J Biomol NMR (2014) 60:241–264
DOI 10.1007/s10858-014-9871-x

ARTICLE

Improvements to REDCRAFT: a software tool for simultaneous characterization of protein backbone structure and dynamics from residual dipolar couplings

Mikhail Simin · Stephanie Irausquin · Casey A. Cole · Homayoun Valafar

Received: 7 August 2014 / Accepted: 30 October 2014 / Published online: 18 November 2014

Abstract Within the past two decades, there has been an increase in the acquisition of residual dipolar couplings (RDC) for investigations of biomolecular structures. Their use however is still not as widely adopted as the traditional methods of structure determination by NMR, despite their potential for extending the limits in studies that examine both the structure and dynamics of biomolecules. This is in part due to the difficulties associated with the analysis of this information-rich data type. The software analysis tool REDCRAFT was previously introduced to address some of these challenges. Here we describe and evaluate a number of additional features that have been incorporated in order to extend its computational and analytical capabilities. REDCRAFT's more traditional enhancements integrate a modified steric collision term, as well as structural refinement in the rotamer space. Other, non-traditional improvements include: the filtering of viable structures based on relative order tensor estimates, decimation of the conformational space based on structural similarity, and forward/reverse folding of proteins. Utilizing REDCRAFT's newest features we demonstrate de-novo folding of proteins 1D3Z and 1P7E to within less than 1.6 Å of the corresponding X-ray structures, using as many as four RDCs per residue and as little as two RDCs per residue, in two alignment media. We also show the successful folding of a structure to less than 1.6 Å of the X-ray structure using $\{C^{i-1}–N^i$, $N^i–H^i$, and $C^{i-1}–H^i\}$ RDCs in one alignment medium, and only $\{N^i–H^i\}$ in the second alignment medium (a set of data which can be collected on deuterated samples). The program is available for download from our website at http://ifestos.cse.sc.edu.

Keywords REDCRAFT · Structure · Dynamics · RDC · Dipolar · Computational

M. Simin · S. Irausquin · C. A. Cole · H. Valafar (✉)
Department of Computer Science and Engineering, University of South Carolina, Columbia, SC 29208, USA
e-mail: homayoun@cec.sc.edu

Introduction

Traditional experimental methods for protein structure determination include X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. The two approaches have been utilized extensively, contributing 90,358 and 10,537 biomolecular structures respectively, to the Protein Databank (PDB; as of Jul, 2014) (Berman et al. 2000). While both methods have been instrumental, each exhibits certain limitations in exploring the protein structure-space. For instance, despite coding for more than 30 % of the human genome, membrane proteins have been very poorly represented (Opella et al. 2002) in the PDB. Furthermore, no novel fold families [as classified by CATH (Orengo et al. 1997) or SCOP (Murzin et al. 1995)] have been submitted to the PDB since 2010. Specific to NMR spectroscopy, such limitations are the result of a heavy dependence on internuclear distances that can be obtained from the nuclear overhauser effect (NOE) (Wuthrich 1986). While distance based approaches have been very successful in macromolecular structure determinations, they are not universally applicable to all classes of molecular structures. Non-globular proteins and membrane bound/associated proteins can be cited as such examples, as their recalcitrant nature is often exacerbated by the requirements imposed by NOE experiments. Residual dipolar couplings (RDCs) are an alternative type of data obtained by NMR spectroscopy. Although their role in

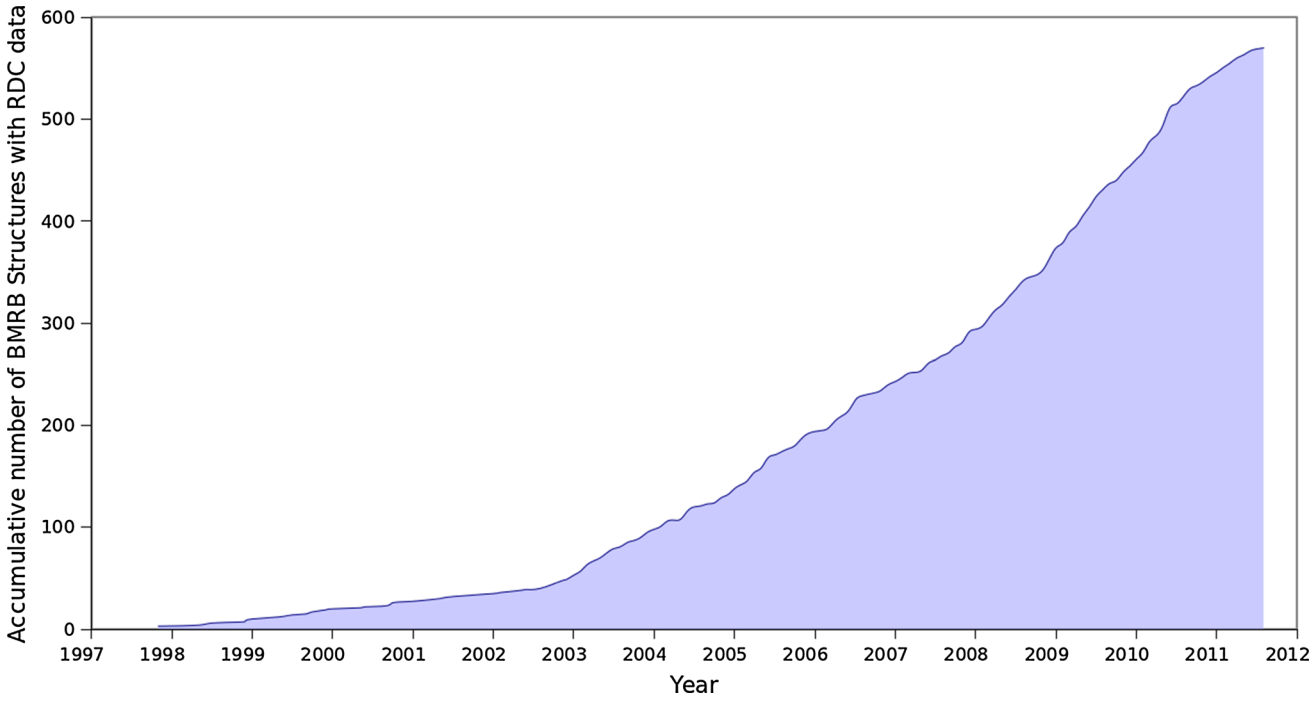

Fig. 1 Accumulative number of RDC data depositions in the Biological Magnetic Resonance Bank (BMRB) as a function of time

structure determination has been limited to providing only minor additional restraints relative to the large number of distance-based NOE (Prestegard et al. 2001; Valafar et al. 2005; Bryson et al. 2008) restraints, more recent developments (Dosset et al. 2001; Tian et al. 2001b; Valafar et al. 2005; Prestegard et al. 2005) have demonstrated the potential for a reversal in the role of these data types which could possibly absolve NMR spectroscopy from its reliance upon NOE restraints.

Historically, the use of RDCs has been impeded mainly by data acquisition and data analysis. The introduction of various alignment media (Prestegard et al. 2004), combined with improvements in instrumentation and pulse sequences (Prestegard et al. 2004; Banci et al. 2010), have significantly reduced the experimental limitations in obtaining RDCs. Moreover, a number of distinct advantages exhibited by orientational constraints (compared to traditional distance-based constraints) have further contributed to progress in RDC data acquisition (Tian et al. 2001b; Valafar et al. 2005; Prestegard et al. 2005; Park et al. 2009). Consequently, the amount of acquired RDC data has increased precipitously over the past few years as evidenced by submissions to the Biological Magnetic Resonance Bank (BMRB; refer to Fig. 1) (Doreleijers et al. 2003; Ulrich et al. 2008). However, the full potential of RDC data has not yet been realized because of challenges associated with their analysis.

REDCRAFT (Bryson et al. 2008) was previously introduced as a new, alternative approach to protein structure calculation explicitly from orientational restraints. In this report we present improvements to the original REDCRAFT software, which have improved the reliability and robustness of the results produced. Our reported exercises utilize the new features to demonstrate protein structure characterizations with as little as two RDC data per residue from two alignment media. The latest version of the RECRAFT software package can be obtained freely at http://ifestos.cse.sc.edu.

Methods

In order to illustrate REDCRAFT's most recent improvements, and facilitate a more informed discussion, we begin this section by providing an overview of RDCs as it relates to the presented work. This is followed by a discussion of other software programs currently available for structure determination from RDC data. Finally, we conclude with a summary of REDCRAFT which includes a detailed list of its newest features as well as a description of our software testing and validation procedures.

Residual dipolar couplings

RDCs have been observed as early as 1963 (Saupe and Englert 1963) and have been acquired for a number of structure determination studies including small molecules (Thiele 2008; Kummerlöwe and Luy 2009), carbohydrates

(Tian et al. 2001a; Azurmendi and Bush 2002; Azurmendi et al. 2002; Adeyeye et al. 2003), nucleic acids (Al-Hashimi et al. 2000a; Tjandra et al. 2000; Vermeulen et al. 2000; Al-Hashimi et al. 2002a, b) and proteins (Cornilescu et al. 1999; Fowler et al. 2000; Tian et al. 2001b; Andrec et al. 2001; Clore and Bewley 2002; Bertini et al. 2003; Assfalg et al. 2003). RDCs arise from the interaction of two magnetically active nuclei in the presence of the external magnetic field of an NMR instrument (Tolman et al. 1995; Tjandra et al. 1996; Prestegard et al. 2000; Bax et al. 2001). This interaction is normally reduced to zero due to the isotropic tumbling of molecules in their aqueous environment. However, the introduction of partial order to the molecular alignment by minutely limiting their isotropic tumbling will reintroduce the finite RDC interactions. The resulting RDCs are measured relatively easily and represent an abundant source of precise and informative data which includes the relative orientation of different internuclear bonds within the alignment frame.

In order to motivate the use of RDC data as the main source of constraints for protein structure determination, we first begin by discussing the theoretical aspects of RDC analysis. This is followed by a summary of their utility, which includes establishing the theoretical limits of RDC data requirements and a comparison of RDCs to NOEs.

Theory

The physical principles (Saupe and Englert 1963; Cavanagh et al. 2007) that lead to manifestation of RDCs, and methods inducing alignment of biological macromolecules, have been fully described previously (Tjandra et al. 1997; Bax and Tjandra 1997; Prestegard et al. 2000, 2004). Here we briefly review those components utilized by REDCRAFT. In addition, we limit our discussion to nuclei with spin quantum number of ½, and refer to the formula in Eq. 1 from which all mathematical derivation of the RDC interactions (for a pair of spin ½ nuclei) begin. In this equation, μ_0 is the magnetic permeability of free space, γ_i and γ_j are gyromagnetic ratios of the interacting nuclei, h is Planck's constant, r is the distance separating nuclei i and j, and θ is the angle between the magnetic field of the NMR device and a vector connecting atoms i and j.

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

It is important to note that the RDC value D_{ij} (reported in units of Hz) is a function of the time-dependent angle $\theta(t)$ averaged over time t, as represented by the angular brackets in Eq. 1. This time averaging phenomenon may account for molecular motions due to: natural bond vibrations, internal dynamics, or overall tumbling of the molecule in the solution state. Mathematical transformation (Saupe and Englert 1963) of Eq. 1 can produce a computationally friendlier formulation of the RDC phenomenon, as shown in Eq. 2. In this representation of the RDC interaction, v signifies the normalized orientation of the interacting vector, s_{ij} denotes the ijth element of the Saupe order tensor matrix, S_{ii} represents the principle order parameters, and $\boldsymbol{R}$ symbolizes the rotation matrix which relates the molecular frame to the principal alignment frame. The remaining constants have been subsumed into a single constant, D_{max}.

$$\begin{aligned} D &= D_{max}\bar{v}^T \cdot \begin{pmatrix} s_{xx} & s_{xy} & s_{xz} \\ s_{yx} & s_{yy} & s_{yz} \\ s_{zx} & s_{zy} & s_{zz} \end{pmatrix} \cdot \bar{v} \\ &= D_{max}\bar{v}^T \cdot R \cdot \begin{pmatrix} S_{xx} & 0 & 0 \\ 0 & S_{yy} & 0 \\ 0 & 0 & S_{zz} \end{pmatrix} \cdot R^T \cdot \bar{v} \end{aligned} \tag{2}$$

Within recent years, various methods of obtaining the order tensor have appeared in the literature (Clore et al. 1998; Losonczi et al. 1999; Warren and Moore 2001; Dosset et al. 2001; Valafar and Prestegard 2003, 2004; Zweckstetter 2008; Miao et al. 2008; Mukhopadhyay et al. 2009; Fahim et al. 2013; Schmidt et al. 2013). These diverse approaches exhibit some advantages over other existing methods, such as: highly accurate estimation of order tensors, estimation of order tensor or order parameters in the absence of a structure, relative order tensor estimation from unassigned RDCs, and reconstruction of interacting vectors in space from unassigned RDCs. However, the most reliable method of obtaining an order tensor is from using an assigned set of RDCs to a high-resolution structure.

