montelione GT. Protein NMR Structures Refined with Rosetta Have Higher Accuracy Relative to Corresponding X-ray Crystal Structures. *J Am Chem Soc.* 2014; 136:1893–1906. [PubMed: 24392845]
18. Sharma M, Yi M, Dong H, Qin H, Peterson E, Busath DD, Zhou HX, Cross TA. Insight into the mechanism of the influenza A proton channel from a structure in a lipid bilayer. *Science.* 2010; 330:509–12. [PubMed: 20966252]
19. Cady SD, Schmidt-Rohr K, Wang J, Soto CS, Degrado WF, Hong M. Structure of the amantadine binding site of influenza M2 proton channels in lipid bilayers. *Nature.* 2010; 463:689–92. [PubMed: 20130653]
20. Tang M, Nesbitt AE, Sperling LJ, Berthold DA, Schwieters CD, Gennis RB, Rienstra CM. Structure of the disulfide bond generating membrane protein DsbB in the lipid bilayer. *J Mol Biol.* 2013; 425:1670–1682. [PubMed: 23416557]
21. Wang S, Munro RA, Shi L, Kawamura I, Okitsu T, Wada A, Kim SY, Jung KH, Brown LS, Ladizhansky V. Solid-state NMR spectroscopy structure determination of a lipid-embedded heptahelical membrane protein. *Nat Methods.* 2013; 10:1007–12. [PubMed: 24013819]
22. Park SH, Das BB, Casagrande F, Tian Y, Nothnagel HJ, Chu M, Kiefer H, Maier K, De Angelis AA, Marassi FM, Opella SJ. Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature.* 2012; 491:779–83. [PubMed: 23086146]
23. Murray DT, Li C, Gao FP, Qin H, Cross TA. Membrane protein structural validation by oriented sample solid-state NMR: diacylglycerol kinase. *Biophys J.* 2014; 106:1559–69. [PubMed: 24739155]
24. Alford RF, Koehler Leman J, Weitzner BD, Duran AM, Tilley DC, Elazar A, Gray JJ. An Integrated Framework Advancing Membrane Protein Modeling and Design. *PLoS Comput Biol.* 2015; 11:e1004398. [PubMed: 26325167]
25. Software: The PyMOL Molecular Graphics System, Version 1>8. Schroedinger LLC; at <<https://www.schrodinger.com/suites/pymol>>
26. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr.* 2010; 66:12–21. [PubMed: 20057044]
27. Hubbard SJ, Thornton JM. Naccess, Computer Program. 1993
28. Rosato A, Tejero R, Montelione GT, Chiu W, Wagner G. Quality assessment of protein NMR structures. *Curr Opin Struct Biol.* 2013; 23:715–724. [PubMed: 24060334]
29. Montelione GT, Nilges M, Bax A, Güntert P, Herrmann T, Richardson JS, Schwieters CD, Vranken WF, Vuister GW, Wishart DS, Berman HM, Kleywegt GJ, Markley JL. Recommendations of the wwPDB NMR Validation Task Force. *Structure.* 2013; 21:1563–1570. [PubMed: 24010715]
30. Everett JK, Tejero R, Murthy SBK, Acton TB, Aramini JM, Baran MC, Benach J, Cort JR, Eletsky A, Forouhar F, Guan R, Kuzin AP, Lee HW, Liu G, Mani R, Mao B, Mills JL, Montelione AF, Pederson K, Powers R, Ramelot T, Rossi P, Seetharaman J, Snyder D, Swapna GVT, Vorobiev SM, Wu Y, Xiao R, Yang Y, Arrowsmith CH, Hunt JF, Kennedy MA, Prestegard JH, Szyperski T, Tong L, Montelione GT. A community resource of experimental data for NMR/X-ray crystal structure pairs. *Protein Sci.* 2016; 25:30–45. [PubMed: 26293815]
31. Braun T, Leman JK, Lange OF. Combining Evolutionary Information and an Iterative Sampling Strategy for Accurate Protein Structure Prediction. *PLoS Comput Biol.* 2015; 11:1–20.
32. Martin JW, Zhou P, Donald BR. Systematic solution to homo-oligomeric structures determined by NMR. *Proteins Struct Funct Bioinforma.* 2015; 83:651–661.
33. Orozco M. A theoretical view of protein dynamics. *Chem Soc Rev.* 2014; 43:5051–5066. [PubMed: 24709805]
34. Kovermann M, Rogne P, Wolf-Watz M. Protein dynamics and function from solution state NMR spectroscopy. *Q Rev Biophys.* 2016; 49:e6. [PubMed: 27088887]
35. Narayanan C, Bafna K, Roux LD, Agarwal PK, Doucet N. Applications of NMR and computational methodologies to study protein dynamics. *Arch Biochem Biophys.* 2017; doi: 10.1016/j.abb.2017.05.002
36. Kay LE. New Views of Functionally Dynamic Proteins by Solution NMR Spectroscopy. *J Mol Biol.* 2016; 428:323–331. [PubMed: 26707200]

