J Biomol NMR (2008) 42:49–58
DOI 10.1007/s10858-008-9263-1

ARTICLE

Investigation of the utility of selective methyl protonation for determination of membrane protein structures

Steve C. C. Shih · Ileana Stoica · Natalie K. Goto

Received: 14 May 2008 / Accepted: 21 July 2008 / Published online: 2 September 2008

Abstract Polytopic α-helical membrane proteins present one of the final frontiers for protein structural biology, with significant challenges causing severe under-representation in the protein structure databank. However, with the advent of hardware and methodology geared to the study of large molecular weight complexes, solution NMR is being increasingly considered as a tool for structural studies of these types of membrane proteins. One method that has the potential to facilitate these studies utilizes uniformly deuterated samples with protons reintroduced at one or two methyl groups of leucine, valine and isoleucine. In this work we demonstrate that in spite of the increased proportion of these amino acids in membrane proteins, the quality of structures that can be obtained from this strategy is similar to that obtained for all α-helical water soluble proteins. This is partly attributed to the observation that NOEs between residues within the transmembrane helix did not have an impact on structure quality. Instead the most important factors controlling structure accuracy were the strength of dihedral angle restraints imposed and the number of unique inter-helical pairs of residues constrained by NOEs. Overall these results suggest that the most accurate structures will arise from accurate identification of helical segments and utilization of inter-helical distance restraints from various sources to maximize the distribution of long-range restraints.

Keywords Polytopic α-helical membrane proteins · Selective methyl protonation · Structure calculation

Introduction

Recent developments in membrane protein structural biology have led to increases in membrane protein structure deposition rates that have been characterized as exponential (White 2004). However, the total number of unique membrane protein folds is still very low, comprising less than 1% of the PDB, with the rate of structure production fluctuating from year to year (Fleishman and Ben-Tal 2006). The relatively small numbers of membrane protein structures highlights the need for an inclusive approach to structure determination that can capitalize on the advantages offered by the range of available technologies. To this end, solution NMR is being increasingly considered for the study of membrane protein targets of growing size and complexity. This has proven particularly successful for β-barrel folds, in part due to the relative ease of sample production for these types of samples (Tamm and Liang 2006). More recent efforts have focused on the application of solution NMR to the more commonly occurring class of membrane protein that are comprised primarily of multiple transmembrane (TM) helices (Columbus et al. 2006; Hu et al. 2007; Poget et al. 2006, 2007; Tian et al. 2007).

Polytopic helical membrane proteins present a unique challenge to structure determination by solution NMR since they tend to be larger than the water-soluble protein structures that are routinely solved by this technique

Steve C. C. Shih and Ileana Stoica have contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s10858-008-9263-1) contains supplementary material, which is available to authorized users.

S. C. C. Shih · I. Stoica · N. K. Goto (✉)
Department of Chemistry, University of Ottawa, 10 Marie Curie, Ottawa, ON, Canada K1N 6N5
e-mail: ngoto@uottawa.ca

(Gerstein 1997). In addition, integration of these proteins into the detergent micelle required to provide a membrane-mimetic environment further increases the molecular weight, typically by 30 kDa or more. This can be further exacerbated by oligomerization, particularly when the minimal functional unit is not monomeric, as illustrated in a large proportion of existing membrane protein structures. These factors make it necessary to use methods specifically geared toward the study of large complexes, including relaxation-optimized NMR experiments in combination with specific isotope labeling schemes (Sanders and Sonnichsen 2006; Tugarinov et al. 2004). An established approach often applied to large water-soluble proteins uses samples that are uniformly deuterated at all non-exchangeable sites except for one or two of the methyl groups of Leu, Val and Ile (Goto et al. 1999; Tugarinov et al. 2006; Tugarinov and Kay 2004). These selectively methyl-protonated samples eliminate a significant source of relaxation that can otherwise prevent the acquisition of spectra for large complexes, while at the same time allowing a critical source of structural information to be accessed via NOEs involving these protonated methyl groups. This labeling scheme has permitted global folds to be obtained for a number of large water-soluble proteins (e.g. the 27 kDa ubiquitin conjugating enzyme (Merkley and Shaw 2004), the 82 kDa malate synthase (Tugarinov et al. 2005)), as well as the OmpX β-barrel protein in detergent (Fernandez et al. 2004). Since protein–detergent complexes for polytopic helical membrane proteins will also tend to exceed 50 kDa, methyl protonation appears to be an attractive approach for the study of these molecules by solution NMR.

One of the potential limitations for the selective methyl protonation strategy that had been observed for water-soluble proteins is that primarily α-helical proteins usually give rise to a relatively small number of long-range NOE restraints, and therefore less accurate structures are obtained compared to those for other structural classes (Gardner et al. 1997). Although this may suggest that the study of polytopic helical membrane protein structure will suffer similar limitations, membrane proteins differ significantly from water-soluble proteins in ways that could improve the quality of fold that is obtained. Specifically, these proteins contain a larger proportion of hydrophobic amino acids, including those that would be labeled by the standard selective methyl protonation approach (Liu et al. 2002). In addition, the environmental constraints imposed by the anisotropy of the native lipid bilayer give rise to structural features that are unique to this class of proteins. For example, the average TM helix length is longer than the length of helices in soluble proteins (Gerstein 1997), with a smaller range of crossing angles between interacting helices (Bowie 2005), and protein interiors tend to be more tightly packed compared to those of water-soluble proteins (Eilers et al. 2000). It is possible that these distinctions could lead to differences in the ability of the selective methyl protonation approach to generate global folds for polytopic helical membrane proteins.

Given the differences between polytopic helical membrane proteins and their water-soluble counterparts we were interested in evaluating the utility of selectively methyl protonated samples for membrane protein structure determination by solution NMR. Understanding the advantages and limitations of this approach prior to sample production was of particular interest for us since lower yields are typically obtained for these types of proteins, making this labeling scheme an especially costly proposition. For these reasons we have performed a series of structure calculations using simulated NOE datasets for representative membrane protein structures taken from the PDB. Results from this work demonstrate that the inclusion of increasing numbers of intra-helical NOEs from selectively methyl protonated samples did not improve structure accuracy, while the number of unique inter-helical restraints was a critical factor controlling accuracy. Consequently, the selective methyl protonation strategy has the potential to give rise to global folds for membrane proteins with a quality that are, at best, comparable to those obtained with α-helical water-soluble proteins. However, our results suggest that the accurate identification of TM helices and the acquisition of a uniformly-distributed sample of inter-helical restraints will be the most important factors for the determination of accurate membrane protein structures.

