Available online at www.sciencedirect.com

ScienceDirect

Journal of Magnetic Resonance 191 (2008) 322–334

JMR
Journal of
Magnetic Resonance

www.elsevier.com/locate/jmr

REDCRAFT: A tool for simultaneous characterization of protein backbone structure and motion from RDC data

Michael Bryson [a], Fang Tian [b], James H. Prestegard [b], Homayoun Valafar [a,*]

[a] *University of South Carolina, Department of Computer Science and Engineering, 315 Main Street, Columbia, SC 29208, USA*
[b] *University of Georgia, Complex Carbohydrate Research Center, 315 Riverbend Rd, Athens, GA 30602, USA*

Received 7 November 2007; revised 23 December 2007
Available online 16 January 2008

Abstract

REDCRAFT, a new open source software tool that accommodates the analysis of RDC data for simultaneous structure and dynamics characterization of proteins is presented in this article. Simultaneous consideration of structure and motion is believed to be necessary for accurate representation of the solution-state of a protein. REDCRAFT is designed to primarily utilize RDC data from multiple alignment media in two stages. During Stage-I, a list of possible torsion angles joining any two neighboring peptide planes is ranked based on their fitness to experimental constraints; in Stage-II, a dipeptide fragment is extended by addition of one peptide plane at a time. The algorithm adopted by REDCRAFT is very efficient and can produce a structure for an 80 residue protein within two hours on a typical desktop computer. REDCRAFT exhibits robustness with respect to noise and missing data. REDCRAFT describes the overall alignment of the molecule in the form of an order tensor matrix and is capable of identifying peptide fragments with internal dynamics. Identification of the location of internal motion will permit a more accurate structural representation. Experimental data from two proteins as well as simulated data are presented to illustrate the capabilities of REDCRAFT in both structure determination and identification of the dynamical regions.

Keywords: Simultaneous; Backbone; Protein; Structure; Motion; Residual; Dipolar; Dynamics

1. Introduction

Structural elucidation, including complete structures of individual domains of proteins as well as the assembly of biomolecular complexes, is often a requisite step in understanding fundamental physiological processes, or in the design of drugs to combat a disease. Therefore, the development of methods leading to rapid, cost-effective structure elucidation is an important task. In addition, it is important to develop methods that can simultaneously deal with internal motion in these assemblies. Motions on a physiologically relevant time scale have been suggested to play an important role in the biological function [6,7]. Traditionally, characterization of inter-molecular dynamics has been separated from structure elucidation, increasing the cost and time of these studies. Furthermore, conceptually, it is difficult to separate structure from dynamics since observables used for structure determination are perturbed by motion, and therefore any attempt at structure elucidation that disregards the dynamics (or vice versa) may produce faulty results [8]. In conventional high resolution NMR the set of experiments required to assess the dynamic properties of a molecule are often disjoint from the set of experiments required for the structural elucidation. Therefore there is ample room for improvement in NMR-based structural and dynamic analysis. The recent reintroduction of residual dipolar couplings (RDC) [1,2] to the biomolecular NMR provides one opportunity for this improvement, provided the appropriate analysis tools are available. The success of REDCRAFT in structure determination of protein backbone structure has been previously demonstrated

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

1090-7807/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2008.01.007

[48,49]. In this report we describe the algorithm of this analysis in detail while introducing its ability in identification of internal motion during the course of structure determination.

REDCRAFT (Residual Dipolar Coupling Residue Assembly and Filtering Tool) is a new open source analysis tool that accommodates the analysis of RDC data for simultaneous structure characterization and identification of dynamics of proteins and polypeptides. Innate properties of RDC data combined with our proposed analysis (REDCRAFT) provides the features listed below:

- A simplified force field and therefore energy landscape.
- True *De novo* structure determination without any a-priori knowledge of the secondary structural elements; only reliance on the ideal structure of a peptide plane is used.
- Simultaneous structure determination and identification of internal motion.
- Robustness with respect to error and missing data.
- Practical computational time complexity.

Since the task of protein structure determination is a computationally demanding one, REDCRAFT has been implemented with a Linux Cluster in mind. Users of this software can perform relatively superficial analysis on a desktop computer and then easily move to a more thorough search on a Linux Cluster. The complete software binary, source code, and manuals are available for public access via the web at http://ifestos.cse.sc.edu. REDCRAFT is distributed with interfacing tools to provide REDCAT [4] input files, and XPLOR-NIH [5] constraint files for further refinement.

2. Theoretical background and preliminaries

2.1. Residual dipolar coupling

Residual dipolar coupling (RDC) had been observed as early as 1963 [9] in a nematic liquid crystal environment, but a number of recent applications [1,2,10–12] have ignited their widespread use in characterization of biomolecules. More specifically, RDCs have been used in studies of carbohydrates [13,14], nucleic acids [15,16], and proteins [14,17–22].

RDCs arise from the interaction of two magnetically active nuclei in the presence of the external magnetic field of a NMR spectrometer [1–3,23]. This interaction is normally reduced to zero in simple aqueous solutions due to the isotropic tumbling of molecules. The introduction of partial order by slight molecular alignment will result in nonzero RDC observables. This partial order can be introduced by exploiting the inherent magnetic anisotropic susceptibility of the molecule [1], incorporating artificial tags with high magnetic anisotropic susceptibility [24], or using a liquid crystalline media [25]. Once restored, RDCs can be measured relatively easily and represent an abundant source of highly precise information on parameters such as the relative orientations of different inter-nuclear ‘bonds’ within the molecule or internal motion. Eq. (1) describes the time average of the RDC interaction between a pair of spin 1/2 nuclei as observed through contributions to the splitting of resonances in the absence of spin decoupling.

$$D_{ij} = \frac{-\mu_0 \gamma_i \gamma_j h}{(2\pi r)^3} \left\langle \frac{3\cos^2(\theta_{ij}(t)) - 1}{2} \right\rangle \tag{1}$$

Here, D_{ij} denotes the residual dipolar coupling in units of Hz between nuclei i and j, γ_i and γ_j are nuclear gyromagnetic ratios, r is the internuclear distance (fixed for directly bonded atoms) and $\theta_{ij}(t)$ is the time dependent angle of the internuclear vector with respect to the external magnetic field. The angled brackets signify the time average of the quantity. When a sufficient number of RDC data are assigned to specific locations in a known structure, the elements of the order tensor can be obtained [4,12,26–28]. Conversely, given a structure and the elements of the order tensor, values expected for various RDCs can be calculated easily.

RDCs play an increasingly important role in the NMR structure determination because of their unique advantages over the traditional NOE data [29]. However, structure determination primarily based on RDC data requires new programs that operate in fundamentally different ways from those that use NOE data. Some of these have been put forward [21,22,30–34]. Information richness and complexity of RDC data however, continue to be a challenge for the analysis tools in existence today. Additionally, development of new experimental methods of acquiring RDC data with improved accuracy and precision necessitates a parallel pursuit of information extraction methods.

2.2. Molecular frame, alignment frame, and order tensor

Traditionally, upon successful determination of a structure, its atomic coordinates are described within some arbitrary coordinate system. The selection of a coordinate system is inconsequential since this structure is independent of any rotation or displacement within this frame. This arbitrarily selected coordinate system is referred to as the “molecular frame” (MF). On the other hand, since RDC data are capable of describing the preferred alignment of the molecule, a more descriptive frame can be selected in which the structure of the molecule of interest is described in the appropriate orientation. Here we define this more descriptive frame as the “principal alignment frame” (PAF).

Alignment properties of a molecule can be described in the form of an order tensor matrix (OTM). Reformulation of Eq. (1) in a matrix form clearly collects and defines the order tensor matrix as shown in Eq. (2). Here X is the normalized coordinates of the vector between the pair of interacting nuclei and S is a 3×3 order tensor matrix encapsulating the alignment properties of the molecule.