37. Chen, K. & Tjandra, N. in 47–67 (Springer Berlin Heidelberg, 2011). doi:10.1007/128_2011_215
38. Van Horn WD, Kim HJ, Ellis CD, Hadziselimovic A, Sulistijo ES, Karra MD, Tian C, Sönnichsen FD, Sanders CR. Solution nuclear magnetic resonance structure of membrane-integral diacylglycerol kinase. *Science*. 2009; 324:1726–9. [PubMed: 19556511]
39. Li D, Lyons JA, Pye VE, Vogeley L, Aragão D, Kenyon CP, Shah STA, Doherty C, Aherne M, Caffrey M. Crystal structure of the integral membrane diacylglycerol kinase. *Nature*. 2013; 497:521–4. [PubMed: 23676677]
40. Voss NR, Gerstein M. 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res*. 2010; 38:W555–W562. [PubMed: 20478824]
41. Sheffler W, Baker D. RosettaHoles: rapid assessment of protein core packing for structure prediction, refinement, design, and validation. *Protein Sci*. 2009; 18:229–39. [PubMed: 19177366]
42. Eilers M, Patel AB, Liu W, Smith SO. Comparison of helix interactions in membrane and soluble alpha-bundle proteins. *Biophys J*. 2002; 82:2720–36. [PubMed: 11964258]
43. Gimpelev M, Forrest LR, Murray D, Honig B. Helical packing patterns in membrane and soluble proteins. *Biophys J*. 2004; 87:4075–86. [PubMed: 15465852]
44. Zhang SQQ, Kulp DWW, Schramm CAA, Mravic M, Samish I, DeGrado WFF. The Membrane- and Soluble-Protein Helix-Helix Interactome: Similar Geometry via Different Interactions. *Structure*. 2015; 23:1–15. [PubMed: 25565099]
45. Hildebrand PW, Rother K, Goede A, Preissner R, Frömmel C. Molecular packing and packing defects in helical membrane proteins. *Biophys J*. 2005; 88:1970–7. [PubMed: 15556989]
46. Pabuwal V, Li Z. Comparative analysis of the packing topology of structurally important residues in helical membrane and soluble proteins. *Protein Eng Des Sel*. 2009; 22:67–73. [PubMed: 19054790]
47. Rose A, Theune D, Goede A, Hildebrand PW. MP:PD—a data base of internal packing densities, internal packing defects and internal waters of helical membrane proteins. *Nucleic Acids Res*. 2014; 42:D347–51. [PubMed: 24194596]
48. Cournia Z, Allen TW, Andricioaei I, Antonny B, Baum D, Brannigan G, Buchete NV, Deckman JT, Delemotte L, del Val C, Friedman R, Gkeka P, Hege H-C, Hénin J, Kasimova MA, Kolocouris A, Klein ML, Khalid S, Lemieux J, Lindow N, Mahua Roy Selent J, Tarek M, Tofoleanu F, Vanni S, Urban S, Wales DJ, Smith JC, Bondar A-N. Membrane Protein Structure, Function and Dynamics: A Perspective from Experiments and Theory. *J Membr Biol*. 2015; :611–640. DOI: 10.1007/s00232-015-9802-0 [PubMed: 26063070]
49. Joh NH, Oberai A, Yang D, Whitelegge JP, Bowie JU. Similar energetic contributions of packing in the core of membrane and water-soluble proteins. *J Am Chem Soc*. 2009; 131:10846–7. [PubMed: 19603754]
50. Acharya R, Carnevale V, Fiorin G, Levine BG, Polishchuk AL, Balannik V, Samish I, Lamb Ra, Pinto LH, DeGrado WF, Klein ML. Structure and mechanism of proton transport through the transmembrane tetrameric M2 protein bundle of the influenza A virus. *Proc Natl Acad Sci U S A*. 2010; 107:15075–80. [PubMed: 20689043]
51. Cross TA, Sharma M, Yi M, Zhou HXX. Influence of solubilizing environments on membrane protein structures. *Trends Biochem Sci*. 2011; 36:117–25. [PubMed: 20724162]
52. Coey AT, Sahu ID, Gunasekera TS, Troxel KR, Hawn JM, Swartz MS, Wickenheiser MR, Reid RJ, Welch RC, Vanoye CG, Kang C, Sanders CR, Lorigan GA. Reconstitution of KCNE1 into lipid bilayers: Comparing the structural, dynamic, and activity differences in micelle and vesicle environments. *Biochemistry*. 2011; 50:10851–10859. [PubMed: 22085289]
53. Xu Y, Wang Y, Meng XY, Zhang M, Jiang M, Cui M, Tseng GN. Building KCNQ1/KCNE1 channel models and probing their interactions by molecular-dynamics simulations. *Biophys J*. 2013; 105:2461–73. [PubMed: 24314077]
54. Gofman Y, Shats S, Attali B, Haliloglu T, Ben-Tal N. How Does KCNE1 Regulate the Kv7.1 Potassium Channel? Model-Structure, Mutations, and Dynamics of the Kv7.1-KCNE1 Complex. *Structure*. 2012; 20:1343–1352. [PubMed: 22771213]
55. Kang C, Tian C, Soennichsen FD, Smith JA, Meiler J, George AL, Vanoye CG, Hak JK, Sanders CR. Structure of KCNE1 and implications for how it modulates the KCNQ1 potassium channel. *Biochemistry*. 2008; 47:7999–8006. [PubMed: 18611041]

