

## Simultaneous structure and dynamics of a membrane protein using REDCRAFT: Membrane-bound form of Pf1 coat protein

Paul Shealy<sup>a</sup>, Mikhail Simin<sup>a</sup>, Sang Ho Park<sup>b</sup>, Stanley J. Opella<sup>b</sup>, Homayoun Valafar<sup>a,\*</sup>

<sup>a</sup> Department of Computer Science & Engineering, University of South Carolina, 315 Main Street, Columbia, SC 29208, United States

<sup>b</sup> Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0307, United States

### ARTICLE INFO

#### Article history:

Received 11 May 2010

Revised 18 July 2010

Available online 30 July 2010

#### Keywords:

REDCRAFT

Simultaneous

Backbone

Protein

Structure

Motion

Residual dipolar coupling

Dynamics

### ABSTRACT

A strategy for simultaneous study of the structure and internal dynamics of a membrane protein is described using the REDCRAFT algorithm. The membrane-bound form of the Pf1 major coat protein (mbPf1) was used as an example. First, synthetic data is utilized to validate the simultaneous study of structure and dynamics with REDCRAFT using dihedral restraints and backbone N–H RDCs from two different alignments. Subsequently, the validated analysis is applied to experimental data and confirms that REDCRAFT produces meaningful structures from sparse RDC data. Furthermore, simulated data from a two-state jump motion is used to illustrate the necessity for simultaneous consideration of structure and dynamics. Disregarding internal dynamics during the course of structure determination is shown to produce an average-state that is not related to the two intermediate states. During the analysis of RDC data from the dynamic model, REDCRAFT appropriately identifies the region separating the static and dynamic domains of the protein. Finally, analysis of experimental data strongly suggests the existence of internal motion between the amphipathic and the transmembrane helices of the membrane-bound form of the protein. The ability to perform fragmented structure determination of each domain without a *a priori* assumption of the order tensors allows an independent determination of the order tensors, which yields a more comprehensive description of protein structure and dynamics and is particularly relevant to the study of membrane proteins.

Published by Elsevier Inc.

### 1. Introduction

Nuclear Magnetic Resonance (NMR) structure determination of proteins has traditionally relied largely on Nuclear Overhauser Effect (NOE) data, which yield short-distance restraints between atoms in close proximity. In recent years, however, residual dipolar couplings (RDCs) have played an increasingly important role in structure determination of protein structures by NMR, due in part to their ability to orient distant portions of a protein with respect to a common frame of reference. Furthermore, for some classes of proteins, such as membrane proteins, NOEs are an insufficient source of structural restraints and have largely been replaced by RDCs.

RDCs have additional value for use in structure determination. Their sensitivity to internal motions on picosecond to millisecond time scales [1–3] is valuable in elucidating the internal dynamics of a protein. Studies of dynamics have traditionally been performed in separate steps; the protein's structure is determined under the assumption of static positions of the atoms using a fixed

order tensor, and only later are its motions characterized. The structure determination protocols based on the assumption of molecular rigidity produce a single structure based on data that are perturbed by internal dynamics. The degree of similarity between the static model of a protein structure and many conformations of a dynamic model is not always clear and merits further investigation. A more appealing, rigorous approach is simultaneous treatment of both the structure and dynamics of a protein. The REDCRAFT algorithm [4] provides this capability. Here we analyze the membrane-bound form of Pf1 coat protein as an example of a membrane protein using REDCRAFT. We demonstrate REDCRAFT's ability to determine this protein's structure and internal dynamics in the complete absence of NOE data by utilizing only RDCs and TALOS torsion angle constraints. We define *internal motion* as dynamics of domains or rigid fragments of a macromolecule relative to each other. The membrane-bound form of Pf1 coat protein consists of two  $\alpha$ -helices separated by a short loop. We further show that REDCRAFT provides strong evidence for the existence of internal motion between the two helices, and that the helices exhibit markedly different order tensors when treated separately. Finally, we compare the REDCRAFT structure to structures determined under the static assumption. The complete software binary,

\* Corresponding author.

E-mail address: [homayoun@cse.sc.edu](mailto:homayoun@cse.sc.edu) (H. Valafar).

source code and manuals are available for public access via the web at <http://ifestos.cse.sc.edu>. REDCRAFT is distributed with tools to provide REDCAT [5] input files and Xplor-NIH [6] constraint files for further refinement.

## 2. Theoretical background

### 2.1. Residual dipolar coupling

Residual dipolar coupling (RDCs) data have been used extensively in recent years for structural studies of a broad range of macromolecules, including globular proteins [7–13], membrane proteins [14,15], nucleic acids [16–20], and carbohydrates [21–24]. They have also been used to provide insights into the internal dynamics of protein structures [3,25,26]. A thorough description of RDCs can be found in Section S1.

The RDCs for a protein can be collected into a single formulation as shown

$$\begin{bmatrix} x_1^2 - z_1^2 & y_1^2 - z_1^2 & 2x_1y_1 & 2x_1z_1 & 2y_1z_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^2 - z_n^2 & y_n^2 - z_n^2 & 2x_ny_n & 2x_nz_n & 2y_nz_n \end{bmatrix} \begin{bmatrix} S_{xx} \\ S_{yy} \\ S_{xy} \\ S_{xz} \\ S_{yz} \end{bmatrix} = \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix} \quad (1)$$

In this equation,  $x_i$ ,  $y_i$  and  $z_i$  are the Cartesian coordinates of the normalized  $i$ th internuclear vector and  $r_i$  is the  $i$ th RDC. The Saupe order tensor matrix  $S = [S_{xx} \ S_{yy} \ S_{xy} \ S_{xz} \ S_{yz}]^T$  [27] provides information about the anisotropy introduced by partial alignment [28–30]. This symmetric, traceless,  $3 \times 3$  matrix contains five degrees of freedom. Given the vectors and collected RDCs, an order tensor can be estimated for the sample in an alignment medium [5,31]. Substituting this estimate into Eq. (1) yields a set of back-computed RDCs  $c_1, \dots, c_n$ . The RMSD between these back-computed RDCs and the experimental RDCs provides a measure of fitness between the RDCs and the assigned vectors:

$$\text{RMSD} = \sqrt{\frac{1}{m} \sum_{i=1}^m \frac{1}{n} \sum_{j=1}^n (r_{ij} - c_{ij})^2} \quad (2)$$

where  $m$  is the number of alignment media. Under practical conditions the experimental error should provide an upper bound to the RMSD score shown in Eq. (2). An RMSD score that exceeds the magnitude of the experimental error has several possible causes such as: mis-assigned RDCs, a severe violation of acceptable peptide geometry, or the existence of internal motion. The analysis presented here assumes properly assigned RDCs and no severe violations of peptide geometry; high RMSD values are attributed solely to the presence of internal motions.

Some proteins exhibit a modular structure with several rigid fragments. These fragments may either be domains or sections of a single domain. When these fragments undergo motion relative to one another, each fragment will yield a separate order tensor describing its average orientation and strength of alignment [32–34]. Elucidation of these individual order tensors yields information about the relative alignments of these average fragment orientations. It is also possible for order tensors from different fragments to yield a complete description of the internal dynamics of a protein on biologically relevant time scales [32,33,35].

