# AssignFit: A program for simultaneous assignment and structure refinement from solid-state NMR spectra

Ye Tian [a,b], Charles D. Schwieters [c], Stanley J. Opella [b], Francesca M. Marassi [a,*]

[a] Sanford Burnham Medical Research Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA
[b] Department of Chemistry and Biochemistry, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0307, USA
[c] Division of Computational Bioscience, Building 12A, Center for Information Technology, National Institutes of Health, Bethesda, MD 20892-5624, USA

## ARTICLE INFO

## ABSTRACT

AssignFit is a computer program developed within the XPLOR-NIH package for the assignment of dipolar coupling (DC) and chemical shift anisotropy (CSA) restraints derived from the solid-state NMR spectra of protein samples with uniaxial order. The method is based on minimizing the difference between experimentally observed solid-state NMR spectra and the frequencies back calculated from a structural model. Starting with a structural model and a set of DC and CSA restraints grouped only by amino acid type, as would be obtained by selective isotopic labeling, AssignFit generates all of the possible assignment permutations and calculates the corresponding atomic coordinates oriented in the alignment frame, together with the associated set of NMR frequencies, which are then compared with the experimental data for best fit. Incorporation of AssignFit in a simulated annealing refinement cycle provides an approach for simultaneous assignment and structure refinement (SASR) of proteins from solid-state NMR orientation restraints. The methods are demonstrated with data from two integral membrane proteins, one α-helical and one β-barrel, embedded in phospholipid bilayer membranes.

## 1. Introduction

Methods for NMR structure determination typically rely on obtaining resonance assignments by establishing correlations between neighboring atoms, followed by measuring a series of restraints (e.g. distances, orientations) for each assigned site, that are used in structure determination by simulated annealing. In recent years, methods for simultaneous assignment and structure refinement (SASR) have been developed for both solution NMR and solid-state NMR.

For solution NMR, several such methods rely on backbone residual dipolar couplings (RDCs) measured from weakly aligned samples, in combination with backbone chemical shifts, to define and connect structured fragments of a protein in a sequence-specific manner [1], to obtain backbone resonance assignments from a known protein structure [2], or to determine the three-dimensional arrangement of protein–protein complexes from the pre-determined structures of the individual components [3,4]. Alternatively, it is possible to generate low-resolution structures of globular proteins by fitting unassigned NMR data (e.g. chemical shifts, NOEs, RDCs) to computationally predicted structural models using a Monte Carlo procedure [5]. Finally, methods have been developed to compute realistic spatial proton distributions for proteins in solution, solely from experimental NOE data with minimal assignments [6–8].

For solid-state NMR, the direct correlation between protein structure and the orientation-dependent dipolar coupling (DC) and chemical shift anisotropy (CSA) frequencies, measured in samples with uniaxial order [9–11], provides a method for SASR based on minimizing the difference between the experimentally observed spectral frequencies and the frequencies back-calculated from a structural model. Because such solid-state NMR spectra display full, or near full, magnitudes of the DC and CSA, the order tensor is known a priori, and their interpretation is significantly facilitated. The SASR approach relieves the burden of having to obtain near complete resonance assignments prior to structure determination: resonance assignments are obtained as a side product of fitting a structural model to the NMR data, but is not a prerequisite for structure determination.

Uniaxial order can be achieved by either inducing sample alignment relative to the magnetic field ($B_o$), as in oriented sample (OS) solid-state NMR [12,13], or by exploiting the inherent uniaxial rotation of a protein relative to an internal principal axis in a non-aligned sample (e.g. [14–16]). Since the direction of order is fixed by the sample geometry, the resulting NMR frequencies provide not only precise internal restraints for structure determination, but also relative restraints that enable the structure to be positioned in the context of the alignment medium. This is

particularly useful for membrane proteins in lipid bilayers where structure determination also yields the three-dimensional position of the protein within the membrane [17–21]. For membrane proteins embedded in phospholipid bilayer membranes, the direction of order is determined by the membrane preparation, which can consist of planar lipid bilayers supported on glass or aligned magnetically, or of spherical vesicles where the protein undergoes rotational diffusion around the lipid bilayer normal (n).

In the first applications of SASR to α-helical membrane proteins, the $^1H/^{15}N$ separated local field (SLF) spectra obtained from combinations of selectively $^{15}N$-labeled (by residue type) and uniformly $^{15}N$-labeled (all residues) proteins were assigned by comparison with the spectra calculated from a structural model, and the assigned experimental frequencies were then used to either directly calculate backbone dihedral angles [18], or as orientation restraints in a simulated annealing protocol [22], to obtain a final membrane-oriented structure consistent with the data. Alternatively, an algorithm has been described to build structural models from random assignments of SLF data and comparison of the data back-calculated from each structural model with the experimental data [23]. Furthermore, a method based on graph theory has been developed to simultaneously obtain structure and assignment of $^1H/^{15}N$ SLF spectra [24]. These two approaches were developed specifically for α-helical proteins, although they should also be applicable to other regular secondary structures.

In a recent application of SASR to a β-barrel outer membrane protein, the $^1H/^{15}N$ SLF spectrum of $^{15}N$-Phe-labeled OmpX in magnetically oriented lipid bilayers was assigned through an iterative approach where each of the possible peak assignment combinations was tested for its ability to provide $^1H/^{15}N$ DC and $^{15}N$ CSA orientation restraints, consistent with the proper spatial orientation of the crystal structure within the membrane and with its associated back-calculated spectrum [25]. Although powerful, this type of analysis can quickly evolve into a complicated problem when the number of assignment permutations to be tested is very large, since for $n$ number of peaks there are $n!$ assignment permutations. For example, there are 5040 (7!) ways to assign the 7 Phe peaks in the SLF spectrum of selectively $^{15}N$-Phe-labeled OmpX and, while the task can be alleviated by further subdividing the spectrum into separate sets of peaks according to their H/D exchange [25], or other properties, such simplifications are not always possible.

Here we present a computer program, AssignFit, developed within the XPLOR-NIH package [26], that greatly facilitates the SASR process. Unlike the first applications of SASR to α-helical [18] or β-barrel [25] membrane proteins, where the potential assignment permutations were generated by hand and analyzed with the aid of home-developed FORTRAN code, AssignFit generates all permutations computationally and tests them for best fit to the data.

## 2. Results and discussion

### 2.1. AssignFit and the SASR protocol

The SASR calculation cycle starts with an initial structural model and a set of unassigned DCs and CSAs for a particular residue type (e.g. from a selectively labeled sample). Each SASR cycle consists of generating optimal residue-specific assignments for the input data, and then using the assigned DC and CSA restraints to refine the structural model, which provides the input for the next SASR cycle where a new set of DCs and CSAs are assigned. The cycles are continued until all DC and CSA restraints are assigned and the resulting structure is consistent with the entire data set (Fig. 1A).