56. Khatib F, Cooper S, Tyka MD, Xu K, Makedon I, Popovic Z, Baker D, Players F. Algorithm discovery by protein folding game players. *Proc Natl Acad Sci U S A*. 2011; 108:18949–53. [PubMed: 22065763]
57. Tyka MD, Keedy DA, André I, Dimaio F, Song Y, Richardson DC, Richardson JS, Baker D. Alternate states of proteins revealed by detailed energy landscape mapping. *J Mol Biol*. 2011; 405:607–18. [PubMed: 21073878]
58. Sanders CR, Sönnichsen F. Solution NMR of membrane proteins: Practice and challenges. *Magn Reson Chem*. 2006; 44:24–40.
59. Kim HJ, Howell SC, Van Horn WD, Jeon YH, Sanders CR. Recent advances in the application of solution NMR spectroscopy to multi-span integral membrane proteins. *Prog Nucl Magn Reson Spectrosc*. 2009; 55:335–360. [PubMed: 20161395]
60. Laganowsky A, Reading E, Allison TM, Ulmschneider MB, Degiacomi MT, Baldwin AJ, Robinson CV. Membrane proteins bind lipids selectively to modulate their structure and function. *Nature*. 2014; 510:172–5. [PubMed: 24899312]
61. Shapovalov MV, Dunbrack RL. Statistical and conformational analysis of the electron density of protein side chains. *Proteins*. 2007; 66:279–303. [PubMed: 17080462]
62. Leaver-Fay A, O'Meara MJ, Tyka M, Jacak R, Song Y, Kellogg EH, Thompson J, Davis IW, Pache RA, Lyskov S, Gray JJ, Kortemme T, Richardson JS, Havranek JJ, Snoeyink J, Baker D, Kuhlman B. Scientific benchmarks for guiding macromolecular energy function improvement. *Methods Enzymol*. 2013; 523:109–43. [PubMed: 23422428]
63. Pielak RM, Schnell JR, Chou JJ. Mechanism of drug inhibition and drug resistance of influenza A M2 channel. *Proc Natl Acad Sci U S A*. 2009; 106:7379–84. [PubMed: 19383794]
64. Stouffer AL, Acharya R, Salom D, Levine AS, Di Costanzo L, Soto CS, Tereshko V, Nanda V, Stayrook S, DeGrado WF. Structural basis for the function and inhibition of an influenza virus proton channel. *Nature*. 2008; 451:596–9. [PubMed: 18235504]
65. Zhou Y, Cierpicki T, Jimenez RHF, Lukasik SM, Ellena JF, Cafiso DS, Kadokura H, Beckwith J, Bushweller JH. NMR solution structure of the integral membrane enzyme DsbB: functional insights into DsbB-catalyzed disulfide bond formation. *Mol Cell*. 2008; 31:896–908. [PubMed: 18922471]
66. Inaba K, Murakami S, Nakagawa A, Iida H, Kinjo M, Ito K, Suzuki M. Dynamic nature of disulphide bond formation catalysts revealed by crystal structures of DsbB. *EMBO J*. 2009; 28:779–791. [PubMed: 19214188]
67. Patzelt H, Simon B, terLaak A, Kessler B, Kühne R, Schmieder P, Oesterhelt D, Oschkinat H. The structures of the active center in dark-adapted bacteriorhodopsin by solution-state NMR spectroscopy. *Proc Natl Acad Sci U S A*. 2002; 99:9765–70. [PubMed: 12119389]
68. Schobert B, Cupp-Vickery J, Hornak V, Smith SO, Lanyi JK. Crystallographic Structure of the K Intermediate of Bacteriorhodopsin: Conservation of Free Energy after Photoisomerization of the Retinal. *J Mol Biol*. 2002; 321:715–726. [PubMed: 12206785]
69. Gautier A, Mott HR, Bostock MJ, Kirkpatrick JP, Nietlispach D. Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy. *Nat Struct Mol Biol*. 2010; 17:768–74. [PubMed: 20512150]
70. Gordeliy VI, Labahn J, Moukhametzianov R, Efremov R, Granzin J, Schlesinger R, Büldt G, Savopol T, Scheidig AJ, Klare JP, Engelhard M. Molecular basis of transmembrane signalling by sensory rhodopsin II–transducer complex. *Nature*. 2002; 419:484–487. [PubMed: 12368857]
71. Reckel S, Gottstein D, Stehle J, Löhr F, Verhoeven MK, Takeda M, Silvers R, Kainosho M, Glaubitz C, Wachtveitl J, Bernhard F. Solution NMR Structure of Proteorhodopsin. *Angew Chem Int Ed Engl*. 2011; doi: 10.1002/anie.201105648
72. Ran T, Ozorowski G, Gao Y, Sineshchekov OA, Wang W, Spudich JL, Luecke H. Cross-protomer interaction with the photoactive site in oligomeric proteorhodopsin complexes. *Acta Crystallogr D Biol Crystallogr*. 2013; 69:1965–80. [PubMed: 24100316]
73. Choi G, Landin J, Galan JF, Birge RR, Albert AD, Yeagle PL. Structural studies of metarhodopsin II, the activated form of the G-protein coupled receptor, rhodopsin. *Biochemistry*. 2002; doi: 10.1021/bi025507w