The RDC advantage

The emergence of RDCs as an alternative means of studying the structure and dynamics of macromolecules has prompted many to question the differences between RDC and NOE data types with regard to structure calculation. Comparing the effectiveness of each data type is challenging, due to their vastly different information contents. While it has been informally noted by the community that RDC data report better fitness to X-ray structures than that of structures determined by NOE data, this observation is unexpected since NOE based structures are designed to study the protein in the same environment as the acquired RDCs. Yet, any discrepancies existing between RDC solution-state and X-ray derived structures can be attributed to the influences of crystal packing forces and the desiccated crystalline environment. Similar explanations, substantiating any observed differences between RDC-based and NOE-based molecular

structures are lacking, as are investigations which fully explore the sensitivity of structure calculation between the two data types. Therefore to meaningfully examine the nature of differences between RDC and NOE -based structure determination, we resort to an exercise which uses simulated data. In this exercise we utilize the 56 residue immunoglobulin binding domain of streptococcal protein G (PDB ID 1GB1) and the 114 residue Ubiquitin/UIM fusion protein (PDB ID 2KDI). For each protein, computational RDC restraints for the backbone N–H and C_α–H_α vectors were generated in two alignment media using the structure and two arbitrarily selected order tensors with typically observed order parameters. The inclusion of RDCs from two alignment media is necessary in order to avoid the inversion degeneracies of RDCs (Al-Hashimi et al. 2000b). NOE distances were calculated for each protein from the published structures for the pair of atoms with experimentally observed NOEs. To better emulate experimental conditions, computed RDC and NOE restraints were randomly altered with uniformly distributed noise (within ± 1 Hz in the case of RDCs and up to 1 Å in the case of NOEs). Following generation of computed distances and RDC restraints from the native structures, three thousand derivative structures were generated from their corresponding original structure by randomly altering backbone dihedral angles. This process resulted in a population of structures that differed from their respective original structure in the ranges of $\sim$0–30 Å (in the case of 1GB1) and $\sim$0–23 Å (in the case of 2KDI), with structural similarity measured over the backbone atoms as root mean square deviation (BB-RMSD). Each structure in the ensemble was refined by allowing only side-chain modification to better fit the distance restraints. This refinement step was necessary to allow conformational changes of side-chains to compensate for changes in backbone structure, and did not affect the RDC fitness since all RDCs originated from the backbone atoms (which were fixed). RDC and NOE fitness data resulting from the generated structures were then normalized (by dividing by the maximum observed value) and plotted as a function of normalized BB-RMSD (shown in Fig. 2). The normalization of BB-RMSD is necessary in order to remove the influence of protein size from BB-RMSD variations. In this exercise we used protein 1GB1 as our reference protein and normalized BB-RMSDs of 2KDI via scaling by a normalization factor of $\sqrt{(114/56)}$. The resulting figure allows for inferences in sensitivity to be compared between RDCs and NOEs, while providing a number of conclusions related to structure calculation strategies. First, this figure suggests that backbone N–H and C_α–H_α RDCs may be sufficient to obtain the backbone structure of a protein. Second, it reveals that as the calculated structure approaches the actual structure, NOEs tend to plateau (lose sensitivity) while RDCs become more sensitive. Therefore, NOEs may be indiscriminate probes when operating in the range of 0–3.5 Å from the actual structure. In contrast, the use of RDC data may very well provide structures within less than 1.0 Å from the actual structure and provides evidence for the observation noted previously, which identifies RDC data as having a better fitness to X-ray structures than that of NOEs. Yet another conclusion resulting from Fig. 2, relates to the manner in which traditional approaches to structure calculation commence their structure calculations. The traditional approaches initiate their calculations with an extended structure, which are then allowed to alter the initial structure to better fit the experimental constraints. In the case of RDC data, it is clear that these traditional approaches operate in the least sensitive region where structural fitness to the RDC data has already reached its saturation point. Therefore the search of the conformational space is not optimally assisted by RDCs until arriving to within 10 Å of the actual structure. The NOE fitness profile, on the other hand, shows a complementary behavior where structural sensitivity is relatively acute at farther distances, and diminishes in the vicinity of the native structure.

Within the context of our work we would like to differentiate between "high-resolution" and "complete" structure determinations. Based on our definitions high-resolution structure determination refers to the precision by which atomic positions are described in space (for a structure or a portion of a structure), while complete structure determination refers to the atomic description of the entire protein or a complex of proteins. The RDC based results shown in Fig. 2, demonstrate partial structure determination—the possibility of backbone structure determination without the need for side-chain structure determination. This approach confers a number of advantages, further discussed in the "Results and discussion" section, and has the potential to address some of the longstanding challenges in NMR spectroscopy. In conclusion, the results of our exercise illustrated in Fig. 2 demonstrate that RDCs may be an indispensable source of data in high-resolution structure determination by NMR spectroscopy, and also highlight some of the difficulties associated with protein structure calculation from RDC data (particularly when data is scarce). These findings encourage the examination of strategies currently available for structure determination using RDC data and provide the incentive for programs designed explicitly for RDC analysis; topics which are further discussed in the two subsections that follow.

RDC information content of {C′–N, N–H, and C′–H} set

In section "An exploration of the minimum data requirement" of this report we present results for structure

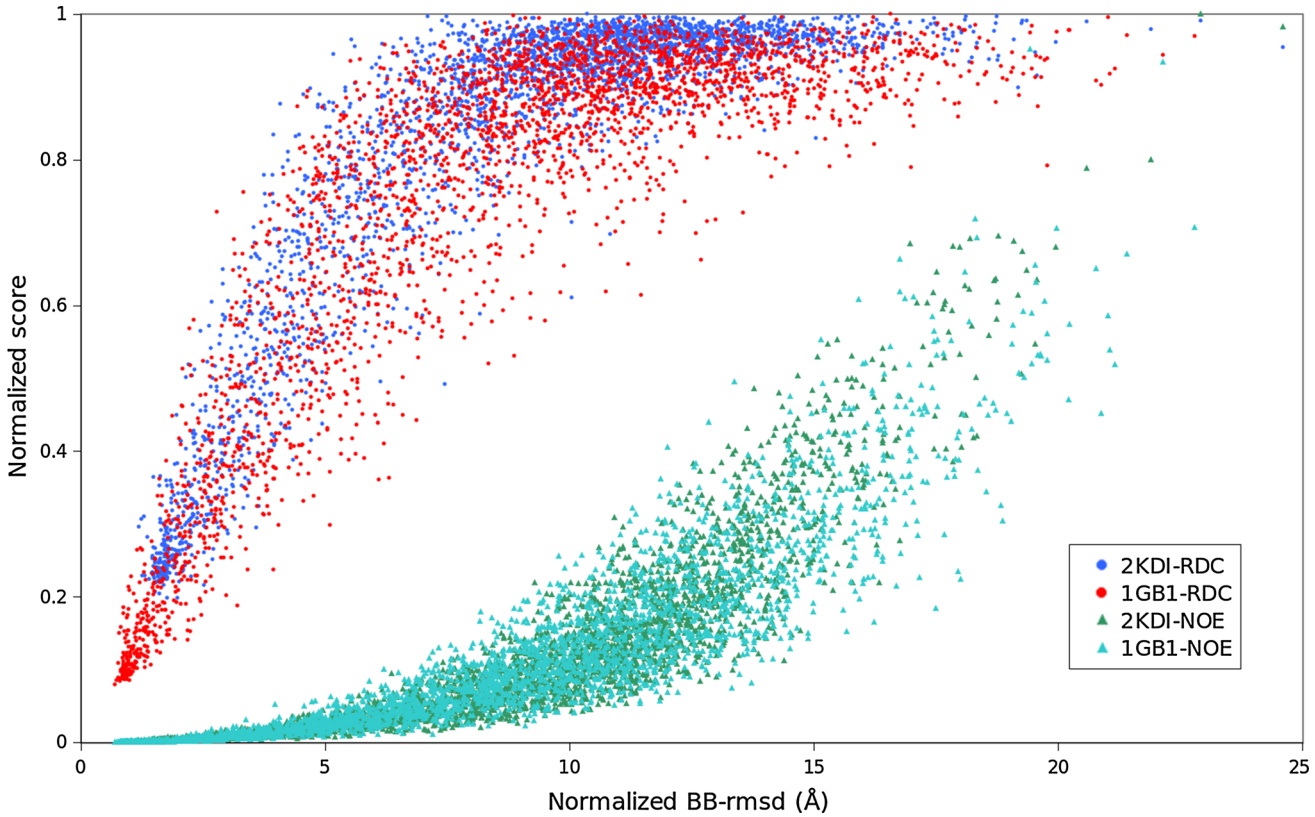

Fig. 2 Comparing the sensitivity of structure calculations between RDC and NOE data. RDC and NOE data, incorporating uniformly distributed noise, was simulated for 1GB1 and 2KDI proteins. Three thousand derivative structures were then generated from the corresponding original structures by randomly altering backbone dihedral angles. RDC (*blue* and *red circles*) and NOE (*green* and *cyan triangles*) fitness data resulting from the generated structures were then plotted as a function of BB-RMSD comparisons of each generated structure to the corresponding original structure

calculation of protein 1D3Z using RDC data originated from the set: {C′–N, N–H, and C′–H}. Given the planar arrangement of these three vectors, their independent information content becomes questionable. It is therefore appropriate to discuss the theoretical basis of this set of RDC data. While it is clear that these three vectors are linearly dependent in a three dimensional Cartesian space, their RDC information content has not been properly explored. Here we utilize Singular Value Decomposition (SVD) to ascertain the linear independence of these three vectors in the system of linear equations in the form of $\boldsymbol{Ax} = \boldsymbol{b}$. Our theoretical exploration proceeds in both the three dimensional Cartesian space and the five dimensional RDC space (related to the spherical harmonic space) (Valafar and Prestegard 2004; Schmidt et al. 2013). The coordinates of these three vectors in a three dimensional Cartesian space are shown in Eq. 3, while the coordinates of the five dimensional representation of the three vectors is shown in Eq. 4. The translation mechanism between Cartesian and Spherical Harmonic representation of a vector is shown in Eq. 5.

$$\begin{pmatrix} \vec{C}N \\ \vec{N}H \\ \vec{C}H \end{pmatrix}_{R_3} = \begin{pmatrix} 1.325 & 0.159 & 0 \\ 0.603 & -0.811 & -0.002 \\ 1.928 & -0.652 & -0.002 \end{pmatrix} \tag{3}$$

$$\begin{pmatrix} \vec{C}N \\ \vec{N}H \\ \vec{C}H \end{pmatrix}_{R_5} = \begin{pmatrix} 0.986 & 0.0142 & 0.237 & 0 & 0 \\ 0.356 & 0.644 & -0.958 & -0.0024 & 0.0032 \\ 0.897 & 0.103 & -0.607 & -0.0019 & 0.00063 \end{pmatrix} \tag{4}$$

$$\begin{pmatrix} x & y & z \end{pmatrix}_{Cartesian} \Leftrightarrow \begin{pmatrix} x^2 - z^2 & y^2 - z^2 & 2xy & 2xz & 2yz \end{pmatrix}_{SpherialHarmonic} \tag{5}$$

The use of SVD to establish the information content and linear independence of entries within a system of linear equations is well established (Press et al. 2003). SVD can provide the condition-number of a matrix that can be used

to quantify the singularity of a matrix by observing the ratio of the largest eigenvalue to the smallest eigenvalue (Greshenfeld 1998; Press et al. 2003). Similarly, the same ratio can be used to quantify the degree of orthogonal information content of each entry of a matrix. The condition-number of each entry can vary from 1 to ∞, indicating the maximum to null information content, respectively. Eigenvalues and the corresponding condition-number for the Cartesian and Spherical Harmonic representation of the three vectors are shown in Eqs. 6 and 7, respectively. Note that the condition-numbers for the case of Cartesian representation consist of 1.0 (indicating maximal information content), 2.22 and $5e^{14}$ (indicating no information content). Comparatively, the condition numbers for the same three vectors represented in the Spherical Harmonic space are 1.0, 1.55 and 6.15 indicating that all three vectors contain useful and independent information content. This is the fundamental reason why structure determination from a set of three planar vectors is therefore possible and we demonstrate the success of structure determination from these three sets of data in subsection "An exploration of the minimum data requirement" of the "Results and discussion" section.

$$\begin{aligned} \Lambda &= \begin{pmatrix} 1.5791 & 0 & 0 \\ 0 & 0.71160 & 0 \\ 0 & 0 & 2.6963\text{e}{-15} \end{pmatrix} \Rightarrow \eta \\ &= \begin{pmatrix} 1.0 & 0 & 0 \\ 0 & 2.22 & 0 \\ 0 & 0 & 5\text{e}{+14} \end{pmatrix} \end{aligned} \tag{6}$$

$$\begin{aligned} \Lambda &= \begin{pmatrix} 1.59445 & 0 & 0 & 0 & 0 \\ 0 & 1.03018 & 0 & 0 & 0 \\ 0 & 0 & 0.25913 & 0 & 0 \end{pmatrix} \Rightarrow \eta \\ &= \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1.55 & 0 & 0 & 0 \\ 0 & 0 & 6.15 & 0 & 0 \end{pmatrix} \end{aligned} \tag{7}$$

Existing software for structure determination from RDC data

Until recently, RDCs have been limited to providing only minor additional restraints relative to the large number of distance-based NOE (Prestegard et al. 2001; Valafar et al. 2005; Bryson et al. 2008) restraints required for structure determination. Therefore nearly all of the previously existing NMR data analysis software have been modified, to a certain extent, to allow for the use of RDC data in their structure calculation protocols. Programs currently available for calculating protein structures using RDCs fall into one of two categories: general protein structure calculation techniques, and methods designed specifically to exploit the mathematical properties of RDCs. Popular protein structure calculation tools include Xplor-NIH (Schwieters et al. 2003), CNS (Brunger et al. 1998), Gromacs (Hess et al. 2008) and Cyana (Güntert et al. 1997). Although these methods are powerful in instances with large heterogeneous data sets, in the case where RDCs are used as the only restraint and a good starting structure is not available, they are susceptible to entrapment in local minima and usually never recover. While a Monte Carlo approach to starting structures may be utilized, this is rarely feasible without depending on less reliable information (such as secondary structure assignment or torsion angle restraints).

