

# Requirements on Paramagnetic Relaxation Enhancement Data for Membrane Protein Structure Determination by NMR

Daniel Gottstein,<sup>1</sup> Sina Reckel,<sup>1</sup> Volker Dötsch,<sup>1</sup> and Peter Güntert<sup>1,2,3,\*</sup>

<sup>1</sup>Institute of Biophysical Chemistry and Center for Biomolecular Magnetic Resonance

<sup>2</sup>Frankfurt Institute of Advanced Studies

Goethe University Frankfurt, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany

<sup>3</sup>Graduate School of Science and Engineering, Tokyo Metropolitan University, 1-1 Minami-ohsawa, Hachioji, Tokyo 192-0397, Japan

\*Correspondence: [guentert@em.uni-frankfurt.de](mailto:guentert@em.uni-frankfurt.de)

DOI 10.1016/j.str.2012.03.010

## SUMMARY

Nuclear magnetic resonance (NMR) structure calculations of the  $\alpha$ -helical integral membrane proteins DsbB, GlpG, and halorhodopsin show that distance restraints from paramagnetic relaxation enhancement (PRE) can provide sufficient structural information to determine their structure with an accuracy of about 1.5 Å in the absence of other long-range conformational restraints. Our systematic study with simulated NMR data shows that about one spin label per transmembrane helix is necessary for obtaining enough PRE distance restraints to exclude wrong topologies, such as pseudo mirror images, if only limited other NMR restraints are available. Consequently, an experimentally realistic amount of PRE data enables  $\alpha$ -helical membrane protein structure determinations that would not be feasible with the very limited amount of conventional NOESY data normally available for these systems. These findings are in line with our recent first de novo NMR structure determination of a heptahelical integral membrane protein, proteorhodopsin, that relied extensively on PRE data.

## INTRODUCTION

Integral membrane proteins constitute around one-third of all proteins encoded by the genomes of prokaryotic and eukaryotic cells, playing important roles in transport and signal transduction. However, structural studies of membrane proteins are still difficult because of their hydrophobic nature. In the case of soluble proteins the structure determination by solution nuclear magnetic resonance (NMR) relies on the almost complete assignment of the backbone and side-chain <sup>1</sup>H, <sup>13</sup>C, and <sup>15</sup>N resonances and on a dense network of nuclear Overhauser effect (NOE) based distance measurements needed to calculate the three-dimensional structure. Recently, membrane protein structures in detergent micelles have also been solved on the basis of NOE restraints by solution NMR, such as the heptahel-

ical membrane protein sensory rhodopsin II (Gautier et al., 2010) and several 16–32 kDa  $\beta$ -barrel membrane proteins (Fernández et al., 2004; Hiller et al., 2008; Liang and Tamm, 2007; Renault et al., 2009). However, the large effective size of the proteomicelle leads to broad resonance lines and spectral overlap, which often makes a purely NOE-based approach impractical for integral membrane proteins and calls for other sources of conformational data. This is particularly true for  $\alpha$ -helical membrane proteins, where the amide moieties in the backbone often show only short-range intrahelical NOEs and deuteration eliminates most side-chain information required to derive long-range NOEs. In this case, back-protonation of methyl groups has been employed, which is, however, less effective than for soluble proteins because many of the hydrophobic side chains face outward. As a result, for hardly any  $\alpha$ -helical integral membrane protein more than 100 long-range NOEs could be collected, resulting in structures of medium quality. Paramagnetic relaxation enhancement (PRE) induced by paramagnetic spin labels (Gaponenko et al., 2000; Gillespie and Shortle, 1997; Kosen, 1989; Liang et al., 2006) has long been recognized as an approach for obtaining long-range conformational restraints that can complement NOEs, which are limited to distances of up to 5 Å. The spin labels produce distance-dependent line broadening in the NMR spectra that can be translated into distance restraints (Battiste and Wagner, 2000; Iwahara et al., 2007). This method has been applied successfully to several  $\alpha$ -helical membrane proteins, for which between 20 and 1,144 PRE restraints were collected (Table 1) (Berardi et al., 2011; Kang et al., 2008; Maslennikov et al., 2010; Page et al., 2009; Reckel et al., 2011; Roosild et al., 2005; Sobhanifar et al., 2010; Teriete et al., 2007; Van Horn et al., 2009; Zhou et al., 2008). Compared with NOEs, PRE restraints have the advantage to cover longer distances. On the other hand, they are less precise and require the preparation of several spin-labeled protein samples. The precision of PRE distance information is primarily limited by the intrinsic flexibility of the spin labels, typically MTSL (methanethiosulfonate), which is attached to the protein via a disulfide bond to an artificially introduced cysteine residue. Recently, more rigid, but also more bulky, disulfide-linked spin labels have been proposed (Fawzi et al., 2011). The range of distances for which PREs can yield quantitative information is about 10–25 Å; for shorter distances, the relaxation by the spin-label bleaches out the signals, while for longer

**Table 1. Helical Membrane Proteins Solved with PRE Restraints**

Protein <sup>a</sup>	PDB	Year	Residues	Restraints <sup>b</sup>					PRE-Based Distance Limits	
				NOE	H-bond	Dihedral	RDC	PRE	(Å) <sup>c</sup>	
Mistic	1YGM	2005	118	573/29	42	346	—	487	<11, 12–15, 16–19, >20	
FXYD1	2JO1	2007	72	447/32	17	41	55	20	<15, 15–21 ± 1.5, >21	
KCNE1	2K21	2008	138	273/0	36	225	20	464	<18, 18–23 ± 4, >23	
DsbB	2K73, 2K74	2008	183	446/39	97	295	337	1,144	<16, 16–21 ± 2, >21	
DAGK	2KDC	2009	3 × 121	335/0	65	185	67	208	2–19, 15–21 ± 4, >21	
Rv1761c	2K3M	2009	151	—	36	210	218	162	<15, 15–21 ± 3, >21	
ArcB	2KSD	2010	115	72/0	31	72	—	291	<11, 12–15, 16–19, > 20	
QseC	2KSE	2010	186	—	28	84	—	295	<11, 12–15, 16–19, > 20	
KdpD	2KSF	2010	107	—	56	144	—	845	<11, 12–15, 16–19, > 20	
Presenilin-1 CTF	2KR6	2010	176	71/0	50	178	—	508	<12, 12–20 ± 4, >20	
UCP2	2LCK	2011	303	—	—	454	470	452	<16, 1–17, 8–18, 9–19, 10–20, 11–21, 12–22, 13–23, 14–100	
Proteorhodopsin	2L6X	2011	235	376/87	133	392	81	1,006	<13, 13–15 ± 4, >20	

Based on the database “Membrane Proteins of Known Structure Determined by NMR” compiled by Dror E. Warschawski (<http://www.drorlist.com/nmr/MPNMR.html>) as of August 26, 2011.