74. Okada T, Sugihara M, Bondar AN, Elstner M, Entel P, Buss V. The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *J Mol Biol.* 2004; doi: 10.1016/j.jmb.2004.07.044
75. Li D, Stansfeld PJ, Sansom MSP, Keogh A, Voageley L, Howe N, Lyons JA, Aragao D, Fromme P, Fromme R, Basu S, Grotjohann I, Kupitz C, Rendek K, Weierstall U, Zatsepin NA, Cherezov V, Liu W, Bandaru S, English NJ, Gati C, Barty A, Yefanov O, Chapman HN, Diederichs K, Messerschmidt M, Boutet S, Williams GJ, Marvin Seibert M, Caffrey M. Ternary structure reveals mechanism of a membrane diacylglycerol kinase. *Nat Commun.* 2015; 6:10140. [PubMed: 26673816]
76. Hagn F, Wagner G. Structure refinement and membrane positioning of selectively labeled OmpX in phospholipid nanodiscs. *J Biomol NMR.* 2015; 61:249–60. [PubMed: 25430058]
77. Vogt J, Schulz GE. The structure of the outer membrane protein OmpX from *Escherichia coli* reveals possible mechanisms of virulence. *Structure.* 1999; 7:1301–9. [PubMed: 10545325]
78. Marassi FM, Ding Y, Schwieters CD, Tian Y, Yao Y. Backbone structure of *Yersinia pestis* Ail determined in micelles by NMR-restrained simulated annealing with implicit membrane solvation. *J Biomol NMR.* 2015; 63:59–65. [PubMed: 26143069]
79. Yamashita S, Lukacik P, Barnard TJ, Noinaj N, Felek S, Tsang TM, Krukonis ES, Hinnebusch BJ, Buchanan SK. Structural insights into Ail-mediated adhesion in *Yersinia pestis*. *Structure.* 2011; 19:1672–82. [PubMed: 22078566]
80. Hwang PM, Choy WY, Lo EI, Chen L, Forman-Kay JD, Raetz CRH, Privé GG, Bishop RE, Kay LE. Solution structure and dynamics of the outer membrane enzyme PagP by NMR. *Proc Natl Acad Sci U S A.* 2002; 99:13560–5. [PubMed: 12357033]
81. Cuesta-Seijo JA, Neale C, Khan MA, Moktar J, Tran CD, Bishop RE, Pomès R, Privé GG. PagP Crystallized from SDS/Cosolvent Reveals the Route for Phospholipid Access to the Hydrocarbon Ruler. *Structure.* 2010; 18:1210–1219. [PubMed: 20826347]
82. Cierpicki T, Liang B, Tamm LK, Bushweller JH. Increasing the accuracy of solution NMR structures of membrane proteins by application of residual dipolar couplings. High-resolution structure of outer membrane protein A. *J Am Chem Soc.* 2006; doi: 10.1021/ja0608343
83. Pautsch A, Schulz GE. High-resolution Structure of the OmpA Membrane Domain. *J Mol Biol.* 2000; 298:273–282. [PubMed: 10764596]
84. Horst R, Stanczak P, Wüthrich K. NMR polypeptide backbone conformation of the *E. coli* outer membrane protein W. *Structure.* 2014; 22:1204–9. [PubMed: 25017731]
85. Hong H, Patel DR, Tamm LK, van den Berg B. The outer membrane protein OmpW forms an eight-stranded beta-barrel with a hydrophobic channel. *J Biol Chem.* 2006; 281:7568–77. [PubMed: 16414958]
86. Liang B, Tamm LK. Structure of outer membrane protein G by solution NMR spectroscopy. *Proc Natl Acad Sci U S A.* 2007; 104:16140–5. [PubMed: 17911261]
87. Korkmaz-Özkan F, Köster S, Kühlbrandt W, Mäntele W, Yildiz Ö. Correlation between the OmpG Secondary Structure and Its pH-Dependent Alterations Monitored by FTIR. *J Mol Biol.* 2010; doi: 10.1016/j.jmb.2010.06.015
88. Hiller S, Garces RG, Malia TJ, Orekhov VY, Colombini M, Wagner G. Solution structure of the integral human membrane protein VDAC-1 in detergent micelles. *Science (80-).* 2008; 321:1206–1210.
89. Ujwal R, Cascio D, Colletier JP, Faham S, Zhang J, Toro L, Ping P, Abramson J. The crystal structure of mouse VDAC1 at 2.3 Å resolution reveals mechanistic insights into metabolite gating. *Proc Natl Acad Sci U S A.* 2008; 105:17742–7. [PubMed: 18988731]