Materials and methods

Structure coordinate files were taken for five different membrane proteins for this study: the homodimeric transmembrane domain of human glycophorin A (MacKenzie et al. 1997); a fragment of the human glycine receptor containing TM segments 2 and 3 (Ma et al. 2005); the glycerol-conducting facilitator channel from *E. coli* (Fu et al. 2000); the light-driven chloride pump halorhodopsin (Kolbe et al. 2000) and the mitochondrial ADP/ATP carrier (Pebay-Peyroula et al. 2003). For the NMR structures GpA and GlyR, the first model from each ensemble was chosen as the representative structural model for this study.

NOE restraints were simulated for H^N–H^N, H^N–CH_3, and CH_3–CH_3 pairs between residues in α-helical regions for each membrane protein structure. Following typical observable distances from NOESY experiments on selectively methyl protonated samples (Venters et al. 1995) only H^N–H^N distances less than 5.0 Å and H^N–CH_3, and CH_3–CH_3 distances less than 6.0 Å were taken as NOE

restraints. Methyl distances involved only Val, Leu, and Ile $\delta 1$ methyl protons, and were measured to their pseudo-atom positions. Distance restraints were divided into 3 categories (1.8–3.0 Å, 3.0–4.0 Å, and 4.0–5.0 Å or 6.0 Å for methyl protons) and then each group was assigned to the corresponding NOE restraint bin (namely 1.8–4.0 Å, 2.8–5.2 Å, and 3.2–6.7 Å respectively) following a similar study for soluble proteins (Gardner et al. 1997). Backbone dihedral restraints for α-helices were used in all the simulations with canonical helix values of $\varphi = -57°$ and $\psi = -47°$, similar to values used in previous studies (Alexandrescu 2004; Gardner et al. 1997; Park et al. 2003). (φ, ψ) dihedral angle ranges of either (±2.5°, ±5.0°) for strong restraints or (±20°, ±30°) for more conventional restraints were used, along with the default energy constant set in CNS (Brunger et al. 1998).

Using the simulated data for each membrane protein, structures were calculated using a torsion-angle molecular dynamics simulated annealing protocol implemented in the program CNS (Stein et al. 1997). All calculations were performed starting from extended structures that were subjected to a high temperature (50,000 K) torsion angle annealing stage for 1000 15-fs steps using CNS default values for NOE and dihedral angle force constants (specifically, $k_{NOE} = 150$ and $k_{DA} = 100$ kcal Å^{-2}). Structures then underwent a slow torsion angle molecular dynamics cooling stage from 50,000 K to 2000 K for 10,000 15-fs steps using the same k_{NOE}, and with k_{DA} brought up to 200. The same constants were used for the following 3000 5-fs steps of slow Cartesian cooling from 2000 K to 300 K, before subjecting the structures to standard Powell minimization with $k_{NOE} = 75$ and $k_{DA} = 400$ kcal Å^{-2}. Default CNS values for the repulsive van der Waals term were used throughout. For the larger proteins in the set (namely GlpF, Hrh and AMC) it was necessary to increase the number of cooling steps to 55,000 with a smaller time-step of 5-fs for the torsion angle phase, and to 40,000 steps for the Cartesian phase in order to reach convergence. This change is in line with previous studies that have shown that larger proteins tend to require a larger number of cooling steps, particularly when an extensive set of NOEs are unavailable (Choy et al. 2001; Fossi et al. 2005).

From the torsion angle SA protocol, 100 structures were generated and the 10 lowest energy structures were used to represent the final ensemble. The accuracy of structures was determined from the average of the rmsd between each member of the ensemble and the target structure originally used to generate the set of structural restraints for backbone atoms in helical parts of the protein. The precision of the ensemble was also determined for backbone helical atoms as the average rmsd from the mean structure. In general, generated structures exhibited small deviations from ideal geometry (e.g. mean rms deviations from idealized; bond geometries 0.0006–0.0009 Å; angle geometries 0.24°–0.33°, and; improper geometries 0.08°–0.10°), no NOE violations >0.5 Å, and a low overall ensemble energy (e.g. 30–175 kcal/mol).

Results

In order to examine the quality of structure that can be obtained for polytopic α-helical membrane proteins using selectively methyl labeled samples, representative structures were chosen from the PDB, and a simulated set of NMR data was constructed from each structure. At the time that this study was initiated there were ~120 polytopic membrane proteins structures available (White 2004), 66 of which were comprised of α-helical transmembrane segments representing 23 non-homologous protein families. Since smaller membrane proteins will be more feasible targets for determination by solution NMR only those proteins with fewer than 300 residues were included in the study. This limited the selection to just three representatives; namely the 7 TM segment G-protein coupled receptors, the 6 TM segment facilitator/aquaporin family and the 6 TM segment mitochondrial carrier family. One high-resolution representative was used from each family, specifically the 1.8 Å structure of halorhodopsin (Hrh, Kolbe et al. 2000), the 2.2 Å structure of the glycerol facilitator (GlpF, Fu et al. 2000), and the 2.2 Å structure ATP/ADP mitochondrial carrier (AMC, Pebay-Peyroula et al. 2003). We also added two simpler systems that had been determined by solution NMR; the well-characterized single-spanning homodimer glycophorin A (GpA) determined in dodecyl phosphocholine (DPC) micelles, (MacKenzie et al. 1997) and a two TM segment fragment from the human glycine receptor (GlyR) determined in trifluoroethanol (Ma et al. 2005). Overall, the database used for our study included four monomeric polytopic membrane protein structures and one homodimeric monotopic structure (Fig. 1).