### 2.2. Simultaneous structure and motion with REDCRAFT

REDCRAFT [4,36,37] is an RDC data analysis tool that is capable of simultaneous structure determination and identification of internal motion. RDCs play an increasingly important role in NMR structure determination because of their unique advantages

over traditional NOE data [38]. Structure determination based primarily on RDC data requires algorithms that operate in fundamentally different ways from those that use NOE data, and several new programs have been proposed for this purpose [7,12,39–43]. However, the rich information content and complexity of RDC data continues to challenge current analysis tools. Furthermore, identification of the mobile domains or fragments of macromolecules is critical not only for the study of the molecule's internal motion, but also for accurate structure determination. Other software packages like Xplor-NIH [6], CNS [44], or CYANA [45] are capable of using RDCs during the process of structure determination, however, they rely on an accurate *a priori* estimate of the order parameters ( $S_{xx}$ ,  $S_{yy}$ ,  $S_{zz}$  or  $Da/R$ ). The estimate is often obtained from an unrefined NOE based structure, although recent developments have enabled the calculation of order tensors in the absence of assignment or an initial structure [14,46]. This process for structure determination uses a single order tensor that applies to the whole molecule, which imposes the assumption that the entire protein is rigid. This approach may perturb portions of the structure with different dynamics with respect to remainder of the molecule. Structural deviations may not be confined just to the mobile regions, since the rigid fragments of a protein may not even be oriented correctly with respect to one another [33]. REDCRAFT provides the ability to determine a protein's structure while recognizing fragments that are mobile with respect to each other. Once the dynamic fragments of a protein are identified, each of them can be individually subjected to structure determination. This process allows each fragment to be characterized using a separate order tensor that describes its anisotropic alignment.

REDCRAFT's approach to structure determination proceeds in two stages: Stage-I and Stage-II. During Stage-I a list of all possible torsion angles between any two adjoining peptide planes is pruned (using data such as scalar couplings or Ramachandran constraints) and ranked based on structural fitness to the RDC data. Stage-II of REDCRAFT extends a given fragment of size  $N$  peptide planes (initially a single peptide seed) by addition of one peptide plane at a time. All backbone torsion angles retained from Stage-I are considered for the new peptide plane, yielding a large number of candidate structures. Typically the 2000–10,000 structural candidates that best fit the RDC data are propagated forward for the next iteration of adding a peptide plane. The number of structures propagated forward is called the *search depth*. This process provides a sufficiently diverse population of conformers to prevent entrapment in local minima.

Stage-II provides the unique ability to identify points of internal motion during the process of structure extension. REDCRAFT reports the fitness between the given RDCs and the best structure at each residue; the metric used to quantify this fitness is the RMSD between the experimental and back-computed RDCs. The plot of RDC fitness as a function of residue number is called the *dynamic profile*. Analyzing the dynamic profile provides insight into the quality of the final structure and the likelihood that the data represent a rigid structure. A dynamic profile of REDCRAFT analysis may consist of a number of distinct phases. A typical dynamic profile for a static structure consists of three distinct phases. During the first phase the REDCRAFT score (also the RDC fitness as described in Eq. (2)), will be zero while the system remains undetermined (less than 5–7 RDCs). The score gradually increases as a function of fragment size as the initial fragment elongates to include a sufficient number of RDCs (>7 RDCs). Finally, the increasing score will stabilize with a value near that expected from experimental error. Final scores slightly higher than the expected values are not diagnostic and are artifacts of the algorithm. In contrast, for a protein with internal motions, the dynamic profile will show a marked increase in RDC fitness at the residue separating two fragments that are moving relative to one another. The actual residue

at which motion occurs may not be the residue at which the increase is observed. Several residues after the point of motion may need to be incorporated for the effect of motion to become recognizable in the dynamic profile. This lag is due to the number of RDC data points per residue. A portion of a dynamic profile with decreasing score may be indicative of a region with exceptionally high-quality of data or a region with sparse data.

### 2.3. Pf1 major coat protein

Pf1 is a filamentous bacteriophage that infects *Pseudomonas aeruginosa*. The major coat protein of Pf1 contains 46 residues and plays two major roles in the viral lifecycle. During infection, newly synthesized copies of the protein are stored within the host cell's membrane prior to virus assembly, which occurs at the cell membrane. After the virus particles are extruded from the membrane and assembled, the coat protein forms the capsid surrounding the DNA and constitutes the bulk of the virus particles. Static structures have been previously determined for the membrane-bound form (PDB: 2KLK) and the viral coat form (PDB: 1PJF). In both cases, the structure of the major coat protein consists of two  $\alpha$ -helices connected by a short stretch of residues with irregular conformations. The major difference between these two structures is the orientation of the N-terminal amphipathic helix. Here we examine the membrane-bound form of the Pf1 major coat protein, which we refer to as *mbPf1*. In the membrane-bound form of the protein the N-terminal helix contacts the membrane surface (the *amphipathic* helix) and the C-terminal helix is inserted into the membrane (the *transmembrane* helix). The main question that we investigate here is what effects do the amphipathic helix's motions relative to the transmembrane helix have on the structure determined for the protein in micelles by solution NMR spectroscopy.

### 2.4. Simulated RDC data

Simulated RDC data are often crucial in establishing the fundamentals of any algorithmic approach. Simulated RDC data are used to validate REDCRAFT's ability to identify internal dynamics for *mbPf1*. Although this has been established in a general context [4], it is possible that any observed evidence of internal dynamics for *mbPf1* is an anomaly related to the special nature of this protein's structure. The structure of *mbPf1* is non-globular, consisting of two  $\alpha$ -helical regions joined by a loop. In  $\alpha$ -helices, backbone N–H vectors exhibit a near-parallel orientation and may provide a less than optimal number of independent data values. The non-globular structure of *mbPf1* (with its apparently perpendicular helices) may lead to a condition that is problematic for REDCRAFT; anomalous features of the dynamic profile may resemble that of internal dynamics. Structure determination of this protein using simulated data can help to eliminate the structural characteristics of this protein as the source of any observed internal dynamics. We therefore utilize simulated RDC data from a sample *mbPf1* structure under static and dynamic conditions to illustrate and validate the performance of REDCRAFT on this system.

The two order tensors shown in Table 1 were used to generate simulated RDC data. These two order tensors produce a range of

RDCs very similar to the experimental data. The first model from the ensemble of previously determined structures for this protein was used to generate simulated data. A derivative structure was created to model a two-state motion. The second state of the simulated motion was derived from the first state by altering the  $\varphi$  angle of the 19th residue by 60°. This corresponds to a motion in the amphipathic helix, as shown in Fig. 1. This model of motion is similar to that speculated for *mbPf1* in lipid environments. MOLMOL [47] was used for manipulation of the protein structure. The dynamic averaging function of REDCAT was used to calculate RDC data for the simulated two-state jump motion. The final calculated RDC data for both the static and dynamic cases were altered through the addition of  $\pm 0.5$  Hz of uniformly distributed noise; this level of noise corresponds to the previously reported experimental error. The simulated motion influences the overall observed alignment of the molecule and manifests itself as an alteration of the simulated order tensors. Table 2 lists the values for the effective principal order parameters obtained after simulation of motion and addition of  $\pm 0.5$  Hz of error. Minor differences between the values listed in Tables 1 and 2 are due to the addition of noise, while major differences are due to the effect of internal motion.