Programs such as Meccano (Hus et al. 2001), RDC-analytic (Zeng et al. 2009), and others (Delaglio et al. 2000; Wang and Donald 2004) apply a more systematic approach, fitting protein structures to experimentally determined RDCs. RDC-analytic requires N–H and C_α–H_α residual dipolar couplings in one alignment medium in addition to secondary structure assignment, sparse NOEs and hydrogen bonding restraints. The use of only one alignment precludes it from being robust to experimental error, and it is often infeasible to predict secondary structure or collect the needed NOEs and hydrogen bond restraints to perform the final assembly. In general, RDC-Analytic is used as only one part of the larger RDC-Panda protocol which uses a sizable amount of NOE data and other information to compute protein folds and is not intended as an RDC-only method. Meccano requires an extensive set of RDCs, {N–H, C–N, C–H, C_α–C, C_α–H_α, and C_α–C_β}, collected in two alignment media. It can be expensive and time consuming to perform the labeling and assignment of resonances for that many RDC vector types, and in cases where data is missing, Meccano has to perform a local optimization over phi (φ) and psi (ψ) torsion angles to produce the best fitting structure.

In summary, of the software packages currently in use and previously described, none provide a comprehensive analysis of structure and dynamics of proteins using a minimal set of RDC data. Therefore, despite their increasingly important role in NMR structure determination, RDCs continue to be used as supplementary information during the refinement stage where the expense of collecting as many as 20 RDC restraints per residue has been undertaken (Ulmer et al. 2003). This finding, inspired the creation of REDCRAFT (REsidual Diploar Coupling Residue Assembly and Filtering Tool) and has motivated the improvements reported in the section that follows.

REDCRAFT

REDCRAFT (Valafar et al. 2005; Prestegard et al. 2005; Bryson et al. 2008; Shealy et al. 2010) is designed for structure determination primarily from orientational

restraints and it sets itself apart from other approaches in a number of ways. REDCRAFT deploys a new and more powerful search mechanism that is significantly different from traditional optimization techniques, allowing it to achieve the same level of performance as other algorithms while using less data. REDCRAFT is able to accomplish this improvement by analyzing a given molecule through rigid peptide planes which adhere to strict peptide geometry. Therefore, all variations within a molecule are described through backbone torsion angles (φ, ψ). In terms of computation, this significantly reduces the number of variables, translating to a reduction in the dimensionality of the search space and improvements in program execution time. It also contributes to the program's robustness, allowing for fragmented study of protein structures when segments of data are erroneous or missing. Additionally, REDCRAFT's ability to utilize RDCs from backbone atoms allows for simultaneous structure elucidation (backbone only) and assignment of data (Tian et al. 2001b) as well as the concurrent study of structure and motion (Shealy et al. 2010). Moreover, applications of REDCRAFT have been demonstrated using aqueous (Tian et al. 2001b; Valafar et al. 2005; Prestegard et al. 2005; Bryson et al. 2008) and membrane (Shealy et al. 2010) proteins with as little as two RDCs per residue, or one RDC per residue when combined with backbone torsion angle constraints. Yet another advantage afforded by the publicly available REDCRAFT software package is its development using a sound Object Oriented (OO) programming paradigm that is easily extendable. This lends itself well to encapsulation of physical and biophysical properties of proteins since construction of a Polypeptide object, from more fundamental Residue and Atom objects, directly reflects the natural process of protein polymerization. It also allows for better program source code readability and more efficient program development, while further contributing to improvements in execution time.

In this section we begin by briefly presenting the core REDCRAFT (Bryson et al. 2008) search algorithm. This is followed by a detailed list of the newest features designed to address the limitations of the previous version. We then include a disclosure of our recently incorporated software engineering strategies and conclude with a description of our testing and validation procedures.

REDCRAFT's computational core

The Computational core of REDCRAFT has been extensively discussed previously (Bryson et al. 2008), here we present a succinct summary of its overall algorithm. REDCRAFT's approach to structure determination utilizes standard peptide geometries, performing a search over all possible combinations of φ and ψ torsion angles at each residue in two stages (Stage-I and Stage-II). In Stage-I, a list of all possible torsion angles adjoining any two neighboring peptide planes is pruned and ranked based on structural fitness to the RDC data. Initial pruning of the local torsion angles can be based on any combination of the following: scalar coupling data, secondary structure prediction programs, or Ramachandran constraints. The ranking of local geometries is accomplished based on the RDC fitness quantified by the RMSD between the back-calculated and experimental RDC data as reported previously (Valafar and Prestegard 2004; Bryson et al. 2008; Shealy et al. 2010). Stage-I concludes by creating a pruned and ranked list of torsion angles for every residue of the protein. Under ideal conditions, the final structure of the protein would correspond to the top torsion angle at each position. Due to practical conditions however, such as data acquisition errors and structural noise, this is hardly ever the case. The descent of the globally optimal torsion angle in the ranked list of local geometries necessitates further analyses, which are subsequently conducted in Stage-II.

Stage-II of REDCRAFT extends a given fragment of size N peptide planes (initially a single dipeptide seed) one peptide plane at a time, iteratively. All surviving geometries of the seed fragment are exhaustively extended by the incoming residue. Every combination of the new extended structures is then ranked based on fitness to the RDC data. Normally the top 2,000–10,000 (out of $\sim$10,000,000) structural candidates with best fitness to the RDC data are propagated for extension by one additional amino acid, with the remaining conformations being discarded. This process maintains a sufficiently diverse population of conformations to prevent entrapment into any one local minimum, and facilitates identification of the point of internal motion, which is detected from the resulting dynamic-profiling information (Bryson et al. 2008; Shealy et al. 2010).

New features of REDCRAFT

The functionality and performance of REDCRAFT has been enhanced through the addition of a number of features which we present in this section. A number of other additional features (not discussed here) have also been incorporated and their descriptions are available via user documentation at http://ifestos.cse.sc.edu.

Programming hooks Considering the complexities of NMR data, and the vast array of software that can be used for their analyses, a critical aspect in our development of the REDCRAFT package was to incorporate a flexible means of interacting with other programs. We refer to this formal interaction mechanism as a "hook." Communication between any hook and the core REDCRAFT engine is

Table 1 List of commands that can be used in development of a "hook"

Command	Description
TRUNCATE	Instructs REDCRAFT to empty the current candidate list
ADD	Instructs REDCRAFT to add a structure to the candidate list. Following this command REDCRAFT expects an appropriate number of angles (φ, ψ) consistent with the fragment size. An RDC fitness value is placed following the list of angles
COMMENT	Any content followed by this command will be printed to the screen as soon as it is received. This is useful for printing additional information for the user, as well as debugging the refinement code

facilitated by three commands (listed in Table 1). The use of hooks make it possible for the execution thread of REDCRAFT to be extended to include the execution of any script or collection of scripts. A hook can be executed at every stage of REDCRAFT's operation, or based on a set of rules (i.e. invoked at arbitrary residues). The following are just a few examples of how extensions to REDCRAFT can be implemented: to filter the intermediate structure of simulated data based on structural homology to an existing structure, for further evaluation by inclusion of experimental data other than RDCs, for minimization by programs such as Xplor-NIH or CNS.

Refinement by Levenberg–Marquardt minimization REDCRAFT reduces the degrees of freedom required in describing a protein's structure by its representation in the rotamer space. As a result, search for the optimal structure proceeds over the set of all possible backbone torsion angles and not in the Cartesian space. REDCRAFT further simplifies the solution space by converting a continuous rotamer space into a discrete space (typically in increments of 10°), in order to maintain computational tractability. Although this conversion allows for a quick and robust search over the space of all plausible torsion angles, searching only those structural candidates confined to discrete representations raises some concern as to its potential impact in discovering the real structure. In order to overcome these snap-to-grid choices, REDCRAFT's calculation has been extended through an unconstrained minimization hook (shown in Display 1) which performs a periodic structure refinement (indicated by the user) using the Levenberg–Marquardt algorithm (LMA) (Levenberg 1944). The minimization routine can be configured by two main parameters: the top number of structures to be minimized, and the frequency of minimization. The former parameter dictates how many of the top surviving conformations should be subjected to the minimization routine, while the latter parameter governs the frequency (indicated in number of residues) by which the structures are minimized. The frequency of minimization is normally influenced by data richness in order to prevent structural modifications in under-determined regions. The minimization routine accepts an initial set of backbone torsion angles, and returns a modified list of torsion angles that improve the RDC fitness but are no longer bound by any discretization of the search space. The LMA objective function is computed through the least squares function, which is identical (and therefore compatible) to REDCRAFT's RDC fitness evaluation metric. This iterative procedure does not require a static order tensor because an optimal order tensor is calculated at every iteration of the LMA. In addition to RDC fitness, this implementation of the LMA also considers the steric collision penalty (described as the next new feature) to prevent structures with severe steric collisions. The minimized structure does not replace any one of the previously computed candidates. Instead, the structure provided by the LMA is added to the candidate list through the use of the "ADD" command. The fragment extension procedure of REDCRAFT (Stage-II) is subsequently conducted after receiving the additional refined structures.

```
[Refinement]
script1=./minimize.prl
[/Refinement]
```

Display 1 REDCRAFT definition of the refinement hook that utilizes the external perl program "minimize.prl."

Backbone–backbone collision detection When considering an ensemble of structures computed using only experimental data, some structures may produce unacceptably high natural forces such as improper angles, or steric collisions. REDCRAFT resolves improper backbone torsion angles in two ways, first by adhering to acceptable Ramachandran dihedral space, and second by calculating a modified steric collision term. The steric collision term is evaluated based on a 12–6 Lennard-Jones (L–J) Equation, where the potential energy is calculated based only on the C_α atom of the last residue represented by an exaggerated VDW radius. This modification is adopted in order to reduce the computational impact of this calculation, with regard to the final outcome. Since the objective of REDCRAFT's steric collision detection is to eliminate unlikely structural candidates (and not for precise calculation of internal forces), an approximated force can be used in place of the traditional L–J forces. Furthermore, a complete VDW energy term cannot be computed since REDCRAFT excludes side-chain atoms. The use of only backbone C_α atoms has proven to be a viable structural filtration tool (Chakraborty et al. 2013) for detection of obvious collisions that lead to implausible conformations. In addition, REDCRAFT's unique assembly algorithm allows for a

further simplification of the steric force by only considering the most recently appended residue—a modification made possible by the iterative nature of REDCRAFT's fragment assembly. Since the previously appended residues have undergone an identical filtration mechanism in previous iterations, any surviving candidates will not contain a collision aside from the last residue. Thus, for the last residue j in a fragment, the L–J approximation of the Van der Waals collision is defined as shown in Eq. 8.

$$LJ_j = 4\varepsilon \sum_{i=1,j-1} \left(\sigma/r_{ij}\right)^{12} - \left(\sigma/r_{ij}\right)^{6} \tag{8}$$

In this equation, fixed values of 0.0903 and 3.81 are used for parameters ε and σ, respectively. The selected value of σ represents an exaggerated VDW radius for the C_α atom in order to produce a L–J value of 0 for two adjacent C_α atoms that define the two proximal corners of a peptide plane (when in trans conformation). The parameter r_{ij} represents the Euclidean distance between two C_α atoms of the ith and jth residues. Activation of the steric collision evaluation is specified within REDCRAFT's configuration file as shown in Display 2. The value that is indicated, following the key "LJ_Threshold," represents the cut-off threshold for the steric collision violations. In this example, any conformation that exhibits an L–J violation in excess of 10 will be strictly eliminated from the pool of viable geometries. A leading hash character ("#") will disable the L–J evaluation term. Due to the strict elimination policy, it is recommended that a liberal L–J threshold value be used. Furthermore, due to the highly nonlinear nature of the 12–6 L–J term, selection of threshold values in the range of [10–10^6] may become functionally equivalent.

```
[Run_Settings]
Run_Type=new
Start_Residue=24
Stop_Residue=52
Media_Count=2
Data_Path="."
RDC_File_Prefix=wRDC
Default_Search_Depth=200
LJ_Threshold = 10
[/Run_Settings]
```

Display 2 Steric collision activation and threshold term within REDCRAFT's configuration file

Order tensor-based filtering of structures Although REDCRAFT has been designed to conduct structure determination in the absence of any information related to the alignment of the target protein, it has been modified to leverage a priori knowledge of order tensors to better guide the structure calculation process. Recent developments have enabled estimation of relative order tensors when RDC data are available from multiple alignment media (Miao et al. 2008; Mukhopadhyay et al. 2009; Fahim et al. 2013) in the absence of a structure. When the estimated order tensors (absolute or relative) are available, REDCRAFT incorporates an additional scoring penalty based on the similarity between the computed order tensor from each conformational candidate and the estimated order tensors. The underlying principle that enables this approach is that there exists an associated order tensor to every hypothetical structure. Therefore structure calculation can be assisted based on the observed order tensors. The scoring mechanism for the order tensor fitness is shown in Eq. 9. In this equation, M denotes the total number of alignment media, s_{ij}^m and $\hat{s}_{ij}^m$ represent the ijth element of the computed or estimated order tensors of the mth alignment medium respectively. The function $\mu(.)$ in Eq. 9 represents an activation function (or step function) that is 0 for all negative arguments and 1 for all positive arguments. Using this definition of a step function, θ serves as the activation threshold that triggers a transition from 0 to 1. The parameter ω indicates the weighted contribution of this potential term to the overall score of REDCRAFT. The final value of this penalty is converted to units of Hz for an equivalent N–H interacting vector.

$$E_{OT} = 24350 \cdot \omega \cdot \mu\left(\sqrt{\frac{1}{M}\sum_{m=1..M}\sum_{ij=\{x,y,z\}}\left(s_{ij}^m - \hat{s}_{ij}^m\right)^2} - \theta\right) \cdot \left(\sqrt{\frac{1}{M}\sum_{m=1..M}\sum_{ij=\{x,y,z\}}\left(s_{ij}^m - \hat{s}_{ij}^m\right)^2} - \theta\right) \tag{9}$$

Display 3 illustrates the configuration options available for incorporation of order tensor estimates in structure calculation with REDCRAFT. The portion of the configuration file that is dedicated to order tensor estimates is embedded within the tags [OTEstimation] and [/OTEstimation]. Within this block, all lines with a leading hash character are treated as a comment. Lines starting with the tags *S1* and *S2* (*S?* For any additional alignment media) will signify the estimated values of order tensors for each alignment tensor followed by five parameters that represent s_{xx}, s_{yy}, s_{xy}, s_{xz} and s_{yz} elements of the corresponding order tensors. The threshold value of θ is identified by the tag "Tolerance" and the relative weight of this potential term ω is indicated by the parameter "Weight". The tag "Estimation_Range" allows for a flexible way of controlling the range of residues where this term should be engaged during the structure calculation of REDCRAFT. For instance, the example provided in Display 3 (when uncommented) will include the contribution of this potential term during the evaluation of the given structure at residues 5–25, 40 and 42.

```
[OTEstimation]
# syntax for OrderTensorEstimation is S?=Sxx Syy Sxy Sxz Syz
#S1=1.989e-04 3.879e-4 0 0 0
#S2=-3.174e-04 1.543e-4 3.132e-04 1.757e-04 5.076e-04
#Tolerance=1.0
#Weight=1.0
#Estimation_Range=5-25,40,42
[/OTEstimation]
```

Display 3 A portion of the REDCRAFT configuration file allowing for customizable filtration of structural candidates based on one or several user defined order tensors

Proper canonicalization of order tensors is critical for meaningful comparison of two (or more) order tensors. Canonicalization of absolute or relative order tensors (Miao et al. 2008; Mukhopadhyay et al. 2009) is necessary due to the degenerate nature of order tensors, and their dependency on a defined molecular frame. The following canonicalization protocol has been implemented to eliminate orientational degeneracies. As the first step, all order tensors (computed internal to REDCRAFT and provided as constraints) are expressed in the principal alignment frame of the first order tensor (*S1*, also referred as the anchor medium). The remaining transformations are designed to eliminate ambiguities that arise from relative order tensors and have been described previously (Mukhopadhyay et al. 2009).