AssignFit is an integral part of the SASR process. Starting with a structural model and a set of DC and CSA data, AssignFit generates
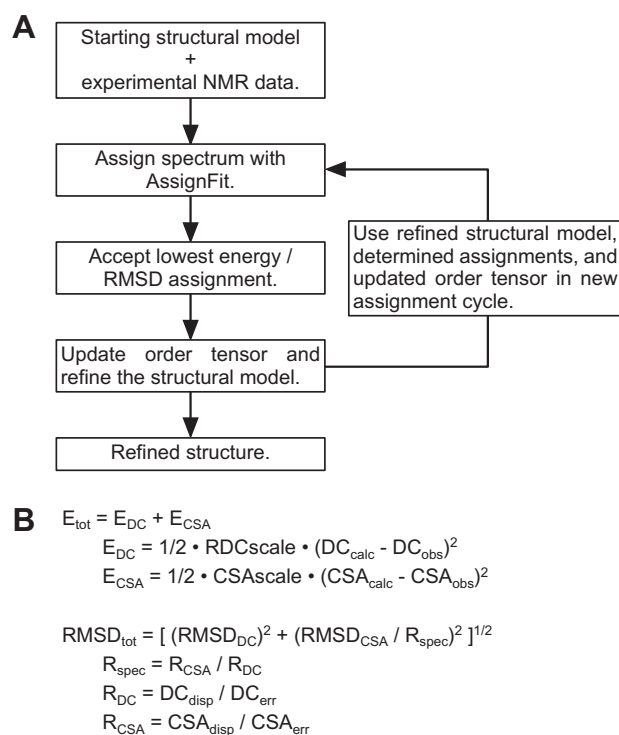


**A**

Starting structural model + experimental NMR data.

Assign spectrum with AssignFit.

Accept lowest energy / RMSD assignment.

Update order tensor and refine the structural model.

Refined structure.

Use refined structural model, determined assignments, and updated order tensor in new assignment cycle.

**B**
$$E_{tot} = E_{DC} + E_{CSA}$$
$$E_{DC} = 1/2 \cdot RDCscale \cdot (DC_{calc} - DC_{obs})^2$$
$$E_{CSA} = 1/2 \cdot CSAscale \cdot (CSA_{calc} - CSA_{obs})^2$$

$$RMSD_{tot} = [ (RMSD_{DC})^2 + (RMSD_{CSA} / R_{spec})^2 ]^{1/2}$$
$$R_{spec} = R_{CSA} / R_{DC}$$
$$R_{DC} = DC_{disp} / DC_{err}$$
$$R_{CSA} = CSA_{disp} / CSA_{err}$$

**Fig. 1.** SASR protocol. (A) Flowchart representation of SASR using AssignFit. (B) Energy and RMSD functions calculated by AssignFit to evaluate the results.

all possible assignment permutations and calculates the corresponding molecular alignment, the atomic coordinates reoriented in the alignment frame, and the associated set of NMR frequencies, which are then compared with the experimental data for best fit. For each possible assignment the optimal orientation of the input structural model is determined using conjugate gradient minimization of a weighted sum of energies ($E_{tot}$) corresponding to DC and CSA terms (Fig. 1B). The assignments are sorted by the weighted sum of RMSDs for both DC and CSA ($RMSD_{tot}$), and the lowest RMSD assignments are saved for further analysis.

While the orientation determination is quite fast, the number of possible assignments grows combinatorially such that exhaustive calculation of these orientations becomes impractical when there are more than about 11 residues of a particular type to be assigned. In this case, it may be possible to split the data points into smaller subsets grouped by other known properties, such as $^2H$ exchange or dynamics. Moreover, SLF spectra provide a natural separation of frequencies according to the orientation of their corresponding structural domain (e.g. transmembrane versus membrane surface helices), thus separating the data into smaller subgroups that can be handled more readily by the program. For example, the cross-polarization matching condition in SLF experiments can be adjusted to selectively couple only those residues in transmembrane helices. Finally, graph-theoretic approaches, such as those used in [8], completely avoid the combinatorial growth problem, and are straightforward to implement in our current protocol when required by large system size.

AssignFit uses the modules for variable tensor (varTensorTools), RDC potential (rdcPotTools), and CSA potential (csaPotTools) available in XPLOR-NIH [26,27]. The relevant parameters and associated energy ($E_{tot}$) and RMSD ($RMSD_{tot}$) functions are described below and in Fig. 1B.

For solid-state NMR studies, the magnitude and orientation of the alignment tensor are known, or can be closely estimated, from the sample geometry. In this case, AssignFit can be performed by

specifying and fixing the values of the axial alignment parameter ($D_a$) and rhombicity (Rh), so that only the orientation of the input structure relative to the principal alignment axis (e.g. the lipid bilayer normal) is varied. Alternatively, AssignFit could be applied to cases where the tensor is not known (e.g. data from weakly aligned samples) by allowing the values of $D_a$ and Rh to vary. Selecting this option results in AssignFit performing singular value decomposition for each set of assignment permutation to calculate

the optimum values of $D_a$ and Rh, together with molecular orientation. The application of this approach for solution NMR alignment tensor determination, based on RDCs assigned solely by residue type and a known protein structure, has been demonstrated previously for the structural analysis of protein–protein complexes [28]. In that case a pre-refined structure was used to determine the values of $D_a$ and Rh, which were then used to position the protein in the context of the complex.
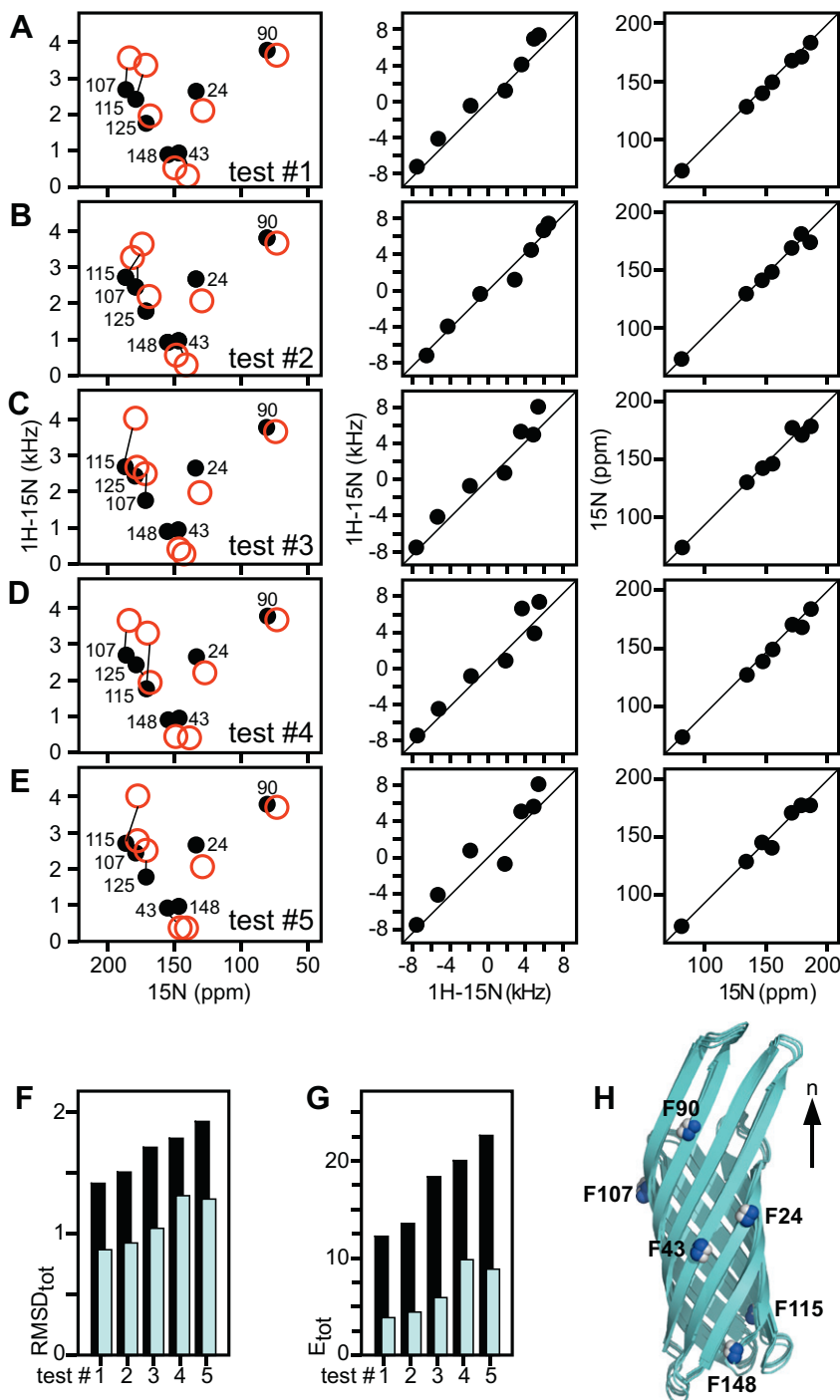