### 2.5. Experimental RDC data

Isotopically labeled Pf1 coat protein was prepared as described previously [48]. NMR samples were prepared by dissolving the purified protein in 400  $\mu$ l of 100 mM deuterated DHPC (1,2-dihexyl-sn-glycero-3-phosphocholine; [www.isotope.com](http://www.isotope.com)) containing 10% (v/v)  $^2\text{H}_2\text{O}$ , at pH 6.7. Weakly aligned samples for RDC measurement were prepared either by soaking the micellar sample into a dried 6% polyacrylamide gel overnight, with the length of the gel restricted to 22 mm from an initial length of 30 mm, or by adding fd bacteriophage at a final concentration of 28 mg/ml [46], yielding two independent alignments of the protein.

NMR experiments were performed on a Bruker DRX 600 MHz spectrometer equipped with a 5 mm triple-resonance cryogenic probe and a z-axis gradient at 40 °C.  $^1\text{H}$ - $^{15}\text{N}$  IPAP-HSQC [46] spectra obtained on isotropic and weakly aligned samples were used to measure the  $^1\text{H}$ - $^{15}\text{N}$  RDCs. The NMR data were processed using the programs NMRPipe/NMRDraw and NMRView [49].

### 2.6. Dihedral constraints

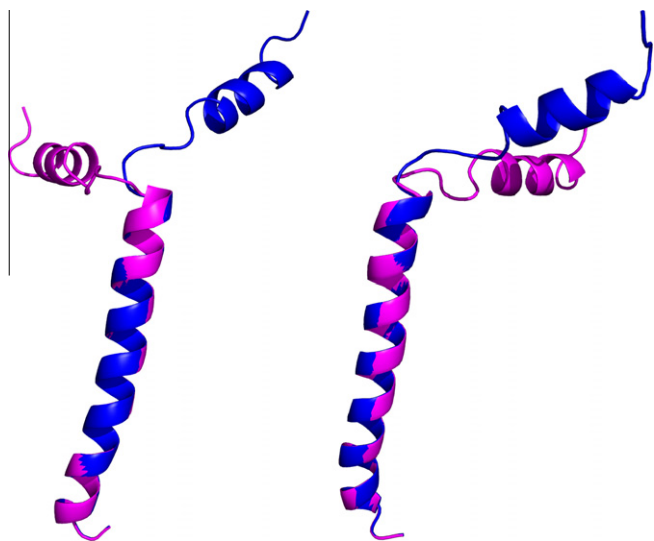
Dihedral angle restraints were derived from TALOS [11] analysis of the chemical shifts measured by solution NMR from isotropic micelle samples and from Dipolar wave analyses [50,51] of  $^1\text{H}/^{15}\text{N}$  RDCs with weakly aligned micelle samples. The helical regions of the protein included those residues for which the RMSD between the experimental values and the data fit to helix was less than the experimental error (0.5 Hz for RDCs).

Typical dihedral restraints (shown in Fig. 2a) based on Ramachandran plots were used for residues 14–19 that constitute the loop region of this protein. An alternative dihedral restraint space that was used for glycines located in the loop region is shown in Fig. 2b.

**Table 1**

Order tensors used to simulate RDC data from two alignment media. These order tensors were used for both the static and dynamic cases.

Alignment medium	$S_{xx}$	$S_{yy}$	$S_{zz}$	$\alpha$	$\beta$	$\gamma$	$D_a$	$R$
M1	3.00E–004	5.00E–004	–8.00E–004	0	0	0	–9.74	0.17
M2	–4.00E–004	–6.00E–004	1.00E–003	–40	–50	60	12.18	0.13



**Fig. 1.** The simulated two-state motion of mbPf1, with the two states in blue (original model) and purple (second state model), as shown from two perspectives. The  $\phi$  angle of the 19th residue is modified by  $60^\circ$  to produce the second model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

### 3. Results

#### 3.1. Simulated data and static structure

Simulated RDC data generated from the static structure and the dihedral restraints were converted to the REDCRAFT format [4]. REDCRAFT was used to calculate the structure with a search depth of 2000. Fig. 3 provides the dynamic profile for these data, showing the REDCRAFT score as a function of the number of residues in a

fragment. The REDCRAFT score reflects the fitness of the RDC data to the calculated structure, with a smaller value indicating a better fitness. The final structure calculated by REDCRAFT clearly satisfies the expected experimental error. The lack of a sharp increase in the REDCRAFT score, and the fact that the final REDCRAFT score is below the experimental error, clearly demonstrate that there are no points of internal motion.

The REDCRAFT structure and the original structure from which the RDC data were generated are shown in Fig. 4. These two structures exhibit less than  $0.5 \text{ \AA}$  of difference over the two helical regions (residues 1–13 and 20–45). The structure of the loop region is under-determined due to an inadequate number of structural constraints, therefore several indistinguishable conformers of this region exist. Despite the lack of data for the loop region, the helices are oriented correctly. This same phenomenon was observed with the previously determined NMR structure ensemble (PDB: 2KLV); when the transmembrane helices from the NMR ensemble are superimposed, the amphipathic helices have the correct orientation but do not overlap. The back-calculated order parameters from the calculated structure are shown in Table 3. Note the degree of similarity between these order parameters and those used to simulate the data. These findings agree with theoretical expectations.

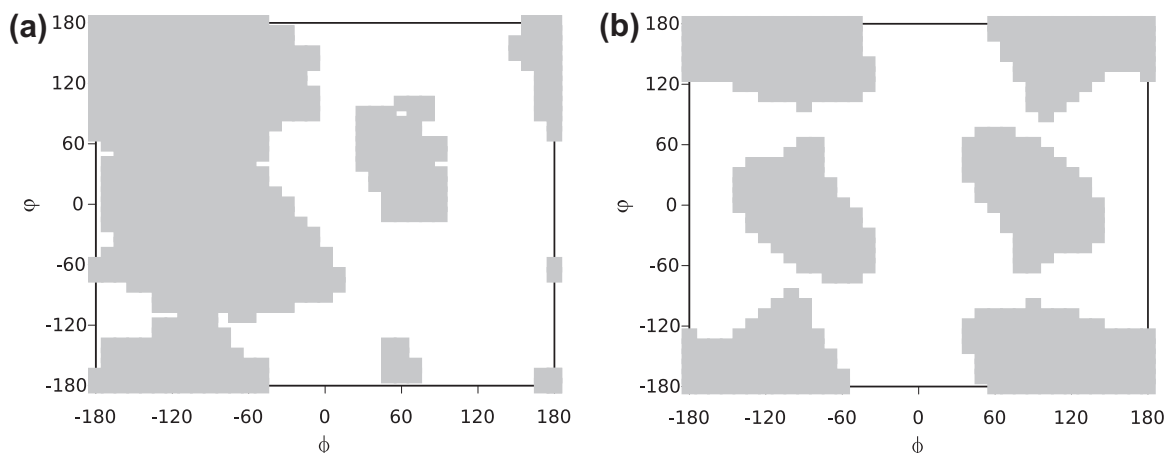
#### 3.2. Simulated data and internal dynamics