Jacobi [35] transformation of this symmetric and traceless matrix can isolate two important types of information. The three elements of the resulting diagonal matrix (S_{xx}, S_{yy}, and S_{zz}) are referred to as the principal order parameters (POM) and reflect the strength of alignment along each of the principal axes x, y, and z within the PAF. The R matrix contains information about the orientational relationship between any arbitrary MF and the PAF in terms of three Euler angles, α, β, and γ. Analysis of these two parts will allow assessment of the strength of alignment for various parts of a molecule (and therefore their relative motions) and the preferred direction of alignment with respect to an arbitrarily selected MF. This order matrix can also be used to calculate theoretical RDC observables for any additional parts of a structure.

$$D \propto X^{\mathrm{T}} \times \begin{bmatrix} s_{xx} & s_{xy} & s_{xz} \\ s_{xy} & s_{yy} & s_{yz} \\ s_{xz} & s_{yz} & s_{zz} \end{bmatrix} \times X$$

$$= X^{\mathrm{T}} \times R \times \begin{bmatrix} S_{xx} & 0 & 0 \\ 0 & S_{yy} & 0 \\ 0 & 0 & S_{zz} \end{bmatrix} \times R^{\mathrm{T}} \times X \tag{2}$$

$$R = R_z(\alpha) R_y(\beta) R_z(\gamma) \tag{3}$$

Various methods for obtaining the order tensor matrix describing alignment of a subject protein have appeared in the literature within recent years [4,28]. While these methods provide robust and reliable means of obtaining the order matrix, they require an existing structure and assignment of the RDC data to specific sites. REDCRAFT is unique in that it provides a simultaneous description of the order tensor and the molecular structure.

The dependence of an order tensor on the overall motion of a rigid entity has been mathematically illustrated previously [1,36,37]. While it is relatively straightforward to determine the perturbation of a given tensor by a well defined internal motion, the reverse is much more difficult to accomplish and is the subject of numerous investigations [38,39]. REDCRAFT is able to identify regions that exhibit internal motion and allows structural characterizations of these regions with different alignment tensors, although it will not provide a full description of the nature of the internal motion. Isolated structure characterization of different regions of the same protein will provide a more accurate and meaningful characterization of the structure while allowing for further study of exact nature of the internal motion through comparison of the two order tensors [36].

2.3. Order tensor solution

The orientation of any rigid molecular entity is embedded within the OTM as shown in Eqs. (2) and (3) above. This information can be extracted from a linear system of equations shown in Eq. (4). The independent elements of the order tensor matrix (OTM) are represented here by a reduced set of s_{ij} elements using the symmetric and traceless properties of the OTM. The Cartesian representation of each individual interaction vector is denoted by (x_i, y_i, and z_i) while the experimentally determined dipolar couplings and error in measurements are denoted by D_i and ε_i, respectively. $D_{\max}$ is the maximum observable dipolar coupling for a particular pair of nuclei at 1 Å separation and r is the actual separation of the two interacting atoms. Note that our representation of the RDC interaction slightly differs from those previously presented in the literature. Usually, the RDC interaction is described in terms of the direction-cosines of the interacting vectors and exhibits a r^{-3} dependence on the distance of the interacting pair of atoms. Our representation is more computationally friendly since evaluation of trigonometric terms is costly. During the course of structure calculation, it is typical for the set of equations presented in Eq. (4) to be calculated on the order of billions of times.

$$\begin{cases} (y_1^2 - x_1^2)s_{yy} + (z_1^2 - x_1^2)s_{zz} + 2x_1y_1s_{xy} + 2x_1z_1s_{xz} + 2y_1z_1s_{yz} = (D_1 \pm \varepsilon_1)r^5/D_{\max} \\ \vdots \\ (y_n^2 - x_n^2)s_{yy} + (z_n^2 - x_n^2)s_{zz} + 2x_ny_ns_{xy} + 2x_nz_ns_{xz} + 2y_nz_ns_{yz} = (D_n \pm \varepsilon_n)r^5/D_{\max} \end{cases} \tag{4}$$

The system of equations described in Eq. (4) can be solved using singular value decomposition (SVD) [35]. This method can efficiently find the best OTM solution for an under-determined or over-determined system [4,28,35,40]. This order tensor matrix can then be used to back calculate RDCs as described previously [1,4,28]. During the assembly of peptide planes, the fitness of any trial geometry can be established by obtaining an RMSD error between the experimentally collected and the back calculated RDCs. This score can be interpreted as the fitness of a proposed structure to the experimental data.

2.4. Experimental data

REDCRAFT is designed to perform the task of backbone structure determination of a protein primarily based on RDC data. The current version of REDCRAFT can accommodate as many as six RDC data per peptide plane as shown in Fig. 1. Here RDCs illustrated in green represent pieces of data that can be collected via the N–H amide group for any peptide plane (except one involving the proline nitrogen). Yellow and red data will not be available, or are omitted for peptide planes that are at the C or N ter-

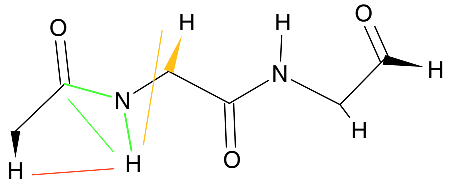

Fig. 1. Illustration of RDC data collected for our experiments. Green data are those that can be collected for all residues of a fragment. The red and yellow data will be ignored for the beginning and last residues of a fragment, respectively. (For interpretation of the references in color in this figure legend, the reader is referred to the web version of this article.)

mini of a fragment under study. The maximum number of RDC data available for analysis of a fragment of size *n* amino acids in the current REDCRAFT implementation is $6n-3$. The data reported for glycines are slightly different from other amino acids. This is due to the fact that glycines possess two H_α protons. Experimentally the sum of RDCs ($D_{C\alpha1-H\alpha1} + D_{C\alpha2-H\alpha2}$, $D_{H\alpha1(i-1)-HN} + D_{H\alpha2(i-1)-HN}$, and $^3J_{HN-H\alpha1} + {}^3J_{HN-H\alpha2}$) can be measured more accurately from the separation of the outermost peaks in a four line multiplet while two inner peaks are often overlapped.

REDCRAFT is capable of analyzing RDC data collected from multiple alignment media. Data from multiple alignment media serve to resolve a number of potential problems [43]. The first problem inherent to the RDC observable is its non-uniform sensitivity to the orientation of vectors in the alignment frame. Combining data from a number of independent alignment media could assist in resolving this problem since the same vector located in the insensitive region of the first alignment is likely to fall into a more sensitive region of the second alignment. The second problem which is also inherent to the nature of the dipolar interaction is its insensitivity to inversion about each of the principal axes of alignment tensor. This will cause degeneracies in allowed torsions. This phenomenon is especially problematic when the C_α–C′ bond coincides with a principal axis. A 180° inversion about this bond can change an α-helical like geometry to a β-strand like geometry without violating any of the experimental constraints.

In addition to RDCs, REDCRAFT is capable of utilizing other types of data such as distance constraints or three bond scalar coupling $^3J_{HN-H}$ [41,42] data. A small set of coincidentally acquired NOE data can restrict translational degrees of freedom for fragments by utilizing programs such as XPLOR-NIH [5] as demonstrated before [48,49]. The amino acid designation in REDCRAFT has been simplified to include only three classes based on side chain characteristics, glycine, proline or alanine (anything that is not a glycine or proline is considered to be an alanine). This is possible since our investigation is focused on the structure determination of protein backbone and the data utilized are originated from the backbone atoms of the protein. Finally, the current version of REDCRAFT requires a-priori knowledge of RDC assignments.

2.5. Treatment of error

Since SVD analysis places equal weighting on each RDC entry, proper treatment of the experimental error is necessary. Proper treatment of error is especially important since RDC data sets utilized by REDCRAFT are heterogeneous in the following two ways:

1. RDC data sets span different ranges depending on the gyromagnetic ratios of the two interacting nuclei and their separation distances. For example, the maximum observable RDCs for C_α–H_α and C′–N differ by more than an order of magnitude.
2. RDCs are collected from distinctly different experiments with different errors due to varying sensitivities, spectral resolutions, etc.