For each protein in this set, a simulated set of H^N–H^N, methyl–H^N, and methyl–methyl distance restraints were generated, assuming that non-exchangeable protons were only available at Val, Leu and Ile ($\delta 1$) methyl positions. As summarized in Table 1, the theoretical maximum number of NOEs/helical residue that could be obtained for selectively protonated samples is ~3–4, with the largest number of NOEs/residue being obtained for the 7-TM helix protein Hrh (5 NOEs/helical residue). As would be expected due to the relatively high density of amide protons in the α-helical backbone the majority of these were shorter-range intra-helical restraints.

Since intrahelical NOEs only provide information for secondary structure the effect of these NOEs on the accuracy of the overall fold should be indirect and distinct from

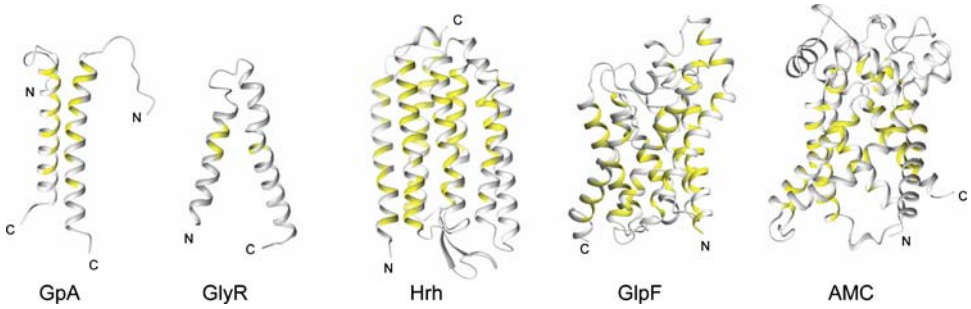

Fig. 1 Structures used to simulate NMR datasets. Locations of amino acids that could give rise to inter-helical NOEs are highlighted in yellow. Note that GlpF is shown in an orientation such that the two half-helices that meet in the membrane are in the central front part of the bundle

Table 1 Membrane proteins and simulated data used

Protein	No. of TM segments	No. of helical residues	No. of simulated NOEs (maximum)	
			Intra-helical	Inter-helical
GpA	1 × 2	54	140	24
GlyR	2	47	199	9
AMC	6	168	641	99
Hrh	7	188	758	195
GlpF	6[a]	181	638	145

[a] This fold also contains two smaller helices with termini that come together in the hydrophobic core of the membrane

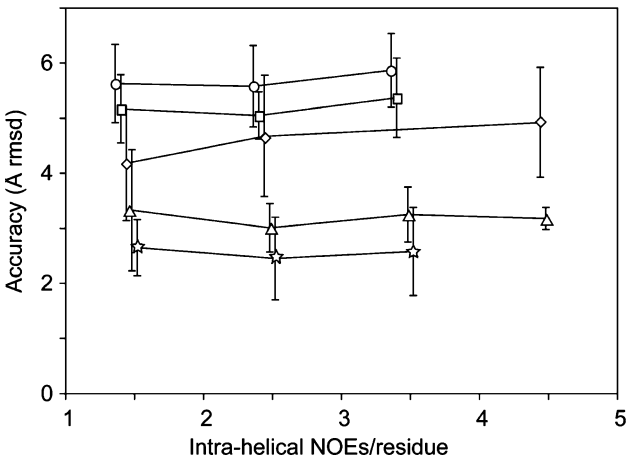

Fig. 2 Investigation of the influence of intra-helical NOEs on structure accuracy. The average accuracy of each ensemble calculated with the indicated number of intra-helical NOEs/residue are shown for GpA (stars), Hrh (triangles), GlyR (diamonds), GlpF (squares) and the AMC (circles). Error bars denote the standard deviation in the rmsds to the target structure

that of inter-helical NOEs. In order to differentiate between the affect of these two classes of NOE restraints on structure accuracy, a series of structures were calculated with the maximum possible number of available inter-helical NOEs and decreasing numbers of intra-helical NOEs for each membrane protein in the sample. The results of these calculations are shown in Fig. 2, which yield average backbone rmsds to the target structure ranging from 2.6 Å (GpA) to 5.9 Å (AMC) when all possible intra-helical NOEs were included in the calculation. In line with previous observations with water-soluble proteins (Gardner et al. 1997), structure accuracy did not correlate with protein size. However, structure accuracy did also not seem to be affected by the number of intra-helical NOEs that were included in the simulated datasets, with ensembles generated with 1.5 intra-helical NOEs/residue being just as accurate as those for structures calculated with the full set of intra-helical NOEs, which constituted a 2- to 3-fold increase in NOE number. These results suggested that intra-helical NOE restraints from selectively methyl protonated samples were not helping to define these membrane protein folds.

In contrast with intra-helical NOEs, inter-helical NOEs for these membrane proteins exclusively provide long-range restraints, and would be expected to significantly influence structure quality. To investigate the magnitude of this effect for selectively methyl protonated membrane proteins, these calculations were performed with a decreasing number of inter-helical NOEs. As shown in Fig. 3 for GpA, Halor, GlpF and AMC, decreasing the number of inter-helical NOEs by 50% increased the rmsd from the target structure by 1–2 Å. In addition, this trend was conserved regardless of the number of intra-helical NOEs used, indicating that the number of long-range restraints is a dominant factor in determining structure quality.