REDCRAFT was used to determine the structure of the protein with a two-state motion (shown in Fig. 1) using time-averaged RDC data and dihedral restraints with a search depth of 20,000. The dynamic profile of the REDCRAFT analysis is shown in Fig. 5. Note the significant jump in the RDC fitness between residues 22 and 23. The resulting score is well above that of the experimental error, which indicates the presence of internal dynamics. This observed jump is due to the lack of a single order tensor that can consistently describe the alignment of the entire protein. The

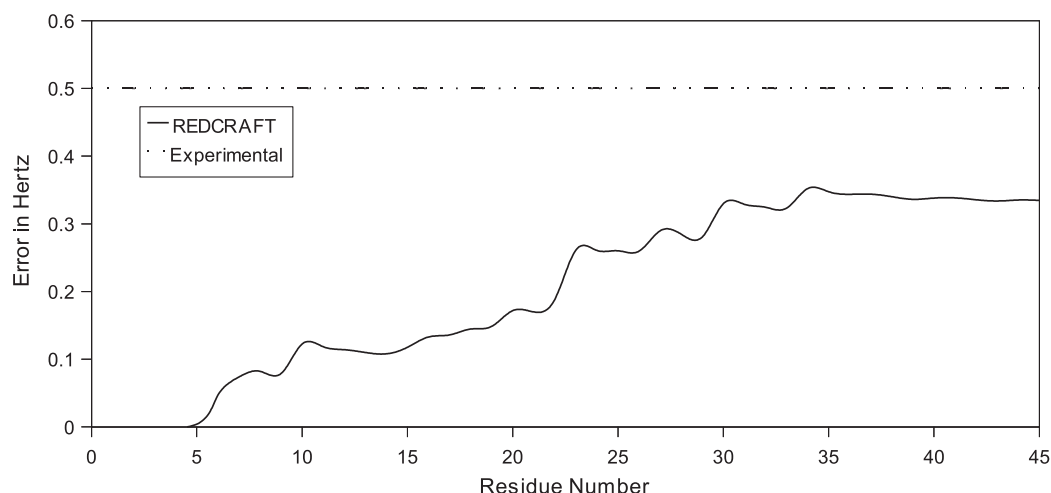
**Table 2**

Order tensors reported from the simulated RDC data from two regions of mbPf1 under the assumption of two rigid fragments that undergo motion relative to one another. Residues 1–13 are simulated to undergo motion with respect to the static region residues 20–45. Residues from the loop region are excluded.

Residues	Alignment medium	$S_{xx}$	$S_{yy}$	$S_{zz}$	$D_a$	$R$
1–13	M1	$3.78\text{E}-06$	$4.89\text{E}-04$	$-4.93\text{E}-04$	-6.00	0.66
	M2	$4.95\text{E}-05$	$5.43\text{E}-04$	$-5.92\text{E}-04$	-7.21	0.56
20–45	M1	$3.01\text{E}-04$	$4.99\text{E}-04$	$-8.00\text{E}-04$	-9.74	0.17
	M2	$-4.01\text{E}-04$	$-5.97\text{E}-04$	$9.98\text{E}-04$	12.15	0.13



**Fig. 2.** Ramachandran based dihedral constraints for a non-glycine (a) or a glycine (b) residue for use as REDCRAFT torsion angle constraints. Gray blocks represent the allowed torsion angles for consideration during REDCRAFT computations. The solution space for glycine differs due to its unique flexibility.



**Fig. 3.** The dynamic profile of REDCRAFT for the static structure mbPf1 using simulated data. REDCRAFT's computed error stays consistently below the expected experimental error. This is indicative of a well determined structure.



**Fig. 4.** Structure from REDCRAFT (green) determined under the assumption of molecular rigidity, as compared to that of the actual structure (blue), shown from two different perspectives. The two structures are aligned by the transmembrane helix and display near parallel alignment of amphipathic helices. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 3**  
Order parameters from the final REDCRAFT static structure.

Medium	$S_{xx}$	$S_{yy}$	$S_{zz}$	$D_a$	$R$
M1	3.13E-04	5.51E-04	-8.64E-04	-10.52	0.18
M2	-4.12E-04	-6.19E-04	1.03E-03	12.55	0.13

extensive search depth of 20,000 eliminates a shallow search depth as the origin of the observed increase in RDC fitness. The final structure that was produced by REDCRAFT is shown in green in Fig. 6. The structures in blue and purple correspond to the conformations of the two-state dynamic structure. Ignoring the dynamics in the analysis produces severely misaligned secondary structural elements. Once the location of the internal dynamics in the sequence has been identified, the structure of each fragment can be obtained independently without introducing further discrepancies. The dynamic profile indicates that the two fragments consist of residues 1–13 and 21–45. Residues 14–18 were ignored due to lack

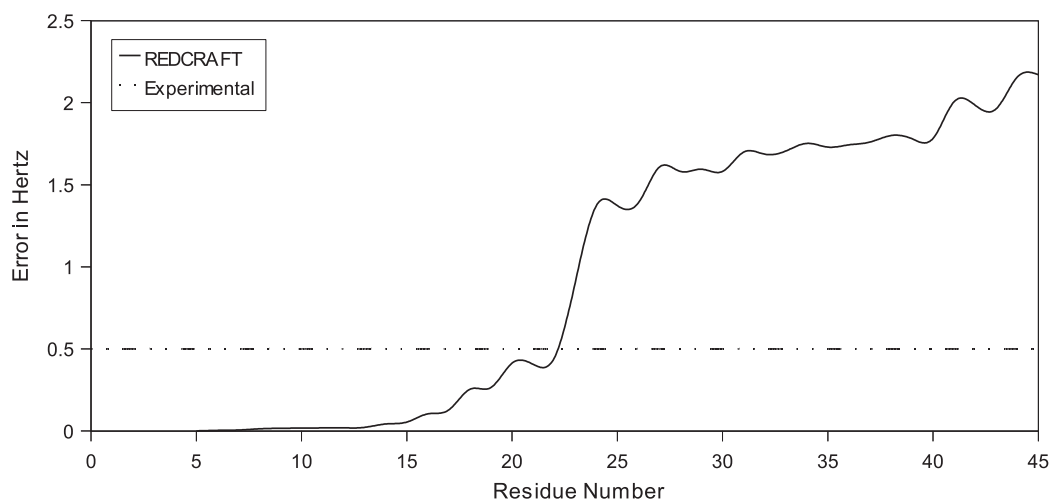
of dihedral restraints. Individual study of each fragment by REDCRAFT produced dynamic profiles with no unexpected internal anomalies.

