Journal of Biomolecular NMR **29:** 223–242, 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

Lincong Wang[a] & Bruce Randall Donald[a,b,c,*,**]
[a]*Dartmouth Computer Science Department,* [b]*Dartmouth Chemistry Department, and* [c]*Dartmouth Department of Biological Sciences, Hanover, NH 03755, U.S.A.*

Received 26 June 2003; Accepted 11 February 2004

Key words: algorithms for protein structure determination, conformational search, exact solutions for backbone dihedral angles, exact solutions for internuclear vectors, global fold determination, high-throughput NMR methods, protein kinematics, residual dipolar couplings, structural genomics, systematic search

Abstract

We have derived a quartic equation for computing the direction of an internuclear vector from residual dipolar couplings (RDCs) measured in two aligning media, and two simple trigonometric equations for computing the backbone (ϕ, ψ) angles from two backbone vectors in consecutive peptide planes. These equations make it possible to compute, *exactly* and *in constant time*, the backbone (ϕ, ψ) angles for a residue from RDCs in two media on any *single* backbone vector type. Building upon these exact solutions we have designed a novel algorithm for determining a protein backbone substructure consisting of α-helices and β-sheets. Our algorithm employs a systematic search technique to refine the conformation of both α-helices and β-sheets and to determine their orientations using exclusively the angular restraints from RDCs. The algorithm computes the backbone substructure employing very sparse distance restraints between pairs of α-helices and β-sheets refined by the systematic search. The algorithm has been demonstrated on the protein human ubiquitin using only backbone NH RDCs, plus twelve hydrogen bonds and four NOE distance restraints. Further, our results show that both the global orientations and the conformations of α-helices and β-strands can be determined with high accuracy using only two RDCs per residue. The algorithm requires, as its input, backbone resonance assignments, the identification of α-helices and β-sheets as well as sparse NOE distance and hydrogen bond restraints.

Abbreviations: NMR – nuclear magnetic resonance; RDC – residual dipolar coupling; NOE – nuclear Overhauser effect; SVD – singular value decomposition; DFS – depth-first search; RMSD – root mean square deviation; POF – principal order frame; PDB – protein data bank; SA – simulated annealing; MD – molecular dynamics.

Introduction

The increasing gap between the speeds of DNA sequencing and protein structure determination requires the development of efficient algorithms to compute structures as accurately as possible using only a minimum number of restraints obtainable rapidly by experimental techniques. One way to achieve this is to develop algorithms whose key components are analytic expressions computable *in constant time.*[1] Here we present such an algorithm for determining a protein backbone structure using global angular (orientational) restraints on internuclear vectors derived from backbone residual dipolar couplings (RDCs) measured in two aligning media (Tjandra and Bax, 1997;

*To whom correspondence should be addressed: 6211 Sudikoff Laboratory, Dartmouth Computer Science Department, Hanover, NH 03755, U.S.A. E-mail: brd@cs.dartmouth.edu
**This work is supported by the following grants to B.R.D.: National Institutes of Health (R01 GM 65982), National Science Foundation (IIS-9906790, EIA-0102710, EIA-0102712, EIA-9818299, EIA-0305444 and EIA-9802068), and the John Simon Guggenheim Foundation.

Tolman et al., 1995). The RDCs can be recorded and assigned much faster than nuclear Overhauser effects (NOEs) required by traditional NMR structure determination methods. Months of time can be required to assign a sufficient number of NOEs, especially those involving sidechain protons, to compute an accurate NMR structure. Therefore, RDCs are better suited for developing high-throughput structure determination methods. For example, algorithms have been designed to compute a protein fold with RDCs alone or with RDCs plus other restraints such as chemical shifts or sparse NOEs (Andrec et al., 2001; Delaglio et al., 2000; Fowler et al., 2000; Giesen et al., 2003; Hus et al., 2001; Rohl and Baker, 2002; Tian et al., 2001). These methods require several sets of RDCs in one or two media and use a fragment replacement method (Andrec et al., 2001; Delaglio et al., 2000; Rohl and Baker, 2002), rely on heuristic algorithms and molecular dynamics (Clore et al., 1999; Giesen et al., 2003; Hus et al., 2001), or employ RDCs only for orienting ideal helices (Fowler et al., 2000). The fragment replacement method relies heavily on statistics from the Protein Databank (PDB) (Berman et al., 2000). A different search technique, systematic search, has been used successfully to determine the structure of a tripeptide by solid-state NMR (Rienstra et al., 2002). A systematic search is a search over all possible conformations (solutions) that employs a provable pruning strategy that guarantees pruned conformations need not be considered further. In this paper we demonstrate that by combining systematic search with exact solutions for computing, first, the directions of an NH vector, then, (ϕ, ψ) angles in constant time for a single residue, it is possible to compute a backbone substructure consisting of α-helices and β-sheets using only RDCs in two media on a single backbone vector type plus very sparse distance restraints. Further, our algorithm uses only the averages for the backbone (ϕ, ψ) angles from the PDB and does not rely on molecular dynamics.

Theoretical background

In this section we outline the derivation of a quartic equation used to compute the orientation of an internuclear unit vector from RDCs in two media, and two simple trigonometric equations used to compute the backbone (ϕ, ψ) angles of residue i given one unit vector in peptide plane i and another unit vector in plane $i + 1$. These unit vectors are computed from the quartic equation. Interested readers can refer to Appendices A, B for the full details of derivation of the quartic and trigonometric equations. Then, we describe a recursive strategy for computing, *consecutively*, the (ϕ, ψ) angles of all the residues of a structural fragment.