Figure 1. Membrane protein pairs determined by both NMR and Xray crystallography
 Proteins are ordered by size, in α -helical and β -barrel categories. The columns from left to right show the NMR ensemble, the top NMR model, the overlay between NMR and crystal structure (red – NMR; gray - Xray), and the crystal structure. Black arrows in DAGK, OmpX, OmpA, and OmpG identify bends in the structures, mostly through loop regions that wrap around the micelle in which the structure was determined.

Figure 2. Membrane protein pairs determined by both NMR and X-ray crystallography
 Proteins are ordered by size, in α -helical and β -barrel categories. The columns from left to right show the NMR ensemble, the top NMR model, the overlay between NMR and crystal structure (red – NMR; gray – X-ray), and the crystal structure. Black arrows in DAGK, OmpX, OmpA, and OmpG identify bends in the structures, mostly through loop regions that wrap around the micelle in which the structure was determined.

Figure 3. Membrane protein pairs determined by both NMR and X-ray crystallography
 Proteins are ordered by size, in α -helical and β -barrel categories. The columns from left to right show the NMR ensemble, the top NMR model, the overlay between NMR and crystal structure (red – NMR; gray – X-ray), and the crystal structure. Black arrows in DAGK, OmpX, OmpA, and OmpG identify bends in the structures, mostly through loop regions that wrap around the micelle in which the structure was determined.

Figure 4. Backbone RMSDs within the NMR ensemble are an indicator of convergence
 Ranges of backbone RMSDs between the top NMR model and all other models in the ensemble. RMSDs are computed over all residues (black), and over the TM region (gray) for the α -helical and β -barrel proteins ordered with increasing size and separated by a vertical line. Long, ‘flexible’ C-termini of DsbB (19 residues) and sensory rhodopsin (21 residues) were removed for these calculations. The average backbone RMSD is 4.0 Å over all residues and 1.0 Å for the TM region. The PDB file for bovine rhodopsin only contains a single model and therefore an RMSD cannot be calculated (* asterisk in Figure).

structures are shown in black, for the NMR structures in gray. The hollow triangles deviate from the general trend that crystal structures typically have smaller bend angles, i.e. a straighter TM region, than the NMR structures. (C) Crystal structures typically have better stereochemistry. We use MolProbity scores as a measure of stereochemistry for the crystal structures (black) and the NMR structures (gray) – lower scores indicate better stereochemistry. The overall MolProbity score combines all individual scores (see Supplemental Table 2) and is a measure of structural quality. The score itself is normalized such that it provides the resolution of the crystal structure at which those values would be expected²⁶. Hence, if the MolProbity score is lower than the crystallographic resolution, the quality of the structure is better than the average structure at this resolution. Hollow triangles deviate from the general trend that crystal structures have lower (i.e. better) scores than NMR structures. (D) Crystal structures are typically more tightly packed than NMR structures. We used Naccess to compute the per-residue accessible surface area (ASA) relative to an extended chain, as a measure of packing density. 100% means completely exposed and 0% means completely buried. The per-residue rASA were averaged over the entire protein. Crystal structures (black) and NMR structures (gray) with hollow triangles deviate from the general trend that crystal structures are typically more tightly packed with a smaller average rASA than NMR structures.