Three sets of calculated order tensors are shown in Table 4. The first set of order parameters are based on the full structure (residues 1–45). Determining the structure of the entire protein while disregarding dynamics can produce incorrect order parameters. The second and third entries in Table 4 correspond to the order parameters that are produced after fragmented analysis of the data. Note the degree of similarity between the order parameters obtained from the corresponding fragments in Table 4 and the order tensors used to generate the data (shown in Table 2).

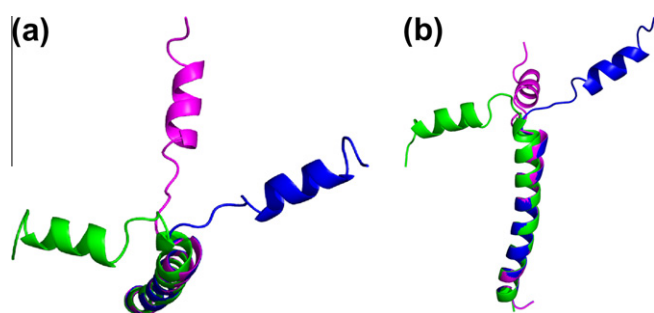
### 3.3. Experimental data

Previously reported [14,48] experimental RDC data and dihedral restraints were analyzed using REDCRAFT. The first experimental data were analyzed under the assumption of structural rigidity by producing a structure that spans residues 4–45. The initial three residues were not included due to insufficient data. The final REDCRAFT structure is shown in Fig. 7. The green structure in this figure is the REDCRAFT result, and the ensemble of 15 red structures is the previously determined mbPf1 result. The ensemble was calculated with Xplor-NIH using the same data and restraints as in this exercise, along with the estimates of the two order tensors determined from unassigned RDCs by the 2D-RDC method [14,52]. The REDCRAFT structure exhibited a backbone RMSD of 1.8–2.7 Å (Fig. 7a) with respect to the ensemble. This is within the 0.3–2.7 Å range observed between the ensemble structures. The main contributor to the structural variation is the loop region, which is poorly restrained. Superimposing the transmembrane region of this protein results in modest positional differences in the amphipathic region as shown in Fig. 7b. It is clear that although the structure of the loop region is poorly defined, enough information is present to restrain the orientation of the two helices with respect to each other.

The dynamic profile from REDCRAFT is shown in Fig. 8. This profile contains a clear indication of internal dynamics between the two helices. The RDC fitness demonstrates that the alignments of the two helices are described by two distinctly different order tensors. Individual study of each fragment by REDCRAFT produced dynamic profiles with final errors below the expected error. Three sets of order parameters were obtained in a manner similar to the previous section and are listed in Table 5.



**Fig. 5.** Dynamic profile from REDCRAFT using simulated data for the structure with internal dynamics. Motion is simulated at residue 19 and precedes a sudden jump in the REDCRAFT score, where it exceeds the experimental error. For mobile domains no single consistent order tensor exists, so the error grows during the analysis of these regions.



**Fig. 6.** The REDCRAFT structure (in green) computed using data from a protein undergoing internal dynamics is compared to the structures used to simulate a two-state jump (blue and purple). The structures are shown from different viewpoints in (a) and (b). These structures were aligned using the transmembrane helix. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Based on the observed dynamic profile in Fig. 8, we conclude that there are two distinct rigid fragments in mbPfl, corresponding to residues 4–13 and 20–45. The structure of each domain was determined separately using REDCRAFT. The first fragment, shown in Fig. 9a, has an excellent fit to the ensemble of previously reported structures, with a backbone RMSD of 0.33–0.98 Å with the ensemble compared to 0.05–1.45 Å among the ensemble. The second fragment, shown in Fig. 9b, also exhibits a reasonable fit to the ensemble, with an RMSD of 0.73–0.94 Å to the ensemble compared to 0.05–0.84 Å among the ensemble. It is important to note that helical structures with the backbone dihedral restraints utilized here can exhibit as much as 1.8 Å of deviation measured over the backbone atoms. Given the lack of dihedral constraints and exper-

imental data from the residues in the loop region, it is not possible to determine a unique structure for the portion of the protein between the two helices.

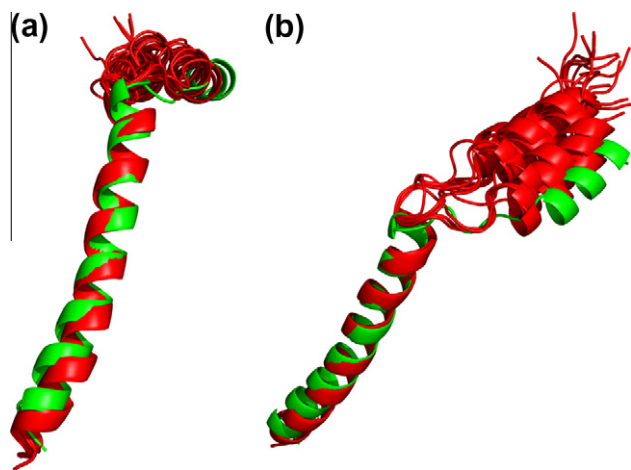
#### 4. Discussion

It is feasible to determine the structure of a membrane protein with sparse experimental data. The procedure described here is both reliable and complete. Assuming ideal geometries, a 46-residue protein can be described with 45 sets of backbone dihedral angles ( $\varphi$ ,  $\psi$ ) (the first and last torsion angles are inconsequential to the position of the backbone atoms). This presents a problem with 90 degrees of freedom. REDCRAFT's approach to structure determination allows the order tensors for the two alignment media to freely float without any *a priori* assumptions regarding their orientation or order parameters. This introduces ten additional degrees of freedom, five for each order tensor. Thus, our formulation of the problem can be summarized with 100 degrees of freedom, but there are only 90 experimental data. This is clearly an under-determined problem, and with no other constraints, there are a number of random-coil structures that consistently satisfy the RDC data. However, based on our observations, the dihedral restraints have reduced the effective degrees of freedom sufficiently to enable a meaningful structure determination. Furthermore, both REDCRAFT and Xplor-NIH produce structures with finite RDC fitness scores close to the experimental error. This provides additional evidence that the problem structure may not be under-determined, since that would produce an RDC fitness score of zero. An additional factor is the geometrical relationship between the N–C $\alpha$  and C $\alpha$ –C' bonds. If these two bonds were perpendicular to one another, a given set of torsion angles would have the maximum number of

**Table 4**

Order parameters from different regions of mbPfl under the assumption of motion with simulated data. Data is presented separately for the full protein and for the two rigid fragments that undergo motion with respect to one another. Residues from the mobile loop region are omitted for the fragment analysis.