In the case of GlyR there were very few inter-helical NOEs to begin with (9 inter-helical NOEs in total), and the poor accuracy of the ensemble did not further suffer from the removal of 50% of the restraints in this case. As shown in Fig. 1, this was likely due to the non-uniform distribution of inter-helical NOEs for GlyR, which were all concentrated in the central region of the interface between the two TM helices. This poor distribution of inter-helical restraints for GlyR would help account for the relatively low accuracy obtained for this small protein in our simulated structure calculations. This extreme example highlights the important role that NOE distributions play, since the small inter-nuclear distances reported by this restraint require that NOEs be uniformly distributed to

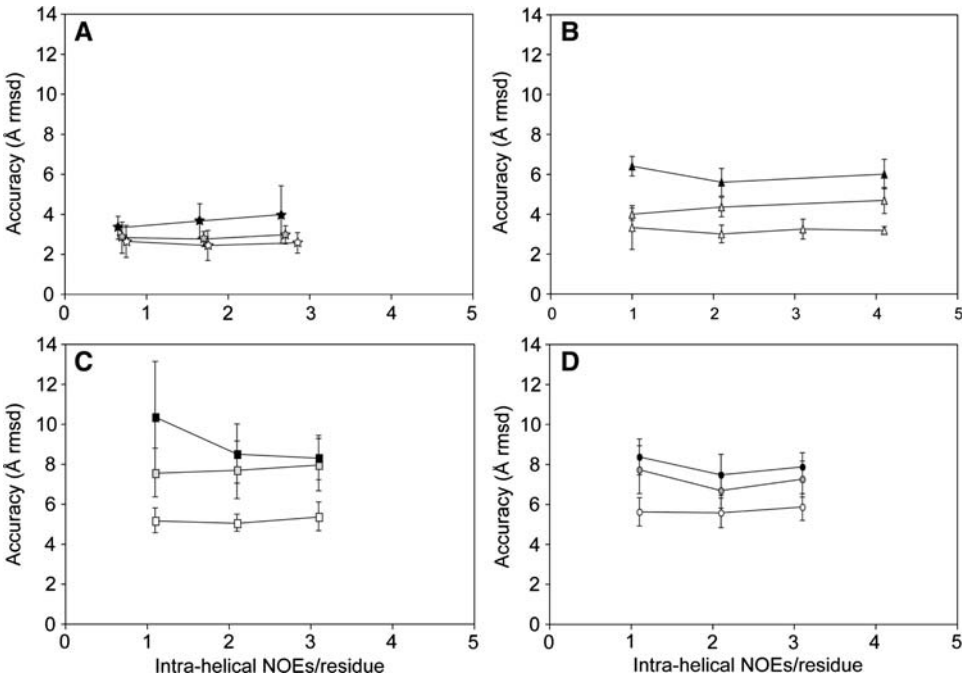

Fig. 3 Inter-helical NOE impact on structure accuracy. Symbols are as indicated in Fig. 2, with the number of inter-helical NOEs being the maximum possible number (open symbols) or a reduced set with 50% (grey) or 25% (black) of the original dataset. Lines are drawn between points to guide the eye

accurately determine the protein fold. As shown in Fig. 4, this was one of the most important factors in determining structure quality for our sample of membrane proteins. For each protein tested, there was a linear relationship between the number of unique inter-helical NOEs (defined as a unique pair of restrained amino acid residues) and the accuracy of each structure. These results demonstrate that the most reliable way to improve structure accuracy for selectively methyl protonated α-helical membrane protein samples is to maximize the number of unique inter-helical NOEs.

Given the apparent insensitivity of the structure accuracy to the number of intra-helical NOEs, we were interested to determine whether the removal of these restraints would produce comparable structures. For this purpose, the previous series of structure calculations with full and reduced sets of inter-helical NOEs were repeated with no intra-helical NOEs. As shown in Fig. 4, for GpA, GlyR, Hrh and AMC the accuracy of structures determined with no intra-helical NOEs were indistinguishable from those determined with the maximum number of intra-helical NOEs. Similar results were also obtained for GlpF when larger numbers of inter-helical restraints were available. In this case there was a significant difference in accuracy in the presence and absence of intra-helical NOEs when only 0.2 inter-helical restraints/residue were available. However, the GlpF structure is unique in our dataset for the presence of two short α-helices that meet in the membrane core to form a 2-segment TM span (Fig. 1). In contrast, for all the other membrane proteins in our sample that did not contain short helices as an integral part of the structural core, the additional information provided by intra-helical NOEs from selectively methyl-labeled proteins did not yield any improvements in accuracy.

The lack of influence of intra-helical NOEs on the structure of these membrane proteins indicated that dihedral angle restraints were dominating the secondary structure quality in these calculations. All structures calculated in these series were determined with canonical helix dihedral angle restraints ($\varphi, \psi = -57°, -47°$) bound by a relatively conservative range of values ($\pm 20°$ and $\pm 30°$ for φ and ψ respectively). Given the importance of the dihedral angle restraint in defining helical secondary structure elements, we also tested narrower dihedral angle restraint boundaries (2.5° and 5.0° for φ and ψ, respectively) using the complete set of inter-helical NOEs. As shown in Fig. 5, all the ensembles that were calculated with narrower dihedral angle restraints showed a statistically significant improvement in structure accuracy relative to those determined with broader dihedral angle restraints. The best improvements were obtained for AMC and GlyR, which saw the rmsd decrease by 0.8 Å in both cases, while Halor and GlpF yielded smaller improvements (0.1–0.2 Å). Underlying this was an increase in accuracies at the single helix level, even though many dihedral angle values in the target structures differed significantly from the range allowed in these calculations (Supplementary Fig. 1). Overall these results indicate that once the helical elements have been identified, more narrowly defined dihedral angle restraints tend to improve the accuracy of calculated structures, even when actual values fall outside the range accepted in the calculations.

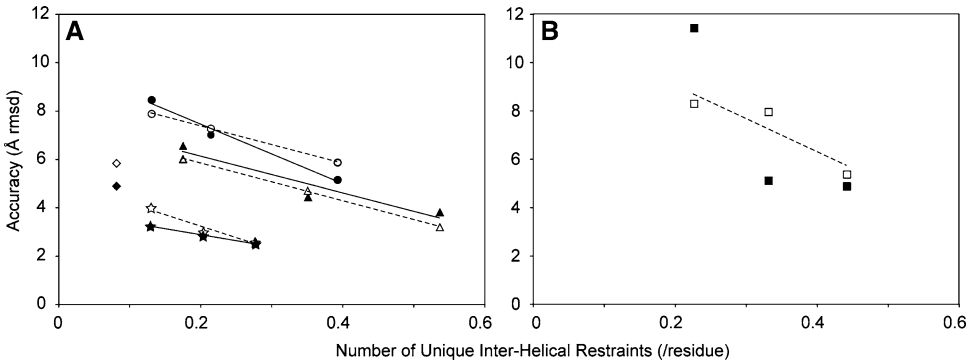

Fig. 4 Structure accuracy versus the number of unique inter-helical restraints. Average accuracies are shown for ensembles calculated either with (*open symbols*) or without (*closed symbols*) intra-helical NOEs as a function of the number of unique pairs of amino acid residues that have one or more inter-helical NOEs holding them together during the structure calculation. A linear correlation ($R^2 > 0.9$) between the ensemble accuracy and the number of unique restraints per helical residue was found both with (*solid line*) and with no (*dashed line*) intra-helical NOE restraints for GpA (*stars*), Hrh (*triangles*) and AMC (*circles*). GlpF showed a weaker correlation ($R^2 = 0.85$), and only when intra-helical restraints were also included