Figure 6. Both NMR and crystal structures contain clashes

Per-residue Rosetta full-atom repulsive scores as an indicator of clashes are mapped onto the protein structures of two examples; the entire dataset is shown in Supplemental Figure 5. The upper panel shows DsbB, the lower panel *Yersinia pestis* OmpX, with the NMR structure on the left and the crystal structure on the right. Residues in gray are clash-free, with increasing clash severity from yellow to orange to red (full-atom repulsive scores <1 REU in gray, 1-10 REU in yellow, 10-100 REU in orange, and >100 REU in red). The

results show that NMR structures typically have more clashes, however they are less severe. Crystal structures have fewer clashes, but they are of higher severity.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 7. Effect of high-resolution refinement in a membrane bilayer onto structural quality

The top NMR models were subjected to Rosetta high-resolution refinement with a membrane score function to minimize discrepancies between the NMR and crystal structures. Structural features arising from NMR structure determination (see Results and Discussion) and/or membrane mimetics could be reduced: Refinement straightens out the helices in DAGK and the ‘flexible’ loop in *Yersinia pestis* OmpX that wraps around the imaginary micelle in which the structure was determined.

Figure 8. High-resolution refinement improves the quality of NMR structures

Correlation plots of structural measures before and after refinement of the NMR structure, using the top-scoring model, for the 14 membrane proteins in the dataset. Gray shaded areas denote improvements toward the crystal structure. (A) Backbone RMSD of full-length structures compared to the crystal structure. Refined NMR structures are more similar to crystal structures than unrefined ones. The outlier is bovine rhodopsin, which unfolds during the simulation due to a low-quality starting structure (Figure 2 A.6). (B) In most cases, bend angles decrease, straightening the TM region in the membrane bilayer. (C) Overall MolProbity score decreases (improves) due to an improvement in clashes, Ramachandran dihedrals and rotamers. The outlier belongs to *E. coli* OmpX, whose clash score and hence MolProbity score worsens during refinement. (D) Relative ASA averaged over all residues indicating tighter packing of refined models. The outlier is again bovine rhodopsin, which unfolds during the simulation due to a low-quality starting structure.