These problems can be rectified through appropriate scaling. The scaling factor intended to alleviate the first problem is simply derived from the ratio of the quantities D_{max}/r^3 using N–H as the internal reference. The distance between two interacting pair of atoms such as H_α and HN may depend on the local structure of a protein and can be calculated during the course of structure calculation. Simple scaling based on the mentioned factor should be adequate for proper treatment of inhomogeneous sets of RDC data under ideal conditions. However under more pragmatic situations, RDC data acquired from different NMR experiments may not adhere to similar acquisition conditions such as signal to noise and digital resolution, resulting in different errors for different measurements. Our implementation of REDCRAFT enables this degree of freedom by reporting independent set of errors not only across different experiments but also for an individual datum. Therefore, another scaling factor is introduced to alleviate the second problem by normalizing the estimate of experimental error supplied along with each RDC to the N–H RDC error. The final scaling factor (S_n) can be calculated by combining the above two scaling factors as shown in Eq. (5), where X denotes observables other than N–H RDCs and ε′ denotes the estimated experimental error for observation X after internal normalization to the N–H observables. The final scaling factor S_n is independent of *r* and can therefore be applied to non-bonded atom pairs.

$$\begin{aligned} S_n &= \left(\frac{D^{NH}_{max} \times r_X^3}{D^{X}_{max} \times r^3_{NH}}\right) \times \left(\frac{\varepsilon_{NH}}{\varepsilon'_{X}}\right) \\ &= \left(\frac{D^{NH}_{max} \times r_{X}^3}{D^{X}_{max} \times r^3_{NH}}\right) \times \left(\frac{\varepsilon_{NH}}{\varepsilon_X \times \frac{D^{NH}_{max} \times r^3_X}{D^X_{max} \times r^3_{NH}}}\right) = \frac{\varepsilon_{NH}}{\varepsilon_X} \end{aligned} \qquad (5)$$

REDCRAFT calculations assume standard covalent geometries of peptide planes and the absence of motion within these planes. Violations of these assumptions can introduce additional modeling or structural noise. Variations of N–H bond lengths by just 0.02 Å, for example, would introduce a 6% error. Errors due to a uniform local motion would be absorbed into magnitudes of order parameters. However, it has been documented that local oscillations of N–H bonds are larger than those for C–N bonds [44,45] and could potentially introduce additional errors. To compensate for these modeling errors we often set generous estimates of error (10% of the range of couplings observed) even when measurements are more precise. This error can be increased or decreased based on the quality of observed data.

Orientational sensitivity of RDC data can also be cited as a component influencing the outcome of the analysis of RDC data. For example, 1 Hz experimental error would translate to a much smaller spatial uncertainty if the vector of interest is oriented close to 54.7° from all three principle axes (the magic angle) versus if the vector is parallel with any of the three principal axes of the alignment. This problem is difficult to address because of the lack of a priori knowledge of orientation of an interaction vector within the PAF. RDC data from multiple alignment media is anticipated to mitigate the effect of this type of variation since a vector that falls in insensitive region of one alignment medium may be likely to fall in a more sensitive direction of the other alignment media.

3. REDCRAFT algorithm

REDCRAFT takes a fundamentally different approach to structure determination of macromolecules compared to the ones in existence. The general approach adopted by REDCRAFT, avoids the exponential complexity of the protein folding problem by eliminating a large number of disallowed geometries quickly. For example, elimination of a single torsion angle at one residue within a fragment of size P peptide planes can eliminate a total of $N^{2(P-2)}$ geometries, where N is the number of possible geometries at each residue. REDCRAFT operates in two distinct stages: Stage-I and Stage-II. During Stage-I, a list of all possible torsion angles joining any two neighboring peptide planes is first pruned (using data such as scalar couplings and Ramachandran space) and then ranked based on structural fitness. Stage-II of REDCRAFT extends a given fragment of size N peptide planes (initially a dipeptide seed) by addition of one peptide plane at a time. The planes must be oriented in a way to satisfy the RDC data; in effect this restricts possibilities for ϕ and ψ angles connecting the planes. If the entire protein can be assembled in this way, an accurate backbone structure is produced. However, in practice the structure elucidation of the protein backbone is accomplished through the assembly of fragments resulting from natural termination points such as prolines or loops (due to severe lack of RDC data), or presence of internal motion. Proper orientation of different fragments in space with respect to each other can be obtained by superimposing the alignment tensors for each fragment. A minimum set of NOEs and other fragment connectivity restraints can be used to translate individual fragments into positions appropriate for a good representation of the structure of the whole protein. These two stages are discussed in detail as follows.

3.1. Stage-I—constructing initial local geometry lists

The first stage of REDCRAFT aims to generate and rank the list of possible torsion angles connecting two adjacent peptide planes based on the observed experimental data. The initial list can be prepared in observation of information such as *a-priori* knowledge of the secondary structural elements or torsion angle constraints [20,46]. In the absence of these information, the current version generates an exhaustive list of all possible torsion angles between each set of connected peptide planes in 10° steps. This will give rise to an initial list consisting of 1296 (36 × 36) distinct sets of torsion angles for each dipeptide plane. Then this initial list may be further reduced using the Ramachandron filter that eliminates torsion angles that are not consistent with the Ramachandran space defined for a given amino acid type. The acceptable space for each of the currently supported amino acid types is based on previous results [47]. In addition to the Ramachandran statistics, when $^3J_{\mathrm{HN-H\alpha}}$ scalar couplings are available, they are used to further eliminate unfit phi torsion angles according to the Karplus equation [42].

Stage-I is concluded by ranking the surviving list of torsion angles for each dipeptide plane. Here, local geometries are ranked based on the RMSD score calculated between the experimental and back-calculated RDC data. The geometry exhibiting the best agreement to the RDC data appears as the best ranked structure while the geometry exhibiting the worst score is placed at the bottom of the list. Although these lists normally contain a varying number of surviving geometries (nearly 100 entries) for different dipeptides, to facilitate further discussion we will assume this number to be constant and denoted by N. These lists of acceptable torsion angles between each pair of neighboring residues are carried to the Stage-II calculation. Stage-I analysis is fast and can be conducted on the order of minutes in application to a 100 residue protein on a typical desktop computer. This stage is also very flexible and users can design and integrate custom filters based on their specific data. The operational flowchart of the Stage-I is shown in Fig. 2a.

3.2. Stage-II—extension of a seed peptide

Under ideal circumstances, the top entry of each list generated from Stage-I could be used to describe structure of the unknown protein. However in practice, due to the presence of experimental error and structural noises, the top entry is almost always not the globally optimal structure. Good quality data will ensure the ascent of the correct structure toward the top of this list. Increasing level of noise will descend the correct structure toward the bottom of this list. The search for the globally fittest geometry is therefore the subject of the Stage-II and constitutes the most computationally intensive portion of this program.

The operational flowchart of Stage-II is illustrated in Fig. 2b. Since it is practically impossible to consider every combination of torsion angles for fragments larger than five residues REDCRAFT operates through iterative addition of a peptide plane to the end of a seed fragment (initially a dipeptide unit). Addition of a peptide plane ($i + 1st$ peptide plane), which can assume N distinct conformations from the analysis of Stage-I to a fragment with top M candidates will create a total of $M \times N$ candidate fragments. The structural quality of a candidate is determined by calculating the agreement of experimental RDCs to the