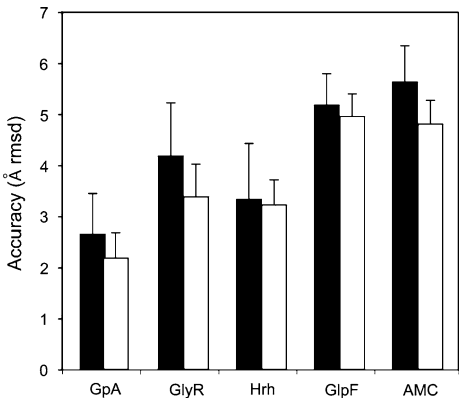

Fig. 5 Impact of dihedral angle restraints on structure accuracy. Average accuracies for ensembles calculated with a complete set of inter-helical NOEs and (φ, ψ) set to ($-57°$, $-47°$) with bounds of ($\pm 20°$, $\pm 30°$) or ($\pm 2.5°$, $\pm 5°$) are shown with black and white bars, respectively. Differences in ensemble accuracy arising from the use of different dihedral angle bounds were found to be statistically significant with $p > 0.01$ for all structures except for Hrh, ($p > 0.05$)

Discussion

While it is generally accepted that a larger number of NOEs give better quality structures, the results from our study demonstrate that the accuracies of these structures are not improved by the intra-helical restraints that could be obtained from selectively methyl protonated samples. This reflects the fact that the intra-helical NOEs were all short or medium-range, with ~90% involving backbone amide protons. Of these, only 15% were between amides separated by more than 2 amino acid residues (i, i + 3) with the remainder being sequential or separated by just one residue. Although the inclusion of these NOEs should reduce the range of potential secondary structures, they do not exclusively define an α-helical conformation, since a variety of helix and turn structures are equally consistent with the same pattern of NOEs (Wuthrich 1986). While this structural ambiguity has the potential to be resolved by NOEs involving protonated methyl groups, these comprised a small part of the datasets and, perhaps more significantly, they were not uniformly distributed along the length of the helix, allowing the formation of distorted helices that nonetheless satisfy available NOE restraints. In light of these factors it is not surprising that inclusion of NOEs involving methyl protons did not improve the accuracy of the helical secondary structure and that dihedral angle restraints were of greater importance in the simulations.

One interesting finding that arose from this study is that the imposition of stronger dihedral angle restraints could improve the accuracy of the overall fold by a small but significant amount, even though the values themselves were not accurate. This parallels a similar observation that arose from a previous study evaluating the ability of a small set of hydrophobicity-based restraints to accurately determine structures for water-soluble proteins (Alexandrescu 2004). Although the application of 10° bounds around canonical dihedral angles often did not include the actual values adopted by the target structures, this approach not only gave rise to more accurate structures, but was also required for convergence. Interestingly, the imposition of rigid dihedral angle restraints was also used in the original structure determination for GpA ($\pm 0°$ bounds with a large force constant) even though helix-defining NOEs were also available for this fully-protonated sample (MacKenzie et al. 1997). Very narrowly defined helical dihedral angle restraints have similarly been used to impose helices for membrane protein structures when only orientational restraints were available (Howell et al. 2005; Lee et al. 2003). Our results provide some validation for this

approach, particularly when few inter-helical restraints are available.

Clearly the utility of dihedral angle restraints hinges on the correct identification of the helical segment in these proteins. This can be facilitated by hydropathy-based sequence analysis, although even the best methods tend to be in error for just over 2 turns of the helix at the termini (Cuthbertson et al. 2005). These methods are also unable to identify irregular structures such as re-entrant loops and the half-helices that are part of the GlpF structure. However, NMR data available from deuterated samples provides a highly reliable identification of helical structures, with characteristic backbone carbon secondary chemical shifts and strong NOEs between sequential amide residues being easily detected in helical structures. Other indications of helical structure can also be provided by the presence of slowly exchanging amide protons, and characteristic periodicities of residual dipolar couplings as seen in dipolar waves (Mesleh et al. 2002). Overall the combination of this data should allow accurate delineation of helical segments in membrane proteins, as has been demonstrated for the potassium channel (Chill et al. 2006). In addition, more accurate dihedral angles values might be accessible through the use of structure-based databases of secondary chemical shifts, as is done in TALOS (Cornilescu et al. 1999). However, TALOS in particular has a small but finite possibility for making incorrect dihedral angle assignments ($\sim$2–7%), and gives predictions with ranges that are substantially larger than the narrow range used in our study. For these reasons, it is not clear that further improvements in structure accuracy will be attainable, beyond those obtained with general helical dihedral angle values.

In line with expectations, one of the most important factors determining structure quality for these membrane protein structures was the inter-helical NOE since it is an inherently long-range restraint. While the correlation between the number of inter-helical NOEs and structure accuracy was weak, it was strongly apparent when considering the number of unique pairs of residues restrained by an inter-helical NOE. This highlights the importance of obtaining a diverse sample of long-range restraints for accurate structure determination. Variation in the distributions of inter-helical restraints also provides some explanation for the range of accuracies obtained for the different membrane protein folds in our dataset. In particular, less accurately defined structures contained a number of regions that would not give rise to inter-helical NOE restraints in a selectively methyl protonated sample. For example, at least 13–15% of residues in both AMC and GlpF would be more than 5 Å away from residues involved in inter-helical NOEs (Fig. 1). In contrast, the inter-helical interface of GpA has the potential to give rise to NOEs along the entire length of the interface. Hrh showed a similarly uniform distribution of inter-helical NOEs with only one restraint-poor region that was mainly isolated to a peripheral part of the structure comprised of regular helical structures.