Table 1

Membrane protein pairs that have both a crystal and NMR structure.

NMR structure										Crystal structure							#mut	
abbrev	protein	α/β	TMs	PDBID	year	chain	mod	nres	organism	membrane mimetic	PDBID	year	chain	reso	nres	organism		membrane mimetic
M2	influenza A virus M2 proton channel	α	4	2KIH ⁶³	2009	ABCD	10	152	<i>Influenza A</i>	DHPC micelles?	3BKD ⁶⁴	2008	ABCD	2.1	100	<i>Influenza A</i>	OG micelles	3
DsbB	disulfide bond protein B	α	4	2K73 ⁶⁵	2008	A	20	183	<i>E coli</i>	DPC micelles	2ZUQ ⁶⁶	2009	A	3.3	148	<i>E coli</i>	DHPC micelles	4
BactRh	bacteriorhodopsin	α	7	1R84 ⁶⁷	2002	A	12	232	<i>Halobacterium salinarum</i>	DDM micelles	1MOL ⁶⁸	2002	A	1.5	222	<i>Halobacterium salinarum</i>	LCP with OG	0
SensRh	sensory rhodopsin	α	7	2KSY ⁶⁹	2010	A	30	247	<i>Natronomonas pharaonis</i>	DHPC micelles	1H2S ⁷⁰	2002	A	1.9	225	<i>Natronomonas pharaonis</i>	LCP with OG	4
ProtRh	proteorhodopsin	α	7	2L6X ⁷¹	2011	A	20	235	<i>Proteobacterium EBAC31A08</i>	diC7PC micelles	4KNF ⁷²	2013	A	2.6	231	<i>Proteobacterium HOT5BPR</i>	DMPC/CHAPSO bicelles	51
BovRh	bovine rhodopsin	α	7	1LN6 ⁷³	2002	A	1	309	<i>Bos taurus</i>	none (DMSO solvent)	1UI9 ⁷⁴	2004	A	2.2	348	<i>Bos taurus</i>	heptyl thioglucoside micelles	0
DAGK	diacylglycerol kinase	α	9	2KDC ³⁸	2009	ABC	16	363	<i>E coli</i>	DPC micelles	4UXX ⁷⁵	2015	ABC	2.7	336	<i>E coli</i>	LCP with DM	4
OmpX-EC	Outer Membrane Protein X	β	8	2M07 ⁷⁶	2013	A	20	148	<i>E coli</i>	DPC micelles	1QJ8 ⁷⁷	1999	A	1.9	148	<i>E coli</i>	C8POE/C8E4 micelles?	1
OmpX-Yp	Outer Membrane Protein X	β	8	2N2M ⁷⁸	2015	A	10	156	<i>Yersinia pestis</i>	DePC micelles	3QRC ⁷⁹	2011	A	1.9	153	<i>Yersinia pestis</i>	C8E4 micelles	0
PagP	Outer Membrane Enzyme PagP	β	8	1MM5 ⁸⁰	2002	A	20	170	<i>E coli</i>	OG micelles	3GP6 ⁸¹	2010	A	1.4	155	<i>E coli</i>	SDS micelles	0
OmpA	Outer Membrane Protein A	β	8	2GE4 ⁸²	2006	A	10	177	<i>E coli</i>	DPC micelles	1QJP ⁸³	2000	A	1.7	137	<i>E coli</i>	C8E4 micelles	6
OmpW	Outer Membrane Protein W	β	8	2MHL ⁸⁴	2014	A	20	192	<i>E coli</i>	30-Fos micelles	2F1V ⁸⁵	2006	A	2.7	182	<i>E coli</i>	C8E4 micelles	0
OmpG	Outer Membrane Protein G	β	14	2IQY ⁸⁶	2007	A	10	280	<i>E coli</i>	DPC micelles	2X9K ⁸⁷	2010	A	2.2	278	<i>E coli</i>	OG micelles	2
VDAC	Voltage-dependent anion channel	β	19	2K4T ⁸⁸	2008	A	20	285	<i>Homo sapiens</i>	LDAO micelles	3EMN ⁸⁹	2008	X	2.3	283	<i>Mus musculus</i>	DMPC/CHAPSO bicelles	4