Fragment	Medium	$S_{xx}$	$S_{yy}$	$S_{zz}$	$D_u$	$R$
Residue 1–45	M1	5.27E–05	5.06E–04	–5.59E–04	–6.81	0.54
	M2	–1.58E–04	–4.97E–04	6.55E–04	7.98	0.35
Residue 1–13	M1	5.95E–06	5.07E–04	–5.13E–04	–6.24	0.61
	M2	9.44E–05	4.99E–04	–5.93E–04	–7.22	0.45
Residue 23–45	M1	3.51E–04	5.03E–04	–8.54E–04	–10.40	0.12
	M2	–4.22E–04	–7.28E–04	1.15E–03	14.00	0.18



**Fig. 7.** The structure of mbPf1 computed by REDCRAFT under the assumption of a static structure, using experimental data, is shown from different viewpoints in (a) and (b). The REDCRAFT structure (in green) is superimposed with the previously determined ensemble of NMR structures (in red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

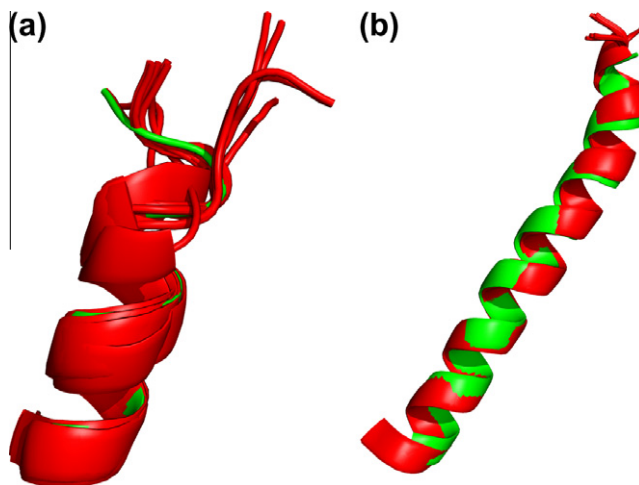
degrees of independence. However, the tetrahedral  $C^\alpha$  carbon imposes a bond angle of  $109.5^\circ$ , which renders two adjacent degrees of freedom somewhat dependent. This has been widely used in the calculation of protein structures determined by oriented sample (OS) solid-state NMR [55–57]. Finally, it is important to note that the dihedral restraints originally obtained from TALOS resulted from experimental data that are not directly acknowledged in this context. Based on this, the effective number of data points exceeds the total degrees of freedom of the problem.

Fig. 8 provides clear evidence to support the existence of internal dynamics in the residues between the transmembrane and the amphipathic helices. Although this feature of REDCRAFT has previously been demonstrated with other proteins and is theoretically sound, it could have been possible for membrane proteins, like the example mbPf1 coat protein, to yield an anomalous dynamic profile. The observed anomalies could be due to the none-globular nature of this protein, which has a small number of secondary structural elements, and orthogonal  $\alpha$ -helices. These new anomalies could have potentially produced dynamic profiles that falsely indicate internal motion. We investigated mbPf1 with synthetic

**Table 5**

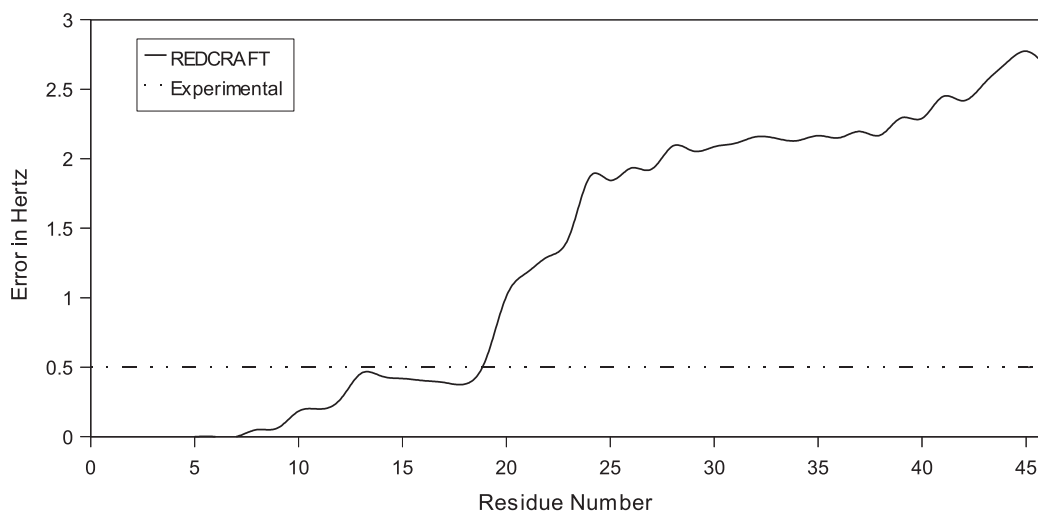
Principal order parameters reported by REDCRAFT from different regions of mbPf1 using experimental data. When treated separately, regions 4–13 and 20–45 clearly exhibit different order parameters.

Residues	Medium	$S_{xx}$	$S_{yy}$	$S_{zz}$	$D_a$	$R$
4–45	M1	1.42E–04	7.25E–04	–8.67E–04	–10.56	0.45
	M2	–2.12E–04	–2.77E–04	4.89E–04	5.95	0.09
4–13	M1	1.35E–04	4.85E–04	–6.20E–04	–7.55	0.38
	M2	–1.35E–04	–2.80E–04	4.15E–04	5.05	0.23
20–45	M1	2.70E–04	7.02E–04	–9.72E–04	–11.83	0.30
	M2	3.45E–04	9.03E–04	–1.25E–03	–15.22	0.30



**Fig. 9.** Structure of fragments 4–13 and 20–45 determined individually using experimental data are shown in (a) and (b) respectively. The REDCRAFT structure is shown in green, with the previously reported NMR structures in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

data to eliminate this as a possibility. If the observed anomaly in the dynamic profile is due to a global or local structural feature and not internal dynamics, then the same phenomenon should have been observed with synthetic data from a static structure. Furthermore, the dynamic profile observed with synthetic data from a dynamic structure is remarkably similar to the profile observed with the experimental data. Finally, a potential source of



**Fig. 8.** The dynamic profile of REDCRAFT in application to experimental data. The observed increase in error is consistent with that of simulated motion.

the observed jump in the dynamic profile could be overly restrained dihedral angles. Although not shown here, our studies included expanded dihedral restraints for the helical regions by as much as  $\pm 20^\circ$  beyond the original restraints. These restraints lead to a nearly identical dynamic profile, with a maximum RDC fitness of 2.4 Hz versus 2.7 Hz in the original analysis.

Additional evidence supports the existence of internal dynamics in mbPf1. The experimental heteronuclear  $^1\text{H}$ - $^{15}\text{N}$  NOEs for the backbone amide sites of Pf1 coat protein in SDS micelles have shown that the residues in the terminal regions as well as in the loop connecting the helices are more mobile than the other residues [53]. Recent NMR data also provide additional support for the existence of internal dynamics. The backbone amide signal intensity profile in the HSQC spectra was analyzed for the mbPf1 in isotropic bicelles by altering the  $q$  parameter, which indicates the molar ratio of long-chain to short-chain phospholipids. In micelles, with  $q = 0$ , resonances from both helices are readily observed. Increasing  $q$  leads to decreasing intensities from resonances in the transmembrane helix; these resonances are no longer observed at  $q = 1$  [48]. These experimental observations further strengthen our finding of motion in the amphipathic helix.