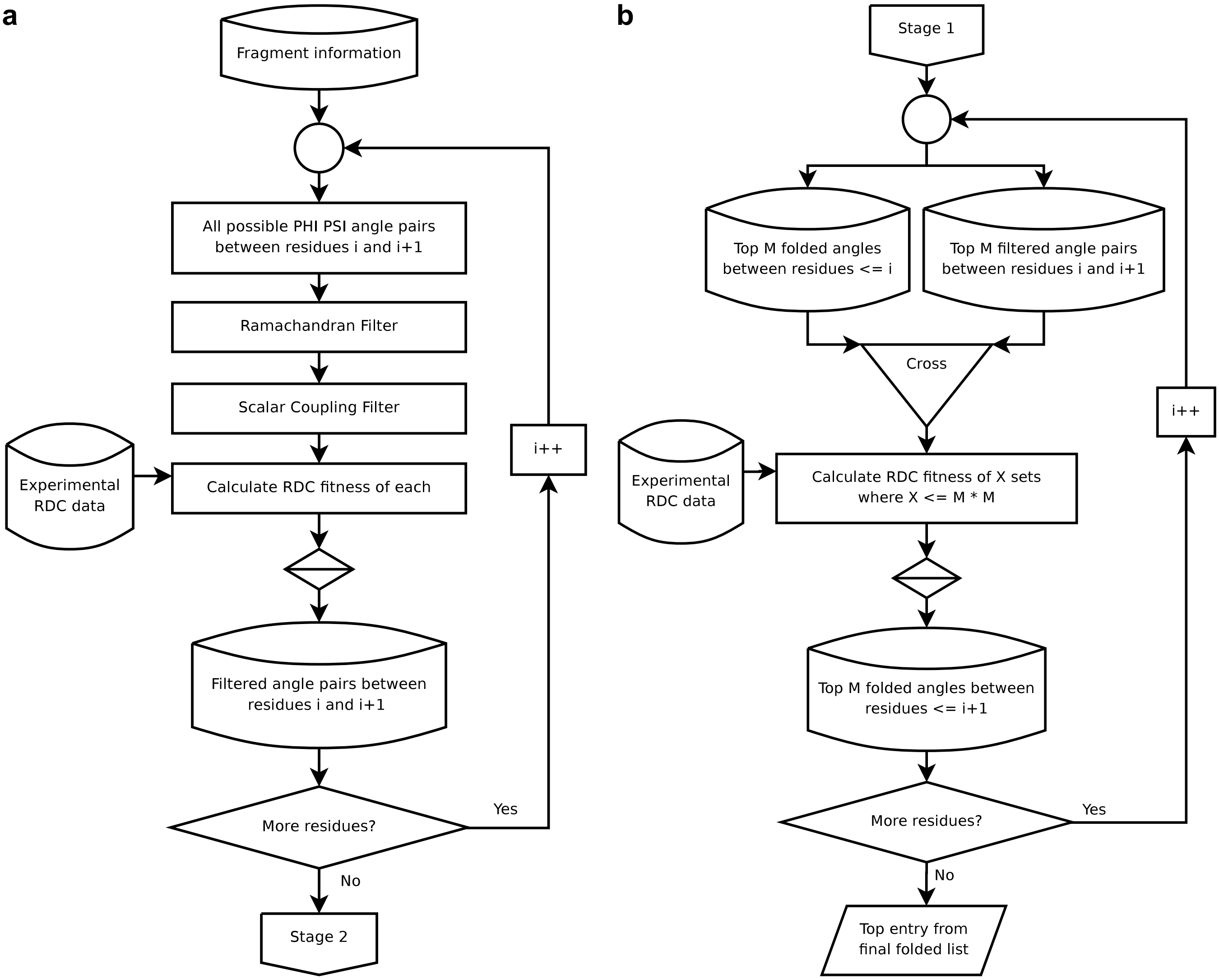
a
Fragment information
All possible PHI PSI angle pairs between residues i and i+1
Ramachandran Filter
Scalar Coupling Filter
Experimental RDC data
Calculate RDC fitness of each
i++
Filtered angle pairs between residues i and i+1
More residues?
Yes
No
Stage 2
b
Stage 1
Top M folded angles between residues <= i
Top M filtered angle pairs between residues i and i+1
Cross
i++
Experimental RDC data
Calculate RDC fitness of X sets where X <= M * M
Top M folded angles between residues <= i+1
More residues?
Yes
No
Top entry from final folded list

Fig. 2. Operational flow chart for the (a) stage 1 and (b) stage 2 of REDCRAFT analysis.

back-calculated RDCs using the best order tensor solution obtained as described before [4,28], and then it is used to rank a new list of candidates for the fragment of size $i+1$ peptide planes. By considering only the M most favorable sets of angles for the section of the fragment that has already been evaluated, an exponentially large set of unfavorable geometries are eliminated. A large M will result in a more thorough and exhaustive search of possible geometries at the cost of longer execution time, while a small M will constitute a superficial search with very short computation time. The depth of search is normally selected to be a number less than 1000 for good quality data and 1000–10,000 for relatively noisy data. Practically, addition of a peptide plane will often create in excess of 1,000,000 structures for evaluation, of which, the top 1000–10,000 survivors (depending on the search depth) propagate to the next round of expansion. Fragment extension continues iteratively as shown in Fig. 2b until arrival of a termination condition. This algorithm renders REDCRAFT much more immune to errors and missing data. Initially, the true structure may not be the most optimal point, but is carried forward as a viable solution during the early rounds. During fragment expansion, additional data will elevate the true structure from an early suboptimal geometry to the global optimal solution. It is important to reiterate that sources of error are not only the experimental uncertainties, but also include our assumptions of standard geometries of peptide planes and the absence of motion within these planes. Deviations from these assumptions can introduce additional modeling errors.

It is appropriate and necessary to study the topography of the RDC penalty landscape to justify our optimization strategy. Here we use data from Rubredoxin because this protein produces data of highest quality. Combining RDC data from peptide planes 2 and 3, an energy landscape can be constructed as a function of the connecting torsion angles. Fig. 3a illustrates the RMSD score landscape between the back-calculated and experimentally col-

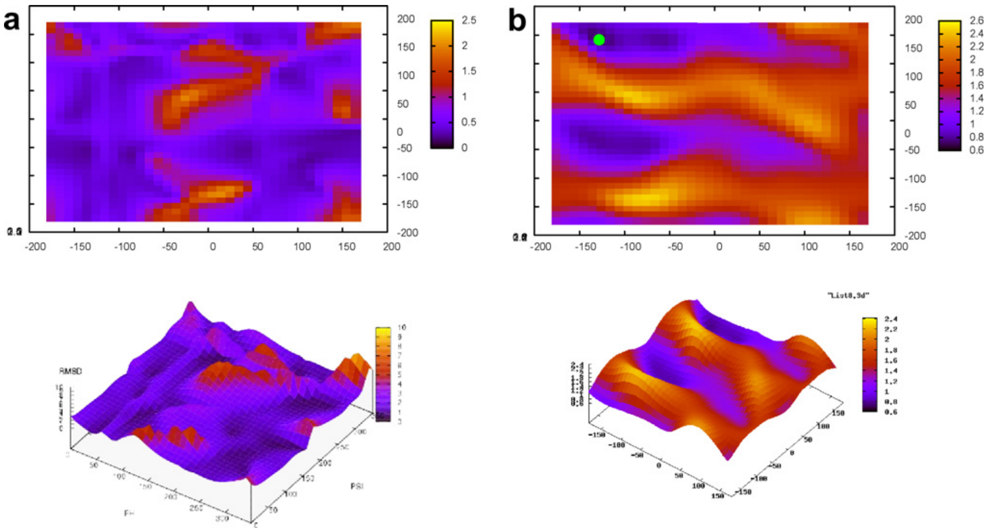

Fig. 3. Residual dipolar coupling energy landscape for torsion angles of the second residue (a) using data from peptide planes 2 and 3, (b) using data from peptide planes 2–8.

lected data for any given combination of the torsion angles. The torsion angles corresponding to the solution structure of this protein is shown with a green dot in Fig. 3b. Fig. 3b illustrates the same energy (RMSD score) landscape as Fig. 3a using RDC data from peptide planes 2–8, where all torsion angles except the first set are fixed. It is clear that a meaningful penalty landscape for torsion angles connecting peptide planes 2 and 3 emerges only after the inclusion of data from five additional peptide planes. Premature determination of the torsion angles joining these two peptide planes based on local data is at the risk of severe error. This phenomenon can be observed with even the most meticulously collected data and it therefore necessitates implementation of the Stage-II of our search.

4. Results

Data from two previously reported proteins are presented here to illustrate REDCRAFT's features. The first protein consists of 53 residues and is a Zn-substituted, triple mutant (W3Y, I23V, and L32I) of Rubredoxin (PDB code 1RWD) from *Pyrococcus furiosus*. The second protein, *PF1061* (PDB code 1SF0) is also from *P. furiosus* and is a 9 kDa protein that was previously not well annotated, and had a nearest structurally characterized homolog with only 31% sequence identity. For more information regarding the details of biological findings, sample preparation, data collection and analysis for these two proteins, please refer to the following publications [14,48,49].