Bidirectional folding/fragmented study Consistent with the natural order of protein synthesis, REDCRAFT's default mode of structure determination is from N to C terminus. However under some circumstances, the challenges of protein structure determination from RDC data are mitigated by reversing the direction of protein folding. For example, it is generally believed that the end termini of proteins are more likely to exhibit internal dynamics, or produce a lesser number of accurate RDC data. Therefore to reduce the effect of internal dynamics or data sparsity on structure determination of the core regions, it is better to start the structure determination from a more advantageous location. Once a starting point has been selected, structure determination toward the C-terminus region can continue as the default mode while structure calculation of the N-terminal end of the protein is better facilitated by the use of reverse-folding. Another example of reverse folding can be cited in relation to simultaneous study of structure and dynamics. As reported previously (Bryson et al. 2008; Shealy et al. 2010; Valafar et al. 2012), observance of an anomalous increase in the Dynamic-Profile of REDCRAFT may be indicative of the onset of internal dynamics. A comprehensive approach for simultaneous study of structure and dynamics would require the means of identifying the ending location of internal dynamics. In this context, forward-folding of the protein will assist in identifying the onset of internal dynamics, while reverse folding will assist in identifying the ending location. Finally, comparison of results obtained from forward-folding and reverse-folding can act as an independent means of validating the reliability of structure calculation.

The reverse-folding of REDCRAFT can be invoked by reversing the start and stopping residue numbers in the configuration file. Panels (a) and (b) of Display 4 illustrate REDCRAFT's configurations that correspond to forward-folding and reverse-folding, respectively.

```
[Run_Settings]
Run_Type=new
Start_Residue=1
Stop_Residue=25
Media_Count=2
Data_Path="."
RDC_File_Prefix=wRDC
Default_Search_Depth=1000
LJ_Threshold=50.0
[/Run_Settings]
```

(a)

```
[Run_Settings]
Run_Type=new
Start_Residue=25
Stop_Residue=1
Media_Count=2
Data_Path="."
RDC_File_Prefix=wRDC
Default_Search_Depth=1000
LJ_Threshold=50.0
[/Run_Settings]
```

(b)

Display 4 REDCRAFT configurations for examples of **a** forward-folding and **b** reverse-folding

Decimation of structures Another feature unique to REDCRAFT is its ability to carry forward multiple structural candidates for further elongation throughout its computations. At any residue, REDCRAFT may evaluate as many as ~10,000,000 conformations for fitness to RDCs, and select only the top candidates (normally 1,000–10,000) as suitable for further extension. This ensemble-selection feature of REDCRAFT (as opposed to a single candidate) renders the algorithm more robust with respect to missing or noisy data. This mechanism of selection maintains only a small portion of the examined population (normally 1,000 out of 10,000,000 and less than 1 % of the population) and eliminates the remaining 99 % of the conformational population. However under certain conditions, the globally optimal structure descends well below a generous depth search. In such instances, the optimal structure will be eliminated from further elongation, which may result in a significant departure in discovery of the native structure. This problem can potentially be eliminated by selecting a manageable number of structures from the pool of eliminated structures for further extension. The proper selection of exempted structures is critical for the effectiveness of this approach, and can vary anywhere from simplistic random selection to clustering of structures. Although a number of conceivable clustering criteria (such as BB-RMSD similarity, order tensor similarity, or similarity based on backbone torsion angles) may be considered, nearly all of them are inappropriate based either on computational complexity, incompatibility with RDC data, or other pragmatic limitations. Here we report a new means of clustering the terminal structures for the purposes of conformational space decimation.

REDCRAFT's decimation algorithm takes advantage of three virtual atoms that it uses to accelerate the task of fragment extension. More specifically, a fragment of size i residues will carry the atomic coordinates of three virtual atoms {N, H^N, and C_α} for the next residue (residue $i + 1$). These three atoms share the unique property of belonging to the same peptide plane defined by atoms {C_α, C, and O} of residue i, and their Cartesian coordinates can be used to efficiently calculate the proper location of the next residue (as illustrated in Fig. 3). Therefore the Cartesian coordinates of these three atoms uniquely influence the extension of the current fragment. The decimation process calculates the membership of each structure based on the atomic coordinates of these three virtual atoms as shown in Eq. 10. In this equation C_n represents the nth conformation under consideration, the symbol $\lfloor \cdot \rfloor$ denotes the floor operator, and {N_j, H_j^N, and C_α} represent the jth Cartesian coordinate of each corresponding atom. The "Membership" function computes the mapping between the Cartesian coordinates of each structure and a nine-dimensional hypercube as a function of resolution R that is defined by the user. All structures that map to the same hypercube are considered to be members of the same cluster and are therefore represented by the member with best fitness to the RDC data. The representative conformation is selected from each cluster to be included for further elongation and evaluation during the next stage of REDCRAFT's elongation. This selection mechanism allows for resurrection of a structural conformation that may prove to be globally optimal when evaluated at later stages of elongation. The size of each hypercube, and therefore the degree of decimation of the terminal structure can be easily adjusted by the resolution parameter R in Eq. 10. The portion of REDCRAFT's configuration file that relates to decimation properties is shown in Display 5.

$$Membership(C_n) = \left\{ \begin{array}{c} \lfloor N_x/R \rfloor \\ \lfloor N_y/R \rfloor \\ \lfloor N_z/R \rfloor \\ \lfloor H_x^N/R \rfloor \\ \lfloor H_y^N/R \rfloor \\ \lfloor H_z^N/R \rfloor \\ \lfloor C_x^\alpha/R \rfloor \\ \lfloor C_y^\alpha/R \rfloor \\ \lfloor C_z^\alpha/R \rfloor \end{array} \right\} \quad (10)$$

```
[Decimation_Settings]
Cluster_Sensitivity=3
Score_Threshold=0.6
Decimation_Ranges=3, 4-6, 9
[/Decimation_Settings]
```

Display 5 REDCRAFT's configuration segment, as it relates to the decimation feature

Useful perl processing scripts The REDCRAFT software package includes an array of useful preprocessing, conversion, and evaluation scripts. A detailed description of all scripts can be found in the /scripts subdirectory distributed with the REDCRAFT package (and additionally in the REDCRAFT manual provided on our website http://ifestos.cse.sc.edu/REDCRAFT/documentation) and includes a list of individual scripts, their function, and input/output requirements. In this section, we highlight those scripts for which we have reported results:

- *multiweight.prl* Normalizes the set of RDCs across different alignment media and different interacting nuclei. This is accomplished by calculating scaling factors that normalize all RDC data based on the N–H vectors of the first alignment media. With this added, weighting the ranges across all data sets become comparable in magnitude.
- *Fragments.prl* Allows the user to explore the completeness of data by providing a report of the number of

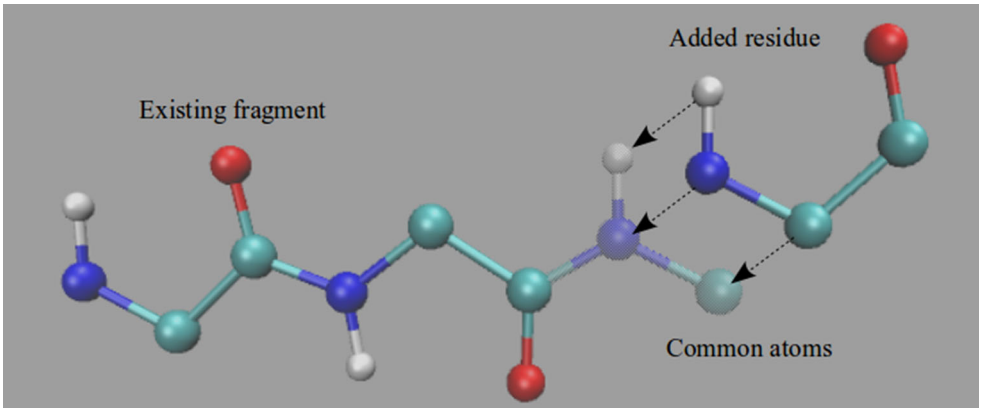

Fig. 3 Visualizing the common atoms that define a peptide plane in order to extend an existing fragment by one residue. REDCRAFT accelerates the task of fragment extension by carrying the atomic coordinates of three virtual atoms {N, H^N, and C_α} belonging to the same peptide plane but to the next amino acid. The coordinates of these three virtual atoms are defined from the atoms {C_α, C, and O} of the last residue in the fragment. Color coding of atoms: N (*blue*), H^N (*white*), C_α (*cyan*), O (*red*), C (*cyan* connected to O)

RDCs present at each peptide plane spanning all data sets. This becomes especially helpful in the context of a fragmented study, where it may not be immediately obvious where to split a protein.

- *CalcBBRMSD.prl* Provides an easy way to observe the progress of a REDCRAFT run. It works by aligning and comparing each .pdb file created by REDCRAFT to a reference protein. It displays two different scores—the RDC fitness score and the BB-RMSD.
- *molan* Provides a number of useful calculations including: back-calculation of order tensors (for OTM-based filtering of structures), evaluation of structural fitness to RDC data, calculation of the modified L-J collision, and of particular interest, back-calculation of the backbone {N, C′, and H} Residual Chemical Shift Anisotropy (RCSA) values. Chemical shift data provide complementary information to the traditional distance constraints and have had a distinct impact in the field through programs such as ShiftX (Neal et al. 2003) and TALOS (Shen et al. 2009). Chemical shift data have also been used to improve the quality of computationally modeled structures (Shen et al. 2008; van der Schot et al. 2013). Utilization of RCSA, as an extension to the Chemical Shift data can be very useful during the course of structure calculation since they are freely available as part of the RDC data acquisition. While RCSA data are generally considered to be of low quality with regard to high-resolution structure determination, recent developments have allowed for more accurate acquisition of RCSA data (Liu and Prestegard 2010). As a result, the development of structural filtering tools (similar to the modified L–J term) based on RCSA data may be easily performed. The calculation of RCSA values is effortlessly facilitated through the Object Oriented Programming paradigm of REDCRAFT, since the Residue class contains information regarding the orientation of the three principal axes of the RCSA. REDCRAFT creates individual tensor objects associated with each residue that is appropriately oriented based on the protein structure. With the available Order Tensor, which can be back-computed for optimal RMSD fitness of available RDC data, RCSA computation becomes relatively straight forward based on the formulation presented by Liu and Prestegard (Liu and Prestegard 2010). The atom-specific values of the chemical shift order parameters are shown in Table 2, which have been previously determined experimentally (Cornilescu and Bax 2000).

Software engineering

REDCRAFT (Bryson et al. 2008; Shealy et al. 2010) originated as a prototype algorithm encapsulating a number of various languages, libraries, compiled and scripted codes. It has since undergone a significant alteration, including a rewrite of its core computational engine in object oriented C++ design. REDCRAFT has also been redesigned to take advantage of multi-core modern computing environments when available. To that end, REDCRAFT accommodates parallel computer architectures by utilizing OpenMP API and distributes its computational

Table 2 Average RCSA tensor constants

Atom	σ_{11} (ppm)	σ_{22} (ppm)	σ_{33} (ppm)
$^{13}C'$	−75	−12	87
^{15}N	−108	46	63
$^{1}H^N$	−6	0	6

operations across as many available computing cores as possible.