<sup>a</sup>Proteins: Mistic (Roosild et al., 2005); FXYD1 (Teriete et al., 2007); KCNE1 (Kang et al., 2008); DsbB (Zhou et al., 2008); DAGK, diacylglycerol kinase (Van Horn et al., 2009); Rv1761c (Page et al., 2009); ArcB, QseC, KdpD (Maslennikov et al., 2010); Presenilin-1 C-terminal fragment (Sobhanifar et al., 2010); UCP2, uncoupling protein 2 (Berardi et al., 2011); Proteorhodopsin (Reckel et al., 2011).

<sup>b</sup>Number of conformational restraints used in the structure calculation, as reported in the structural statistics table of the corresponding publication. Where two numbers are given for NOEs, the first refers to the total number of NOE restraints, the second to long-range restraints between residues *i* and *j* with  $|i - j| \geq 5$ . H-bond, hydrogen bond distance restraints. Dihedral, dihedral angle restraints. RDC, residual dipolar coupling orientational restraints. PRE, paramagnetic relaxation enhancement distance restraints.

<sup>c</sup>PRE-based distance limits: < *x*, upper bound of *x* Å; *x*–*y* ± *d*, lower and upper bounds with an error margin of *d* Å for distances in the range of *x* to *y* Å; > *y*, lower bound of *y* Å.

distances, the effect becomes undetectably weak. Therefore, the nature of the conformational information from PREs differs significantly from that of NOEs and different approaches, in particular regarding the size of the error bounds, have been used to incorporate PRE-derived distance restraints into the structure calculation of helical membrane proteins (Table 1). A systematic study of the impact of PRE-derived distances on membrane protein structure determinations by NMR is necessary to assess the possible accuracy of PRE-derived protein structures. Since for each spin-label position one or several isotope-labeled samples have to be prepared, it is important to estimate the minimum number of spin labels required to obtain a structure of given quality. Studies on the use of PRE restraints in NMR structure calculations have been conducted, e.g., for large soluble proteins (Battiste and Wagner, 2000),  $\beta$ -barrel membrane proteins (Liang et al., 2006), protein-ligand (Constantine, 2001), and protein-DNA (Iwahara et al., 2004) complexes, as well as for denatured proteins (Gillespie and Shortle, 1997).

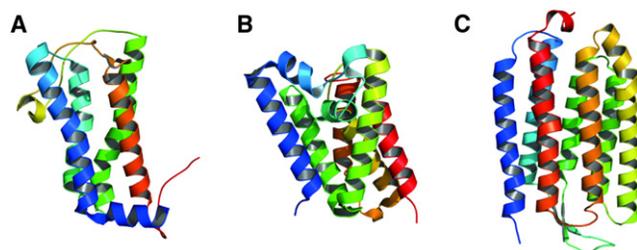
## RESULTS AND DISCUSSION

Here, we present a systematic evaluation of PRE-derived distance restraints for structure calculations of  $\alpha$ -helical membrane proteins with limited or no long-range NOE information. In particular, we investigated the impact of the precision of PRE-derived restraints and of the number and the location of spin labels on the quality of the structure. The evaluation was based on simulated structural information derived from the solution NMR structure of DsbB (Protein Data Bank [PDB]

2K73) (Zhou et al., 2008), an oxidoreductase, as well as the crystal structures of the monomeric proteins GlpG (PDB 2IC8) (Wang et al., 2006), a member of the rhomboid protease family, and the light-driven chloride pump halorhodopsin (HR; PDB 1E12) (Kolbe et al., 2000). Thus, we used data sets for integral membrane proteins with four, six, and seven transmembrane helices (Figure 1).

Data sets of simulated PRE, NOE, hydrogen bond, and torsion angle restraints were generated from the regularized reference structures with the program CYANA (see Experimental Procedures) (Güntert et al., 1997). The data sets were designed to reflect the difficulties of NMR data collection for helical membrane proteins. The basic set of restraints used for each calculation consisted of restraints for short-range backbone  $H^N-H^N$  NOEs,  $\phi$  and  $\psi$  backbone torsion angles predictable from backbone chemical shifts using the program TALOS+ (Shen et al., 2009), backbone hydrogen bonds in the  $\alpha$ -helical regions as defined in the PDB entries, and a small subset of the expected side-chain NOEs between Ala, Leu, and Val methyl groups (Table 2; Kainosho et al., 2006; Tugarinov et al., 2006). We assumed that because of spectral overlap and insensitivity only 10% of all methyl-methyl NOEs expected based on spatial proximity can be assigned (see Experimental Procedures).

These basic data sets were complemented by PRE distance restraints for which distances were measured from the  $C^\beta$  atom of the spin-labeled residue to amide protons (Table 2). If not denoted otherwise, distances smaller than 13 Å received an upper limit of 13 Å, distances greater than 20 Å a lower limit of 20 Å, and distances between 13 and 20 Å upper and lower



**Figure 1. Integral Membrane Proteins Used for the Evaluation of the PRE Restraints**

(A) DsbB, four transmembrane helices.

(B) GlpG, six transmembrane helices.

(C) Halorhodopsin (HR), seven transmembrane helices.

limits equal to the distance  $\pm 4$  Å (Battiste and Wagner, 2000). Since the often severe peak overlap in the spectra of  $\alpha$ -helical membrane proteins favors the use of selectively labeled samples, PRE-derived distance restraints were simulated assuming selective  $^{15}\text{N}$  labeling of the Gly, Ser, Ala, Phe, Thr, and Leu residues (Reckel et al., 2008; Trbovic et al., 2005). Spin-label sites were introduced at helix ends, since loop regions are often too flexible for meaningful distance measurements and the attachment of spin labels inside a helix can be difficult due to shielding by the micelle.