Computation of vector orientations

The equations for NH RDCs measured in two media can be written as (Saupe, 1968):

$$D_{NH} = D_{\max}(S_{xx}x^2 + S_{yy}y^2 + S_{zz}z^2), \quad (1)$$

$$D'_{NH} = D_{\max}(S'_{xx}x'^2 + S'_{yy}y'^2 + S'_{zz}z'^2), \quad (2)$$

where $D_{\max}$ is the dipolar interaction constant, D_{NH}, an NH RDC value in medium 1, S_{xx}, S_{yy} and S_{zz} the three diagonal elements of a diagonalized Saupe matrix $\mathbf{S}$ (the alignment tensor) for medium 1. A Saupe matrix (a 3×3 traceless and symmetric matrix) specifies the ensemble-averaged anisotropic orientation of a molecule in the laboratory frame. x, y and z are, respectively, the x, y, $z-$components of an internuclear NH unit vector $\mathbf{v} = (x, y, z)$ in a principal order frame (POF) which diagonalizes $\mathbf{S}$. D'_{NH} and S'_{xx}, S'_{yy}, S'_{zz} of $\mathbf{S}'$ are the corresponding variables for medium 2 and $\mathbf{v}' = (x', y', z')$ is the same NH unit vector in a POF for medium 2. $\mathbf{v}$ and $\mathbf{v}'$ are related by

$$\mathbf{v}' = \mathbf{R_{12}}\mathbf{v}, \quad (3)$$

where $\mathbf{R}_{12}$ is a relative rotation matrix between the two POFs of medium 1 and 2. From Equations (1–3) we have derived a quartic equation satisfied by x^2:

$$f_4u^4 + f_3u^3 + f_2u^2 + f_1u + f_0 = 0, \quad (4)$$

$$u = 1 - 2\left(\frac{x}{a}\right)^2,$$

where the coefficients f_0, f_1, f_2, f_3, f_4 and a are computed from $\mathbf{S}$, $\mathbf{S}'$ and $\mathbf{R}_{12}$ which, in turn, can be computed from the alignment tensors as detailed in the *Computation of alignment tensors* section. Full expressions for these coefficients are provided in Appendix A. From a given x, y can be computed directly from Equation (1) by solving a quadratic equation. Due to symmetry in Equations (1, 2) the number of real solutions for $\mathbf{v}$ is at most 8 (Figure 4A), in agreement with what has been found previously by other methods (Al-Hashimi et al., 2000; Ramirez and Bax, 1998).

Table 1. Distance restraints derived from NOEs. The distance restraints between H^N or H_α and methyl groups are extracted from the NOE table of the restraint file with PDB ID 1D3Z (Cornilescu et al. 1998) for ubiquitin and are converted into the distances between the two nuclei H^N and C_α or between C_α and C_α for convenience. A value of 8.0 Å is assigned to each NOE-derived distance using a value of 5.0 Å for upper bound and 3.0 Å for pseudoatom correction

Residue No.	Nucleus	Residue No.	Nucleus	Distance
26	C_α	3	C_α	8.0 Å
27	H^N	43	C_α	8.0 Å
26	C_α	15	C_α	8.0 Å
30	H^N	15	C_α	8.0 Å

Computation of dihedral angles for a single residue

If the directions of any two vectors $\mathbf{v}_i$ and $\mathbf{v}_{i+1}$ in consecutive peptide planes i and $i+1$ are known we then showed that the intervening backbone angles (ϕ_i, ψ_i) can be computed from the following two simple trigonometric equations:

$$\sin(\phi_i + a_1) = b_1, \qquad \sin(\psi_i + a_2) = b_2, \quad (5)$$

where a_1 and b_1 are computed from the six backbone angles (Figure 1 and Table 4) between two consecutive residues and $\mathbf{v}_i$ and $\mathbf{v}_{i+1}$, while a_2 and b_2 are computed from the six angles, $\mathbf{v}_i$ and $\mathbf{v}_{i+1}$ and a solution for ϕ_i. The full expressions for a_1, b_1, a_2 and b_2 are computed from backbone kinematics, based on 8 rotations that transform a coordinate system (Figure 1) defined in peptide plane i to an isomorphic system in plane $i+1$, and are given in Appendix B. In the present paper, $\mathbf{v}_i$ and $\mathbf{v}_{i+1}$ are, respectively, the NH unit vectors for residue i and $i+1$, which are both obtained by solving Equation 4. Equation 5 can be solved exactly for $\sin\phi_i$ and $\sin\psi_i$ because Equation 5 is equivalent to two quadratic equations (Appendix B). There are 2 possible solutions for (ϕ_i, ψ_i) angles from Equation 5 for each orientation of the peptide plane i. However, there will be an infinite number of such solutions if the orientation of the peptide plane i is not known.

Computation of (ϕ, ψ) *angles for all the residues of a structural fragment*

As stated in the previous section, only after the peptide plane i is known can we obtain a finite number of solutions for the (ϕ_i, ψ_i) angles (up to $8 \times 8 \times 2$: 8 solutions are possible for each NH vector, and, given 2 consecutive NH vectors, 2 solutions are possible for the intervening (ϕ, ψ) angles). In fact, given the first peptide plane all the possible (ϕ, ψ) angles of a fragment can be computed by exploiting the following two observations:

1. A peptide plane i with respect to a POF for medium 1 can be determined by its NC_α vector and an NH vector from Equation 5.
2. A unique NC_α vector for the peptide plane $i+1$ can be computed from the NC_α vector of the peptide plane i and (ϕ_i, ψ_i).

Therefore, given the NC_α vector of the first peptide plane (called the first NC_α vector for brevity) of a fragment we can use Equations 4, 5 to compute, *consecutively*, all the possible discrete (ϕ, ψ) solutions for all the residues of the fragment. Moreover, such a scheme is ideally suited for employing a depth-first search $(DFS)^2$ strategy to select a best set of (ϕ, ψ) angles, out of all possible such sets, for the fragment based on some score function. The method for computing the first NC_α vector of a fragment will be detailed in the *Computation of an optimal first* NC_α *vector* section.

The algorithm

Building upon our newly derived Equations 4, 5 and the two observations in the previous section, we have developed a novel algorithm for determining a 3-dimensional backbone substructure consisting of the α-helices and β-sheets using the following input: assigned backbone NH RDCs in two media, identified α-helices and β-sheets with known hydrogen bonds (H-bonds) between paired strands and very sparse NOE distance restraints. Our algorithm is divided into three stages (Figure 2): (I) *computation of alignment tensors and an optimal first* NC_α *vector*, (II) *DFS-based refinement of secondary structure elements (both* α*-helices and* β*-strands)* and (III) *backbone structure determination*. The alignment tensors and the first optimal NC_α vector from stage I are used in stage II for computing all the backbone (ϕ, ψ) angles of a structure element using RDCs alone. Their relative positions are determined in stage III using H-bonds and NOE restraints. The interested reader can refer to Appendix C for the tree data structure used for DFS and a pseudocode of the algorithm for stage II.