The Object Oriented abstraction of protein folding assists in capturing the biophysical and biochemical aspect of protein folding. A programming environment, that directly reflects the physical aspect of a problem, helps in producing readable code that is easily maintainable and able to be extended for adaptation by other research groups. For instance, Display 6 illustrates the C++ code that is required to produce a tri-alanine peptide. This segment of a code is easy to read and replicate. Creation of protein structures from torsion angles using ideal peptide geometry can be a very easy exercise. The recent modifications to REDCRAFT's source code were implemented to increase the ease of development for users with little software engineering experience. More examples are available at http://ifestos.cse.sc.edu/REDCRAFT/documentation.

```
Polypeptide triala();
triala.appendAminoAcid("ALA", -60, -40);
triala.appendAminoAcid("ALA", -65, -45);
triala.appendAminoAcid("ALA", -62.5, -42.5);
triala.writePDB("triala.pdb");
```

Display 6 An example illustrating implementation of a tri-alanine peptide with backbone torsion dihedrals of (−60, −40), (−65, −45), and (−62.5, −42.5) using REDCRAFT's Object Oriented Programming libraries

Testing and validation

In this section we describe our software testing approach, which has been designed to both validate and demonstrate some of REDCRAFT's newest features. The general overview of our strategy consists of utilizing data that has been computationally generated from structures obtained from the Protein Data Bank (PDB) (Berman et al. 2000), and when possible, by using experimental data obtained from the Biological Magnetic Resonance Bank (BMRB) (Doreleijers et al. 2003; Ulrich et al. 2008). In order to avoid the previously mentioned inversion degeneracies of RDCs, all our evaluations (using either simulated or experimentally collected data) have utilized RDC data from two alignment media (Al-Hashimi et al. 2000b). The protein structures used during our testing and validation of REDCRAFT, and any additional information, are included in the subsections that follow.

Simulated data

The use of simulated data allows one to easily test the usability of new concepts within REDCRAFT while excluding the numerous unknowns (such as quality and quantity of the RDC data) that are often associated with experimental data. Our simulated RDC data were generated for the 83 residue FADD protein (PDB ID 1A1Z) (Table 3).

Simulation of RDC values for an arbitrary atomic vector requires an assumed order tensor. An order tensor (Saupe and Englert 1963) can be expressed via a 3×3 matrix, or by providing principal order parameter values S_{xx}, S_{yy}, and S_{zz} and rotational Euler angles α, β, γ. If RDC data are available along with a corresponding structure, REDCAT (Valafar and Prestegard 2004) provides a method of obtaining a best fit order tensor; it may also produce RDC values, provided a structure and an order tensor. For our experiments with two simulated alignment media the order tensors listed in Table 3 were used.

Simulated RDC data may also be modified to include the addition of simulated error or noise. Unless specified otherwise, all simulated RDCs are accompanied by simulated noise, a uniform random change in the RDC value in the range of ± 1 Hz, and contain the most complete set of data (6 RDCs per residue).

Experimental data

The application of experimental data is beneficial in both testing and illustrating the utility of REDCRAFT's newest features under more realistic conditions. Here we utilized experimental RDC data from the 56 residue third IgG-binding domain of Protein G (1P7E) and a 76 residue human Ubiquitin protein (1D3Z). While a complete set of RDC data from several alignment media have been deposited to the BMRB database, we have only utilized RDC data from two alignment media in order to establish the success of REDCRAFT under sparse data conditions.

Using experimental data obtained from the BMRB, relative order tensors were calculated using the online 2D-RDC (Mukhopadhyay et al. 2009) tool, and subsequently incorporated into REDCRAFT's order-tensor-filtering tool (one of the newest features demonstrated in this report). Traditionally, attainment of an order tensor has been possible only in the presence of RDCs assigned to a relatively high-resolution structure. These approaches are not useful in the context of structure determination due to their circular dependency on a structure, from which to calculate alignment tensors. Emergence of recent technologies (Bansal et al. 2008; Miao et al. 2008; Mukhopadhyay et al. 2009) have enabled estimation of relative order tensors that have led to structure elucidation of proteins in the absence of resonance assignment (Fahim et al. 2013). Availability of relative order tensors can be very beneficial during structure determination of proteins since they can act as very effective structural constraints. We have utilized the estimated order tensors in structure calculation of proteins 1D3Z (Cornilescu et al. 1998) and 1P7E (Ulmer et al.

Table 3 Order Tensor parameters used during simulations

	S_{xx}	S_{yy}	S_{zz}	α (°)	β (°)	δ (°)
S1	-3.00×10^{-4}	-5.00×10^{-4}	8.00×10^{-4}	0	0	0
S2	2.00×10^{-4}	5.00×10^{-4}	-7.00×10^{-4}	−40	−50	60

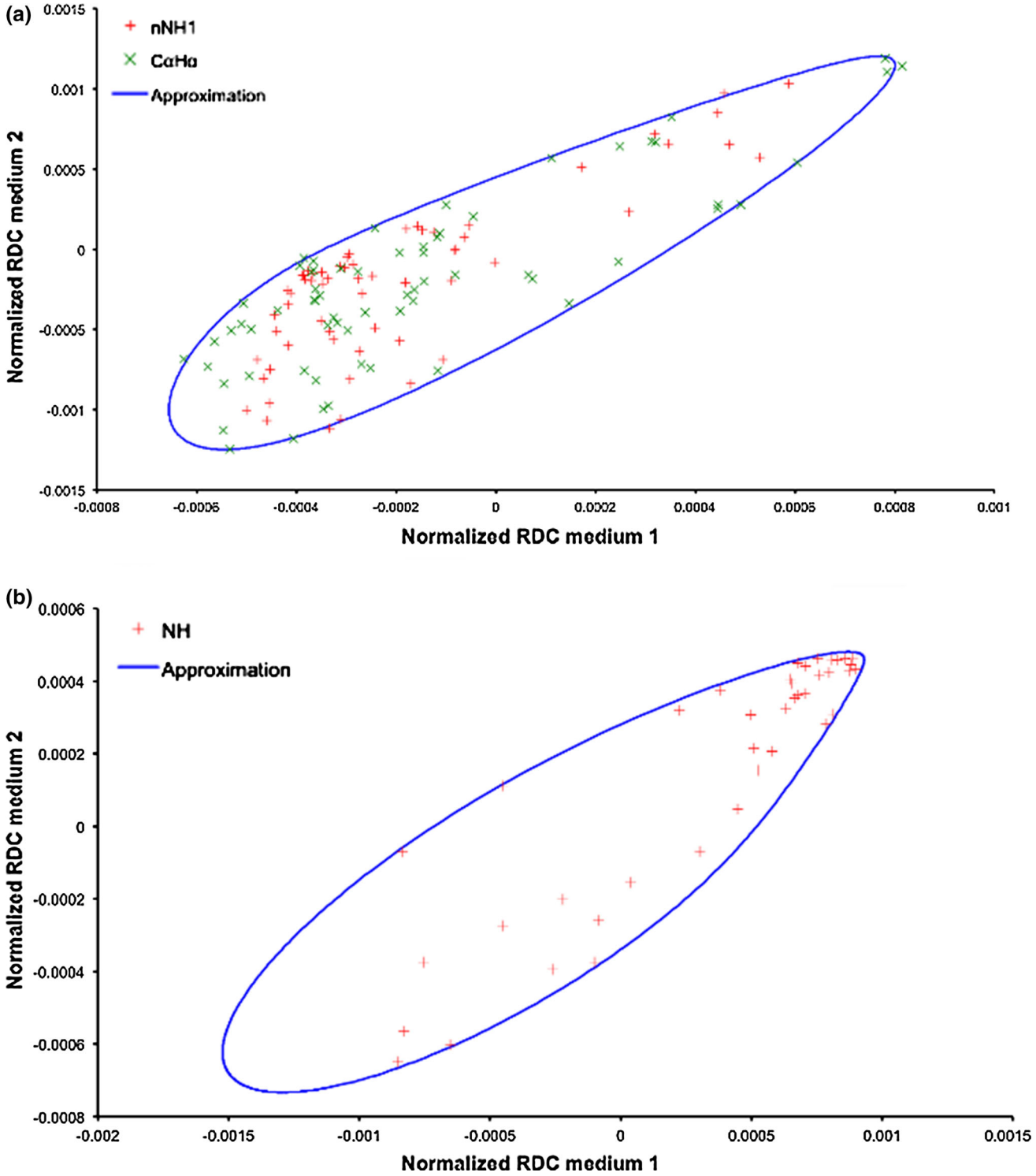

Fig. 4 Order tensors for proteins 1D3Z and 1P7E estimated using experimental RDC from 2 alignment media and the 2D-RDC online tool available at http://ifestos.cse.sc.edu/approx2D/. Maps generated from the resulting 2D-RDC order tensor estimation for **a** 1D3Z and **b** 1P7E, denote hull approximations (*blue line*) with regard to N–H and $C_\alpha H_\alpha$ vector types (denoted as *red crosses* and *green crosses*, respectively)

Table 4 Approximated relative order tensors of 1D3Z and 1P7E from two alignment media using 2D-RDC

PDB-ID	Alignment	S_{xx}	S_{yy}	S_{xy}	S_{xz}	S_{yz}
1D3Z	M1	1.99E−04	3.88E−04	0.00E+00	0.00E+00	0.00E+00
	M2	−3.17E−04	1.54E−04	3.13E−04	1.76E−04	5.08E−04
1P7E	M1	−4.85E−04	−9.51E−04	0.00E+00	0.00E+00	0.00E+00
	M2	−1.30E−04	−4.61E−04	9.24E−06	2.97E−04	1.36E−04

2003). Panels (a) and (b) of Fig. 4 illustrate 2D-RDC maps for proteins 1D3Z and 1P7E, respectively obtained by using the online server http://ifestos.cse.sc.edu/approx2D/. Table 4 lists the relative order tensors obtained from each protein using the 2D-RDC estimation mechanism. Comparison of these order tensors with those obtained from a high-resolution structure and assigned RDCs have been previously reported (Miao et al. 2008; Mukhopadhyay et al. 2009; Shealy et al. 2011).

Computational resources

The order of complexity associated with an analysis method is typically what imposes the size limitation for proteins that can be analyzed by that method. Based on the previously reported computational complexity of REDCRAFT (Bryson et al. 2008), its execution time is confined to within the first and second order polynomial functions. This streamlined operation has allowed all of our experiments to be conducted on typical desktop computers equipped with an *i7* central processing unit, 4 GB RAM, and approximately 100 MB of available storage space. Utilization of such hardware required no more than 2 h of computational time for most experiments. Under sparse data conditions, and depending on the specific configuration of REDCRAFT analysis (extent of decimation and overall depth search), some instances require as much as 12 h of computation time for analysis of the RDC data.

Although 4 GB of memory is sufficient for most applications, additional memory (such as 8 or 16 GB of RAM) is recommended to accommodate deep searches that are required under sparse data conditions. In experiments not reported here, REDCRAFT successfully produced structures for proteins in excess of 300 residues long in less than 24 h of execution time. Therefore the analysis time of REDCRAFT is not the limiting factor in calculation of protein structures from RDC data.

Results and discussion

Here we present our evaluation of REDCRAFT's newest features. Results, and their discussion, are organized in subsections corresponding to the examined feature, and include a brief description of the experiments conducted. For all subsections (except that of "An exploration of the minimum data requirement") we have illustrated the success of each feature by incorporating experiments that compare results obtained when the feature is disabled to those acquired when it is enabled.

Use of programming hooks in Levenberg–Marquardt minimization

The minimization component of REDCRAFT has been implemented as a hook that utilizes the LM minimization algorithm (Greshenfeld 1998; Press et al. 2003). Therefore results shown in this section simultaneously demonstrate the functionality of programming-hooks and structure refinement. The minimization hook has been implemented as an external Perl script (that invokes the LM algorithm). The external program receives the torsion angles that are produced by REDCRAFT's core engine as the starting point of the minimization. The set of torsion angles that correspond to the minimized structure is then communicated back to the REDCRAFT engine. The minimized set of torsion angles is inserted to the appropriate location of REDCRAFT's data structure for storage and future use. Display 7 shows one example set of torsion angles before and after the minimization. Each line terminates with the RDC-RMSD fitness of each structure.

```
Before minimization:
-40 -60 160 160 -70 -70 0.71581
After minimization:
-40 -59.0689 156.926 160.116 -74.9735 -71.193 0.43157
```

Display 7 Results before and after using REDCRAFT's minimization component implemented as a hook, utilizing the LM Algorithm

To demonstrate the efficacy of the LM-minimization, we generated one thousand variations of the Ubiquitin structure (1UBQ). Using the published experimental data, each of the variant structures was subjected to LM-minimization and the RDC-RMSD was recorded for the original and minimized structure. The derivative structures were generated by randomly altering torsion angles of the native structure to within ±5° and their BB-RMSD was calculated with respect to the published structure. Figure 5 displays RDC-RMSD as a function of BB-RMSD before and after

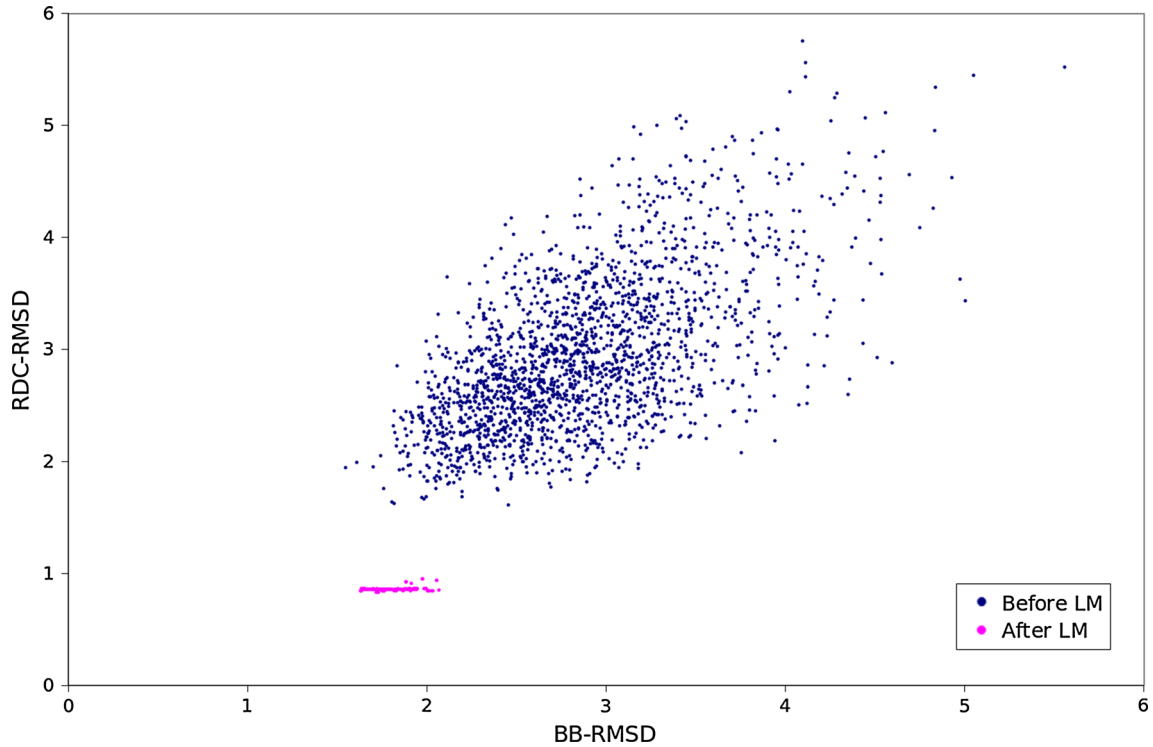

Fig. 5 Demonstrating the efficacy of Levenberg–Marquardt (LM)-minimization. One thousand variations of the 1UBQ structure were generated by randomly altering torsion angles of the native structure. Using the published experimental data, each variant was subjected to LM-minimization. RDC-RMSD was recorded for the original and minimized structures and plotted as a function of BB-RMSD before (*blue circles*) and after minimization (*magenta circles*) for each of the one thousand structures

the minimization for each of the one thousand structures. In this figure blue circles correspond to structures before minimization and magenta circles correspond to the minimized structures. The effect of minimization in improving overall fitness to RDCs and improving BB-RMSD is clearly illustrated in this figure.