On the basis of these data sets, we first wanted to investigate the effect of the error margin of the PRE-derived distances on the structural quality. In these simulations, we assumed a “fully” spin-labeled protein with spin labels attached at both ends of each transmembrane helix. This scenario comprised 8 spin labels for DsbB, 12 for GlpG, and 14 for halorhodopsin. To date, there is no generally accepted approach for dealing with the uncertainty of large distances resulting from PRE measurements. During the structure determination of other membrane proteins, such as the ones listed in Table 1, errors of 3–8 Å for distances in the range of 11–25 Å were assumed (Table 1) to account for the size and flexibility of the disulfide-linked nitroxide spin label and for errors in the analysis. In an early evaluation of PRE restraints for structure determination, Battiste and Wagner showed that PRE-derived distances with an error bound of  $\pm 4$  Å match well with the corresponding distances in the three-dimensional protein structure (Battiste and Wagner, 2000). It has also been argued that the  $r^{-6}$  dependency reduces the distance error in the analysis of PRE data (Battiste and Wagner, 2000; Gillespie and Shortle, 1997). We performed structure calculations for data sets with error bounds for PRE restraints of  $\pm 0.5$ –10 Å, and analyzed the precision and accuracy of the resulting structures (Figure 2). The results show that with a PRE error margin of 0.5 Å the target structure can be obtained with an accuracy of 0.5–0.63 Å backbone rmsd to the reference structure. Surprisingly, however, even with error bounds of up to 10 Å, structures with a backbone rmsd of 1.0–1.6 Å to the reference structure were obtained, which demonstrates that the dependence of the accuracy of the structure on the value of the PRE error margin is relatively weak. In contrast, using only the non-PRE restraints in our data sets yielded structures with rmsds to the reference of above 6 Å for all three proteins, emphasizing the importance of complementary PRE restraints. However, while PRE restraints

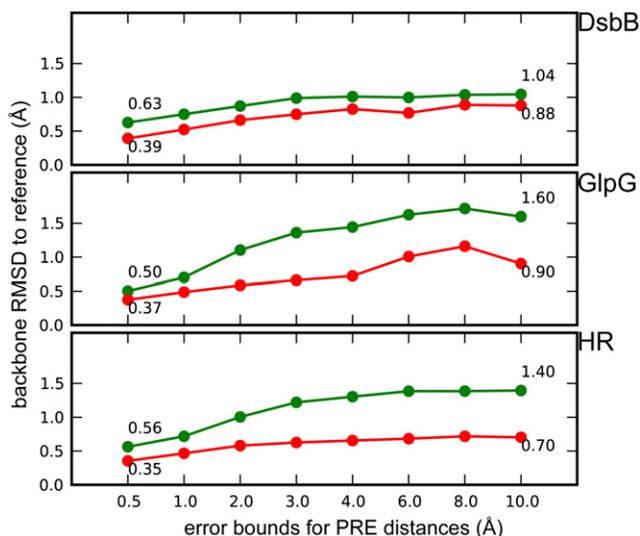
**Table 2. Structural Restraints Generated for the “Fully” Spin-Labeled Target Proteins**

Conformational Restraints	DsbB	GlpG	HR
$\text{H}^{\text{N}}\text{-H}^{\text{N}}$ NOE restraints	279	305	417
$\phi/\psi$ torsion angle restraints	284	301	448
Restrained hydrogen bonds	81	99	138
Methyl-methyl NOE restraints	23	17	25
PRE restraints	814	1,314	2,223
Upper limits 13 Å	96	193	247
Upper/lower limits = distance $\pm 4$ Å	278	540	742
Lower limits 20 Å	440	581	1,234
Total number of restraints	1,481	2,036	3,251

with a small error margin could in principle yield very accurate structures, this accuracy cannot be achieved in practice because of the flexibility of the spin labels and the limited spectral quality of helical membrane proteins. Our results therefore support the finding by Battiste and Wagner suggesting that an error margin of  $\pm 4$  Å for PRE-based distance restraints accounts sufficiently for the experimental limitations of the PRE data (Battiste and Wagner, 2000). A higher value of the error bound could be compensated by increasing the number of restraints collected from spin-label sites evenly distributed throughout the protein.

Another parameter in the interpretation of the PRE effect is the upper distance limit used for residues in close proximity to the spin label. The resonances of these residues are usually broadened beyond detection and no distance calculation based on peak intensity ratios is possible. In previous PRE-based structure determinations different upper distance limits have been employed in this case, ranging from 11 to 19 Å (Table 1). We applied an upper distance limit of 13 Å in our simulations (Battiste and Wagner, 2000). Choosing this value for all true distances  $< 13$  Å in the reference structures implies that for some residues this upper limit may actually be close to the real distance leaving only a small error margin and therefore resulting in an overprecise distance limit for true distances near 13 Å. To exclude that our findings were biased by such overprecise distances, an additional round of simulation applied an upper limit of 15 Å instead of 13 Å for all true distances  $< 13$  Å. The results showed a similar dependence of the structural accuracy and precision on the size of the error bounds albeit with slightly increased rmsd values as would be expected for increased error bounds (see also Figure 3B).

Thus far, the simulations were conducted assuming that all transmembrane helices are spin labeled at both termini. In practice, however, this “fully” labeled state can often not be achieved and a reduced number of spin-labeling sites has to be used, making the number of spin-label sites an important factor. We therefore performed also structure calculations with varying numbers of spin-label sites and also evaluated the impact of long-range NOE information at different stages. The calculations indicate that structures with rmsd values to the reference structure below 3 Å are possible also with not “fully” spin-labeled data sets containing limited PRE data (Figures 3A and 4). Notably, there exists a threshold of approximately one spin label per helix



**Figure 2. Effect of the Error Bounds Applied to PRE Derived Distances**

Structures were calculated using the conformational restraints of Table 2 with increasing error bounds of the PRE-derived distance restraints and the effect on the structural precision (rmsd to the mean coordinates, red) and accuracy (rmsd to reference structure, green) was evaluated.

to obtain correct structures. PRE distance restraints from four, six and seven spin labels for DsbB, GlpG, and halorhodopsin, respectively, provided (together with the other restraints) sufficient information to calculate correct structures of these three  $\alpha$ -helical membrane proteins.