**Fig. 2.** Results for the five lowest energy assignments obtained with XPLOR-NIH AssignFit for the $^1H/^{15}N$ SLF spectrum of $^{15}N$-Phe-labeled OmpX in magnetically aligned bilayers ($n||B_o$). (A)–(E) Five lowest energy assignments showing observed (black) and back-calculated (red) peaks. For each test assignment (#1–5), the corresponding correlations between observed and back-calculated $^1H-^{15}N$ DC (middle) and $^{15}N$ CSA (right) reflect the resulting RMSD (F) and energy values (G). (F) and (G) Total RMSD (F) and energy values (G) obtained for each of the five lowest energy assignments before (black) or after (cyan) structural refinement. (H) Overlay of the OmpX membrane orientations consistent with each of the five lowest energy test assignments obtained from AssignFit. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Both DC and CSA need to be signed $(+/-)$ values in the input files of AssignFit. The CSA signs are obtained experimentally and, while the SLF experiment does not provide direct information about the DC signs, the latter can often be determined from the peak positions in the spectrum [29]. In cases where this is not possible, AssignFit enables peaks with ambiguous signs to be specified and taken into account during the calculation by testing different sign permutations.

The value of $E_{tot}$ reflects the combined difference between the observed and back-calculated DC and CSA, each scaled by the corresponding user-defined force constants (RDCscale, CSAscale). The value of $RMSD_{tot}$ reflects the combined RMSD for DC and CSA, scaled by the relative effective spectral resolution ($R_{spec}$) available in the DC and CSA dimensions (Fig. 1B). $R_{spec}$ is a user-defined parameter that reflects the ratio of the total spectral range available to the resonance line width. Its value can be determined from the observed spectral ranges and associated experimental line widths or errors for the DC and CSA data. The value of $RMSD_{tot}$, scaled in this way, provides an effective parameter for selecting the AssignFit result with best fit to the experimental spectrum. For the spectra of OmpX and fd coat protein examined in this study values of $R_{spec} = 7.5$ and $R_{spec} = 4.0$ were used, respectively; they each reflect experimental CSA spectral ranges of 150 ppm (OmpX) or 80 ppm (fd) with a CSA error of 1 ppm, and DC spectral ranges of 5000 Hz (OmpX, fd) with a DC error of 250 Hz.

## 2.2. AssignFit assignment of the $^1H/^{15}N$ SLF spectrum of the OmpX β-barrel in oriented lipid bilayers

We first used AssignFit to assign the $^1H/^{15}N$ SLF spectrum of $^{15}N$-Phe labeled *Escherichia coli* OmpX in magnetically aligned phospholipid bilayers (n||$B_o$). OmpX forms an eight-stranded transmembrane β-barrel in the outer membrane of *E. coli*. Its structure has been determined by both X-ray crystallography [30] and solution NMR [31] in detergents, and we have determined its orientation in phospholipid membranes using solid-state NMR [25].

Using AssignFit, starting from the crystal structure of OmpX, all 5040 test assignments of the 7 Phe peaks were obtained in 634 s of CPU time (Hp-390t Linux computer, Intel Core i7-970 3.20 GHz six-core processor) without the need to separate the data into two sets according to their H/D exchange properties, as was done previously [25]. The five lowest energy assignments (Fig. 2A–E) all yield calculated spectra that are visually very similar to the experimental spectrum, good correlations between observed and calculated $^{15}N$ CSA and $^1H/^{15}N$ DC frequencies, and very similar molecular orientations of OmpX within the membrane (Fig. 2H). One test assignment with the best correlations (test #1; Fig. 2A) has the lowest values of both $RMSD_{tot}$ (Fig. 2F) and $E_{tot}$ (Fig. 2G). It is identical to our previously determined assignment [25] and yields an identical transmembrane orientation of the β-barrel.

The next four lowest energy results swap the assignments among F107, F115 and F125 (test #2–4; Fig. 2B–D) or between F43 and F148 (test #4–5; Fig. 2D–E), and thus produce poorer correlations between experimental and calculated data, reflected in higher values of $RMSD_{tot}$ and $E_{tot}$. The peaks for F107, F115 and F125 are clustered within a 15 ppm by 2 kHz spectral region, and their corresponding NH bonds have relatively similar orientations. Therefore, it is not surprising that swapping their assignments yields similar molecular orientations. Similarly, the peaks from F43 and F148 are very close in a narrow spectral window (8 ppm by 80 Hz) and their assignments can be swapped with little consequence. In contrast, the peaks from F90 and F24 occupy singular positions in the spectrum, and their assignments are constant in all of the lowest $RMSD_{tot}$ and $E_{tot}$ AssignFit determinations.

Since lipid bilayers magnetically aligned with n||$B_o$ have net axial order parameters in the range of $S_{zz} = 0.9$–0.8, a value of $D_a = 8.5$
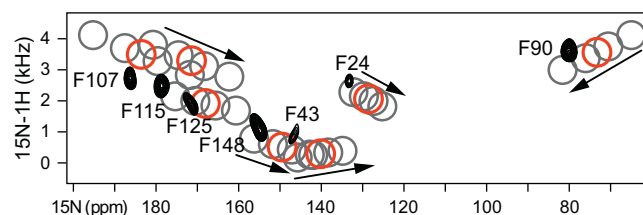


**Fig. 3.** Effect of varying the order parameter $S_{zz}$ on the fit between experimental (black) and AssignFit-calculated (circles) SLF spectra of selectively $^{15}N$-Phe labeled OmpX in magnetically aligned bilayers (n||$B_o$). For each peak, the arrows show the frequency shift observed from $D_a = 10$–7 kHz (including $D_a = 10, 9, 8.5, 8, 7$ kHz). The best fit to the experimental spectrum is obtained for $D_a = 8.5$ kHz (red circles). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

was used to generate the AssignFit results in Fig. 2. Varying $D_a$ from 10 to 7 kHz does not affect the AssignFit result but produces calculated SLF spectra where the DC and CSA frequencies scale as $S_{zz}$ and the SLF peaks move progressively towards the isotropic frequencies with decreasing $D_a$ (Fig. 3). The best fit to the experimental spectrum is observed for $D_a = 8.5$ kHz, the value which also yields the lowest AssignFit values of $RMSD_{tot}$ and $E_{tot}$, and which was found to give the best agreement between experimental and calculated frequencies in the previous study [25].

## 2.3. AssignFit and SASR of the α-helical fd bacteriophage coat protein in oriented lipid bilayers

We next used AssignFit to assign the $^1H/^{15}N$ SLF spectrum of $^{15}N$-selectively labeled fd coat protein in glass-aligned phospholipid bilayers (n||$B_o$). The membrane-bound form of the major pVIII coat protein of filamentous fd bacteriophage resides in the inner membrane of infected *E. coli* before incorporation into virus particles that are extruded through the bacterial cell membrane. The structure of the membrane-bound form has been extensively studied in micelles and lipid bilayers using NMR (reviewed in [18]). The protein has two distinct α-helices: a short amphipathic helix that associates with the membrane surface and a longer hydrophobic helix that traverses the membrane.