As the final demonstration of this feature, we have compared the quality of structure determination with and without the use of LM-minimization. In this exercise we have utilized the protein 1D3Z and its previously published experimental data for the vector set {C–N, N–H, C–H, and C_α–H_α}. The RDC-RMSD of the protein as a function of the fragment size (as it is developed by REDCRAFT) is shown in Fig. 6. Although the difference in RDC-RMSD between the two instances is 0.2 Hz, this little difference is deceiving and could potentially result in as much as 10 Å of backbone deviation (refer to Fig. 2).

Simplified backbone–backbone collision detection

To demonstrate the impact of this structural filtering term we use the computed backbone {N–H and C_α–H_α} RDC data in two alignment media from the protein 1A1Z. The use of minimal data in addition to a shallow depth search (with decimation and minimization features disabled) is intentionally selected in order to force the REDCRAFT engine to fail in structure determination. Under these unfavorably biased conditions, the optimal geometries are ranked below the acceptable thresholds and therefore eliminated from the pool of viable conformations. Figure 7 illustrates the calculated structure of 1A1Z from residues 1–29 (a) without and (b) with the steric collision term. In each panel the red structure corresponds to the REDCRAFT computed structure and the green structure corresponds to the native 1A1Z structure. In this scenario, inclusion of our VDW term resurrects the viable geometries.

Order tensor-based filtering of structures

Given a fixed set of experimental RDC data, any hypothetical structure can be associated with its best order tensor. Therefore the task of structure determination can be guided by limiting the scope of search to structures with viable order tensors. Here, initial relative order tensors can be obtained from assigned or unassigned RDC data and used to guide the structure calculation. We demonstrate this feature in application to synthetic data on 1A1Z and experimental data on 1D3Z.

In application to 1A1Z, we have utilized {N–H and C_α–H_α} RDCs from two alignment media with a depth search

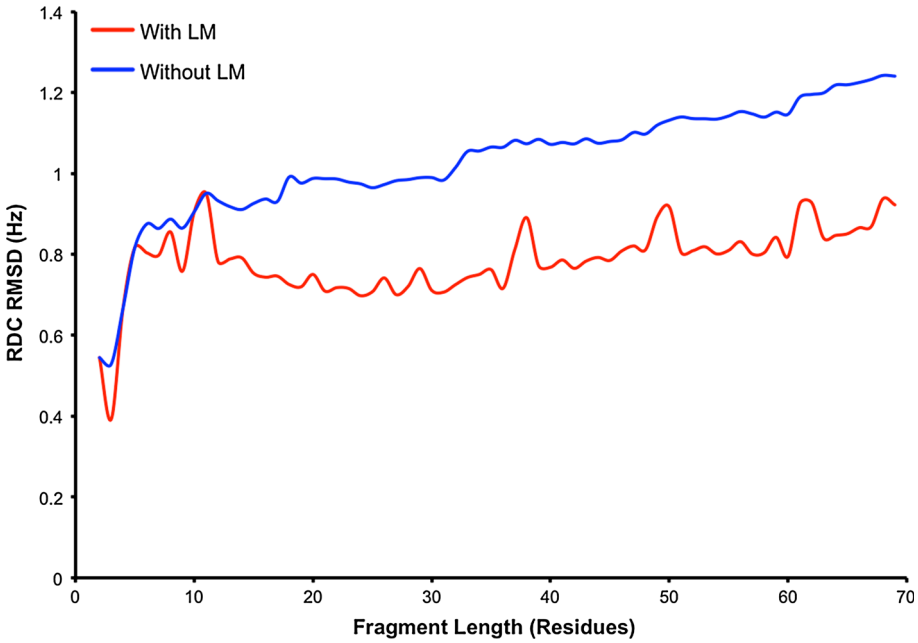

Fig. 6 Comparing the quality of structures determined from experimental data with and without the use of REDCRAFT's Levenberg–Marquardt (LM)-minimization. Experimental data for the 1D3Z protein acquired with 4 RDCs {C–N, N–H, C–H, and C_α–H_α} from 2 alignment media was utilized by REDCRAFT for 1D3Z structure determination. The resulting RDC-RMSD of the protein with LM-minimization enabled (*red line*) and LM-minimization disabled (*blue line*) are plotted as a function of the fragment size (as it is developed by REDCRAFT)

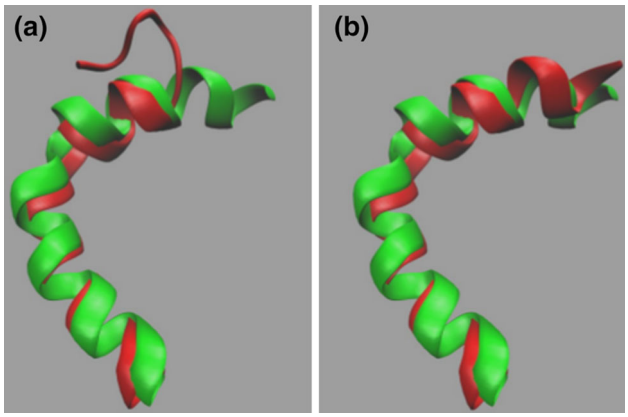

Fig. 7 Demonstrating REDCRAFT's backbone–backbone collision detection with simulated data. Using residues 1–29 of the 1A1Z protein and simulated backbone RDCs {C_α–H_α and N–H} computed from 2 alignment media, 1A1Z structures were generated by REDCRAFT with **a** the steric collision term disabled and **b** the steric collision term enabled. The resulting BB-RMSD's are 3.58 and 0.60 Å, respectively as calculated between the structures computed by REDCRAFT (*red*) and the native 1A1Z structure (*green*). All illustrations created using VMD

of 2,000. The structure calculated by REDCRAFT without the OTM-filter is shown as the red structure in Fig. 8a, and superimposed on the native structure shown in green (BB-RMSD of 12.94 Å). Similarly, the structure calculated by REDCRAFT with the OTM-filter enabled is shown as the red structure in Fig. 8b, and superimposed on the native structure shown in green (BB-RMSD of 1.32 Å).

In application to 1P7E, we have utilized {N–H and C_α–H_α} RDCs from two alignment media with a depth search of 2,000. Sparse data conditions that we have imposed resulted in a number of contiguous residues without any RDC data. Therefore the structure of this protein was calculated in four fragments: residues 1–8, 10–22, 26–39, and 41–52. The structure calculated by REDCRAFT without the OTM-filter is shown as the red structure in Fig. 9a, and superimposed on the native structure shown in green (BB-RMSD range of 0.86–3.91 Å). Similarly, the structure calculated by REDCRAFT with the OTM-filter enabled is shown as the red structure in Fig. 9b, and superimposed on the native structure shown in green (BB-RMSD range of 0.60–2.40 Å).

Bidirectional folding and fragmented structure elucidation

In order to illustrate the functional importance of bidirectional folding, we have utilized experimental data for residues 1–70 of the 1D3Z protein for the vector set {C–N, N–H, and C_α–H_α} collected from two alignment media. Although forward and reverse analysis of this protein by REDCRAFT with all features enabled produces very similar structures, for demonstration purposes we present results from a limited analysis of REDCRAFT. In this scenario, the search depth has been limited to 200 and all other features (including collision detection, decimation and order tensor based filtering of structures) have been disabled. Under these conditions, the forward and reverse-folding of the protein produced significantly different results as shown in Fig. 10. In each panel the red structure corresponds to the REDCRAFT computed structure superimposed on the native X-ray structure 1UBQ (shown

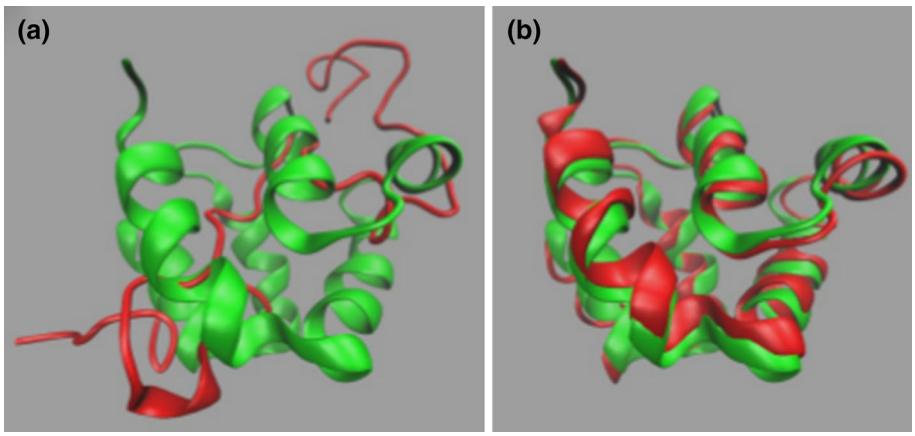

Fig. 8 Demonstrating REDCRAFT's order tensor-based filtering of structures with simulated data. Using residues 1–83 of the 1A1Z protein, simulated backbone RDCs {C_α–H_α and N–H} computed from 2 alignment media and a search depth of 2,000, 1A1Z structures were generated by REDCRAFT with **a** the Order Tensor Matrix (OTM)-filter disabled and **b** the OTM-filter enabled. The resulting BB-RMSD's are 12.94 and 1.32 Å, respectively as calculated between the structure computed by REDCRAFT (*red*) and the native 1A1Z structure (*green*). All illustrations created using VMD

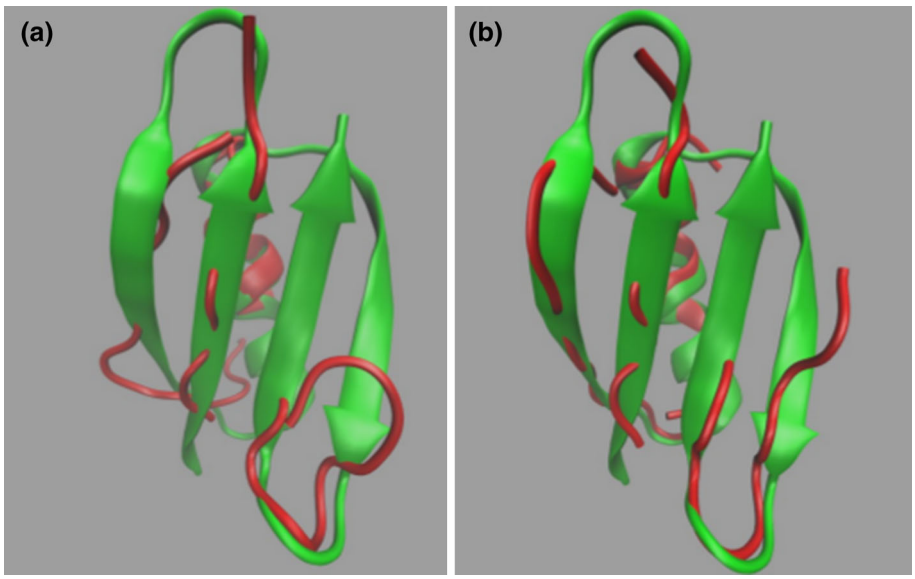

Fig. 9 Demonstrating REDCRAFT's order tensor-based filtering of structures with experimental data. Using residues 1–56 of the 1P7E protein, backbone RDCs {C_α–H_α and N–H} acquired experimentally from 2 alignment media and a search depth of 2,000, 1P7E structures were computed by REDCRAFT in four separate fragments (residues 1–8, 10–22, 26–39, and 41–52 because of gaps in the data) with **a** the OTM-filter disabled and **b** the OTM-filter enabled. BB-RMSD's ranged from 0.86–3.91 and 0.60–2.40 Å, respectively as calculated between the structure computed by REDCRAFT (*red*) and the native 1P7E structure (*green*). All illustrations created using VMD

in green). Under normal conditions, both forward (panel a) and reverse (panel b) folding should produce comparable results with only negligible variations. However, in this particular case, the results are significantly different due to structural ambiguities of residues 60–70 (as identified by a significant jump in the dynamic-profile not shown here). A reverse folding that is originated from residue 61 produces results with better agreement to the forward folding exercise (shown in Fig. 10 panel c). This exercise illustrates the power of bi-directional folding of proteins in more precise identification of structurally anomalous regions. In addition, bi-directional folding can serve as a confirmation of a successful analysis session where REDCRAFT's configuration is adequate for a given set of data.