A known drawback of structure calculation using PRE restraints with large error bounds of  $\pm 4$  Å is the occurrence of structures with different relative orientation of the secondary structure elements, i.e., the occurrence of approximate mirror image topologies that are both compatible with the imprecise restraints (Battiste and Wagner, 2000). Recently, a method for the optimal positioning of spin labels for PRE data collection was presented (Chen et al., 2011). In contrast to our study, Chen and coworkers suggested the minimal number and optimal positioning of spin-label sites to obtain a starting structure for prediction methods. In a simplified model of membrane protein structures, consisting of cylinders in a rhombic assembly, intra-helical distances were derived upon introduction of a paramagnetic label. Based on the models, the PRE distances principally allowed the identification of the correct helix topology of the protein. However, the resulting structures were of low accuracy (up to 6 Å rmsd) when compared to existing structures of the target class of proteins. To determine an accurate structure, additional prediction methods have to be employed. As also mentioned by the authors, PRE data were not sufficient to distinguish between structures with correct and mirrored topology.

To analyze the problem of obtaining incorrect topologies with PRE restraints, we clustered the calculated structures into subgroups based on their pairwise rmsd values (see Experimental Procedures). With at least one spin label per helix, the calculated ensembles were always of low topological ambiguity, as demonstrated by the small number of clusters present in the structure bundles (Figure 3A). In contrast, lower numbers of spin

labels tended to yield erroneous structures. The importance of this threshold of one spin label per helix becomes clear upon removal of the small number of long-range NOEs (Figure 3A, data sets marked by "\*", e.g., 4\* for DsbB, 6\* for GlpG, and 7\* for HR). Without the long-range NOE information, the structural divergence increases, as manifested by larger numbers of clusters in the ensembles. The absence of methyl-methyl NOEs could be compensated by increasing the number of spin labels (Figure 3A). Similar results were also obtained when the upper limit of 13 Å was increased to 15 Å (Figure 3B). While the structural precision is lower, the number of clusters found in the structure bundle remains almost identical.

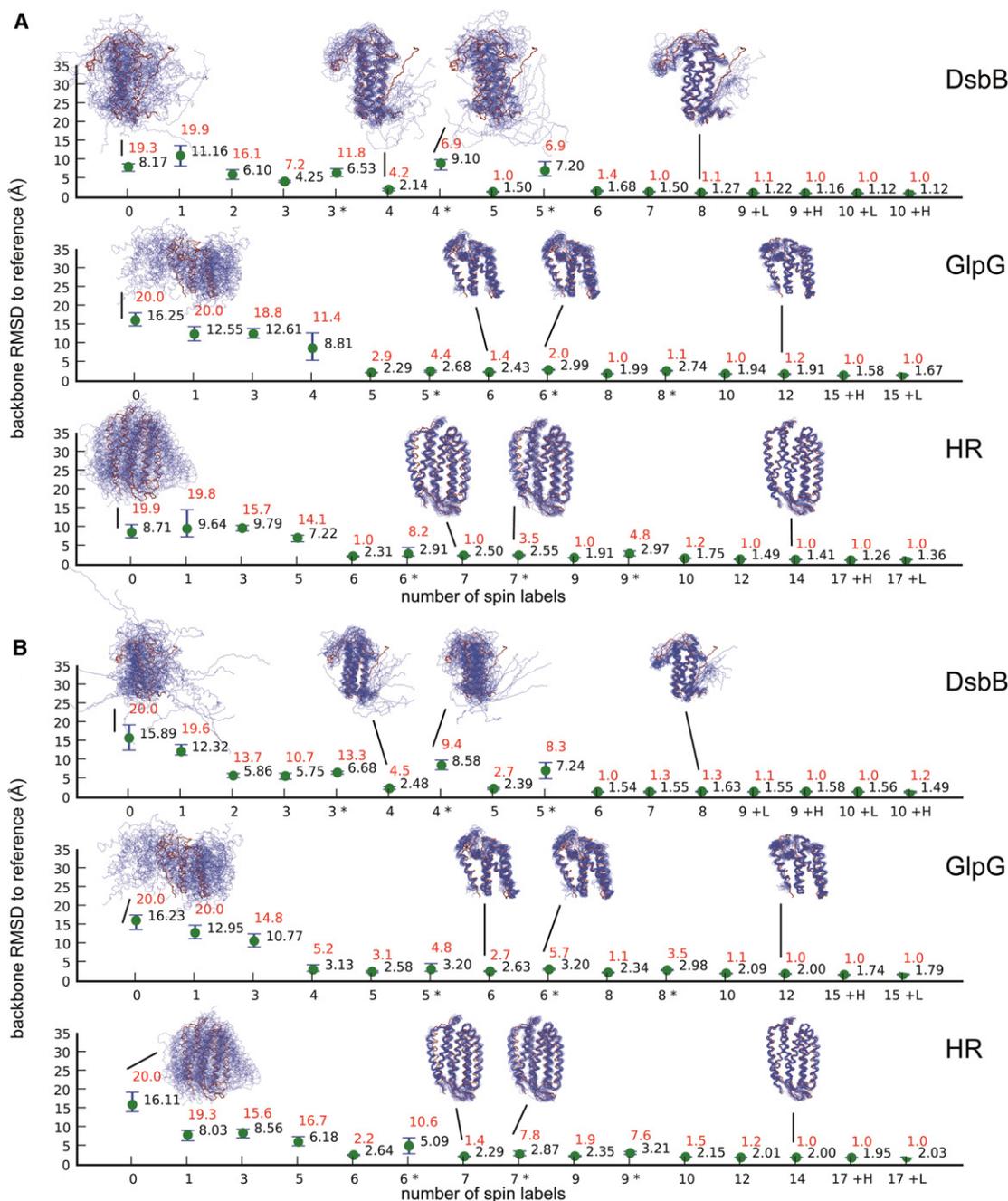
Since the  $\pm 4$  Å error range of the PRE distance restraints is comparable to the diameter of an  $\alpha$  helix, it could be imagined that PRE data might be insensitive to rotations of a helix around its axis, and that structures could result that have the correct topology but severely incorrect rotational orientations of one or several helices, by which side chains buried in the interior would be exposed to the membrane and vice versa. However, a visual inspection of the structures showed that this did not happen for the structures with rmsd to the reference below 3 Å.