Many structure determination methods assume a static structure for the entire protein. However, the impact of this assumption on the final structure when using data collected from a dynamic structure can be extensive. Here we have presented one example of this effect based on synthetic data. Fig. 6 illustrates the simulated two-state jump motion of the amphipathic portion of mbPf1. Computing a structure using simulated data and insisting on the assumption of molecular rigidity has yielded the blue structure shown in Fig. 6. It is important to note that the orientation of the mobile region is strongly affected by internal motion; the amphipathic helix lies well outside the region spanned by the helices in the two states that were used to generate the synthetic data. There is no possibility of an inverted orientation due to our use of RDC data from two independent alignment media [54]. Although the effect of internal dynamics is normally discussed in terms of altered order parameters, a perturbation of the orientational component of the anisotropy is often accompanied by a change in the order parameters. Therefore, by taking internal dynamics into account in the calculations, it is possible to obtain a more accurate structure determination of a membrane protein.

## Acknowledgments

This work was supported by grant number R01GM081793 from National Institutes of Health to Dr. Homayoun Valafar and NIH grants to Dr. Stanley Opella. It utilized the Biomedical Technology Resource for NMR Molecular Imaging of Proteins at the University of California, San Diego, which is supported by Grant P41EB002031.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmr.2010.07.016.

## References

- [1] G. Bouvignies, P. Bernadó, S. Meier, K. Cho, S. Grzesiek, R. Brüschweiler, M. Blackledge, Identification of slow correlated motions in proteins using residual dipolar and hydrogen-bond scalar couplings, *Proc. Natl. Acad. Sci. U. S. A.* 102 (2005) 13885–13890.
- [2] G.M. Clore, C.D. Schwieters, Amplitudes of protein backbone dynamics and correlated motions in a small alpha/beta protein: correspondence of dipolar coupling and heteronuclear relaxation measurements, *Biochemistry* 43 (2004) 10678–10691.
- [3] J.R. Tolman, J.M. Flanagan, M.A. Kennedy, J.H. Prestegard, NMR evidence for slow collective motions in cyanometmyoglobin, *Nat. Struct. Biol.* 4 (1997) 292–297.
- [4] M. Bryson, F. Tian, J.H. Prestegard, H. Valafar, REDCRAFT: a tool for simultaneous characterization of protein backbone structure and motion from RDC data, *J. Magn. Reson.* 191 (2008) 322–334.
- [5] H. Valafar, J. Prestegard, REDCAT: a residual dipolar coupling analysis tool, *J. Magn. Reson.* 167 (2004) 228–241.
- [6] 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.
- [7] M. Andrec, P.C. Du, R.M. Levy, Protein backbone structure determination using only residual dipolar couplings from one ordering medium, *J. Biomol. NMR* 21 (2001) 335–347.
- [8] M. Assfalg, I. Bertini, P. Turano, A.G. Mauk, J.R. Winkler, H. Gray, N-15-H-1 residual dipolar coupling analysis of native and alkaline-K79A *Saccharomyces cerevisiae* cytochrome c, *Biophys. J.* 84 (2003) 3917–3923.
- [9] I. Bertini, C. Luchinat, P. Turano, G. Battaini, L. Casella, The magnetic properties of myoglobin as studied by NMR spectroscopy, *Chem.-Eur. J.* 9 (2003) 2316–2322.
- [10] G.M. Clore, C.A. Bewley, Using conjoined rigid body/torsion angle simulated annealing to determine the relative orientation of covalently linked protein domains from dipolar couplings, *J. Magn. Reson.* 154 (2002) 329–335.
- [11] G. Cornilescu, F. Delaglio, A. Bax, Protein backbone angle restraints from searching a database for chemical shift and sequence homology, *J. Biomol. NMR* 13 (1999) 289–302.
- [12] C.A. Fowler, F. Tian, H.M. Al-Hashimi, J. Prestegard, Rapid determination of protein folds using residual dipolar couplings, *J. Mol. Biol.* 304 (2000) 447–460.
- [13] 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.
- [14] R. Mukhopadhyay, X. Miao, P. Shealy, H. Valafar, Efficient and accurate estimation of relative order tensors from lambda-maps, *J. Magn. Reson.* 198 (2009) 236–247.
- [15] S.J. Opella, F.M. Marassi, Structure determination of membrane proteins by NMR spectroscopy, *Chem. Rev.* 104 (2004) 3587–3606.
- [16] H.M. Al-Hashimi, P.J. Bolon, J.H. Prestegard, Molecular symmetry as an aid to geometry determination in ligand protein complexes, *J. Magn. Reson.* 142 (2000) 153–158.
- [17] H.M. Al-Hashimi, A. Gorin, A. Majumdar, Y. Gosser, D.J. Patel, Towards structural Genomics of RNA: rapid NMR resonance assignment and simultaneous RNA tertiary structure determination using residual dipolar couplings, *J. Mol. Biol.* 318 (2002) 637–649.
- [18] H.M. Al-Hashimi, Y. Gosser, A. Gorin, W.D. Hu, A. Majumdar, D.J. Patel, Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings, *J. Mol. Biol.* 315 (2002) 95–102.
- [19] N. Tjandra, S. Tate, A. Ono, M. Kainosho, A. Bax, The NMR structure of a DNA dodecamer in an aqueous dilute liquid crystalline phase, *J. Am. Chem. Soc.* 122 (2000) 6190–6200.
- [20] A. Vermeulen, H.J. Zhou, A. Pardi, Determining DNA global structure and DNA bending by application of NMR residual dipolar couplings, *J. Am. Chem. Soc.* 122 (2000) 9638–9647.
- [21] J. Adeyeye, H.F. Azurmendi, C.J.M. Stroop, S. Sozhamannan, A.L. Williams, A.M. Adetumbi, J.A. Johnson, C.A. Bush, Conformation of the hexasaccharide repeating subunit from the *Vibrio cholerae* O139 capsular polysaccharide, *Biochemistry* 42 (2003) 3979–3988.
- [22] H.F. Azurmendi, C.A. Bush, Conformational studies of blood group A and blood group B oligosaccharides using NMR residual dipolar couplings, *Carbohydr. Res.* 337 (2002) 905–915.
- [23] H.F. Azurmendi, M. Martin-Pastor, C.A. Bush, Conformational studies of Lewis X and Lewis A trisaccharides using NMR residual dipolar couplings, *Biopolymers* 63 (2002) 89–98.
- [24] F. Tian, H.M. Al-Hashimi, J.L. Craighead, J.H. Prestegard, Conformational analysis of a flexible oligosaccharide using residual dipolar couplings, *J. Am. Chem. Soc.* 123 (2001) 485–492.
- [25] J. Meiler, J.J. Prompers, W. Peti, C. Griesinger, R. Brüschweiler, Model-free approach to the dynamic interpretation of residual dipolar couplings in globular proteins, *J. Am. Chem. Soc.* 123 (2001) 6098–6107.
- [26] W. Peti, J. Meiler, R. Brüschweiler, C. Griesinger, Model-free analysis of protein backbone motion from residual dipolar couplings, *J. Am. Chem. Soc.* 124 (2002) 5822–5833.
- [27] A. Saupe, G. Englert, High-resolution nuclear magnetic resonance spectra of orientated molecules, *Phys. Rev. Lett.* 11 (1963) 462–464.
- [28] J.H. Prestegard, H.M. al-Hashimi, J.R. Tolman, NMR structures of biomolecules using field oriented media and residual dipolar couplings, *Q. Rev. Biophys.* 33 (2000) 371–424.
- [29] N. Tjandra, S. Grzesiek, A. Bax, Magnetic field dependence of nitrogen-proton J splittings in N-15-enriched human ubiquitin resulting from relaxation interference and residual dipolar coupling, *J. Am. Chem. Soc.* 118 (1996) 6264–6272.
- [30] J.R. Tolman, J.M. Flanagan, M.A. Kennedy, J. Prestegard, Nuclear magnetic dipole interactions in field-oriented proteins – information for structure determination in solution, *Proc. Natl. Acad. Sci. U. S. A.* 92 (1995) 9279–9283.
- [31] J.A. Losonczy, M. Andrec, M.W.F. Fischer, J.H. Prestegard, Order matrix analysis of residual dipolar couplings using singular value decomposition, *J. Magn. Reson.* 138 (1999) 334–342.