Previously [18], we showed that the $^1H/^{15}N$ SLF spectrum of fd could be assigned by comparison with the spectra back-calculated from ideal helices oriented in the magnetic field by relying on the characteristic wheel-like patterns (Pisa wheels) observed in the spectra of helical proteins [9–11]. In this first example, each potential assignment permutation was generated by hand and analyzed with the aid of home-developed FORTRAN code. The best assignments were selected based on the ability of the corresponding DC and CSA frequencies to yield the most favorable helical backbone dihedral angles for connected residues. In contrast, AssignFit generates all permutations computationally and automatically tests them for best fit to the experimental data and a reasonable structural model.

Using AssignFit in combination with an ideal helix starting model, the SLF peaks corresponding to the four Val, three Ala, and three Gly residues in the transmembrane helix of the fd coat protein could be assigned quickly, and the accuracy of the assignments could be evaluated quantitatively (Fig. 4). Furthermore, using AssignFit as part of an SASR cycle where the structural model is refined after assigning each set of peaks, increases the assignment confidence by ensuring that each assigned data set is consistent with the next, to produce a final refined structure consistent with the data.

To start the SASR cycle we performed AssignFit for the four Val peaks, using an ideal helix with uniform backbone dihedral angles ($\phi = -60$, $\psi = -45$) for the 50-residue fd coat protein (Fig. 5A). Since there is a single Leu (L41) in the transmembrane helix of the coat protein, assignment of its peak was held fixed in all
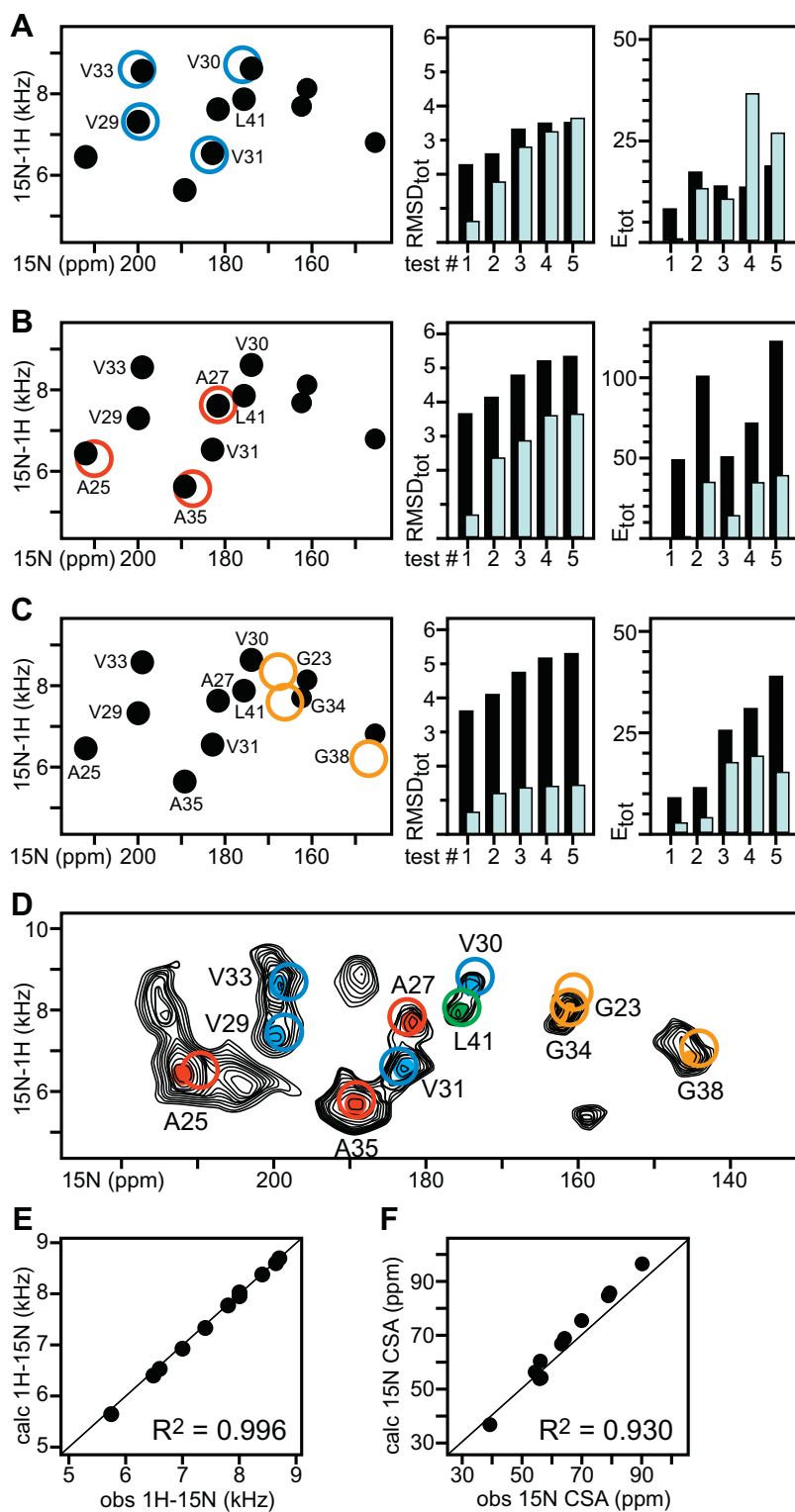
**Fig. 4.** SASR with the $^1H/^{15}N$ SLF spectra of $^{15}N$-Val, Ala, Leu, and Gly labeled fd coat protein (transmembrane domain) in glass-aligned lipid bilayers (n$\|B_o$). In all spectra, filled circles represent experimentally observed peaks. Unfilled circles represent frequencies back-calculated after structural model refinement with the assigned restraints. The peaks are color-coded by amino acid type: Val (blue), Ala (red), Leu (green), and Gly (gold). The spectral region corresponds to the protein's transmembrane domain. AssignFit was implemented with $D_a$ = 10 kHz, $k_{DC}$ = 1 kcal mol$^{-1}$ kHz$^{-2}$, $k_{CSA}$ = 0.01 kcal mol$^{-1}$ ppm$^{-2}$. (A)–(C) Residue specific assignments of the Val, Ala, and Gly peaks were obtained using AssignFit, after fixing the L41 assignment, with: (A) an ideal helix structural model; (B) the structural model refined with Val DC and CSA restraints; (C) the structural model refined with Val, and Ala DC and CSA restrains. For each assignment/refinement cycle, the total RMSD and energy of the five lowest RMSD AssignFit results, obtained before (black) and after (cyan) model refinement, are shown on the right. (D) Experimental SLF spectrum from uniformly $^{15}N$-labeled fd (black) showing peaks from Ala (red), Val (blue), Gly (gold) and Leu (green), assigned with AssignFit. (E) and (F) Correlations between values of the $^1H$–$^{15}N$ DC and $^{15}N$ CSA observed experimentally (obs) and back-calculated after refinement of the starting structural model (calc). The $R^2$ correlation coefficients are listed for each correlation graph. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
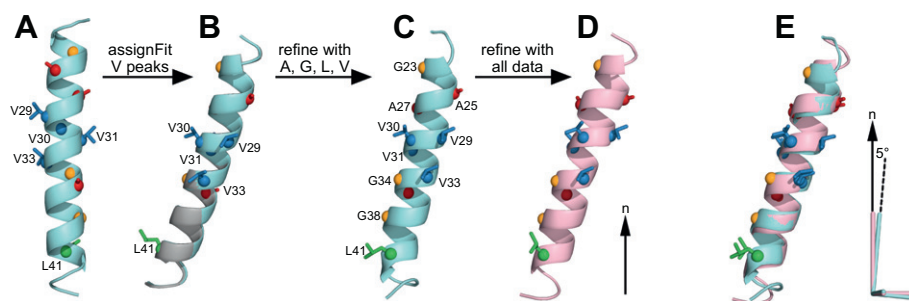
**Fig. 5.** Structural model refinement of the transmembrane domain for the membrane-bound fd-coat protein. Coordinates are oriented in the frame of the lipid bilayer, relative to the lipid bilayer normal (n; arrow). (A) Ideal helix structural model (arbitrary orientation) used to initiate the SASR cycle. (B) Rigid body orientations of the starting structural model obtained for the lowest (cyan) and second lowest (gray) RMSD$_{tot}$ AssignFit assignment of the Val peaks. Swapping the assignments of V29 and V33 has little effect on model orientation. (C) Structure obtained after refinement with the DC and CSA restraints from Leu, Val, Ala and Gly. (D) Structure obtained after refinement with the DC and CSA restraints from all of the previously measured DC and CSA restraints [18]. (E) The AGLV-refined structure (cyan) is aligned to the all-data-refined structure (pink). The cyan and pink axes represent the order tensors for the respective structures. The pink principal axis coincides with the lipid bilayer normal, while the cyan principal axis is off by 5°. Alignment was obtained for the backbone Cα atoms. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