Computation of alignment tensors

The algorithm begins with the computation of Saupe matrices for both media using an ideal helix model

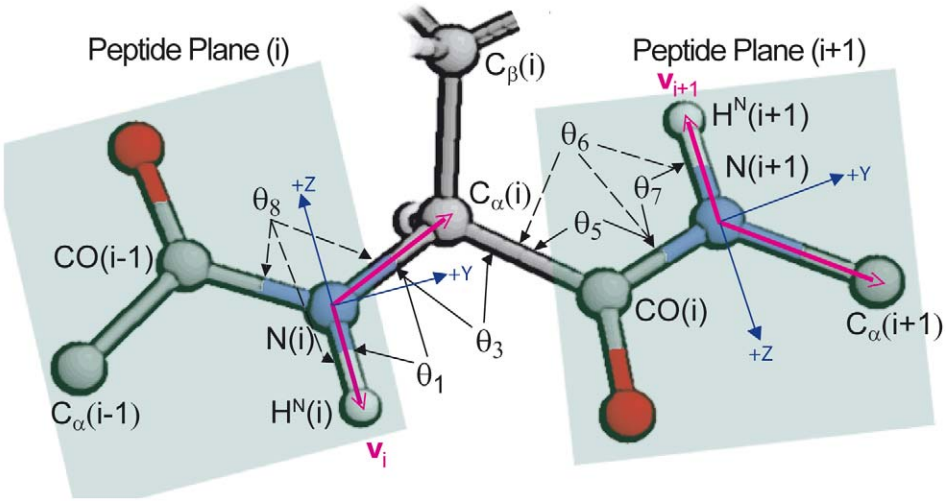

Figure 1. The six backbone angles. The six angles (θ_1,θ_3,θ_5,θ_6,θ_7 and θ_8) required for computing the coefficients of the quartic equation (Equation (5)) and for building α-helix and β-strand models. A plane angle (θ_1, θ_3, θ_5 or θ_7) is labeled with solid arrows pointing to its two defining bond vectors while a dihedral angle (θ_6 or θ_8) is labeled with dashed arrows pointing to its three defining bond vectors. θ_8 is the dihedral angle between the plane defined by the vector from CO($i-1$) to N(i) and the NH vector, and the plane defined by the NH and NC$_\alpha$ vectors. The NH and NC$_\alpha$ vectors are labeled with magenta arrows. +Y and +Z are, respectively, the y-axis and z-axis of a coordinate system defined in peptide plane i and $i+1$. The x-axis is defined based on right handedness.

built with the backbone (ϕ, ψ) = ($-65.3°$, $-39.4°$), the six angles (Figure 1 and Table 4) and standard values for bond lengths (Engh and Huber, 1991). The ($-65.3°$, $-39.4°$) is the average of the backbone (ϕ, ψ) for α-helices over the PDB. The Saupe matrices $\mathbf{S}$ for medium 1 and $\mathbf{S}'$ for medium 2 are computed from the model by a SVD method (Losonczi et al., 1999). They are diagonalized to yield, respectively, three diagonal elements and four possible rotation matrices $\mathbf{R}_1$ for medium 1, and three diagonal elements and four possible matrices $\mathbf{R}_2$ for medium 2. One $\mathbf{R}_1$ (POF) is chosen arbitrarily for medium 1 for all subsequent computations. The relative rotation matrix $\mathbf{R}_{12}$ between the POFs of medium 1 and 2 can be computed from $\mathbf{R}_{12} = \mathbf{R}_2\mathbf{R}_1^{-1}$. The coefficients of Equation (5) are computed from $\mathbf{R}_{12}$ and any two diagonal elements for each medium. Out of sixteen possible relative rotational matrices (i.e. $\mathbf{R}_{12}$) only four represent different orientations and one of them is chosen based on NOE distance restraints (detailed in the *backbone structure determination* section).

In stage I, alignment tensors are refined as follows. First, the helix structure is refined (as described in the *refinement of secondary structure elements* section) using only RDCs (Figure 2). In particular, the helix refinement uses the alignment tensors previously computed from an ideal helix and a first NC$_\alpha$ vector (computed from $\mathbf{R_1}\mathbf{v}_n$ where $\mathbf{v}_n$ is the first NC$_\alpha$ vector of the ideal helix). To improve accuracy, new alignment tensors are then computed by the SVD method using *the refined helix structure*. The new tensors are then used for computing all other secondary structure elements.

Computation of an optimal first NC$_\alpha$ vector

An optimal first NC$_\alpha$ vector is computed using an m residue fragment built with the average (ϕ, ψ) angles for either an α-helix or a β-strand and standard bond lengths and angles (Figure 1 and Table 4) (Engh and Huber, 1991). A grid search for the orientation of the fragment in a POF of medium 1, defined by a Euler rotation matrix $\mathbf{R}(\alpha, \beta, \gamma)$, is performed to minimize the score function $\sigma_{RMS}(\alpha, \beta, \gamma)$:

$$\sigma^2_{RMS}(\alpha,\beta,\gamma) = \frac{1}{m}\sum_{i=1}^{m}\Big((D^c_{NH,i}(\alpha,\beta,\gamma) - D_{NH,i})^2 + (D'^c_{NH,i}(\alpha,\beta,\gamma) - D'_{NH,i})^2\Big), \quad (6)$$

where $D_{NH,i}$ and $D'_{NH,i}$ are, respectively, the experimental RDC values in medium 1 and 2, and $D^c_{NH,i}$ and $D'^c_{NH,i}$ are the corresponding back-computed RDC values using the alignment tensors computed from the stage I and the three Euler angles (α, β, γ). An optimal first NC$_\alpha$ vector of the fragment is computed