The presence of short helices in the core of the structure also appeared to affect accuracy by increasing the potential for restraint-poor secondary structure elements to be present in the calculation. This turned out to be the case for GlpF, since the non-uniform distribution of inter-helical NOEs in reduced datasets led to lower-than-average restraint densities involving these two helices (e.g. 0.1 or 0.2 inter-helical NOEs per residue for these helices versus the average numbers of 0.2 or 0.3, respectively). In reduced datasets for TM helices of more typical lengths ($\sim$20 amino acids), the average number of inter-helical NOEs for each helix typically remained close to the average number overall. Altogether these factors indicate that the accuracy of structure that can be obtained using the selective methyl protonation approach depends on the distribution of Leu/Val/Ile residues and hence ultimately on the sequence and overall fold.

At the outset of this project we were interested in determining whether helical membrane proteins would be more amenable to the selective methyl protonation approach compared to water soluble proteins, particularly since membrane proteins tend to be enriched in the amino acids targeted in this labeling scheme. There are a handful of examples where this method has been applied to all-α water-soluble proteins that can be compared to the results from the current study. These are summarized in Table 2, and generally show a comparable range of accuracies to those obtained for our set of membrane proteins. An explanation for this similarity in accuracy levels is provided by the amino acid composition of polytopic helical membrane proteins. Although it is generally true that helical membrane proteins contain a larger proportion of Leu/Val/Ile residues than water-soluble proteins overall, this difference diminishes when the composition of residues at conserved positions alone is considered (Liu et al. 2002). As shown in Table 3, the proportion of Leu/Val/Ile residues at positionally conserved sites for polytopic membrane protein families (20%) is comparable to their levels in water-soluble proteins (21%). Since conserved residues have a greater tendency to be engaged in structurally important helix–helix packing interactions, the additional Leu/Val/Ile residues found in polytopic helical membrane proteins are more likely to be in lipid-facing locations that would be poor sources of structural information. Moreover, conserved ligand-binding or protein–protein interaction sites can also arise in lipid-facing TM-segment residues, a factor that could lead to even lower proportions of conserved Leu/Val/Ile residues in the protein interior. For the proteins in our dataset there

Table 2 Accuracies of all α-helix water-soluble versus membrane protein structures determined by NMR using selectively methyl protonated samples

Protein[a]	Accuracy (rmsd in Å)[b]	NOEs/helical residue[c]
GpA	2.58 ± 0.80	3.0
GlyR	4.92 ± 0.99	4.4
AMC	5.87 ± 0.67	4.4
Hrh	3.18 ± 0.20	5.1
GlpF	5.37 ± 0.72	4.3
$Stat4_{NT}$ (100)	2.90	4.2
Bcl-xL (96)	3.87 ± 0.42	4.6
Dbl (142)	5.83 ± 0.75	6.4
FPP (228)	7.54 ± 2.18	5.5

[a] The number of helical residues in the protein is shown in brackets for water-soluble proteins. $Stat4_{NT}$ is the N-terminal domain from Stat-4 (PDB accession number of reference structure 1BGF, Vinkemeier et al. 1998); Bcl-xL, anti-apoptotic protein (1LXL, Muchmore et al. 1996); Dbl, the guanine nucleotide exchange factor Dbl homology domain (1BY1, Aghazadeh et al. 1998); FPP, the farnesyl diphosphate synthase (1FPS, Tarshis et al. 1994)

[b] For the membrane proteins from the current study accuracies are reported for ensembles calculated using all simulated NOE data and loose dihedral angle bounds. $Stat4_{NT}$ (Gaponenko et al. 2004) and Bcl-xL (Medek et al. 2000) structures were calculated using actual NOE data taken from selectively methyl protonated samples, with some NOEs from the β-protons of Val and γ-protons of Ile and Leu for Bcl-xL. Dbl (Medek et al. 2000) and FPP (Gardner et al. 1997) structures were determined using simulated datasets. For water-soluble structures dihedral angle restraints were derived from TALOS except for FPP which used $-70° \pm 50°$ and $-50° \pm 50°$ for φ and ψ respectively along with hydrogen bond distance restraints

[c] Maximum number of NOEs/helical residue simulated from the 3D structure using the protocol described in Materials and Methods

Table 3 Percent composition of amino acids targeted for methyl protonation

Amino acid	Water-soluble proteins[a]	TM helices[b]	Positionally conserved sites in TM helices[b]
Leu	9	16	12
Val	7	8	4
Ile	5	10	4
Total	21	34	20

[a] (McCaldon and Argos 1988)

[b] Analysis limited to families of α-helical membrane proteins containing two or more TM helices taken from Pfam that have more than 20 members. (Liu et al. 2002)

appears to be a comparable number of structurally informative Leu/Val/Ile residues to those found in the water-soluble α-helical proteins in Table 2, since the number of simulated NOEs/helical residue obtained using the same protocol are similar between the two classes of protein.

Given the importance of obtaining a well-distributed set of unique inter-helical distance restraints, structures determined from the methyl protonation approach could be considered as a starting point for the identification of restraint-poor regions as targets for further structural refinement. Of particular utility would be a method that is being increasingly utilized to refine structures of membrane proteins that uses paramagnetic relaxation enhancement (PRE) of proton resonances by covalently attached spin labels (Liang et al. 2006; Teriete et al. 2007). In fact, it has been shown for the β-barrel membrane protein OmpA that PRE-derived distance restraints alone (with dihedral angle restraints) can be used to determine structures that are comparable in quality to those obtained with a limited set of long-range NOEs (~90), although eleven different mutants were required to obtain ~320 PRE restraints in this case (Liang et al. 2006). While it is not known how many different spin-labeled samples would be required to generate comparable structures for helical membrane proteins with no inter-helical NOEs, the identification of structurally innocuous spin-labeling sites with global correlation times similar to those of the core structure may not be straightforward in the absence of tertiary structural information. The global folds provided by a selectively methyl labeled sample would not only help identify these sites, but would also reduce the number of spin-labeled samples required to generate higher-resolution structures and allow a rational selection of spin-labeling sites based on the need for restraints.