AssignFit calculations. The lowest RMSD$_{tot}$ and E$_{tot}$ assignment is shown in Fig. 4A (test #1). This result generates a rigid body orientation of the starting model (Fig. 5B) similar to the membrane orientation of fd in the previously determined structure (Fig. 5D; [18]). The next best assignment of the Val peaks (test #2) swaps the labels of V29 and V33. These residues occupy similar positions along the α-helix (Fig. 5) and their peaks have overlapped $^{15}N$ chemical shifts. However, swapping their assignments results in significantly higher values of RMSD$_{tot}$ and E$_{tot}$, enabling the Assign-Fit results to be differentiated. Refinement of the starting model using the Leu41 and Val DC and CSA restrains assigned in test #1, followed by another AssignFit search for the Val assignments, yields a calculated spectrum with excellent fit to the experimental data (Fig. 4A), and corresponding values of RMSD$_{tot}$ and E$_{tot}$ that are much lower and much more differentiated from the other test assignments than those obtained before refinement.

The Leu/Val-refined model was subsequently used in a second SASR cycle where AssignFit was used to assign the Ala peaks while the assignments for L41 and Val were held fixed, and the assigned Ala restrains were used together with those for Leu and Val in a second refinement step. Finally, the resulting Leu/Val/Ala-refined model was used in a third SASR cycle, where AssignFit was used to assign the Gly peaks, and the assigned Gly restrains were used together with those for Leu, Val, and Ala in a third and final refinement step. Note that the peaks for G23 and G34 are overlapped, and their assignment can be swapped with little consequence for model refinement or spectrum back-calculation, as evidenced by the similar values of RMSD$_{tot}$ and E$_{tot}$ observed for test assignments #1 and #2 where this occurs (Fig. 4C).

The resulting structure of the fd transmembrane domain (residues 19–46) determined by SASR using only Leu, Val, Ala and Gly restraints compares very favorably (average backbone RMSD = 1.0 Å) with the structure determined using DC and CSA restraints for all residues [18]. Notably, the three-dimensional membrane orientation, which is an integral part of this structure determination method, is also reproduced very well (axially within 5°; Fig. 5E). Finally, the spectrum that is back-calculated during the last refinement step correlates very well with the experimentally measured spectrum of $^{15}N$-uniformly labeled fd coat protein (Fig. 4D), with excellent correlation coefficients (R$^2$) for both $^1H$–$^{15}N$ DC and $^{15}N$ CSA frequencies (Fig. 4E and F); these values provide a quantitative estimate of the agreement between the refined structure and the experimental data.

The results illustrate two important aspects of the SASR method: (1) structure refinement is obtained simultaneously with resonance assignment and (2) complete resonance assignment is not

required. The inclusion of additional assigned restraints will improve both structural accuracy and precision, but a very reasonable backbone structure can be obtained even with a few gaps in the restraints coverage of the amino acid sequence. The examples described in this paper are for double resonance $^1H/^{15}N$ spectra, however AssignFit and SASR can include both $^1H/^{13}C$ DC and $^{13}C$ CSA data. This would additionally enhance the confidence in assignment selection and improve structural refinement by providing an additional set of restraints.

### 2.4. AssignFit parameter optimization

The values of RMSD$_{tot}$ and E$_{tot}$ calculated for each test assignment of AssignFit depend on the user-defined parameter D$_a$ and on the relative values of the DC and CSA force constants (DCscale, CSAscale). The effects of these parameters are illustrated in Fig. 6, for both OmpX and fd coat protein. Viewing RMSD$_{tot}$ as a function of both D$_a$ and the ratio of the force constants (CSAscale/DCscale) indicates that, for both proteins, the best fit to the experimental data is obtained when D$_a$ = 8.5 kHz and when CSAscale/DCscale = 0.01. For OmpX, the value of D$_a$ = 8.5 kHz is in line with the order parameter expected for magnetically aligned bilayers. However, the fd coat protein incorporated in glass-aligned lipid bilayers would be expected to exhibit higher order, with D$_a$ = 10 kHz, which was the value of D$_a$ used in all AssignFit calculations. Indeed, we observed that upon refinement with additional assignments of Ala and Gly, the optimum value of D$_a$ (yielding the lowest RMSD$_{tot}$) shifted to 10 kHz for fd, while that of OmpX remained constant at 8.5 kHz.

The optimal ratio of the DC and CSA force constants between 0.1 and 0.01 reflects the optimum balance between the effects of DC and CSA restraints in the calculation. It is the region where both DC and CSA contribute equally to the assignment determination. When CSAscale/DCscale > 0.1, the CSA makes a disproportionate contribution; in contrast, when CSAscale/DCscale < 0.1, the DC frequency takes over and controls the result. Both cases have the potential of skewing the AssignFit result towards a poor fit with the experimental data.