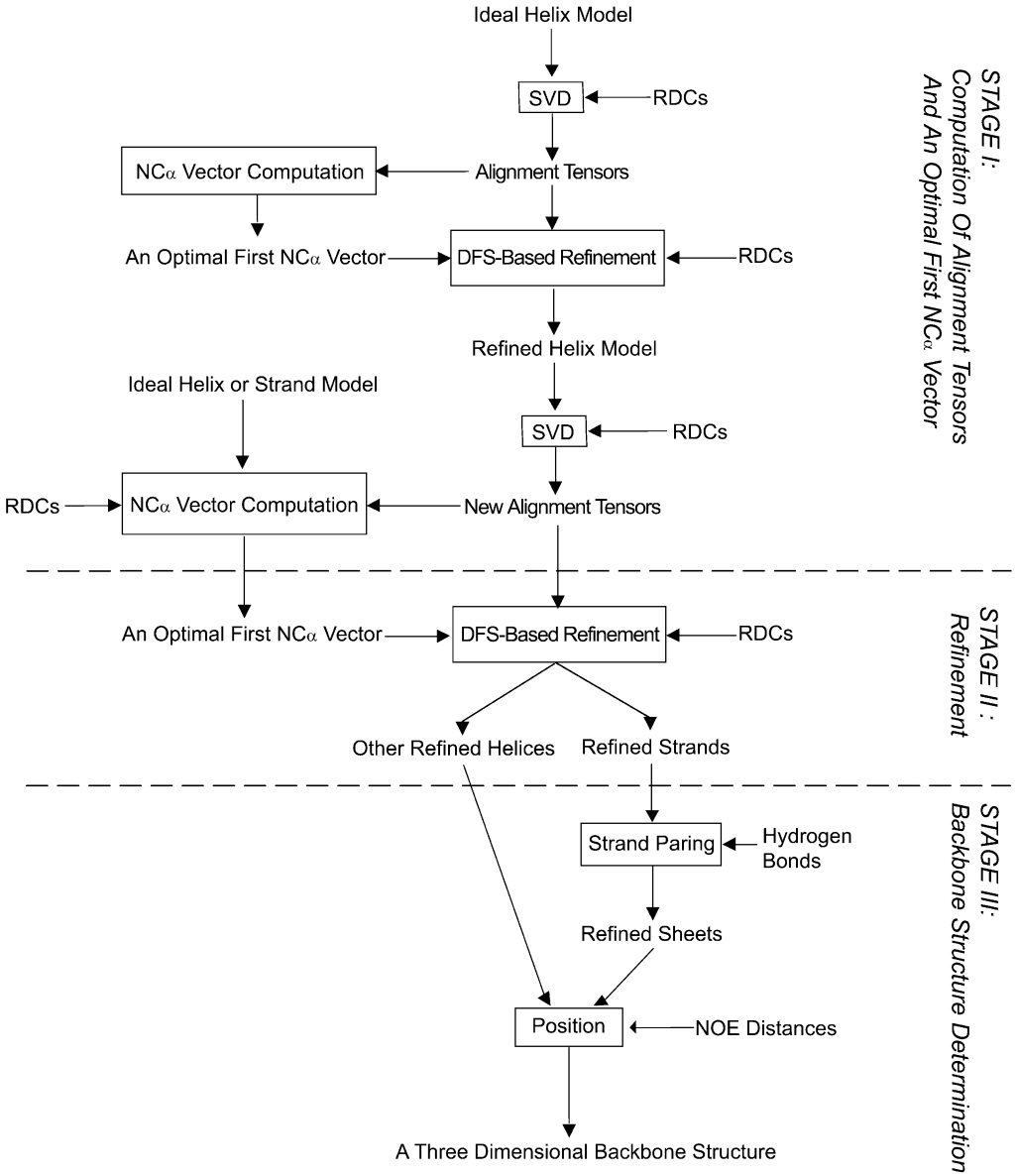

Figure 2. A flow chart illustrating the algorithm. A chart illustrating the three major stages (delimited by two dashed lines) of the algorithm beginning with the input RDC data and ideal helix and strand models to a final three dimensional backbone structure. RDCs refer to NH RDCs measured in two media. As stated in the main text, to improve accuracy, alignment tensors are computed twice in Stage I. DFS refers to depth-first search and SVD refers to singular value decomposition.

as $\mathbf{R}(\alpha, \beta, \gamma)\mathbf{v}_n$ where the rotation matrix $\mathbf{R}(\alpha, \beta, \gamma)$ is the one which minimizes σ_{RMS} and $\mathbf{v}_n$ is the first NC_α vector of the fragment before the minimization. The minimum value σ_{MIN} of σ_{RMS} is the value used in Equation (10) of the *refinement of secondary structure elements* section. The refined fragment will have the correct relative orientation with respect to the structure element chosen to compute the alignment tensors.

Refinement of secondary structure elements

It is well known that given the orientation of the first peptide plane of an m residue secondary structure element (fragment) its conformation can be specified uniquely by a sequence of (ϕ, ψ) angles: $(\phi_1, \psi_1, \phi_2, \psi_2, \cdots, \phi_{m-1}, \psi_{m-1})$, where (ϕ_i, ψ_i) are the backbone dihedral angles for residue i. We call such a sequence a *conformation vector*. The DFS-

based refinement (stage II) is a minimization searching systematically for an NH vector of the first peptide plane (since an optimal first NC_α vector has been determined) and a conformation vector such that the model built from the first peptide plane and the conformation vector has (a) (ϕ, ψ) values as close as possible to the average (ϕ_a, ψ_a) over the PDB for the corresponding secondary structure type and (b) simultaneously fits the experimental RDC data as well as possible. We call such a conformation vector an optimal conformation vector. What we mean by *refinement* here is to optimize both the direction of individual NH vectors and also the (ϕ, ψ) angles of a fragment using only RDCs while leaving the bond lengths and the six angles (Table 1 and Figure 4) fixed. Formally, our algorithm minimizes a score function T_1:

$$\begin{aligned} T_1 \;=\; & \sum_{i=1}^{m-1} \left((\phi_i - \phi_a)^2 + (\psi_i - \psi_a)^2 \right) \\ & + \sum_{i=1}^{m} \left(\left(D^s_{NH,i} - D_{NH,i} \right)^2 \right. \\ & \left. + \left(D'^s_{NH,i} - D'_{NH,i} \right)^2 \right), \end{aligned} \tag{7}$$

for a helix or a β-strand of a sheet chosen to be built first. For the other β-strands of the sheet it minimizes a score function T_2:

$$\begin{aligned} T_2 \;=\; & \sum_{i=1}^{m-1} \left((\phi_i - \phi_a)^2 + (\psi_i - \psi_a)^2 \right) \\ & + \sum_{i=1}^{m} \left(\left(D^s_{NH,i} - D_{NH,i} \right)^2 \right. \\ & \left. + \left(D'^s_{NH,i} - D'_{NH,i} \right)^2 \right) + T_H, \end{aligned} \tag{8}$$