The data from Rubredoxin will be used to illustrate the efficient construction of a peptide backbone structure while examining the effect of missing data for a small fragment. The data from *PF1061* will be used to illustrate simultaneous structure characterization and identification of internal motion. While our analysis is potentially capable of simultaneous assignment of RDC data to sequence specific sites by examining alternate sets of connectivities [14], here we only discuss the analysis of segments with previously established connectivities.

4.1. Robustness to missing data and noise during backbone structure determination

Our first discussion focuses on the topic of missing data by utilizing experimental data for Rubredoxin. Rubredoxin naturally divides into six fragments due to reduced number of experimental RDCs that occur at five prolines in the sequence (due to the absence of an H–N amide group on prolines). In addition, no experimental data was observed for residues 25–27. Here however, we illustrate the immunity of REDCRAFT to small sections of missing data by first partitioning it into two fragments (instead of 6) with each half (residues 2–24 and 28–50) approximately 25 residues in length. This exercise forces REDCRAFT to encounter a number of residues (such as prolines) with at most one RDC datum.

Structure determination of Rubredoxin proceeds normally as expected until the discovery of data-sparse regions such as prolines. Upon addition of a proline residue to the extending fragment, a great structural ambiguity is introduced at the torsion angles because of availability of only the C_α–H_α RDC at this peptide plane (Fig. 4a). However, as the fragment extension proceeds, enough RDC data accumulate past the point of missing data, and ambiguities in geometries are eliminated. The correct geometry is elevated to the top of the surviving list of conformations (Fig. 4b and c).

Integration of these two fragments into a final structure is possible since RDC data can not only determine the local structure of a fragment but also the orientation of each fragment in space. Therefore, once the two fragments are

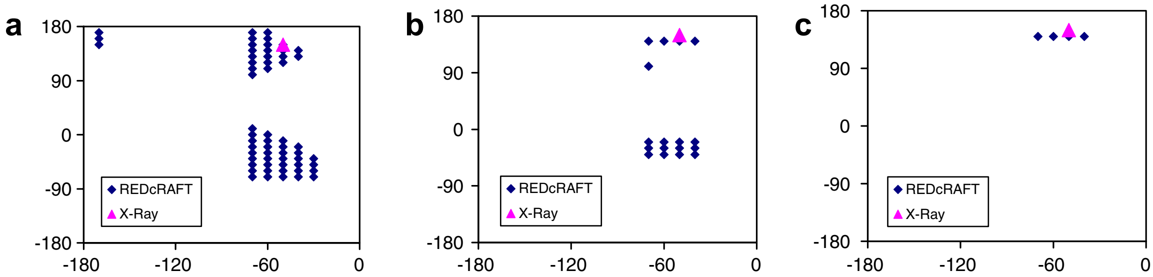

Fig. 4. Solution space produced by REDCRAFT at (a) position P33 with only C_α–H_αRDC, (b) position P33 after considering data from residue D34, (c) position P33 after considering data from residues D34 and D35.

constructed and oriented, the torsion angles of the three connecting residues can in principle be constrained. The local geometries of the three missing residues (25, 26, and 27) between the two fragments were constructed using a Monte Carlo technique (a utility included in REDCRAFT package). Sets of three torsion angles that are consistent with other available data (Ramachandran and scalar coupling filters) were randomly generated. These candidates were then tested to determine if they could connect these two fragments that were oriented in space by the alignment frames computed in REDCRAFT.

The backbone RMSD between 1BRF (a high resolution crystal structure of Rubredoxin) and each of the two fragments (residues 2–24 and 27–50) determined by REDCRAFT was less than 1.9 Å. After minimization within XPLOR-NIH, the final structure exhibited 1.81 Å (Fig. 5) similarity with 1BRF structure. For a more detailed description of the minimization routine please refer to the following work [49].

In order to establish the robustness of REDCRAFT to various levels of noise, we utilize simulated data for a helical protein (1A1Z) with controlled levels of uniformly distributed noise. The structure 1A1Z was used to generate RDC data with two hypothetical alignment tensors that reflect realistic alignments as shown in Table 2. Three noise levels of ±1, ±2, and ±3 Hz were explored. It is important to note that these noise levels only become meaningful when compared to the range of the observed RDC data. For example, a ±2 Hz error (windows size of 4 Hz) constitutes nearly 55%, 15%, 39%, 8%, 33%, and 30% of the total range of observed RDC data originated from C–N, N–H, C–H, C_α–H_α, $H_\alpha(i)$–H(i) and $H_\alpha(i-1)$–H within the first alignment medium. REDCRAFT exhibits adequate immunity to the noise level as indicated in Table 1 below. This table lists the backbone RMSD between the original structure of 1A1Z and the structure obtained by REDCRAFT.

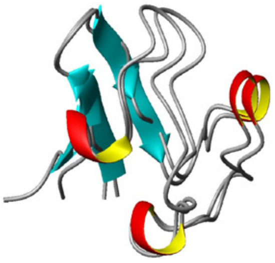

Fig. 5. Cartoon plot of the structure obtained from REDCRAFT (1RWD) and the crystal structure of the wild type protein (1BRF) exhibiting 1.81 Å RMSD over the backbone C_α atoms of residues 1–50.

4.2. Simultaneous characterization of structure and identification of motion

In this section, simulated data and experimental RDCs from a structural genomics target *PF1061* are presented to demonstrate the capabilities of REDCRAFT in performing simultaneous structure characterization and identification of internal motion.

4.2.1. Simulated data

Simulated data can be invaluable in ensuring proper functioning of REDCRAFT. Controlled corruption of data through systematic introduction of noise or effect of motion on the RDC data will allow a proper study of the capabilities and susceptibilities of REDCRAFT. In addition, the utility of simulated data can be instrumental in providing a theoretical explanation to the outcomes of REDCRAFT analyses.

An 83 residue α-protein (PDB code: 1A1Z) has been arbitrarily selected as the subject of our studies. Helical proteins are believed to be more challenging for structure determination by RDC constraints. RDC data were generated for two alignment media with REDCAT [4] using the alignment properties listed in Table 2. The effect of motion on RDC data was emulated using the program REDCAT

Table 1
Performance of REDCRAFT as a function of increasing level of noise in simulated data

	±1 Hz	±2 Hz	±3 Hz
Backbone RMSD to 1A1Z	0.56 Å	0.57 Å	0.6 Å

Table 2
Properties of the two alignment tensors used for simulation of RDC data

	α	β	γ	S_{xx}	S_{yy}	S_{zz}	ζ
Medium-I	10	20	30	2e−4	4e−4	−6e−4	6.11e−4
Medium-II	−45	145	100	−3e−4	−5e−4	8e−4	8.08e−4

with a two state ±15° jump motion about the ψ angle of residue 71. A uniformly distributed random noise of magnitude ±1 Hz was added to the simulated RDCs.

The symbols α, β, and γ in Table 2 correspond to the three Euler angles relating the MF to PAF. S_{xx}, S_{yy}, and S_{zz} denote the three order parameters while the symbol ζ denotes the GDO value as defined before [1,51]. The three order parameters obtained for each fragment can be combined to provide a measure of overall alignment called general degree of order (GDO). A higher GDO is indicative of stronger alignment. Rigid components of a molecule will report similar GDOs while fragments undergoing motion relative to the rest of the molecule will report a GDO with lower magnitude.

The synthetic data were used for structure determination by REDCRAFT. Figs. 6 and 7 illustrate the results of REDCRAFT analysis performed on residues 1–20 and 60–80 of 1A1Z, respectively. In these two figures, (a) shows the RMSD score of the REDCRAFT obtained from comparison of the experimental data to the back calculated RDCs as a function of increasing fragment size, and (b) shows the value of GDO obtained from the best structure as a function of increasing fragment size.