For both OmpX and fd coat protein, the assignment with best fit between structure and experiment displays a minimum in both RMSD$_{tot}$ and E$_{tot}$ at the optimal value of D$_a$ (Fig. 6B and C). However, while RMSD$_{tot}$ is also minimal at the optimal value of CSAscale/DCscale (Fig. 6D), the same is not true for E$_{tot}$, which always reflects the actual values of the force constants used and, thus, always increases with increasing CSAscale and DCscale (Fig. 6E). Therefore, while both RMSD$_{tot}$ and E$_{tot}$ are useful parameters for quantitatively
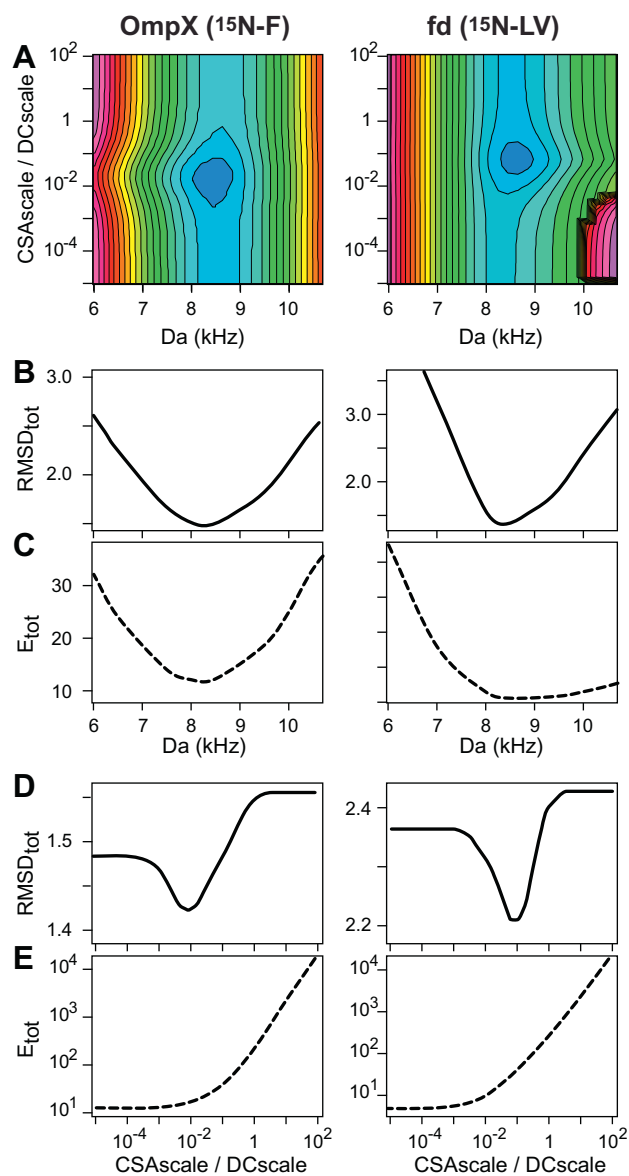
**Fig. 6.** Dependence of AssignFit RMSD$_{tot}$ and E$_{tot}$ on the value of D$_a$ and the relative DC and CSA force constants. Data are shown for the best AssignFit test assignments of the seven Phe peaks of OmpX (left) and of the four Val peaks of fd (right). (A) Contour plots, and (B)–(E) related 2D slices, showing RMSD$_{tot}$ as a function of D$_a$ and the relative force constants for DC and CSA (CSAscale/DCscale; note the log scale). The lowest RMSD$_{tot}$ (blue) is obtained when D$_a \sim$ 8–9 kHz and CSAscale/DCscale = 0.01. The highest RMSD$_{tot}$ (magenta) is obtained for D$_a$ values outside the expected range. (B) Effect of D$_a$ on RMSD$_{tot}$. (C) Effect of D$_a$ on E$_{tot}$ (note the log scale). (D) Effect of CSAscale/DCscale on RMSD$_{tot}$. (E) Effect of CSAscale/DCscale on E$_{tot}$ (note the log scales for both axes). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

assessing confidence in each AssignFit result, RMSD$_{tot}$ is better suited for this purpose, since it always yields a minimum for the best-fit assignment. Indeed, we find that the best results are obtained when we select assignments based on the lowest RMSD$_{tot}$.

## 3. Conclusions

The recent development of bioinformatics methods for predicting protein structure, either by comparative homology modeling based on amino acid sequence similarity to a protein of known structure [32,33], or by *de novo* methods based on database searches for low energy conformations compatible with the target amino acid sequence [34–39], enable structural models of proteins,

including membrane proteins, to be generated. Our interest is not in obtaining the most precise structure from these statistical methods but rather to obtain effective starting structural models that will enable AssignFit and SASR to assign experimental solid-state NMR orientation restraints, which can then be used for structural refinement directed toward the calculation of atomic resolution structures with high accuracy and precision. Furthermore, orientation restraints measured for membrane proteins in lipid bilayer membranes can also be used to refine NMR or crystal structures determined in detergents, to obtain membrane-specific structural information that more closely resembles the native environment [22,25]. Regardless of the provenance of the starting model, even a few DC and CSA measurements can provide effective orientation restraints enabling some structural information to be obtained prior to complete resonance assignment. AssignFit facilitates the SASR process by minimizing the difference between experimentally observed spectral frequencies and the frequencies calculated from a structural model. Its incorporation in the XPLOR-NIH package further facilitates its use in combination with simulated annealing for structure refinement.

## 4. Methods

### 4.1. Computer programs

All calculations were performed with the XPLOR-NIH molecular structure determination package [26]. Molecular structures were analyzed and visualized with Pymol [40]. A set of example scripts and input data files are provided with XPLOR-NIH release 2.29.

### 4.2. Initial structural models

The 50-residue ideal α-helix used as initial model for the fd coat protein was generated from extended random coil coordinates using a high temperature simulated annealing protocol [41] in XPLOR-NIH, restrained by uniform ($\varphi = -65°$, $\psi = -40°$), tight (±0.1°) backbone dihedral angles, imposed with a force constant of 1000 kcal mol$^{-1}$ rad$^{-2}$. The coordinates of the 1.90 Å crystal structure of OmpX [30] were obtained from the Protein Data Bank (PDB code 1QJ8). After adding missing hydrogens, the structure was subjected to Cartesian coordinate Powell minimization as described [25].

### 4.3. AssignFit parameters

The parameters used in the AssignFit calculations are listed in Table 1. User-defined parameters include the values of D$_a$ and Rh, specifying molecular order in terms of the $^1$H/$^{15}$N DC, the values of the force constants for the DC and CSA potentials, and the

**Table 1**
Potential term parameters used for AssignFit calculation of the fd coat protein.

| |
|---|
| *RDCpot (DC restraints)* |
| Force constant = 1 kcal s$^2$ mol$^{-1}$ |
| D$_a$ = 10 kHz |
| Rh = 0 |
| |
| *CSApot ($^{15}$N CSA restraints)*[a] |
| Force constant = 0.01 kcal s$^2$ mol$^{-1}$ |
| DaScale = 20,000 |
| $^{15}$N tensor for non-Gly [51]: $\delta_{iso}$ = 119 ppm, $\delta_{11}$ = −42.3 ppm, $\delta_{22}$ = −55.3 ppm, $\delta_{33}$ = 97.7 ppm, $\beta$ = 17°, $\gamma$ = 0° |
| $^{15}$N tensor for Gly [52]: $\delta_{iso}$ = 105 ppm, $\delta_{11}$ = −41.0 ppm, $\delta_{22}$ = −64.0 ppm, $\delta_{33}$ = 105.0 ppm, $\beta$ = 20°, $\gamma$ = 0° |

[a] Values are reported following the convention $|\delta_{33}| > |\delta_{22}| > |\delta_{11}|$; the angle $\beta$ is between $\delta_{33}$ and the NH bond and the angle $\gamma$ is between $\delta_{22}$ and the axis normal to the peptide plane.