Decimation of structures

Here we present the impact of the decimation feature utilizing both synthetic and experimental data. Using protein 1A1Z, we generated simulated RDC data from {C^{i-1}–N^i,

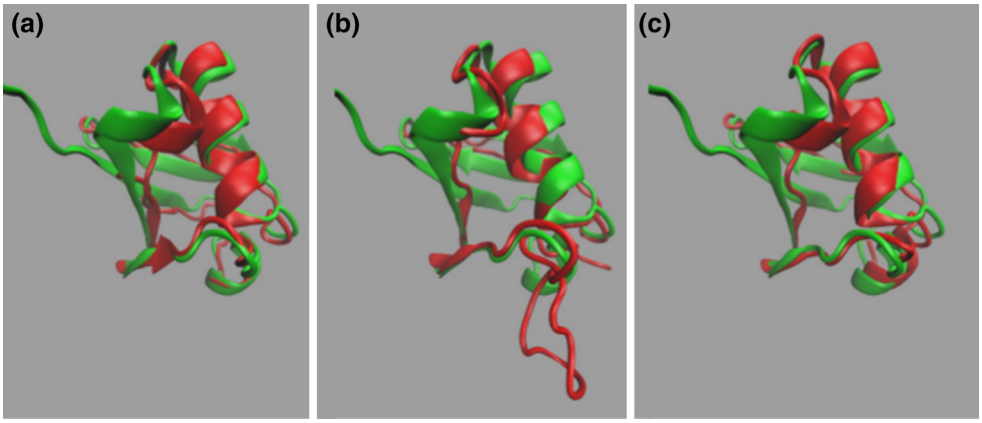

Fig. 10 Establishing the utility of REDCRAFT's bidirectional folding capabilities. Residues 2–70 of the 1D3Z protein and backbone RDCs {C–N, N–H, and C_α–H_α} acquired experimentally from 2 alignment media were utilized. In order to obtain a scenario in which forward and reverse-folding of the 1D3Z protein would produce significantly different results, a limited search depth of 200 was used with all other features (collision detection, decimation, and order tensor-based filtering of structures) disabled. Panels illustrate results for: **a** forward folding of residues 1–69, **b** reverse folding of residues 69–1, and **c** reverse folding of residues 60–1. BB-RMSD results of 1.65, 7.17, and 1.55 Å, were obtained respectively as calculated between the structure computed by REDCRAFT (*red*) and the native crystal structure 1UBQ (*green*). All illustrations created using VMD

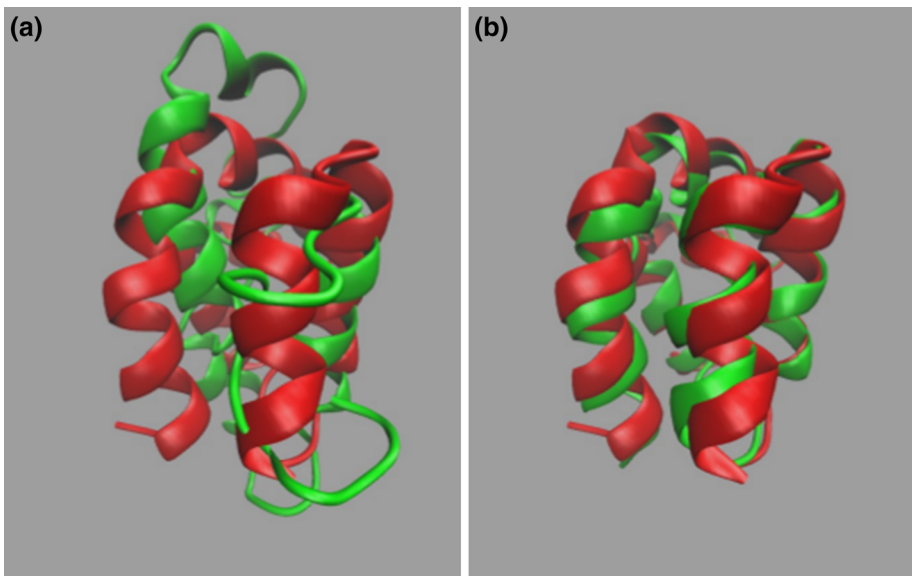

Fig. 11 Demonstrating REDCRAFT's decimation feature with simulated data. Residues 1–82 of the 1A1Z protein and simulated backbone RDCs {C^{i-1}–N^i, N^i–H^i, C^{i-1}–H^i, C^i_α–H^i_α, H^i_α–H^i, and H^{i-1}_α–H^i} computed from 2 alignment media (including uniformly distributed noise in the range of ± 4 Hz) were utilized by REDCRAFT to produce 1A1Z structures with **a** the decimation feature disabled and **b** the decimation feature enabled. The resulting BB-RMSD's are 6.22 and 1.60 Å respectively, as calculated between the structures computed by REDCRAFT (*red*) and the native 1A1Z structure (*green*). All illustrations created using VMD

N^i–H^i, C^{i-1}–H^i, C^i_α–H^i_α, H^i_α–H^i, and H^{i-1}_α–H^i} in two alignment media with uniformly distributed noise in the range of ±4 Hz. Because REDCRAFT successfully recovers the structure of this protein from the given set of data, we limit the search depth to only the top 20 structures. Under these modified conditions, although the calculated structure exhibits some similarities to the native structure, the two deviate by 6.22 Å (shown in Fig. 11a). By simply enabling the structure-decimation feature, in order to compensate for the inadequate search depth of our previous run, we are able to improve the quality of the calculated structure from 6.22 to 1.60 Å (compared to the native structure) as illustrated in Fig. 11b.

Similarly, using experimental data from the protein 1D3Z for vectors {C–N, N–H, and C_α–H_α} in two alignment media, we first conducted a REDCRAFT run with an insufficient depth search in order to obtain a poorly calculated structure. This run was then repeated with the decimation feature

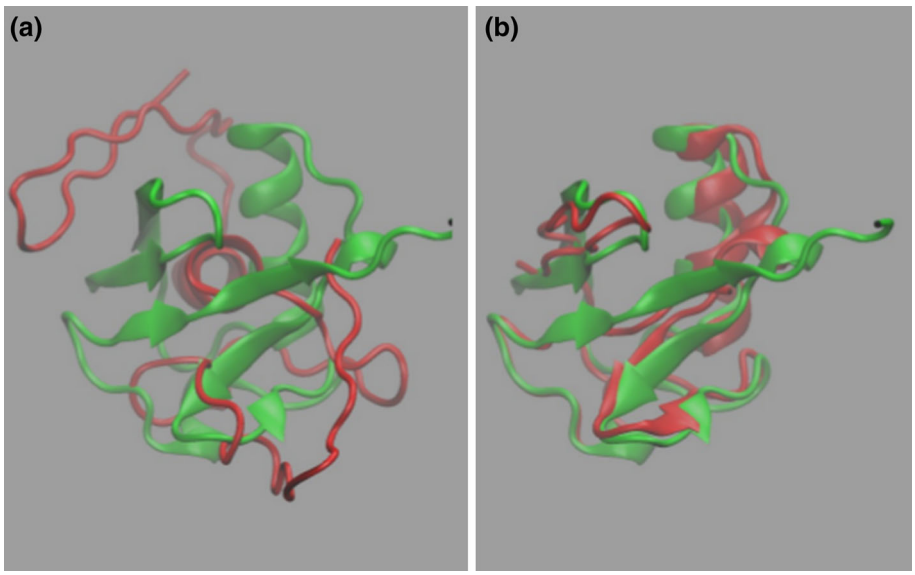

Fig. 12 Demonstrating REDCRAFT's decimation feature with experimental data. Residues 1–69 of the 1D3Z protein with backbone RDCs {C–N, N–H, and C_α–H_α} acquired experimentally from 2 alignment media and an insufficient depth search were utilized by REDCRAFT to produce 1D3Z structures with **a** the decimation feature disabled and **b** the decimation feature enabled. The resulting BB-RMSD's are 13.23 and 1.67 Å, respectively as calculated between the structure computed by REDCRAFT (*red*) and the native crystal structure 1UBQ (*green*). All illustrations created using VMD

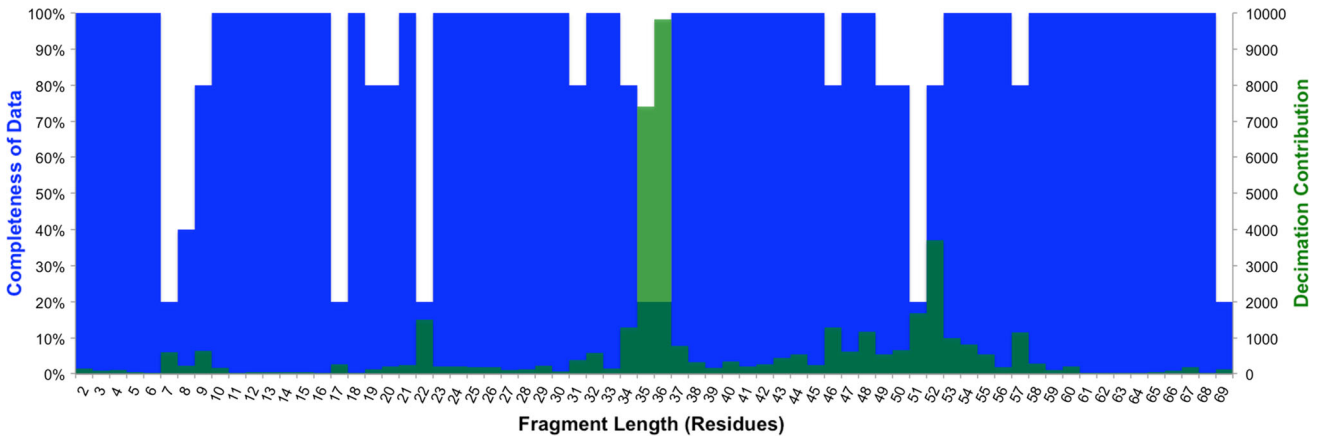

Fig. 13 Illustrating the complementarity between decimation and data-density. Two values, data completeness (*blue*) and contribution of decimation (*green*), were plotted at each residue of the 1D3Z protein. When data is sparse the contribution of decimation increases by increasing the search depth, in order to maintain more sampling of the conformational space

enabled. Results of this exercise are shown in Fig. 12. Panel (a) of this figure illustrates the superimposed backbone atoms of the native structure (X-ray structure 1UBQ) in green and the REDCRAFT computed structure without the decimation feature in red. Panel (b) of this figure illustrates the exact same result but with the decimation feature enabled. The BB-RMSD's between each of the cases and the X-ray structure are 13.23 and 1.67 Å, respectively demonstrating the impact of the decimation feature.

Figure 13 reveals just how the decimation feature is able to compensate for inadequate search depths. In this figure the blue pattern indicates the completeness of the data, and the green pattern indicates the contribution of decimation at each residue. In order to counteract the effect of diminished data-quantity, decimation increases the depth search in order to maintain more sampling of the conformational space. The varying degrees of contribution for a given quantity of data (as seen in residues 17 and 22) can be justified based on the information content of the present RDC data, as well as the spatial restriction of the conformational space at any given residue. Therefore, this figure summarizes the complementarity that exists

Table 5 A summary of REDCRAFT's ability in calculating protein structures with varying quantity and types of RDCs

Structure	Size (residues)	RDC data	BB-RMSD range (Å)	Enabled features
1D3Z	76	$\{C^{i-1}–N^i, N^i–H^i, C^{i-1}–H^i, C^i_\alpha–H^i_\alpha\}$	1.182	LJ, LM, D, d200
		$\{C^{i-1}–N^i, N^i–H^i, C^{i-1}–H^i, C^i_\alpha–H^i_\alpha\}$		
		$\{C^{i-1}–N^i, N^i–H^i, C^i_\alpha–H^i_\alpha\}$	1.622	LJ, LM, D, d200
		$\{C^{i-1}–N^i, N^i–H^i, C^i_\alpha–H^i_\alpha\}$		
		$\{C^{i-1}–N^i, N^i–H^i, C^{i-1}–H^i\}$	1.576	LJ, LM, OTE, d200
		$\{N^i–H^i\}$		
1P7E	56	$\{C^{i-1}–N^i, N^i–H^i, C^i_\alpha–H^i_\alpha\}$	0.634–1.631	F, LM, d1000
		$\{C^{i-1}–N^i, N^i–H^i, C^i_\alpha–H^i_\alpha\}$		
		$\{N^i–H^i, C^i_\alpha–H^i_\alpha\}$	0.783–1.539	F, LJ, LM, d1000
		$\{N^i–H^i, C^i_\alpha–H^i_\alpha\}$		

Key for enabled features: *D* decimation, *d*[200–1,000] depth of the searching area, *F* fragmented study, *LJ* Lennard-Jones term for collision detection, *LM* Levenberg–Marquardt minimization, *OTE* order tensor estimation

between the decimation contribution and the completeness of the data.

An exploration of the minimum data requirement

Here we examine the minimum amount of RDC data required by REDCRAFT for successful structure determination. Table 5 summarizes some examples of structures determined with varying degrees of data completeness. All sets of data utilized in this phase of testing come from experimentally collected RDCs. Results are listed from the least challenging (or most complete set of data with four RDCs per residue) to the most challenging case, and include: the name and size (in residues) of the structure used; the amount and types of RDCs utilized in each of 2 alignment media; BB-RMSD as calculated between the structure produced by REDCRAFT and the actual structure (in the case of sparse data sets, a fragmented study was conducted and the BB-RMSD is reported as a range); and which REDCRAFT features were enabled to produce the reported results.