When the data are originated from a relatively rigid portion of a molecule (as illustrated in Fig. 6), an overall increase in the REDCRAFT score is initially observed as the fragment size grows. Increasing number of the experimental data versus degrees of freedom of the problem is the main contributor to this pattern. As the fragment size continues to grow, this pattern approaches and stabilizes around the expected error. A similar trend is observed for the GDO value. A fragment of sufficiently large size will provide a reasonably accurate GDO value similar to the value used during simulation of the RDC data. Fig. 7, on the other hand, exhibits a completely different trend. From residue 60 to 70, the normal increase in the REDCRAFT score is observed. When extension of the fragment proceeds beyond the point of motion (residue 71), internal discrepancies emerge, indicating an inability to identify one single order tensor to describe the alignment of both portions (residues 60–70 and 71–80). This increase in the REDCRAFT score is at first subtle but becomes signifi-

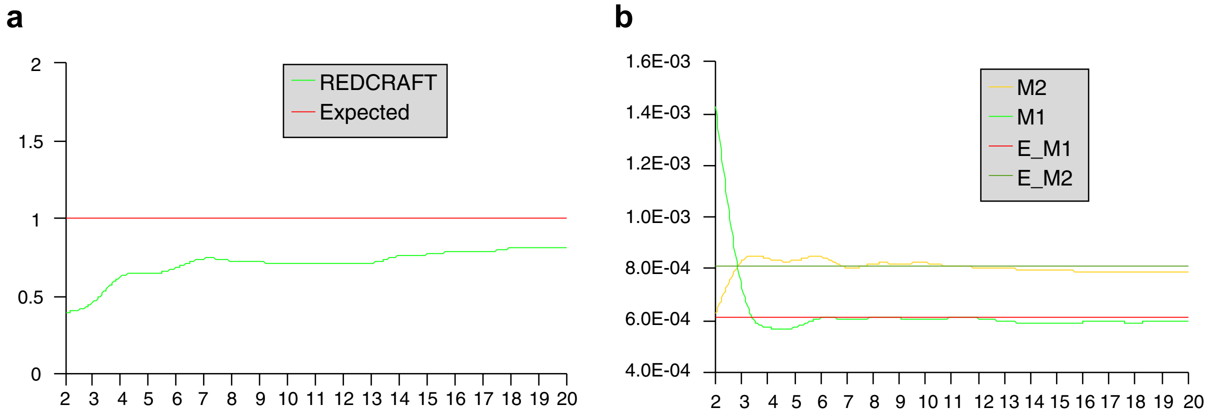

Fig. 6. Results of REDCRAFT for residues 1–20 of 1A1Z (static region). (a) The structural fitness as a function of increasing fragment size. (b) The value of GDO as a function of increasing fragment size. The values M1, M2, E_M1, and E_M2 in (b) corresponds to value from M2, value from M1, Expected M1 and Expected M2 values, respectively.

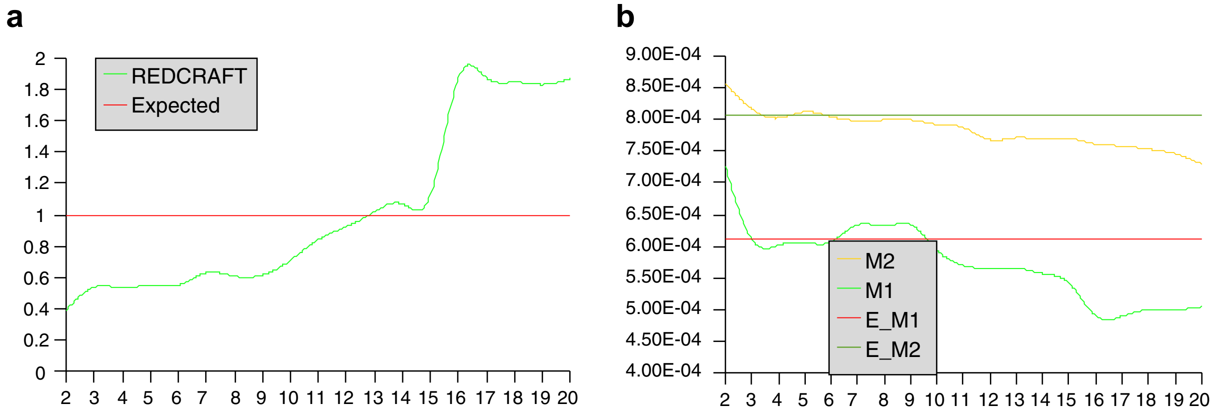

Fig. 7. Results of REDCRAFT from residues 60–80 of 1A1Z (dynamic region). (a) The structural fitness as a function of increasing fragment size. (b) The value of GDO as a function of increasing fragment size. The values M1, M2, E_M1, and E_M2 in (b) corresponds to value from M2, value from M1, Expected M1 and Expected M2 values, respectively.

cantly large after inclusion of more data from the dynamical portion of the fragment. Peptide plane number 12 in Fig. 7a corresponds to residue 71 of the fragment indicating the starting point of the internal dynamics. The lag observed in the significant increase of the REDCRAFT score (at peptide plane 15) can be explained by the need for including sufficient amount of data from the dynamical region in order to clearly quantify internal disagreements.

It is important to note that ignoring the presence of molecular motion can produce a faulty structure as illustrated in this case. Fig. 8a illustrates a superimposition of residues 71–81 of 1A1Z and the structure obtained from REDCRAFT while insisting on extension of the static region past the residue 71. These two structures exhibit a 2.9 Å distance measured over the C_α backbone atoms. Fig. 8b illustrates the same region after structure determination of residues 71–81 in isolation from the rest of the protein. Allowing an independent structural investigation of this region provided the freedom of defining an order tensor appropriate for this section and therefore facilitates provision of a structure without the influence of motion.

4.2.2. Experimental data from PF1061

Complete description of the structure determination protocol using REDCRAFT for this protein has previously been described [49] and the final structure has been submitted to the PDB under the code 1SF0. In summary, the structure of this protein was determined in five distinct fragments. The final structure obtained from RDC analysis by REDCRAFT is shown in Fig. 9 using approximately 5 and 3 RDCs per peptide plane from alignment media 1 and 2, respectively. Fragmented study of this protein was partially due to missing data at certain residues and partially due to existence of internal motion. The order tensor obtained in each of the two media was used in orienting each fragment, and a small set of NOEs were used to translate each fragment into final structure.

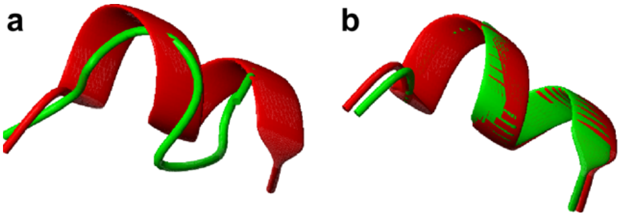

Fig. 8. Structure of residues 71–81 of 1A1Z determined by REDCRAFT. (a) The structure when determined with the remainder of the protein and (b) shows the structure when determined in isolation.

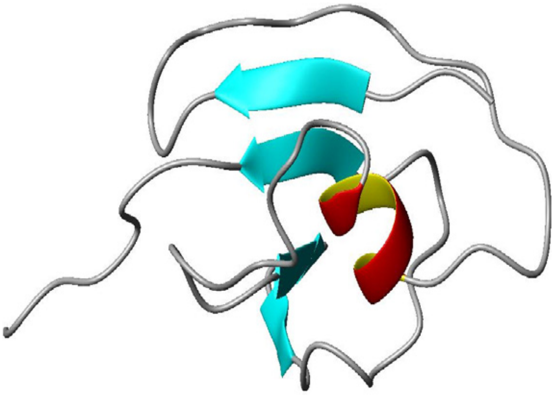

Fig. 9. RDC based structure of PFU-1601 obtained from REDCRFT.