**Table 2**
Potential term parameters used for structure refinement.

| Potential | Force constant | 300 K Dynamics | 300–20 K Simulated annealing |
|---|---|---|---|
| SANI (DC restraints) | (kcal s$^2$ mol$^{-1}$) | 0.1 | 0.1–1.0 |
| DCSA (CSA restraints) | (kcal s$^2$ mol$^{-1}$) | 0.001 | 0.001–0.01 |
| CDIH (predicted dihedral restraints) | (kcal mol$^{-1}$ rad$^{-2}$) | 300 | 300 |
| RAMA (knowledge-based dihedral restraints) | (kcal mol$^{-1}$) | 0.02 | 0.02–2.0 |
| ANG (bond angle) | (kcal mol$^{-1}$ rad$^{-2}$) | 0.4 | 0.4–1.0 |
| IMPR (improper dihedral angle) | (kcal mol$^{-1}$ rad$^{-2}$) | 0.1 | 0.1–1.0 |
| VDW[a] (non-bonded atom–atom repulsion) | (kcal mol$^{-1}$ Å$^{-4}$) | 0.004 | 0.004–4.0 |

[a] Atomic radii are scaled by 0.4 during initial dynamics and minimization and scaled by a value ramped from 0.4 to 0.8 during simulated annealing.

value of $R_{spec}$ specifying the relative spectral resolution used to calculate $RMSD_{tot}$. It is also possible to provide specific values for the magnitudes and molecular orientation of the chemical shift tensor. Previously we noted that residue-specific variations in the $^{15}$N chemical shift tensor are minor compared to the spectral manifestation of molecular orientation, indicating that $^{15}$N chemical shifts as well as dipolar couplings can be useful restraints for structure determination and refinement [18,25]. This is also evidenced by the observation of the characteristic wheel like patterns in the spectra of aligned membrane proteins [9–11].

For solid-state NMR experiments with membrane proteins in either glass aligned (e.g. for fd coat protein) or magnetically aligned (e.g. for OmpX) lipid bilayer samples, the order tensor (S) is set by the sample's liquid crystalline geometry. Order is axially symmetric around the bilayer normal (n), which can be either parallel or perpendicular to the magnetic field ($B_o$), resulting in net axial alignment of $S_{zz} = 1.0$ for n$||B_o$ and $S_{zz} = -0.5$ for n$_\perp B_o$, and null rhombicity (Rh = $2/3(S_{xx} - S_{yy})/S_{zz} = 0$) [42–45].

The axial alignment parameter, $D_a$, incorporates both the values of $S_{zz}$ and of $D_{max}$, the maximum DC expected for a static NH bond of a given bond length ($D_a = 1/2 \cdot S_{zz} \cdot D_{max}$). For an NH bond length of 1.07 Å and corresponding $D_{max} = 20$ kHz, values of $D_a = 10$ kHz and $D_a = -5$ kHz are expected for fully ordered bilayers with n$||B_o$ and n$_\perp B_o$, respectively. Glass-aligned lipid bilayers are highly ordered and typically exhibit an overall order parameter for internal motion close to 1.0, whereas for magnetically aligned bilayers, axially symmetric motional averaging scales $S_{zz}$ by a factor in range from 0.9 to 0.8.

The CSA alignment tensor was normalized to the maximum value of DC by setting the user-defined XPLOR-NIH AssignFit parameter, DaScale, equal to 20 kHz. The CSA for each residue was calculated by subtracting the isotropic $^{15}$N chemical shift frequency ($\delta_{iso}$) from the orientation-dependent chemical shift frequency measured in the solid-state NMR SLF spectra of the aligned protein; similarly, the CSA values calculated in AssignFit or after structural refinement were converted to orientation-dependent frequencies by adding $\delta_{iso}$.

### 4.4. Parameters used for structure refinement

Refinement of OmpX with the DC and CSA restraints for the 7 Phe residues was as described previously [25]. For refinement of the fd coat protein, the DC and CSA energy terms were minimized as described [46,47], after obtaining the appropriate assignments with AssignFit. Additional restraints included loosely imposed (±15°) backbone dihedral angles, derived from the predicted secondary structure, and the torsion angle potential of mean force [48,49]. Finally, energy terms were included to enforce covalent

geometry (bonds, bond-angles and improper dihedral angles) and prevent atomic overlap (via the standard repulsive quartic Van der Waal term).

Structure refinement was performed with a simulated annealing protocol consisting of a 10 ps, variable timestep, torsion angle molecular dynamics [50] phase at a temperature of 300 K, followed by simulated annealing from 300 K to 20 K, in 10 K increments at each step of 2 ps torsion-angle molecular dynamics. Finally, gradient minimization was performed in Cartesian coordinates. The schedule for the force constants is specified in Table 2. A total of 100 structures were calculated and those with covalent or dihedral restraint violations were discarded. The remainder was sorted using experimental, dihedral and covalent energies. The structure with the lowest energy was used as the initial structure in the next SASR cycle.

### References

[1] F. Tian, H. Valafar, J.H. Prestegard, A dipolar coupling based strategy for simultaneous resonance assignment and structure determination of protein backbones, J. Am. Chem. Soc. 123 (2001) 11791–11796.
[2] J.C. Hus, J.J. Prompers, R. Bruschweiler, Assignment strategy for proteins with known structure, J. Magn. Reson. 157 (2002) 119–123.
[3] Y.S. Jung, M. Zweckstetter, Backbone assignment of proteins with known structure using residual dipolar couplings, J. Biomol. NMR 30 (2004) 25–35.
[4] Y.S. Jung, M. Sharma, M. Zweckstetter, Simultaneous assignment and structure determination of protein backbones by using NMR dipolar couplings, Angew. Chem. Int. Ed. Engl. 43 (2004) 3479–3481.
[5] J. Meiler, D. Baker, Rapid protein fold determination using unassigned NMR data, Proc. Natl. Acad. Sci. USA 100 (2003) 15404–15409.
[6] A. Grishaev, M. Llinas, CLOUDS, a protocol for deriving a molecular proton density via NMR, Proc. Natl. Acad. Sci. USA 99 (2002) 6707–6712.
[7] A. Grishaev, M. Llinas, Protein structure elucidation from NMR proton densities, Proc. Natl. Acad. Sci. USA 99 (2002) 6713–6718.
[8] G.A. Bermejo, M. Llinas, Deuterated protein folds obtained directly from unassigned nuclear overhauser effect data, J. Am. Chem. Soc. 130 (2008) 3797–3805.
[9] F.M. Marassi, S.J. Opella, A solid-state NMR index of helical membrane protein structure and topology, J. Magn. Reson. 144 (2000) 150–155.
[10] J. Wang, J. Denny, C. Tian, S. Kim, Y. Mo, F. Kovacs, Z. Song, K. Nishimura, Z. Gan, R. Fu, J.R. Quine, T.A. Cross, Imaging membrane protein helical wheels, J. Magn. Reson. 144 (2000) 162–167.
[11] F.M. Marassi, A simple approach to membrane protein secondary structure and topology based on NMR spectroscopy, Biophys. J. 80 (2001) 994–1003.
[12] S.J. Opella, F.M. Marassi, Structure determination of membrane proteins by NMR spectroscopy, Chem. Rev. 104 (2004) 3587–3606.
[13] R.C. Page, C. Li, J. Hu, F.P. Gao, T.A. Cross, Lipid bilayers: an essential environment for the understanding of membrane proteins, Magn. Reson. Chem. 45 (2007) S2–S11.
[14] F.A. Kovacs, T.A. Cross, Transmembrane four-helix bundle of influenza A M2 protein channel: structural implications from helix tilt and orientation, Biophys. J. 73 (1997) 2511–2517.
[15] S.D. Cady, C. Goodman, C.D. Tatko, W.F. DeGrado, M. Hong, Determining the orientation of uniaxially rotating membrane proteins using unoriented samples: a $^2$H, $^{13}$C, and $^{15}$N solid-state NMR investigation of the dynamics and orientation of a transmembrane helical bundle, J. Am. Chem. Soc. 129 (2007) 5719–5729.
[16] F.M. Marassi, B.B. Das, G.J. Lu, H.J. Nothnagel, S.H. Park, W.S. Son, Y. Tian, S.J. Opella, Structure determination of membrane proteins in five easy pieces, Methods, in press.
[17] R.R. Ketchem, W. Hu, T.A. Cross, High-resolution conformation of gramicidin A in a lipid bilayer by solid-state NMR, Science 261 (1993) 1457–1460.
[18] F.M. Marassi, S.J. Opella, Simultaneous assignment and structure determination of a membrane protein from NMR orientational restraints, Protein Sci. 12 (2003) 403–411.
[19] S.H. Park, A.A. Mrse, A.A. Nevzorov, M.F. Mesleh, M. Oblatt-Montal, M. Montal, S.J. Opella, Three-dimensional structure of the channel-forming trans-

membrane domain of virus protein "u" (Vpu) from HIV-1, J. Mol. Biol. 333 (2003) 409–424.