One obvious source of structural data not included in our calculations is the residual dipolar coupling, which for membrane proteins can be measured with aligned gels (Cierpicki and Bushweller 2004; Ishii et al. 2001), DNA nanostructures (Douglas et al. 2007) or by the attachment of lanthanide-binding motifs (Ma and Opella 2000). In particular, inclusion of multiple types of dipolar couplings as has been done for the homopentamer phospholamban (Oxenoid and Chou 2005) should substantially improve the structure. This has already been demonstrated in calculations with a simulated sparse NOE dataset with five different types of backbone dipolar couplings for a large all-α water-soluble protein (Choy et al. 2001; Mueller et al. 2000). Extrapolation from these results suggests that potential increases in structure quality for helical membrane proteins could be as large as ~1.5–2.0 Å, particularly for the lower resolution structures.

In spite of the impressive technological advances that have been made in the study of large biomolecular complexes by solution NMR, significant challenges still hinder its application to the study of polytopic helical membrane protein structure. Even when samples are available, sequence degeneracy, the high proportion of helical structure and unfavorable dynamic properties of many detergent–protein assemblies impede spectral assignment (Marassi and Opella 2004). Poor spectral dispersion also

makes the assignment of methyl resonances and their respective NOEs particularly difficult, as has been noted for diacyl glycerol kinase (Sanders and Sonnichsen 2006). Based on these factors, the results of our study should help guide the choice of structure determination strategy to pursue. Specifically, when spectra are of a quality that would allow a uniform distribution of inter-helical NOEs to be assigned, useful global folds can be obtained with the selective methyl protonation approach. However, in cases where spectral quality greatly limits the number of unique inter-helical NOEs that can be assigned, other strategies that do not require assignments to be made beyond the backbone atoms (e.g. RDCs, PREs), will likely prove to be more fruitful avenues of structural investigation to explore.

Acknowledgements This work was supported by a Natural Sciences and Engineering Research Council Discovery grant to NKG.

References

Aghazadeh B, Zhu K, Kubiseski TJ, Liu GA, Pawson T, Zheng Y, Rosen MK (1998) Structure and mutagenesis of the Dbl homology domain. Nat Struct Biol 5:1098–1107

Alexandrescu AT (2004) Strategy for supplementing structure calculations using limited data with hydrophobic distance restraints. Proteins 56:117–129

Bowie JU (2005) Solving the membrane protein folding problem. Nature 438:581–589

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 54:905–921

Chill JH, Louis JM, Miller C, Bax A (2006) NMR study of the tetrameric KcsA potassium channel in detergent micelles. Protein Sci 15:684–698

Choy WY, Tollinger M, Mueller GA, Kay LE (2001) Direct structure refinement of high molecular weight proteins against residual dipolar couplings and carbonyl chemical shift changes upon alignment: an application to maltose binding protein. J Biomol NMR 21:31–40

Cierpicki T, Bushweller JH (2004) Charged gels as orienting media for measurement of residual dipolar couplings in soluble and integral membrane proteins. J Am Chem Soc 126:16259–16266

Columbus L, Lipfert J, Klock H, Millett I, Doniach S, Lesley SA (2006) Expression, purification, and characterization of *Thermotoga maritima* membrane proteins for structure determination. Protein Sci 15:961–975

Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 13:289–302

Cuthbertson JM, Doyle DA, Sansom MSP (2005) Transmembrane helix prediction: a comparative evaluation and analysis. Protein Eng Des Sel 18:295–308

Douglas SM, Chou JJ, Shih WM (2007) DNA-nanotube-induced alignment of membrane proteins for NMR structure determination. Proc Natl Acad Sci USA 104:6644–6648

Eilers M, Shekar SC, Shieh T, Smith SO, Fleming PJ (2000) Internal packing of helical membrane proteins. Proc Natl Acad Sci USA 97:5796–5801

Fernandez C, Hilty C, Wider G, Guntert P, Wuthrich K (2004) NMR structure of the integral membrane protein OmpX. J Mol Biol 336:1211–1221

Fleishman SJ, Ben-Tal N (2006) Progress in structure prediction of alpha-helical membrane proteins. Curr Opin Struct Biol 16: 496–504

Fossi M, Oschkinat H, Nilges M, Ball LJ (2005) Quantitative study of the effects of chemical shift tolerances and rates of SA cooling on structure calculation from automatically assigned NOE data. J Magn Reson 175:92–102

Fu DX, Libson A, Miercke LJW, Weitzman C, Nollert P, Krucinski J, Stroud RM (2000) Structure of a glycerol-conducting channel and the basis for its selectivity. Science 290:481–486

Gaponenko V, Sarma SP, Altieri AS, Horita DA, Li J, Byrd RA (2004) Improving the accuracy of NMR structures of large proteins using pseudocontact shifts as long-range restraints. J Biomol NMR 28:205–212

Gardner KH, Rosen MK, Kay LE (1997) Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR. Biochemistry 36:1389–1401

Gerstein M (1997) A structural census of genomes: comparing bacterial, eukaryotic, and archaeal genomes in terms of protein structure. J Mol Biol 274:562–576

Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE (1999) A robust and cost-effective method for the production of Val, Leu, Ile (delta 1) methyl-protonated N-15-, C-13-, H-2-labeled proteins. J Biomol NMR 13:369–374

Howell SC, Mesleh MF, Opella SJ (2005) NMR structure determination of a membrane protein with two transmembrane helices in micelles: MerF of the bacterial mercury detoxification system. Biochemistry 44:5196–5206

Hu J, Qin H, Li C, Sharma M, Cross TA, Gao FP (2007) Structural biology of transmembrane domains: efficient production and characterization of transmembrane peptides by NMR. Protein Sci 16:2153–2165