α/β – α -helical or β -barrel

TMs – number of transmembrane spans

year – year in which structure was determined

mod – number of models in the NMR ensemble

nres – number of residues over all relevant chains

organism – native organism – organism not expressing protein; organism in bold differ between NMR and crystal structure
mut – number of mutations between NMR and crystal structure
Å reso – crystallographic resolution in

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Summary of possible influence of the different effects on the MP structure. Effects that are likely larger have more asterisks.

	<i>bend angle: large distortions</i>	<i>stereochemistry: local geometry</i>	<i>rASA: packing</i>
<i>method: Xray vs. NMR (direct vs. indirect detection)</i>	* bend angles could stem from the fact that loop regions in NMR are less restrained (and we don't know where they really are)	** NMR restraints during refinement can work against rotamer and Ramachandran dihedrals, so those are worse for NMR, but bond lengths and angles are better in NMR structure because they are refined through the energy function in MD package	less long-range information (across β -barrels) in NMR; error bars in restraints might lead to lower packing density
<i>environment: crystal vs. solution</i>	** crystal packing likely restricts loop conformations	*** crystal environment could lock down rotamers and local geometry	*** protein crystallizes in a low-energy conformation with higher packing density
<i>environment: hydrophobic membrane bilayer</i>	unclear: solution is isotropic, whereas the membrane is anisotropic	unclear: membrane bilayer could affect observed rotamers	unclear: helix-helix interactions are closer in MPs but entire MPs are less dense than soluble proteins
<i>environment: membrane mimetic: micelle vs. bicelle vs. LCP etc</i>	*** loop regions can wrap around soluble micelle, this is energetically unfavorable in a bilayer; bending in DAGK comes from soluble micelle: RDC restraints in DAGK match the NMR structure, but not the Xray structure (not a computational problem)	unclear	** micelles likely have a smaller lateral pressure than bilayers
<i>environment: specific lipids/detergents</i>	unclear: chain length could influence bending to minimize hydrophobic mismatch	unclear	unclear