Apart from the number of spin labels, the positioning of the spin label also influences the structural quality. This is true, in particular, if predominantly only one membrane side of the protein is spin labeled, e.g., only cytosolic helix ends (Figure 5). In this case, the resulting PRE restraints are highly redundant, while important structural information on the other membrane side is lost. Testing this unfavorable labeling scheme for all three model proteins demonstrated that the correct global fold could no longer be established, ending in low-quality structures with high rmsd values relative to the reference structure. This effect could partially be compensated by the use of long-range NOEs (Figure 5), demonstrating that even a small number of long range NOEs can significantly improve the structural quality.

In addition to PREs, residual dipolar couplings (RDCs) have been used successfully in the structure determination of  $\alpha$ -helical membrane proteins (Table 1). We therefore also evaluated the impact of additional RDC restraints on the structure calculations with our PRE-based data sets. Structures of halorhodopsin were calculated including in addition to the other data simulated HN RDCs for the  $\alpha$ -helical regions with varying number of spin labels. Using RDCs improved the backbone rmsd values to the reference by about 0.6 Å in the threshold cases of six and seven spin labels. Rmsds of structures calculated with PRE distances derived from more than seven spin labels were improved only by around 0.2 Å. The number of structure clusters identified in the structure ensembles remained effectively the same. These results indicate that RDC data are a valuable additional source of structural information that may serve to refine existing structures with correct topologies that were obtained in predominantly PRE-based structure calculations.

## Conclusions

In summary, our study shows that PRE-derived distance restraints can provide sufficient structural information to accurately obtain the backbone structure of  $\alpha$ -helical integral membrane proteins in cases where only very limited long-range NOE data are available. NMR structures of  $\alpha$ -helical membrane



**Figure 3. Membrane Protein Structures Calculated with Different Sets of Structural Information**

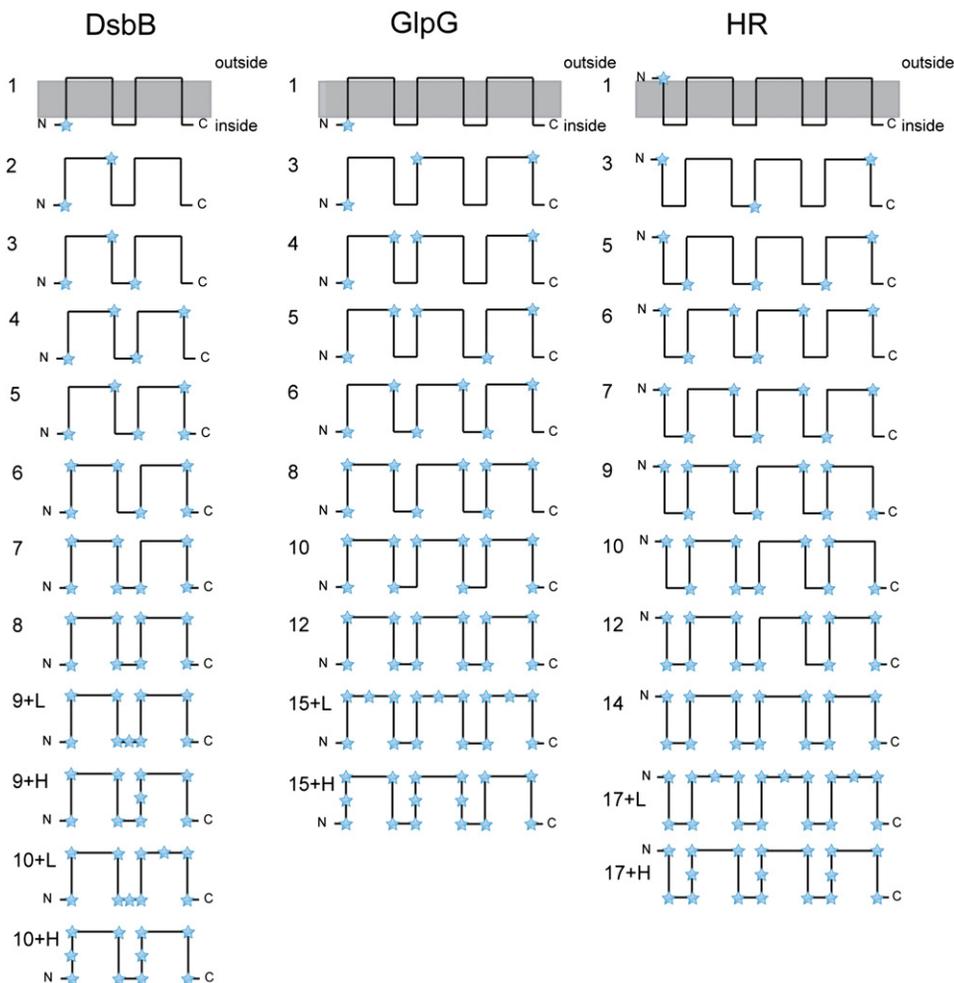
The horizontal axis identifies the data sets by the number of spin labels, and the vertical axis gives the backbone rmsd to the reference structure. Details of the labeling schemes are given in Figure 4. Data sets marked by asterisks excluded long-range methyl-methyl NOEs. Data sets with spin labels in addition to those at the end of the helices are marked by “+H” or “+L” depending on the position of the additional spin label either inside helices or in loop regions, respectively. The range of rmsd values observed in ten independent calculations is shown by the blue bar, with the arithmetic mean given by the green circle. The average number of distinct structural clusters is given in red numbers.

(A) Structures calculated using intrahelical backbone NOE-derived restraints and upper limits of 13 Å for the PRE-derived distances shorter than 13 Å.