where

$$\begin{aligned} & T_H(\phi_1, \psi_1, \cdots, \phi_{m-1}, \psi_{m-1}) = \\ & \sum_{j=1}^{q} \left(\left(H_{L,j} - H_{L,a} \right)^2 + \left(H_{A,j} - H_{A,a} \right)^2 \right), \end{aligned} \tag{9}$$

and where $D_{NH,i}$ and $D'_{NH,i}$ are, respectively, the experimental RDC values in medium 1 and 2, $D^s_{NH,i}$ and $D'^s_{NH,i}$ are the corresponding RDC values sampled from Gaussian distributions simulating the experimental errors (the experimental value and error are, respectively, the mean and variance), $H_{L,j}$ and $H_{A,j}$ are, respectively, the computed H-bond length and angle, and $H_{L,a}$ and $H_{A,a}$ are the corresponding values for an ideal H-bond, and q is the number of H-bonds between the paired strands. During the refinement ϕ_a, ψ_a, $D_{NH,i}$, $D'_{NH,i}$, $H_{L,a}$ and $H_{A,a}$ are treated as constants while $D^s_{NH,i}$, $D'^s_{NH,i}$ and the ϕ_i, ψ_i angles computed from them are treated as variables.

The DFS-based refinement (stage II) is divided into three phases (Figure 3). Phase I is the sampling of RDC values in both media for every residue of the fragment such that Equation (5) will produce a solution for every NH vector and such that

$$\begin{aligned} m\sigma^2_{MIN} \;\geq\; & \sum_{i=1}^{m} \left(\left(D^s_{NH,i} - D_{NH,i} \right)^2 \right. \\ & \left. + \left(D'^s_{NH,i} - D_{NH,i} \right)^2 \right), \end{aligned} \tag{10}$$

where σ_{MIN} is the minimum RMSD between the experimental RDCs and the RDCs back-computed from an ideal helix by the SVD method as detailed in the *computation of alignment tensors* section or from a β-strand as detailed in the previous section. Phase II is the computation of an NH vector for the first residue *and* an optimal conformation vector from the two sets of sampled RDCs together with an optimal first NC_α vector. Phase III is the construction of a model from the NH and NC_α vectors defining the first peptide plane, and the optimal conformation vector. The first two phases are repeated for the number of times equal to the sampling size (see Appendix C for a pseudocode of the algorithm). In phase II, the set of all the plausible conformation vectors are first computed by a depth-first search (DFS) over all the possible combinations of NH orientations from Equation 5. A plausible conformation vector is defined as a vector with all its $m - 1$ (ϕ, ψ) angles in the favorable Ramachandran region for the corresponding secondary structure type. The NC_α vector of residue $i + 1$ is computed from an NC_α vector of residue i and the intervening (ϕ_i, ψ_i) angles. During the DFS every computed (ϕ, ψ) pair is filtered through favorable Ramachandran regions. An optimal conformation vector and its corresponding first NH vector are computed from the set of all the plausible conformation vectors by comparing their scores computed according to either Equation 7 or 8. Except for the β-strand chosen to be refined first the hydrogen bond energy, T_H in Equation 8, between paired strands is also computed as a part of the score as detailed in the next section after a new β-strand has been built using the optimal conformation vector and the first peptide plane defined by the first NH and NC_α vectors.

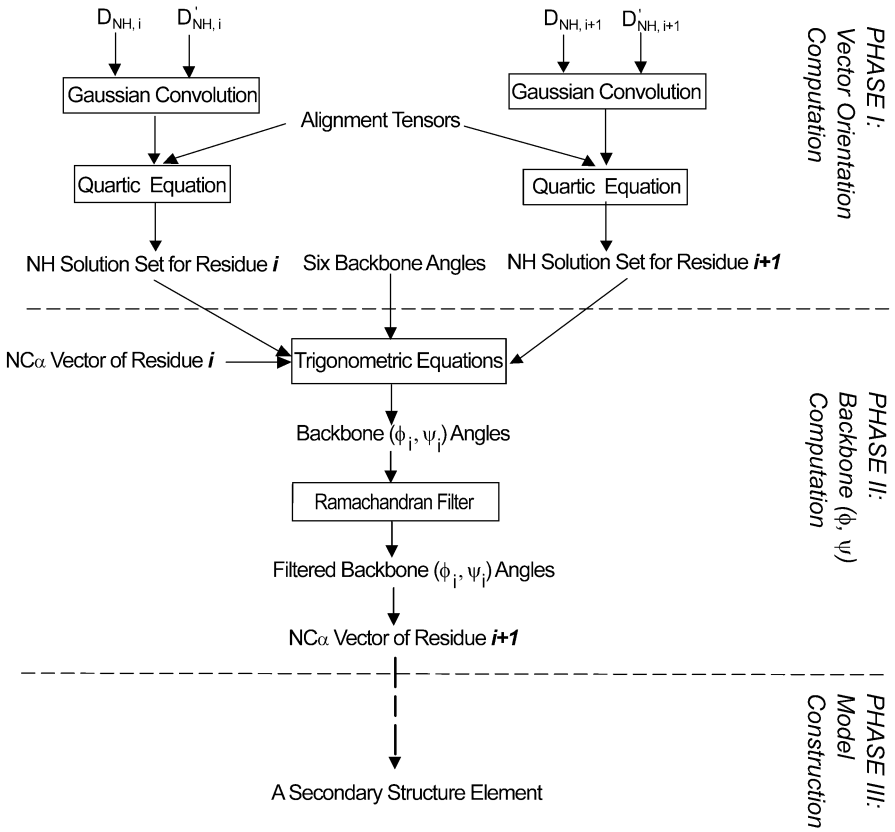

Figure 3. A flow chart of a single step of the DFS-based refinement stage. The three major phases are delimited by two dashed lines. The six angles (Figure 1 and Table 4) are defined in the main text. The quartic and trigonometric equations refer, respectively, to Equations 4, 5. The Ramachandran filter represents filtering through the favorable Ramachandran regions and the Gaussian convolution corresponds to the simulation of the experimental error by a Gaussian distribution with the experimental value ($D_{NH,i}$ or $D'_{NH,i}$) as its mean and the error as its variance. Many such steps are required for refining a secondary structure element as shown by a dashed arrow.