Fig. 10 plots the best RMSD score between the back-calculated and experimental RDC data reported by REDCRAFT for two regions of PF1601. Similar to the simulated data in Fig. 6, as the size of each fragment increases from the starting dipeptide seed, the RMSD score gradually increases due to an increase in the amount of data. For the rigid fragment (blue line in Fig. 10), the RMSD score stabilizes once its size is sufficient for accurate description of the global alignment tensor. Any additional peptide planes do not increase the value of the RMSD score beyond the expected value (determined based on the N–H RDC errors) from that point forward. However, for the fragment consisting of last 15 residues of the C-terminus in red line, a different profile is observed. Initially an increase in the RMSD score proceeds as expected. Addition of the 8th peptide plane, however, significantly increases this score, indicating internal inconsistency of data in defining one commonly agreeable order tensor. As this fragment continues to grow, the RMSD score gradually recovers, since some internal consistency emerges between the peptide planes number 8 and higher. When structure determination of the last 7 residues of this protein proceeded separately from the rest of the protein with more than 50 RDCs, a nicely formed extended structure with total RMSD score of approximately 1 Hz was obtained.

Once an accurate structure is obtained, comparison of the principal order parameters reported by different fragments of the protein can be used to further characterize internal dynamics of the protein as illustrated previously [36]. GDOs for all five fragments of *PF1061* are listed in Table 3. Comparison of the principal order parameters in Table 3 indicates that the first four segments of this protein are nearly rigid with respect to one another. The C-termi-

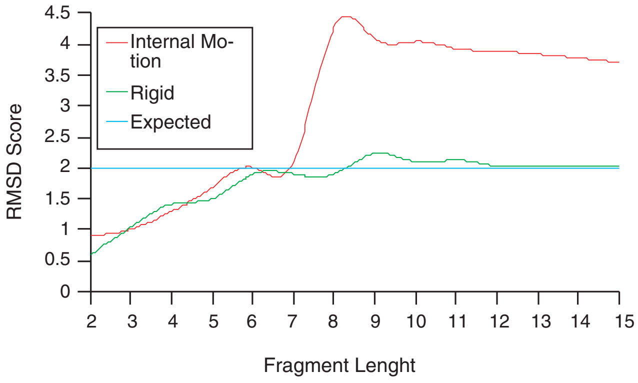

Fig. 10. RMSD between the back-calculated and experimental RDCs for the best geometry reported by REDCRAFT as an increasing fragment size. (For interpretation of the references in color in this figure legend, the reader is referred to the web version of this article.)

Table 3
The best solution order parameter and GDO calculated by REDCAT for each fragment of the PF-1601

Fragment	S_{xx}	S_{yy}	S_{zz}	ζ
1	−2.06 E−04	−5.56 E−04	7.62 E−04	9.66 E−04
2	−1.70 E−04	−5.70 E−04	7.40 E−04	9.49 E−04
3	−1.74 E−04	−4.53 E−04	6.27 E−04	7.93 E−04
4	−3.55 E−04	−4.32 E−04	7.87 E−04	7.88 E−04
5	8.20 E−05	2.11 E−04	−2.93 E−04	3.70 E−04

nal fragment, however, exhibits order parameters that are significantly different from other regions. The reduction in GDO indicates a substantial degree of motion at the C-terminus. The magnitude of motion can be estimated if a wobbling motion in a cone of angle θ is assumed. Using previously published work, θ can be estimated to be approximately 35–50° [36]. Moreover, the observed change in the signs of the order parameters for the C-terminal fragment is reminiscent of substantial internal rotation about an axis perpendicular to the z axis within the PAF. Since dynamics of this segment was identified and isolated from the rest of the molecule, the accuracy in structure determination of different regions of this protein was not compromised. The effect of internal motion of the C-terminal region was neutralized by allowing independent description of an order tensor which reports overall alignment of that fragment, including molecular tumbling and internal motion. Incidentally, it is relevant to mention that this protein was initially very refractory to crystallization and when crystals were obtained, the diffraction patterns revealed structural incoherence.

5. Discussion and conclusion

Elucidation of protein backbone structure is of direct importance within the context of the current protein structure initiative [50,52]. Here it is anticipated that producing a few structures in each of a few thousand protein families can lead to prediction of structures for the remaining members. Because members of a given family may exhibit more than 30% sequence identity, most side chains can be replaced during prediction and it is the backbone structure that is of primary importance. Programs such as REDCRAFT can clearly play an important role in producing these backbone structures.

REDCRAFT is not the only program put forth for the production of protein structures primarily from RDCs and it is therefore appropriate to make comparisons to some of these programs [21,22,30–34,55]. REDCRAFT exhibits the following unique features which become relevant during the task of protein structure determination:

5.1. Simplified search space

REDCRAFT exploits the rich information content of RDC data to limit structure determination to the backbone of the unknown protein, eliminating complexities of the energy landscape contributed by the sidechains. Moreover, the search for a protein structure is conducted by gradually increasing the fragment size, leading to a less complicated energy landscape. Extension of fragments using ideal peptide planes and search for structural conformers in torsion angle space eliminates bond length, bond angles and improper energy terms, further reducing the complexity of the structural energy landscape. Although these energy terms do contribute to the final structure of a protein, it is reasonable to argue that these terms would be more consequential during the last stages of structure refinement.

Discovery of the global optimal point within a complex energy landscape is very difficult and computationally intractable. Although approaches such as simulated annealing have been used to overcome the accidental entrapment in local minima, in general they require repeated minimization sessions. Simplification of the energy landscape employed by REDCRAFT is certainly welcome when faced with this problem.

5.2. Robustness

A number of recent approaches have explored the concept of incremental structure determination and the structural search is conducted in the rotamer space similar to REDCRAFT [22,32]. These programs however, unlike REDCRAFT, either analytically or through a search routine obtain the most optimal torsion angles of the last added residue based on the data originated from that peptide plane. While these "greedy" approaches may intuitively make sense and are computationally friendly, they are highly susceptible to imperfect data. As shown in Section 3.2, the complexity of RDC energy landscape justifies the shortcomings of a "greedy" approach. Careful examination of Fig. 3 can provide the following two important conclusions. First, a number of local minima can be identified that correspond to incorrect geometries. Second, the correct geometry (−130°, 138°) does not even appear as a local minimum for a fortuitous entrapment of gradient descent search methods. This phenomenon is attributed to both experimental and structural noise. Only after appending five additional peptide planes (2–8) a meaningful and significant landscape is observed. Only then, the global minimum point corresponding to the true structure emerges. Any method that proceeds based on a single exact torsion angle would start with a highly inaccurate structure. Progression down the path initiated by a false geometry will certainly lead to an incorrect or suboptimal structure. It is important to note that exhaustive search of structures for a small fragment with seven peptide planes can take two weeks of computation time on a 100 CPU Linux cluster.

5.3. Simultaneous characterization of structure and identification of motion

REDCRAFT possesses the unique ability of simultaneous structure determination and identification of motion.

This is a very important feature especially within the context of dynamical proteins such as membrane proteins where the transmembrane region is likely to experience a relative motion with respect to the aqueous region or signaling proteins [53,54]. Although RDC data provide the information needed to assess the relative motion between different regions, proper fragmentation of the protein into consistent regions will be a prerequisite. Fragmentation of a protein into regions with internally consistent rigidity will effectively remove the influence of overall motion of that fragment from structure determination, resulting in accurate structures and assessment of motion. Any method that disregards presence of motion (of a protein or a fragment of a protein) may potentially produce a faulty and inaccurate structure influenced by inconsistent internal dynamics. Assessment of relative motion using faulty starting conformations will certainly produce incorrect answers.

5.4. De novo structure determination

REDCRAFT's approach to structure determination is truly *De novo*. The course of structure determination is governed only by the observed experimental data and assumption of the standard peptide plane geometry. REDCRAFT does not depend on any knowledge extracted from the currently existing database of protein structures other than the Ramachandran statistics. Furthermore, REDCRAFT does not require an existing model for refinement or include structural information embedded in its force field. Structure determination guided purely by the experimental data is a desirable attribute when dealing with proteins that are underrepresented and may produce proteins with unique structural characteristics.

5.5. Easy customization

REDCRAFT incorporates advanced programming concepts such as class based programming implemented in C++. In addition, the modular cascade of filters provides a very flexible platform for customization of the software. Users of this package are able to design and integrate their own filters easily. As long as the input/output format of these filters adheres to the specification of our software, they can be inserted into any stage of the analysis. Furthermore, additional filters provided by the users can be implemented in the programming language of their choice.