(B) Structures calculated excluding intrahelical backbone NOE-derived restraints and upper limits of 15 Å for the PRE-derived distances shorter than 13 Å.

proteins with an accuracy of about 1.5 Å backbone rmsd are feasible even if the error bounds of the PRE distances are as large as  $\pm 4$  Å. On the other hand, the wide error bounds for PRE restraints limit the maximally attainable accuracy to about

1.0 Å even when using large numbers of spin labels. PRE distance restraints should be collected from at least one spin label per helix, distributed evenly throughout the protein. PRE data thus constitute a valuable source of information for the



**Figure 4. Labeling Scheme Used for the Structure Calculations in Figure 3**

Helices are represented by vertical, loops by horizontal lines for all three  $\alpha$ -helical membrane proteins. Stars indicate spin-label positions for the respective data sets. The orientation of the proteins within the membrane is shown in the upper panel.

structure determination of challenging  $\alpha$ -helical membrane proteins, as we have recently shown with the first de novo NMR structure determination of a heptahelical integral membrane protein, proteorhodopsin (Reckel et al., 2011). PREs data might play an important role for future NMR structure determinations of G protein-coupled receptors.

## EXPERIMENTAL PROCEDURES

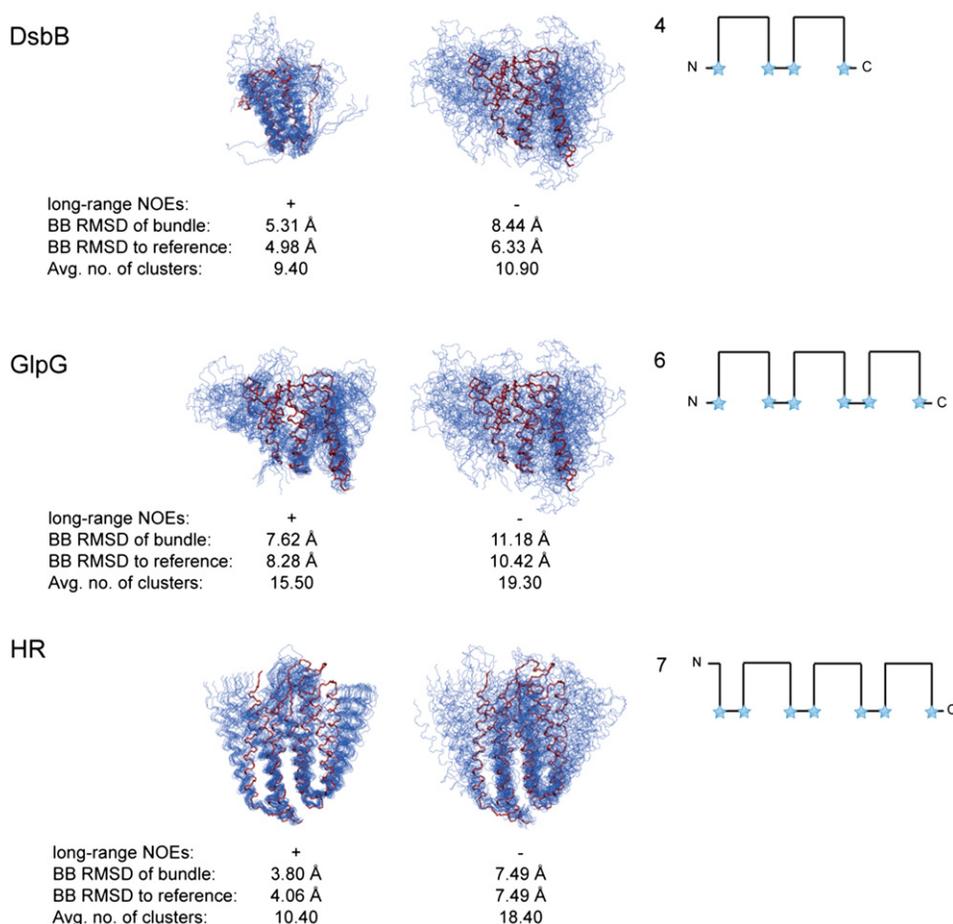
### Regularization of Reference Structures

The DsbB, GlpG, and halorhodopsin structures from the PDB (accession codes 2K73, 2IC8, and 1E12, respectively) were regularized to adhere exactly to the standard geometry (bond lengths, bond angles, planarities) of CYANA (Güntert et al., 1997) in order to exclude any possible bias arising from the use of different force fields in the original refinement and in the CYANA calculations. Regularization was achieved by recalculating the structures in CYANA using restraints with an upper bound of 0.1 Å on the distances between corresponding N, C<sup>α</sup>, and C' atoms in the structured regions of the regularized and original PDB structures (residues 2–97, 112–163 for DsbB, 91–272 for GlpG, and 24–262 for halorhodopsin), and torsion angle restraints with a width of 20° centered around the value of each torsion angle in the reference structure (Gottstein et al., 2012). If the original structure is represented by a structure

bundle, the distance restraints are applied between the regularized structure and the mean coordinates of the structure bundle, and the torsion angle restraints are applied to the mean value of the angle in the individual conformers with a range given by the standard deviation (but not smaller than  $\pm 10^\circ$ ). It is possible to add complementary structural information to the recalculation procedure, e.g., in the case of an NMR structure the original experimental restraints that had been used to calculate it. Using these restraints, the structure was recalculated by the standard torsion angle dynamics simulated annealing protocol, resulting in a close overlay of the target molecule with its reference structure. The regularized coordinates were used for the simulation of NMR restraints and as reference structures for the calculation of rmsds.

### Data Set Generation for Structure Calculation

Experimental NMR restraints were simulated from the regularized reference structures using CYANA. Short-range backbone NOEs were generated as upper limits of 5 Å between the backbone amide protons of residues ( $i, i + 1$ ) and ( $i, i + 2$ ), if the corresponding distance in the reference structure was shorter than 5 Å. Restraints for the backbone torsion angles  $\varphi$  and  $\psi$  were centered at the value measured in the reference structure and applied with a width of 20°. Hydrogen bonds for the  $\alpha$ -helical regions were determined from the HELIX entries of the original crystal or NMR structure PDB files, and restrained by upper and lower limits of  $1.8 \leq d(O_i, H_{i+4}) \leq 2$  Å and



**Figure 5. Unfavorable Spin-Labeling Patterns**

All spin labels were placed at the cytosolic end of the transmembrane helices and simulations were conducted with and without long-range NOE information. The resulting structures are of different quality as indicated by the backbone (BB) rmsd values given below each structure bundle.