- [32] J. Tolman, H. Al-Hashimi, NMR studies of biomolecular dynamics and structural plasticity using residual dipolar couplings, *Annu. Rep. NMR Spectrosc.* 51 (2003) 106–167.
- [33] J.R. Tolman, H.M. Al-Hashimi, L.E. Kay, J.H. Prestegard, Structural and dynamic analysis of residual dipolar coupling data for proteins, *J. Am. Chem. Soc.* 123 (2001) 1416–1424.
- [34] J.R. Tolman, Dipolar couplings as a probe of molecular dynamics and structure in solution, *Curr. Opin. Struct. Biol.* 11 (2001) 532–539.
- [35] J.R. Tolman, A novel approach to the retrieval of structural and dynamic information from residual dipolar couplings using several oriented media in biomolecular NMR spectroscopy, *J. Am. Chem. Soc.* 124 (2002) 12020–12030.
- [36] J.H. Prestegard, K.L. Mayer, H. Valafar, G.C. Benison, Determination of protein backbone structures from residual dipolar couplings, *Methods Enzymol.* 394 (2005) 175–209.
- [37] H. Valafar, K. Mayer, C. Bougault, P. LeBlond, F. Jenney, P. Brereton, M. Adams, J. Prestegard, Backbone solution structures of proteins using residual dipolar couplings: application to a novel structural genomics target, *J. Struct. Funct. Genomics* 5 (2005) 241–254.
- [38] S. Roy, A.G. Redfield, Nuclear overhauser effect study and assignment of D-stem and reverse-hoogsteen base pair proton resonances in yeast transfer Rna-Asp. 9 (1981) 7073–7083.
- [39] F. Delaglio, G. Kontaxis, A. Bax, Protein structure determination using molecular fragment replacement and NMR dipolar couplings, *J. Am. Chem. Soc.* 122 (2000) 2142–2143.
- [40] J.C. Hus, D. Marion, M. Blackledge, Determination of protein backbone structure using only residual dipolar couplings, *J. Am. Chem. Soc.* 123 (2001) 1541–1542.
- [41] Y. 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.
- [42] K. Lindorff-Larsen, R.B. Best, M.A. Depristo, C.M. Dobson, M. Vendruscolo, Simultaneous determination of protein structure and dynamics, *Nature* 433 (2005) 128–132.
- [43] L.C. Wang, B.R. Donald, Exact solutions for internuclear vectors and backbone dihedral angles from NH residual dipolar couplings in two media, and their application in a systematic search algorithm for determining protein backbone structure, *J. Biomol. NMR* 29 (2004) 223–242.
- [44] A. Brunger, P. Adams, G. Clore, W. DeLano, P. Gros, R. Grosse-Kunstleve, J. Jiang, J. Kuszewski, M. Nilges, N. Pannu, R. Read, L. Rice, T. Simonson, G. Warren, Crystallography & NMR system: a new software suite for macromolecular, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 54 (1998) 905–921.
- [45] P. Güntert, Automated NMR structure calculation with CYANA, *Methods Mol. Biol.* 278 (2004) 353–378.
- [46] S.H. Park, W.S. Son, R. Mukhopadhyay, H. Valafar, S.J. Opella, Phage-induced alignment of membrane proteins enables the measurement and structural analysis of residual dipolar couplings with dipolar waves and lambda-maps, *J. Am. Chem. Soc.* 131 (2009) 14140–14141.
- [47] R. Koradi, M. Billeter, K. Wuthrich, MOLMOL: a program for display and analysis of macromolecular structures, *J. Mol. Graphics* 14 (1996) 51–55.
- [48] S. Park, I.F. Marass, D. Black, S. Opella, Structure and dynamics of the membrane-bound form of Pf1 coat protein: implications for structural rearrangement during virus assembly, *Biophys. J.* (2010) 59.
- [49] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR* 6 (1995) 277–293.
- [50] M.F. Mesleh, G. Veglia, T.M. DeSilva, F.M. Marassi, S.J. Opella, Dipolar waves as NMR maps of protein structure, *J. Am. Chem. Soc.* 124 (2002) 4206–4207.
- [51] M.F. Mesleh, S. Lee, G. Veglia, D.S. Thiriot, F.M. Marassi, S.J. Opella, Dipolar waves map the structure and topology of helices in membrane proteins, *J. Am. Chem. Soc.* 125 (2003) 8928–8935.
- [52] X. Miao, R. Mukhopadhyay, H. Valafar, Estimation of relative order tensors, and reconstruction of vectors in space using unassigned RDC data and its application, *J. Magn. Reson.* 194 (2008) 202–211.
- [53] S. Lee, M.F. Mesleh, S.J. Opella, Structure and dynamics of a membrane protein in micelles from three solution NMR experiments, *J. Biomol. NMR* 26 (2003) 327–334.
- [54] H.M. Al-Hashimi, H. Valafar, M. Terrell, E.R. Zartler, M.K. Eidsness, J.H. Prestegard, Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings, *J. Magn. Reson.* 143 (2000) 402–406.
- [55] T.A. Cross, S.J. Opella, Protein structure by solid-state nuclear magnetic resonance. Residues 40–45 of bacteriophage fd coat protein, *J. Mol. Biol.* 182 (1985) 367–381.
- [56] J.R. Quine, M.T. Brennehan, T.A. Cross, Protein structural analysis from solid-state NMR-derived orientational constraints, *Biophys. J.* 72 (1997) 2342–2348.
- [57] A.A. Nevzorov, M.F. Mesleh, T.A. Cross, Structure determination of aligned samples of membrane proteins by NMR spectroscopy, *Magn. Reson. Chem.* 42 (2004) 162–171.