Conclusions

In summary, utilizing both simulated and experimental data, we have demonstrated the success and applicability of some of REDCRAFT's newest features. Our example which implements the LM minimization as a hook reveals the flexibility that is provided when interaction with other programs is allowed, while also exhibiting how the LM minimization feature offers the option of refining structures without replacing any of the previously computed candidate structures. Additionally, our resolution of improper backbone torsion angles illustrates the benefit of including the backbone–backbone collision detection feature in order to eliminate unlikely structural candidates. Furthermore, to accommodate recent developments in the research community which have enabled estimation of relative order tensors when RDC data are available from multiple alignment media, we have made modifications which leverage a priori knowledge of order tensors by incorporating order tensor-based filtering of structures. Moreover, the capabilities of fragmented and/or bidirectional study offer an alternative means of structure determination especially in regions where data is sparse. Decimation of structures is yet another useful feature which provides a new means of reducing the complexity of a large solution space. Lastly and perhaps more importantly, we have explored the minimal data requirement and demonstrated REDCRAFT's success in determining protein structures using as little as two sets of RDC data per residue.

Approaches, such as REDCRAFT, which incorporate partial-structure determination methods are analogous to divide-and-conquer strategies implemented in the computational sciences. In fact, from a computational stand point, divide-and-conquer techniques are often more efficient at solving complex problems as opposed to the alternative task of solving a problem in its entirety. With regard to complete structure determination, it is simpler to envision a strategy involving two incremental steps: backbone structure determination, followed by the packing of side-chains. Although a high-resolution description of the backbone will reduce the degrees of freedom in packing of the side-chains, backbone structure determination from RDC data followed by side-chain packing using either computational methods or NOE data will require less data than the alternative of performing complete structure determination at once. Additional scenarios employing partial-structure determination include protein structures that undergo

conformational changes, where the focus of structure determination is only on a portion of the protein and not the entire protein.

Acknowledgments This work was supported by NIH Grant Numbers 1R01GM081793 and P20 RR-016461 to Dr. Homayoun Valafar.

Conflict of interest The authors declare that they have no conflict of interest.

References

Adeyeye J, Azurmendi HF, Stroop CJM, Sozhamannan S, Williams AL, Adetumbi AM, Johnson JA, Bush CA (2003) Conformation of the hexasaccharide repeating subunit from the Vibrio cholerae O139 capsular polysaccharide. Biochemistry 42:3979–3988

Al-Hashimi HM, Bolon PJ, Prestegard JH (2000a) Molecular symmetry as an aid to geometry determination in ligand protein complexes. J Magn Reson 142:153–158

Al-Hashimi HM, Valafar H, Terrell M, Zartler ER, Eidsness MK, Prestegard JH (2000b) Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings. J Magn Reson 143:402–406

Al-Hashimi HM, Gorin A, Majumdar A, Gosser Y, Patel DJ (2002a) Towards structural Genomics of RNA: rapid NMR resonance assignment and simultaneous RNA tertiary structure determination using residual dipolar couplings. J Mol Biol 318:637–649

Al-Hashimi HM, Gosser Y, Gorin A, Hu W, Majumdar A, Patel DJ (2002b) Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings. J Mol Biol 315:95–102

Andrec M, Du P, Levy RM (2001) Protein backbone structure determination using only residual dipolar couplings from one ordering medium. J Biomol NMR 21:335–347

Assfalg M, Bertini I, Turano P, Grant Mauk A, Winkler JR, Gray HB (2003) 15 N–1H Residual dipolar coupling analysis of native and alkaline-K79A Saccharomyces cerevisiae cytochrome c. Biophys J 84:3917–3923

Azurmendi HF, Bush CA (2002) Conformational studies of blood group A and blood group B oligosaccharides using NMR residual dipolar couplings. Carbohydr Res 337:905–915

Azurmendi HF, Martin-Pastor M, Bush CA (2002) Conformational studies of Lewis X and Lewis A trisaccharides using NMR residual dipolar couplings. Biopolymers 63:89–98

Banci L, Bertini I, Luchinat C, Mori M (2010) NMR in structural proteomics and beyond. Prog Nucl Magn Reson Spectrosc 56:247–266

Bansal S, Miao X, Adams MWW, Prestegard JH, Valafar H (2008) Rapid classification of protein structure models using unassigned backbone RDCs and probability density profile analysis (PDPA). J Magn Reson 192:60–68

Bax A, Tjandra N (1997) High-resolution heteronuclear NMR of human ubiquitin in an aqueous liquid crystalline medium. J Biomol NMR 10:289–292

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

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242

Bertini I, Luchinat C, Turano P, Battaini G, Casella L (2003) The magnetic properties of myoglobin as studied by NMR spectroscopy. Chem Eur J 9:2316–2322

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

Bryson M, Tian F, Prestegard JH, Valafar H (2008) REDCRAFT: a tool for simultaneous characterization of protein backbone structure and motion from RDC data. J Magn Reson 191:322–334

Cavanagh J, Fairbrother WJ, Palmer AG, Rance M, Skelton NJ (2007) Protein NMR Spectroscopy: Principles and Practice, 2nd Edition. Academic Press, London

Chakraborty S, Venkatramani R, Rao BJ, Asgeirsson B, Dandekar AM (2013) Protein structure quality assessment based on the distance profiles of consecutive backbone Cα atoms. F1000Research 2:211

Clore GM, Bewley CA (2002) Using conjoined rigid body/torsion angle simulated annealing to determine the relative orientation of covalently linked protein domains from dipolar couplings. J Magn Reson 154:329–335

Clore GM, Gronenborn AM, Bax A (1998) 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 133:216–221

Cornilescu G, Bax A (2000) Measurement of proton, nitrogen, and carbonyl chemical shielding anisotropies in a protein dissolved in a dilute liquid crystalline phase. J Am Chem Soc 122:10143–10154

Cornilescu G, Marquardt JL, Ottiger M, Bax A (1998) Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. J Am Chem Soc 120:6836–6837

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

Delaglio F, Kontaxis G, Bax A (2000) Protein structure determination using molecular fragment replacement and NMR dipolar couplings. J Am Chem Soc 122:2142–2143

Doreleijers JF, Mading S, Maziuk D, Sojourner K, Yin L, Zhu J, Markley JL, Ulrich EL (2003) BioMagResBank database with sets of experimental NMR constraints corresponding to the structures of over 1400 biomolecules deposited in the Protein Data Bank. J Biomol NMR 26:139–146

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

Fahim A, Mukhopadhyay R, Yandle R, Prestegard JH, Valafar H (2013) Protein structure validation and identification from unassigned residual dipolar coupling data using 2D-PDPA. Molecules 18:10162–10188

Fowler CA, Tian F, Al-Hashimi HM, Prestegard JH (2000) Rapid determination of protein folds using residual dipolar couplings. J Mol Biol 304:447–460

Greshenfeld NA (1998) The Nature of Mathematical Modeling. Cambridge University Press, Cambridge

Güntert P, Mumenthaler C, Wüthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 273:283–298

Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4:435–447

Hus J-C, Marion D, Blackledge M (2001) Determination of protein backbone structure using only residual dipolar couplings. J Am Chem Soc 123:1541–1542

Kummerlöwe G, Luy B (2009) Residual dipolar couplings as a tool in determining the structure of organic molecules. TrAC, Trends Anal Chem 28:483–493

Levenberg K (1944) A method for the solution of certain problems in least squares. Q Appl Math 2:164–168

Liu Y, Prestegard JH (2010) A device for the measurement of residual chemical shift anisotropy and residual dipolar coupling in soluble and membrane-associated proteins. J Biomol NMR 47:249–258

Losonczi JA, Andrec M, Fischer MW, Prestegard JH (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition. J Magn Reson 138:334–342

Miao X, Mukhopadhyay R, Valafar H (2008) Estimation of relative order tensors, and reconstruction of vectors in space using unassigned RDC data and its application. J Magn Reson 194:202–211

Mukhopadhyay R, Miao X, Shealy P, Valafar H (2009) Efficient and accurate estimation of relative order tensors from lambda-maps. J Magn Reson 198:236–247

Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP—a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 247:536–540

Neal S, Nip AM, Zhang HY, Wishart DS (2003) Rapid and accurate calculation of protein H-1, C-13 and N-15 chemical shifts. J Biomol NMR 26:215–240

Opella SJ, Nevzorov A, Mesleh MF, Marassi FM (2002) Structure determination of membrane proteins by NMR spectroscopy. Biochem Cell Biol 80:597–604

Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM (1997) CATH—a hierarchic classification of protein domain structures. Structure 5:1093–1108

Park SH, Son WS, Mukhopadhyay R, Valafar H, Opella SJ (2009) 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 131:14140–14141

Press WH, Teukolsky SA, Vettering WT, Flannery BP (2003) Numerical recipes in C ++: the art of scientific computing (2nd edn) 1 numerical recipes example book (C ++) (2nd edn) 2 numerical recipes multi-language Code CD ROM with LINUX or UNIX single-screen license revised version 3. Eur J Phys 24:329–330

Prestegard JH, Al-Hashimi HM, Tolman JR (2000) NMR structures of biomolecules using field oriented media and residual dipolar couplings. Q Rev Biophys 33:371–424

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

Prestegard JH, Bougault CM, Kishore AI (2004) Residual dipolar couplings in structure determination of biomolecules. Chem Rev 104:3519–3540

Prestegard JH, Mayer KL, Valafar H, Benison GC (2005) Determination of protein backbone structures from residual dipolar couplings. Methods Enzymol 394:175–209

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

Schmidt C, Irausquin SJ, Valafar H (2013) Advances in the REDCAT software package. BMC Bioinform 14:302

Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. J Magn Reson 160:65–73

Shealy P, Simin M, Park SH, Opella SJ, Valafar H (2010) Simultaneous structure and dynamics of a membrane protein using REDCRAFT: membrane-bound form of Pf1 coat protein. J Magn Reson 207:8–16

Shealy P, Liu Y, Simin M, Valafar H (2011) Backbone resonance assignment and order tensor estimation using residual dipolar couplings. J Biomol NMR 50:357–369

Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu G, Eletsky A, Wu Y, Singarapu KK, Lemak A, Ignatchenko A, Arrowsmith CH, Szyperski T, Montelione GT, Baker D, Bax A (2008) Consistent blind protein structure generation from NMR chemical shift data. Proc Natl Acad Sci U S A 105:4685–4690

Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS + : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44:213–223

Thiele CM (2008) Residual dipolar couplings (RDCs) in organic structure determination. Eur J Org Chem 2008:5673–5685

Tian F, Al-Hashimi HM, Craighead JL, Prestegard JH (2001a) Conformational analysis of a flexible oligosaccharide using residual dipolar couplings. J Am Chem Soc 123:485–492

Tian F, Valafar H, Prestegard JH (2001b) A dipolar coupling based strategy for simultaneous resonance assignment and structure determination of protein backbones. J Am Chem Soc 123:11791–11796

Tjandra N, Grzesiek S, Bax A (1996) 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 118:6264–6272

Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A (1997) Use of dipolar H-1–N-15 and H-1–C-13 couplings in the structure determination of magnetically oriented macromolecules in solution. Nat Struct Biol 4:732–738

Tjandra N, Tate S, Ono A, Kainosho M, Bax A (2000) The NMR structure of a DNA dodecamer in an aqueous dilute liquid crystalline phase. J Am Chem Soc 122:6190–6200

Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH, Tolman Flanagan JM, Kennedy MA, Prestegard JHJR (1995) Nuclear magnetic dipole interactions in field-oriented proteins - information for structure determination in solution. Proc Natl Acad Sci U S A 92:9279–9283

Ulmer TS, Ramirez BE, Delaglio F, Bax A (2003) Evaluation of backbone proton positions and dynamics in a small protein by liquid crystal NMR spectroscopy. J Am Chem Soc 125:9179–9191

Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Kent Wenger R, Yao H, Markley JL (2008) BioMagResBank. Nucleic Acids Res 36:D402–D408

Valafar H, Prestegard JH (2003) Rapid classification of a protein fold family using a statistical analysis of dipolar couplings. Bioinformatics 19:1549–1555

Valafar H, Prestegard JH (2004) REDCAT: a residual dipolar coupling analysis tool. J Magn Reson 167:228–241

Valafar H, Mayer K, Bougault C, LeBlond P, Jenney FE, Brereton PS, Adams M, Prestegard JH (2005) Backbone solution structures of proteins using residual dipolar couplings: application to a novel structural genomics target. J Struct Funct Genomics 5:241–254

Valafar H, Simin M, Irausquin S (2012) A Review of REDCRAFT: simultaneous Investigation of Structure and Dynamics of Proteins from RDC Restraints. Annu Reports NMR Spectrosc 76:23–66

Van der Schot G, Zhang Z, Vernon R, Shen Y, Vranken WF, Baker D, Bonvin AMJJ, Lange OF (2013) Improving 3D structure prediction from chemical shift data. J Biomol NMR 57:27–35

Vermeulen A, Zhou H, Pardi A (2000) Determining DNA global structure and DNA bending by application of NMR residual dipolar couplings. J Am Chem Soc 122:9638–9647

Wang L, Donald BR (2004) 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 29:223–242

Warren JJ, Moore PB (2001) A maximum likelihood method for determining D(a)(PQ) and R for sets of dipolar coupling data. J Magn Reson 149:271–275

Wuthrich K (1986) NMR of proteins and nucleic acids. Georg. Fish. Bak. Non-resident Lecturesh. Chem. Cornell University

Zeng J, Boyles J, Tripathy C, Wang L, Yan A, Zhou P, Donald BR (2009) High-resolution protein structure determination starting with a global fold calculated from exact solutions to the RDC equations. J Biomol NMR 45:265–281

Zweckstetter M (2008) NMR: prediction of molecular alignment from structure using the PALES software. Nat Protoc 3:679–690