$2.7 \leq d(O_i, N_{i+4}) \leq 3 \text{ \AA}$ . Side-chain NOEs with an upper bound of 5 Å were generated for random selections of 10% of the distances  $\leq 5 \text{ \AA}$  between methyl groups of valine, alanine, and leucine in the reference structures. The selection of random subsets of 10% of all available distances was done to account for severe spectral overlap in the NOESY spectra of  $\alpha$ -helical membrane proteins. PRE distances were divided into three groups. Spin-labeled residues were not explicitly introduced into the protein sequences used for the structure calculations. Rather, the simulated PRE distances were measured between the  $C^\beta$  atom of potentially labeled residues and the backbone amide protons of  $^{15}\text{N}$  labeled Gly, Ser, Ala, Phe, Thr, and Leu residues in the effective range of the paramagnetic relaxation effects. To account for the size and flexibility of the Cys-MTSL side-chain upper and lower limits were derived from the determined distance in the reference structure with an error bound of  $\pm 4 \text{ \AA}$ , unless noted otherwise, if the distance was in the range between 13 and 20 Å. This range is experimentally amenable to quantitative evaluation. In addition, distances smaller than 13 Å received an upper limit of 13 Å, and distances greater than 20 Å a lower limit of 20 Å (Battiste and Wagner, 2000). For comparison, distances smaller than 13 Å were alternatively modeled also with an upper limit of 15 Å.

#### Structure Calculation

Structure calculations were done using the standard structure calculation protocol implemented in CYANA. One hundred initial conformers with random torsion angle values were subjected to simulated annealing using 20,000 torsion angle dynamics steps, and the 20 lowest target function conformers were selected to represent the NMR structure of the protein. To ensure the

independence of resulting structures from biased starting conformations, all structure calculations were repeated ten times with different random seeds.

#### Analysis and Clustering of Structures

The accuracy of structures was quantified by the rmsd value to the reference structure for the backbone atoms N,  $C^\alpha$ , and  $C'$  in the structured regions of residues 1–37, 41–62, 68–97, 142–163 for DsbB, 91–272 for GlpG, and 24–262 for halorhodopsin. When performing calculations with reduced PRE distance restraint data sets, the resulting ensembles often displayed a lack of convergence resulting in high backbone rmsds to the mean coordinates. In these cases, subgroups of structures in the ensemble often had much lower rmsd values relative to each other than to the entire ensemble. If only a few structures do not show structural resemblance with the rest, the calculated average backbone rmsd does not truly correspond to the convergence of the structure bundles. In extreme cases, a single conformer in an ensemble displayed mirror symmetry to the other conformers. This is a known problem that can arise in structure calculations with imprecise PRE distance restraints. For a better evaluation of the quality of structure ensembles, a clustering method was applied. Pairwise rmsds were calculated among all structures and stored in a  $(20 \times 20)$  matrix. By way of the clustering method UPGMA (unweighted pair group method with arithmetic mean) (Sokal and Michener, 1958) a rooted tree was constructed from the aforementioned rmsd matrix. Thus, it was possible to cluster structures into groups that show rmsd values below a certain threshold value. Clustering was performed as follows: the first structure is, by default, in the first cluster. The second structure is checked for similarity to the first structure. If the rmsd of the two structures is below the

threshold, then structure 2 is added to the first cluster. Otherwise, a new cluster is formed. This procedure is repeated for all structures. If a structure is successfully inserted into a cluster, then the rmsd of this cluster to the reference structure is recalculated as the mean of the rmsds of all structures populating the cluster to the reference. All pairs of structures in a cluster have an rmsd below the threshold value. The threshold value applied was 3 Å. After the clustering process it was possible to classify the quality of resulting structures by the number of clusters in context with the rmsd of the clusters (composed of the mean of all rmsds of the members to the reference). Structures were visualized and inspected using the programs MOLMOL (Koradi et al., 1996), PyMOL (Schrödinger, LLC), and Chimera (Pettersen et al., 2004).

#### ACKNOWLEDGMENTS

P.G. gratefully acknowledges financial support by the Lichtenberg program of the Volkswagen Foundation, the Deutsche Forschungsgemeinschaft, and the Japan Society for the Promotion of Science. V.D. gratefully acknowledges support by the Deutsche Forschungsgemeinschaft (DO545/7-1 and SFB 807) and the Cluster of Excellence Frankfurt (Macromolecular Complexes).