Ishii Y, Markus MA, Tycko R (2001) Controlling residual dipolar couplings in high-resolution NMR of proteins by strain induced alignment in a gel. J Biomol NMR 21:141–151

Kolbe M, Besir H, Essen LO, Oesterhelt D (2000) Structure of the light-driven chloride pump halorhodopsin at 1.8 angstrom resolution. Science 288:1390–1396

Lee S, Mesleh MF, Opella SJ (2003) Structure and dynamics of a membrane protein in micelles from three solution NMR experiments. J Biomol NMR 26:327–334

Liang BY, Bushweller JH, Tamm LK (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

Liu Y, Engelman DM, Gerstein M (2002) Genomic analysis of membrane protein families: abundance and conserved motifs. Genome Biol 3:231–236

Ma C, Opella SJ (2000) Lanthanide ions bind specifically to an added "EF-hand" and orient a membrane protein in micelles for solution NMR spectroscopy. J Magn Reson 146:381–384

Ma DJ, Liu ZW, Li L, Tang P, Xu Y (2005) Structure and dynamics of the second and third transmembrane domains of human glycine receptor. Biochemistry 44:8790–8800

MacKenzie KR, Prestegard JH, Engelman DM (1997) A transmembrane helix dimer: structure and implications. Science 276: 131–133

Marassi FM, Opella SJ (2004) Structure determination of membrane proteins by NMR spectroscopy. Chem Rev 104:3587–3606

McCaldon P, Argos P (1988) Oligopeptide biases in protein sequences and their use in predicting protein coding regions in nucleotide-sequences. Proteins 4:99–122

Medek A, Olejniczak ET, Meadows RP, Fesik SW (2000) An approach for high-throughput structure determination of proteins by NMR spectroscopy. J Biomol NMR 18:229–238

Merkley N, Shaw GS (2004) Solution structure of the flexible class II ubiquitin-conjugating enzyme Ubc1 provides insights for poly-ubiquitin chain assembly. J Biol Chem 279:47139–47147

Mesleh MF, Veglia G, DeSilva TM, Marassi FM, Opella SJ (2002) Dipolar waves as NMR maps of protein structure. J Am Chem Soc 124:4206–4207

Muchmore SW, Sattler M, Liang H, Meadows RP, Harlan JE, Yoon HS, Nettesheim D, Chang BS, Thompson CB, Wong SL, Ng SC, Fesik SW (1996) X-ray and NMR structure of human Bcl-x(L), an inhibitor of programmed cell death. Nature 381:335–341

Mueller GA, Choy WY, Yang DW, Forman-Kay JD, Venters RA, Kay LE (2000) Global folds of proteins with low densities of NOEs using residual dipolar couplings: application to the 370-residue maltodextrin-binding protein. J Mol Biol 300:197–212

Oxenoid K, Chou JJ (2005) The structure of phospholamban pentamer reveals a channel-like architecture in membranes. Proc Natl Acad Sci USA 102:10870–10875

Park SH, Mrse AA, Nevzorov AA, Mesleh MF, Oblatt-Montal M, Montal M, Opella SJ (2003) Three-dimensional structure of the channel-forming trans-membrane domain of virus protein "u" (Vpu) from HIV-1. J Mol Biol 333:409–424

Pebay-Peyroula E, Dahout-Gonzalez C, Kahn R, Trezeguet V, Lauquin GJM, Brandolin R (2003) Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside. Nature 426:39–44

Poget SF, Krueger-Koplin ST, Krueger-Koplin RD, Cahill SM, Shekar SC, Girvin ME (2006) NMR assignment of the dimeric S-aureus small multidrug-resistance pump in LPPG micelles. J Biomol NMR 36:10

Poget SF, Cahill SM, Girvin ME (2007) Isotropic bicelles stabilize the functional form of a small multidrug-resistance pump for NMR structural studies. J Am Chem Soc 129:2432–2433

Sanders CR, Sonnichsen F (2006) Solution NMR of membrane proteins: practice and challenges. Magn Reson Chem 44:S24–S40

Stein EG, Rice LM, Brunger AT (1997) Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. J Magn Reson 124:154–164

Tamm LK, Liang BY (2006) NMR of membrane proteins in solution. Prog Nucl Magn Reson Spect 48:201–210

Tarshis LC, Yan MJ, Poulter CD, Sacchettini JC (1994) Crystal-structure of recombinant farnesyl diphosphate synthase at 2.6-Angstrom resolution. Biochemistry 33:10871–10877

Teriete P, Franzin CM, Choi J, Marassi FM (2007) Structure of the Na, K-ATPase regulatory protein FXYD1 in Micelles. Biochemistry 46:6774–6783

Tian CL, Vanoye CG, Kang CB, Welch RC, Kim HJ, George AL, Sanders CR (2007) Preparation, functional characterization, and NMR studies of human KCNE1, a voltage-gated potassium channel accessory subunit associated with deafness and long QT syndrome. Biochemistry 46:11459–11472

Tugarinov V, Kay LE (2004) An isotope labeling strategy for methyl TROSY spectroscopy. J Biomol NMR 28:165–172

Tugarinov V, Hwang PM, Kay LE (2004) Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins. Annu Rev Biochem 73:107–146

Tugarinov V, Choy WY, Orekhov VY, Kay LE (2005) Solution NMR-derived global fold of a monomeric 82-kDa enzyme. Proc Natl Acad Sci USA 102:622–627

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

Venters RA, Metzler WJ, Spicer LD, Mueller L, Farmer BT (1995) Use of H-1(N)-H-1(N) NOES to determine protein global folds in perdeuterated proteins. J Am Chem Soc 117:9592–9593

Vinkemeier U, Moarefi I, Darnell JE, Kuriyan J (1998) Structure of the amino-terminal protein interaction domain of STAT-4. Science 279:1048–1052

White SH (2004) The progress of membrane protein structure determination. Protein Sci 13:1948–1949

Wuthrich K (1986) NMR of proteins and nucleic acids. Wiley, USA