[20] A.A. De Angelis, S.C. Howell, A.A. Nevzorov, S.J. Opella, Structure determination of a membrane protein with two trans-membrane helices in aligned phospholipid bicelles by solid-state NMR spectroscopy, J. Am. Chem. Soc. 128 (2006) 12256–12267.

[21] M. Sharma, M. Yi, H. Dong, H. Qin, E. Peterson, D.D. Busath, H.X. Zhou, T.A. Cross, Insight into the mechanism of the influenza A proton channel from a structure in a lipid bilayer, Science 330 (2010) 509–512.

[22] C.M. Franzin, P. Teriete, F.M. Marassi, Structural similarity of a membrane protein in micelles and membranes, J. Am. Chem. Soc. 129 (2007) 8078–8079.

[23] A.A. Nevzorov, S.J. Opella, Structural fitting of PISEMA spectra of aligned proteins, J. Magn. Reson. 160 (2003) 33–39.

[24] T. Asbury, J.R. Quine, S. Achuthan, J. Hu, M.S. Chapman, T.A. Cross, R. Bertram, PIPATH: an optimized algorithm for generating alpha-helical structures from PISEMA data, J. Magn. Reson. 183 (2006) 87–95.

[25] R. Mahalakshmi, F.M. Marassi, Orientation of the *Escherichia coli* outer membrane protein OmpX in phospholipid bilayer membranes determined by solid-State NMR, Biochemistry 47 (2008) 6531–6538.

[26] C.D. Schwieters, J.J. Kuszewski, N. Tjandra, G.M. Clore, The XPLOR-NIH NMR molecular structure determination package, J. Magn. Reson. 160 (2003) 65–73.

[27] G.M. Clore, C.D. Schwieters, How much backbone motion in ubiquitin is required to account for dipolar coupling data measured in multiple alignment media as assessed by independent cross-validation?, J Am. Chem. Soc. 126 (2004) 2923–2938.

[28] M. Zweckstetter, Determination of molecular alignment tensors without backbone resonance assignment: aid to rapid analysis of protein-protein interactions, J. Biomol. NMR 27 (2003) 41–56.

[29] J.K. Denny, J. Wang, T.A. Cross, J.R. Quine, PISEMA powder patterns and PISA wheels, J. Magn. Reson. 152 (2001) 217–226.

[30] J. Vogt, G.E. Schulz, The structure of the outer membrane protein OmpX from *Escherichia coli* reveals possible mechanisms of virulence, Structure 7 (1999) 1301–1309.

[31] C. Fernandez, K. Adeishvili, K. Wuthrich, Transverse relaxation-optimized NMR spectroscopy with the outer membrane protein OmpX in dihexanoyl phosphatidylcholine micelles, Proc. Natl. Acad. Sci. USA 98 (2001) 2358–2363.

[32] M.A. Marti-Renom, A.C. Stuart, A. Fiser, R. Sanchez, F. Melo, A. Sali, Comparative protein structure modeling of genes and genomes, Annu. Rev. Biophys. Biomol. Struct. 29 (2000) 291–325.

[33] F.S. Domingues, P. Lackner, A. Andreeva, M.J. Sippl, Structure-based evaluation of sequence comparison and fold recognition alignment accuracy, J. Mol. Biol. 297 (2000) 1003–1013.

[34] D. Baker, A. Sali, Protein structure prediction and structural genomics, Science 294 (2001) 93–96.

[35] R. Srinivasan, G.D. Rose, Ab initio prediction of protein structure using LINUS, Proteins 47 (2002) 489–495.

[36] S.B. Ozkan, G.A. Wu, J.D. Chodera, K.A. Dill, Protein folding by zipping and assembly, Proc. Natl. Acad. Sci. USA 104 (2007) 11987–11992.

[37] V. Yarov-Yarovoy, J. Schonbrun, D. Baker, Multipass membrane protein structure prediction using Rosetta, Proteins 62 (2006) 1010–1025.

[38] P. Barth, J. Schonbrun, D. Baker, Toward high-resolution prediction and design of transmembrane helical protein structures, Proc. Natl. Acad. Sci. USA 104 (2007) 15682–15687.

[39] R. Das, D. Baker, Macromolecular modeling with rosetta, Annu. Rev. Biochem. 77 (2008) 363–382.

[40] W.L. DeLano, PyMol, 2005. <http://www.pymol.org>.

[41] M. Nilges, G.M. Clore, A.M. Gronenborn, Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms. Circumventing problems associated with folding, FEBS Lett. 239 (1988) 129–136.

[42] C.R. Sanders, B. Hare, K.P. Howard, J.H. Prestegard, Magnetically-oriented phospholipid micelles as a tool for the study of membrane-associated molecules, Progr. NMR Spectrosc. 26 (1994) 421–444.

[43] A. Bax, G. Kontaxis, N. Tjandra, Dipolar couplings in macromolecular structure determination, Methods Enzymol. 339 (2001) 127–174.

[44] J.H. Prestegard, C.M. Bougault, A.I. Kishore, Residual dipolar couplings in structure determination of biomolecules, Chem. Rev. 104 (2004) 3519–3540.

[45] A.A. De Angelis, D.H. Jones, C.V. Grant, S.H. Park, M.F. Mesleh, S.J. Opella, NMR experiments on aligned samples of membrane proteins, Methods Enzymol. 394 (2005) 350–382.

[46] G.M. Clore, A.M. Gronenborn, N. Tjandra, Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude, J. Magn. Reson. 131 (1998) 159–162.

[47] R.S. Lipsitz, N. Tjandra, $^{15}$N chemical shift anisotropy in protein structure refinement and comparison with NH residual dipolar couplings, J. Magn. Reson. 164 (2003) 171–176.

[48] J. Kuszewski, A.M. Gronenborn, G.M. Clore, Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases, Protein Sci. 5 (1996) 1067–1080.

[49] J. Kuszewski, A.M. Gronenborn, G.M. Clore, Improvements and extensions in the conformational database potential for the refinement of NMR and X-ray structures of proteins and nucleic acids, J. Magn. Reson. 125 (1997) 171–177.

[50] C.D. Schwieters, G.M. Clore, Internal coordinates for molecular dynamics and minimization in structure determination and refinement, J. Magn. Reson. 152 (2001) 288–302.

[51] C. Wu, A. Ramamoorthy, L.M. Gierasch, S.J. Opella, Simultaneous characterization of the amide $^{1}$H chemical shift, $^{1}$H–$^{15}$N dipolar, and $^{15}$N chemical shift interaction tensors in a peptide bond by 3-dimensional solid-state NMR spectroscopy, J. Am. Chem. Soc. 117 (1995) 6148–6149.

[52] T.G. Oas, C.J. Hartzell, W. Dahlquist, G.P. Drobny, The amide $^{15}$N chemical shift tensors of four peptides determined from $^{13}$C dipole-coupled chemical shift powder patterns, J. Am. Chem. Soc. 109 (1987) 5962–5966.