Received: January 18, 2012

Revised: March 19, 2012

Accepted: March 26, 2012

Published online: May 3, 2012

#### REFERENCES

- Battiste, J.L., and Wagner, G. (2000). Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* 39, 5355–5365.
- Berardi, M.J., Shih, W.M., Harrison, S.C., and Chou, J.J. (2011). Mitochondrial uncoupling protein 2 structure determined by NMR molecular fragment searching. *Nature* 476, 109–113.
- Chen, H.L., Ji, F., Olman, V., Mobley, C.K., Liu, Y.Z., Zhou, Y.P., Bushweller, J.H., Prestegard, J.H., and Xu, Y. (2011). Optimal mutation sites for PRE data collection and membrane protein structure prediction. *Structure* 19, 484–495.
- Constantine, K.L. (2001). Evaluation of site-directed spin labeling for characterizing protein-ligand complexes using simulated restraints. *Biophys. J.* 81, 1275–1284.
- Fawzi, N.L., Fleissner, M.R., Anthis, N.J., Kálai, T., Hideg, K., Hubbell, W.L., and Clore, G.M. (2011). A rigid disulfide-linked nitroxide side chain simplifies the quantitative analysis of PRE data. *J. Biomol. NMR* 51, 105–114.
- Fernández, C., Hilty, C., Wider, G., Güntert, P., and Wüthrich, K. (2004). NMR structure of the integral membrane protein OmpX. *J. Mol. Biol.* 336, 1211–1221.
- Gaponenko, V., Howarth, J.W., Columbus, L., Gasmi-Seabrook, G., Yuan, J., Hubbell, W.L., and Rosevear, P.R. (2000). Protein global fold determination using site-directed spin and isotope labeling. *Protein Sci.* 9, 302–309.
- Gautier, A., Mott, H.R., Bostock, M.J., Kirkpatrick, J.P., and Nietlisbach, D. (2010). Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy. *Nat. Struct. Mol. Biol.* 17, 768–774.
- Gillespie, J.R., and Shortle, D. (1997). Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *J. Mol. Biol.* 268, 170–184.
- Gottstein, D., Kirchner, D.K., and Güntert, P. (2012). Simultaneous single-structure and bundle representation of protein NMR structures in torsion angle space. *J. Biomol. NMR* 52, 351–364.
- Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* 273, 283–298.
- Hiller, S., Garcés, R.G., Malia, T.J., Orekhov, V.Y., Colombini, M., and Wagner, G. (2008). Solution structure of the integral human membrane protein VDAC-1 in detergent micelles. *Science* 321, 1206–1210.
- Iwahara, J., Schwieters, C.D., and Clore, G.M. (2004). Ensemble approach for NMR structure refinement against (1)H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. *J. Am. Chem. Soc.* 126, 5879–5896.
- Iwahara, J., Tang, C., and Marius Clore, G. (2007). Practical aspects of (1)H transverse paramagnetic relaxation enhancement measurements on macromolecules. *J. Magn. Reson.* 184, 185–195.
- Kainosho, M., Torizawa, T., Iwashita, Y., Terauchi, T., Mei Ono, A., and Güntert, P. (2006). Optimal isotope labelling for NMR protein structure determinations. *Nature* 440, 52–57.
- Kang, C.B., Tian, C.L., Sönnichsen, F.D., Smith, J.A., Meiler, J., George, A.L., Jr., Vanoye, C.G., Kim, H.J., and Sanders, C.R. (2008). Structure of KCNE1 and implications for how it modulates the KCNQ1 potassium channel. *Biochemistry* 47, 7999–8006.
- Kolbe, M., Besir, H., Essen, L.O., and Oesterhelt, D. (2000). Structure of the light-driven chloride pump halorhodopsin at 1.8 Å resolution. *Science* 288, 1390–1396.
- Koradi, R., Billeter, M., and Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* 14, 51–55, 29–32.
- Kosen, P.A. (1989). Spin labeling of proteins. *Methods Enzymol.* 177, 86–121.
- Liang, B.Y., and Tamm, L.K. (2007). Structure of outer membrane protein G by solution NMR spectroscopy. *Proc. Natl. Acad. Sci. USA* 104, 16140–16145.
- Liang, B.Y., Bushweller, J.H., and Tamm, L.K. (2006). Site-directed parallel spin-labeling and paramagnetic relaxation enhancement in structure determination of membrane proteins by solution NMR spectroscopy. *J. Am. Chem. Soc.* 128, 4389–4397.
- Maslennikov, I., Klammt, C., Hwang, E., Kefala, G., Okamura, M., Esquivies, L., Mörs, K., Glaubitz, C., Kwiatkowski, W., Jeon, Y.H., and Choe, S. (2010). Membrane domain structures of three classes of histidine kinase receptors by cell-free expression and rapid NMR analysis. *Proc. Natl. Acad. Sci. USA* 107, 10902–10907.
- Page, R.C., Lee, S., Moore, J.D., Opella, S.J., and Cross, T.A. (2009). Backbone structure of a small helical integral membrane protein: a unique structural characterization. *Protein Sci.* 18, 134–146.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612.
- Reckel, S., Sobhanifar, S., Schneider, B., Junge, F., Schwarz, D., Durst, F., Löhr, F., Güntert, P., Bernhard, F., and Dötsch, V. (2008). Transmembrane segment enhanced labeling as a tool for the backbone assignment of  $\alpha$ -helical membrane proteins. *Proc. Natl. Acad. Sci. USA* 105, 8262–8267.
- Reckel, S., Gottstein, D., Stehle, J., Löhr, F., Verhoefen, M.K., Takeda, M., Silvers, R., Kainosho, M., Glaubitz, C., Wachtveitl, J., et al. (2011). Solution NMR structure of proteorhodopsin. *Angew. Chem. Int. Ed.* 50, 11942–11946.
- Renault, M., Saurel, O., Czaplicki, J., Demange, P., Gervais, V., Löhr, F., Réat, V., Piotto, M., and Milon, A. (2009). Solution state NMR structure and dynamics of KpOmpA, a 210 residue transmembrane domain possessing a high potential for immunological applications. *J. Mol. Biol.* 385, 117–130.
- Roosild, T.P., Greenwald, J., Vega, M., Castronovo, S., Riek, R., and Choe, S. (2005). NMR structure of Mistic, a membrane-integrating protein for membrane protein expression. *Science* 307, 1317–1321.
- Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* 44, 213–223.
- Sobhanifar, S., Schneider, B., Löhr, F., Gottstein, D., Ikeya, T., Mlynarczyk, K., Pulawski, W., Ghoshdastider, U., Kolinski, M., Filippek, S., et al. (2010). Structural investigation of the C-terminal catalytic fragment of presenilin 1. *Proc. Natl. Acad. Sci. USA* 107, 9644–9649.
- Sokal, R.R., and Michener, C.D. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin* 38, 1409–1438.

Teriete, P., Franzin, C.M., Choi, J., and Marassi, F.M. (2007). Structure of the Na,K-ATPase regulatory protein FXYP1 in micelles. *Biochemistry* 46, 6774–6783.

Trbovic, N., Klammt, C., Koglin, A., Löhr, F., Bernhard, F., and Dötsch, V. (2005). Efficient strategy for the rapid backbone assignment of membrane proteins. *J. Am. Chem. Soc.* 127, 13504–13505.

Tugarinov, V., Kanelis, V., and Kay, L.E. (2006). Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat. Protoc.* 1, 749–754.

Van Horn, W.D., Kim, H.J., Ellis, C.D., Hadziselimovic, A., Sulistijo, E.S., Karra, M.D., Tian, C.L., Sönnichsen, F.D., and Sanders, C.R. (2009). Solution nuclear magnetic resonance structure of membrane-integral diacylglycerol kinase. *Science* 324, 1726–1729.

Wang, Y.C., Zhang, Y.J., and Ha, Y. (2006). Crystal structure of a rhomboid family intramembrane protease. *Nature* 444, 179–180.

Zhou, Y.P., Cierpicki, T., Jimenez, R.H.F., Lukasik, S.M., Ellena, J.F., Cafiso, D.S., Kadokura, H., Beckwith, J., and Bushweller, J.H. (2008). NMR solution structure of the integral membrane enzyme DsbB: functional insights into DsbB-catalyzed disulfide bond formation. *Mol. Cell* 31, 896–908.