5.6. Efficiency of combinatorial elimination

Finally it is appropriate to discuss the time complexity of REDCRAFT. The computational complexity of the first stage of REDCRAT is simply a linear function of the fragment size and therefore exhibits $O(n)$ time complexity. The computational requirement of this step is insignificant compared to that of the second stage. The results of the first step can be obtained in real time on a typical personal computer. The second stage of the REDCRAFT algorithm is the most time consuming component and exhibits some counterintuitive properties. The time complexity of the second stage can be expressed as: $O[n(NM + NM\,log(NM))]$ where n is the number of peptide planes in the fragment, N is the number of surviving geometries from the Stage-I and M is the search depth. The first term in this expression corresponds to the number of geometries evaluated while the second term corresponds to the time complexity of the sorting algorithm for ranking of different geometries. This time complexity approaches $O(N^2\log(N))$ when $N \approx M$ and $O(M\,log(M))$ for the cases of $M \gg N$. This behavior is very desirable since the time complexity of the problem becomes more efficient as more exhaustive searches are performed.

Acknowledgments

This work was supported by Grant No. GM062407 from the National Institutes of Health (to the Southeast Collaboratory for Structural Genomics) and Grant No. 1R01GM081793 from National Institutes of Health to Dr. Homayoun Valafar.

References

[1] 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.

[2] A. Bax, G. Kontaxis, N. Tjandra, Dipolar couplings in macromolecular structure determination, Nucl. Magn. Reson. Biol. Macromol. Pt. B (2001) 127–174.

[3] J.R. Tolman, J.M. Flanagan, M.A. Kennedy, J.H. Prestegard, Nuclear magnetic dipole interactions in field-oriented proteins—information for structure determination in solution, Proc. Natl. Acad. Sci. USA 92 (1995) 9279–9283.

[4] H. Valafar, J.H. Prestegard, Redcat: a residual dipolar coupling analysis tool, J. Magn. Reson. 167 (2004) 228–241.

[5] 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.

[6] A. Cupane, M. Leone, E. Vitrano, L. Cordone, Structural and dynamic properties of the heme pocket in myoglobin probed by optical spectroscopy, Biopolymers 27 (1988) 1977–1997.

[7] H. Shimada, W.S. Caughey, Dynamic protein structures—effects of pH on conformer stabilities at the ligand-binding site of bovine heart myoglobin carbonyl, J. Biol. Chem. 257 (1982) 1893–1900.

[8] M.A. DePristo, P.I.W. de Bakker, T.L. Blundell, Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography, Structure 12 (2004) 831–838.

[9] A. Saupe, G. Englert, High-resolution nuclear magnetic resonance spectra of orientated molecules, Phys. Rev. Lett. 11 (1963) 462–464.

[10] H.J. Zhou, A. Vermeulen, F.M. Jucker, A. Pardi, Incorporating residual dipolar couplings into the NMR solution structure determination of nucleic acids, Biopolymers 52 (1999) 168–180.

[11] E. de Alba, N. Tjandra, NMR dipolar couplings for the structure determination of biopolymers in solution, Prog. Nucl. Magn. Reson. Spectrosc. 40 (2002) 175–197.

[12] M. Blackledge, Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings, Prog. Nucl. Magn. Reson. Spectrosc. 46 (2005) 23–61.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] M. Assfalg, I. Bertini, P. Turano, A.G. Mauk, J.R. Winkler, H.B. Gray, N-15-H-1 residual dipolar coupling analysis of native and alkaline-k79a *Saccharomyces cerevisiae* cytochrome *c*, Biophys. J. 84 (2003) 3917–3923.

[18] 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.

[19] 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.

[20] 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.

[21] C.A. Fowler, F. Tian, H.M. Al-Hashimi, J.H. Prestegard, Rapid determination of protein folds using residual dipolar couplings, J. Mol. Biol. 304 (2000) 447–460.

[22] 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.

[23] 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.

[24] M. Nitz, M. Sherawat, K.J. Franz, E. Peisach, K.N. Allen, B. Imperiali, Structural origin of the high affinity of a chemically evolved lanthanide-binding peptide, Angew. Chem. Int. Ed. 43 (2004) 3682–3685.

[25] J.H. Prestegard, H. Valafar, J. Glushka, F. Tian, Nuclear magnetic resonance in the era of structural genomics, Biochemistry 40 (2001) 8677–8685.

[26] G.M. Clore, A.M. Gronenborn, A. Bax, A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information, J. Magn. Reson. 133 (1998) 216–221.

[27] P. Dosset, J.C. Hus, D. Marion, M. Blackledge, A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings, J. Biomol. NMR 20 (2001) 223–231.

[28] J.A. Losonczi, 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.

[29] 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, Nucleic Acids Res. 9 (1981) 7073–7083.

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] C.A. Rohl, D. Baker, De novo determination of protein backbone structure from residual dipolar couplings using rosetta, J. Am. Chem. Soc. 124 (2002) 2723–2729.

[35] W. Press, Saul A. Teukolsky, William T. Vetterling, Brian P. Flannery, Numerical Recipes in C, the Art of Scientific Computing, 2002.

[36] 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.

[37] 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.

[38] 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.

[39] J.C. Hus, R. Bruschweiler, Principal component method for assessing structural heterogeneity across multiple alignment media, J. Biomol. NMR 24 (2002) 123–132.

[40] N.A. Greshenfeld, The Nature of Mathematical Modeling, 1998.

[41] J. Cavanagh, W.J. Fairbrother, A.G. Palmer III, N.J. Skelton, Protein NMR Spectroscopy. Principles and Practice, 1995.

[42] M. Karplus, Vicinal proton coupling in nuclear magnetic resonance, J. Am. Chem. Soc. 85 (1963) 2870.

[43] 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.

[44] M. Ottiger, A. Bax, Determination of relative N-HN, N-C′, C-alpha-C′, and C(alpha)-H-alpha effective bond lengths in a protein by NMR in a dilute liquid crystalline phase, J. Am. Chem. Soc. 120 (1998) 12334–12341.

[45] N. Tjandra, A. Bax, Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium, Science 278 (1997) 1111–1114.

[46] R. Kuang, C.S. Leslie, A. Yang, Protein backbone angle prediction with machine learning approaches, Bioinformatics 20 (2004) 1612–1621.

[47] S.C. Lovell, J.M. Word, J.S. Richardson, D.C. Richardson, The penultimate rotamer library, Proteins Struct. Funct. Genet. 40 (2000) 389–408.

[48] 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.

[49] H. Valafar, K. Mayer, C. Bougault, P. LeBlond, F.E. Jenney, P.S. Brereton, M. Adams, J.H. 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.

[50] R.F. Service, Structural genomics—tapping dna for structures produces a trickle, Science 298 (2002) 948–950.

[51] J.R. Tolman, Dipolar couplings as a probe of molecular dynamics and structure in solution, Curr. Opin. Struct. Biol. 11 (2001) 532–539.

[52] Service R. Structural biology—structural genomics, round 2. 307 (2005) 1554–1558.

[53] Y. Cheng, T. LeGall, C.J. Oldfield, J.P. Mueller, Y.J. Van, P. Romero, M.S. Cortese, V.N. Uversky, A.K. Dunker, Rational drug design via intrinsically disordered protein, Trends Biotechnol. 24 (2006) 435–442.

[54] A.K. Dunker, J.D. Lawson, C.J. Brown, R.M. Williams, P. Romero, J.S. Oh, C.J. Oldfield, A.M. Campen, C.M. Ratliff, K.W. Hipps, J. Ausio, M.S. Nissen, R. Reeves, C. Kang, C.R. Kissinger, R.W. Bailey, M.D. Griswold, W. Chiu, E.C. Garner, Z. Obradovic, Intrinsically disordered protein, J. Mol. Graph. Model 19 (2001) 26–59.

[55] J. Wang, J.D. Walsh, J. Kuszewski, Y. Wang, Periodicity, planarity, and pixel (3P): a program using the intrinsic residual dipolar coupling periodicity-to-peptide plane correlation and phi/psi angles to derive protein backbone structures, J. Magn. Reson. 189 (2007) 90–103.