In our implementation the average (ϕ_a, ψ_a) for α-helices and β-sheets are set to be $(-65.3°, -39.4°)$ and $(-120.0°, 138.0°)$, their respective averages over the PDB. The values of $H_{L,a}$ and $H_{A,a}$ for an ideal H-bond are set to be, respectively, 2.90 Å (between the backbone atoms O and N) and 0.0° (the donor, hydrogen and acceptor atoms are collinear) (Stryer, 1994, p. 8), their values for an ideal hydrogen bond. The favorable Ramachandran region for an α-helix is set to be $[-100.0°, -30.0°]$ for φ and $[-90.0°, -15.0°]$ for ψ. For a β-strand the φ range is set to be $[-170.0°, -70.0°]$ and the ψ to be $[80.0°, 180.0°]$.

Pairing of β-strands and computation of H-bond energy

We begin the construction of a β-sheet by pairing a strand A chosen to be refined first with a strand B which is known to be A's partner by secondary structure identification. The strand B is built during the refinement using the optimal conformation vector and the first NH vector. First, the average position of the hydrogen bond partners of A is computed assuming that the hydrogen bonds adopt an ideal configuration. Next, the average position of the atoms of B involved in hydrogen bonding with A is determined. Next, the two average positions are superimposed by translation. Finally, the hydrogen bond distances and angles are computed, and their RMSDs from the ideal values are computed from the overlayed structures according to Equation 9. An entire sheet can be built by successively adding its strands.

Backbone structure determination

The algorithm for positioning the oriented secondary structure elements is similar to that described by Prestegard and coworkers (Fowler et al., 2000). Four

Table 2. Diagonalized Saupe elements. The Ideal Helix and Refined Helix refer to the helix built, respectively, with $(\phi, \psi) = (-65.3°, -39.4°)$ and the optimal conformation vector as detailed in the text. In (a) the comparison is between the ideal and RDC-refined helices and the whole X-ray structure of human ubiquitin while in (b) the comparison is between the three helices: the ideal, RDC-refined and the helix consisting of the ten residues (N25–E34) of the X-ray structure (Vijay-Kumar et al., 1987). RDC_1 and RDC_2 are RDCs of medium 1 and 2, respectively. Protons in X-ray structure were added with the program MOLMOL (Koradi et al. 1996)

		Ideal helix	Refined helix	X-ray structure
Saupe elements (a)	S^1_{xx}	3.23	3.47	4.24
	S^1_{yy}	28.34	26.65	27.42
	S^2_{xx}	8.88	8.55	6.91
	S^2_{yy}	10.06	10.38	11.43
RMSD (b)	Backbone atoms	0.25 Å	0.29 Å	0.0 Å
	(ϕ, ψ) Angles	(10°, 9°)	(5°, 6°)	(0°, 0°)
	RDC_1	0.39 Hz	0.12 Hz	1.43 Hz
	RDC_2	0.46 Hz	0.14 Hz	0.95 Hz

NOE-derived distances (Table 1) between an amide proton and a C_α nucleus or between two C_α nuclei, converted from the NOE restraints between amide or alpha proton and methyl group (Cornilescu et al., 1998), are employed to compute the relative position of the helix (N25–E34) and the single sheet of ubiquitin, which consists of five β-strands: Q2–T7, K11–V17, Q41–A46, K48–E51 and S65–V70. First, the algorithm computes the average d_a of the four NOE distances and the average position of the four nuclei of the helix involved in the NOE interactions. In general, the NOE distances may be different from one another so it is necessary to compute their average. For convenience the vector from the average position of the four nuclei of the helix to the average position of the four nuclei of the sheet is defined as **w**. It has length d_a. Next, a grid search is performed over the orientation of **w** to compute the average position of the four nuclei of the sheet (the position of the helix is fixed here) to minimize the $RMS_{NOE} = \sum_{i=1}^{4} (d_i'^{NOE} - d_i^{NOE})^2$ between the computed distances $d_i'^{NOE}$ and the experimental NOE distance d_i^{NOE}. This procedure is also used to resolve the ambiguities in the relative orientations of the helix and the sheet. Specifically, the relative orientation with the smallest NOE violations (the smallest RMS_{NOE} value) among the four possible orientations is selected. The correct orientation between two paired β-strands can be chosen using the hydrogen bond network between them.

Source of the data

The backbone NH RDC data in two different media (medium 1 and 2 correspond, respectively, to those measured in charged and uncharged bicelles), twelve hydrogen bonds and four NOE restraints of the protein human ubiquitin were extracted from the restraint file with PDB ID 1D3Z (Cornilescu et al., 1998; Ottiger and Bax, 1998) downloaded from the PDB (Berman et al., 2000).

Results and discussion

In this section we first discuss the possible benefits of exact solutions. Then, we demonstrate the successful determination of a backbone substructure (consisting of α-helices and β-sheets) of human ubiquitin with our exact solution-based algorithm using NH RDCs in two media and twelve H-bonds and four NOE restraints.

Equations for computing dihedral angles from RDCs

Our algorithm requires RDCs on a single backbone vector type in two media, which is quite feasible considering that several media are available at present (Al-Hashimi et al., 2000; Chou et al., 2001; Hansen et al., 2000; Ramirez and Bax, 1998). A second set of RDCs has been used, for example, to reduce degeneracy (Al-Hashimi et al., 2000; Ramirez and Bax, 1998) or to improve the precision of NMR determined structures (Clore et al., 1999). However, to our knowledge, there are no known analytic expressions for computing

either internuclear vectors or backbone (ϕ, ψ) angles directly from RDCs when the POFs of medium 1 and 2 are not identical. Previously, numerical fitting has been used to compute NH vector orientations (Barbieri et al., 2002) and 2-dimensional grid-search has been used to determine (ϕ, ψ) angles (Giesen et al., 2003; Tian et al., 2001; Wang et al., 1998) when RDCs are measured on at least three backbone vector types. Scheraga and coworkers (Wedemeyer et al., 2002) have derived a quartic equation for computing vector orientations *when the two POFs are identical*. For the general case when the two POFs are different they suggest using 1-dimensional grid search to compute vector orientations. Griesinger and coworkers (Meiler et al., 2000) have derived an expression for computing pairwise angular restraints between internuclear vectors whose RDCs are measured in *one* medium. In contrast, our equations (Equations 4, 5) are derived to compute, *exactly*, the intervening backbone dihedral angles between two consecutive residues whose NH RDCs have been measured in *two* media. They can not be used to compute the inter-vector angle between two NH vectors. The advantages of exact solution methods have been demonstrated by Cross and coworkers (Bertram et al., 2000) for computing solid-state NMR structures and by Scheraga and coworkers (Wedemeyer et al., 2002) for computing solution NMR structures. The possible benefits of exact methods include:

1. Exact solutions make it possible to characterize the properties of the solutions. For example, the number of solutions from Equation 5 can be 0, 2, 4, 6 or 8, in agreement with the previous results obtained by other methods (Al-Hashimi et al., 2000; Ramirez and Bax, 1998; Wedemeyer et al., 2002) (Figure 4A). Further, the average number of real solutions (as opposed to complex) is close to 4, in agreement with what has been known in algebraic geometry (Kac, 1948). Given the peptide plane i the two possible (ϕ_i, ψ_i) solutions from Equation 5 have very similar ϕ_i values (differing by $<10°$) but with opposite sign if the coefficients of Equation 5 are computed using the averages for the six angles (Table 4). In our implementation these average values were obtained from 23 ultra-high resolution X-ray structures with proton coordinates (Table 4). Consequently, except for glycine, all residues located in regular α-helices and β-sheets have only one (ϕ, ψ) solution in the favorable Ramachandran regions. And the solution with positive ϕ value can be safely discarded for a residue in regular α-helix or β-strand. Further, for most residues in regular α-helices and β-sheets, only a few of the $8 \times 8 \times 2$ possible (ϕ, ψ) solutions will fall into the favorable regions (Figure 5A).
2. With exact solutions it is possible to quantify the contributions to the accuracy of the computed (ϕ, ψ) angles from both the experimental errors in RDCs (Figure 5B) and the statistical distributions of the six angles (Table 4) required as additional restraints in Equation 5. Among the six angles, the dihedral angle θ_6 (Table 4 and Figure 1), which measures the deviation of an NH vector from the peptide plane, has the largest variance (6.0°) while the other five have much smaller variances ($<2.4°$) (Table 4). Our results from exact solutions show that the variation of this angle does not change the ϕ_i value and only changes ψ_i by $<10°$.
3. Exact solutions are expected to be useful for speeding up the structure determination of both proteins and nucleic acids using RDC and/or pseudocontact shift restraints (Kemple et al., 1988) since similar equations can be derived for computing the corresponding vector orientation and dihedral angle. As a demonstration we have shown (see the *Conformation and orientation of secondary structure elements* section) that such exact solutions make it feasible and efficient to search *systematically* through all the possible combinations of (ϕ, ψ) solutions for a secondary structure element to determine a conformation that best satisfies both the experimental RDCs and has backbone (ϕ, ψ) values as close as possible to the PDB averages (Equation 7 or 8).

Computation of alignment tensors

For alignment tensor computation we take the advantage of a priori knowledge about secondary structure elements. Specifically, we select a helix model consisting of residues N25–E34 to be built and refined first for their computation. In general, helices can be identified easily, have less variations in local structure than β-sheets do, are more stable, and their amide protons are less labile than those in loops so the experimental data have smaller errors (Wang et al., 2001). Some concerns (Fowler et al., 2000) have been raised about the accuracy of the computed Saupe matrix because the NH bond vectors are near parallel in a regular helix. However, our results show that the variations in NH orientations in the ideal helix built with (ϕ, ψ) = ($-65.3°$, $-39.4°$) are large enough to

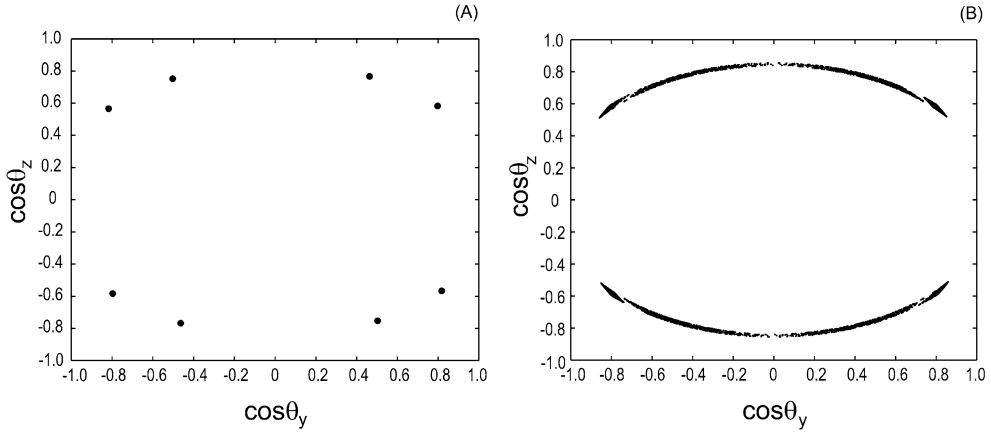

Figure 4. Solutions for NH unit vector direction. The 8 solutions for the NH unit vector of the residue I3 of ubiquitin without (A) and with (B) experimental errors (±1.0 Hz) in RDCs. The x-axis and y-axis are, respectively, the y and z components of the three direction cosines. These solutions are computed assuming a fixed pair of alignment tensors.

Table 3. Backbone RMSDs of fragments. The backbone RMSDs between an RDC-refined fragment consisting of the helix (N25–E34) and a β-strand, and the corresponding fragment from the X-ray structure of ubiquitin (Vijay-Kumar et al., 1987)

β-strand	Q2–T7	K11-V17	Q41–A46	K48–E51	S65–V70
Backbone RMSD	1.25 Å	0.87 Å	0.82 Å	0.52 Å	1.06 Å

yield a reasonably accurate initial tensor estimation. In fact, the diagonal Saupe elements estimated using the ideal helix are quite close to those computed from the entire structure (Table 2). The final set of alignment tensors are computed from an RDC-refined helix model. However, the diagonalized Saupe elements computed from the refined model differ by less than 7 percent from those computed from the ideal helix (Table 2). No further refinement is necessary since the RMSD between the RDCs back-computed from the refined model and the experimental RDCs is <0.14 Hz, very close to the experimental error (Cornilescu et al., 1998; Ottiger and Bax, 1998), and backbone RMSDs between any two of the three helices are rather small: <0.29 Å. Note that the RMSD between the RDCs back-computed from the refined helix and the experimental RDCs is much smaller than the RMSD between the RDCs back-computed from the helix of the X-ray structure (Vijay-Kumar et al., 1987) and the experimental RDCs (Table 2). Our results show that it is possible, first, to construct a helix model which can fit experimental RDCs very well and, then, to compute reasonably accurate alignment tensors from the model. However, our method for computing alignment tensors may not always work well when the helix axis is parallel to the alignment tensor. If this is the case one may use other methods available at present (Clore et al., 1998; Nomura and Kainosho, 2002) for the initial estimation of alignment tensors.

Conformation and orientation of secondary structure elements

It is well known that at least 3 RDCs per peptide plane are required to determine the plane's orientation and subsequently a backbone fold (Quine et al., 1997; Wang et al., 1998). However, our analyses (see the *Theoretical background* section) show that given NH RDCs in two media there will be only a finite number of (ϕ, ψ) solutions and backbone conformations if (a) the orientation of the first NC_α vector is known, (b) the RDCs are perfectly accurate (without experimental errors), (c) the alignment tensors are known, and (d) there are no missing RDCs for the entire protein sequence. Hence, in principle, a backbone structure can be computed by a grid-search for the first NC_α vector followed by a systematic search to select a best conformation out of the finite number of all the possible conformations using a simple score function. However, such a score function must include terms other than the term for RDCs since there are a large (but finite) number of conformations that are consistent with

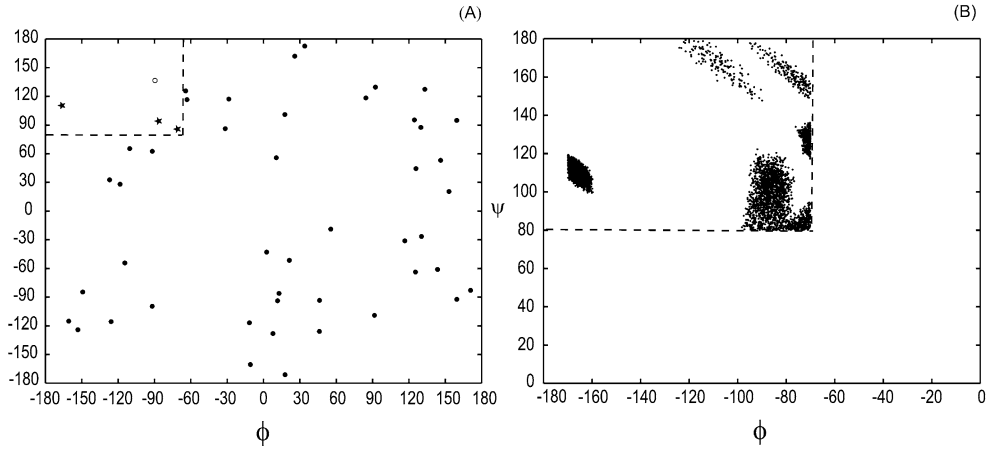

Figure 5. Backbone (ϕ, ψ) solutions. The backbone (ϕ, ψ) solutions between two residues (Q2–I3) of ubiquitin without (A) and with (B) experimental errors ($\pm$1.0 Hz) in RDCs. The x-axis and y-axis are the ϕ and ψ, respectively. The favorable Ramachandran region [$-170.0°$, $-70.0°$] for ϕ and [80.0°, 180.0°] for ψ of β-sheet are delimited by dashed lines. In (B) only the (ϕ, ψ) angles in the favorable region are shown, the ϕ range in the x-axis is from [-180°, 0°] and the ψ range in the y-axis from [0°, 180°]. In (**A**) the three stars indicate (ϕ, ψ) angles in the favorable region and the empty circle represents the (ϕ, ψ) values in the X-ray structure. The 48 solutions (points) come from $6 \times 4 \times 2$, i.e., 6 solutions for the unit NH vector of Q2 and 4 solutions for I3. These solutions are computed assuming a fixed pair of alignment tensors.

2 RDCs per residue. Furthermore, in practice, the experimental RDC value has error (Figures 4B and 5B) and some residues may have only one observable RDC and some may have none. To deal with the requirement for additional terms in the score function, the deficiency in the number of RDC restraints, and also with the experimental errors, we have included a term $((\phi_i - \phi_a)^2 + (\psi_i - \psi_a)^2)$ for backbone (ϕ, ψ) angles (Equations 7, 8) and, for β-sheets only, an additional term T_h for H-bonds (Equation 8). The regularity of α-helices and β-sheets is well-known, and it is relatively straightforward to identify them by NMR. The variances of the backbone (ϕ, ψ) angles are, respectively, about 11° for a helix and less than 25° for a β-sheet. One concern about these terms is that the individual (ϕ, ψ) angles of the computed model for a fragment may be artificially forced to assume the average (ϕ, ψ) values for α-helices or β-sheets. However, the salient feature of our algorithm is that the solution model is computed by searching for an optimal conformation with respect to *all* of its RDCs rather than any individual RDC using score functions (Equations 7, 8) having terms for RDCs as well as (ϕ, ψ) angles and H-bonds. Therefore, an individual dihedral angle of a refined secondary structure element computed by our algorithm may differ from the average value by as much as 29° (Figure 6) and differ from the corresponding value in x-ray structure by more than $>10°$. The same is true for H-bonds: in the final refined β-sheet both the H-bond distances and directions deviate from their ideal values. Nevertheless, the relative orientations between the helix (N25–E34) and each of the five strands of the single sheet, as measured by backbone RMSDs, agree well with those in the X-ray structure (Table 3 and Figure 7). We have not yet tested our algorithm with a long kinked helix though, in principle, our score function should still work. Furthermore, the kink could be identified by comparing the alignment tensors computed from different sets of five consecutive residues (Wang et al., 2001) along the sequence, if dynamics can be ignored. We can then, in principle, divide the kinked helix into independent pieces and refine them independently.

Our algorithm can not only refine α-helices but also β-sheets (both parallel and anti-parallel), which extends fundamentally the previous method (Fowler et al., 2000) targeting only entirely helical proteins and makes our algorithm suitable for determining the folds of a large majority of proteins. Unlike α-helices, β-strands are often twisted in globular proteins so it is important to refine them accurately (using stage II of our algorithm) from RDC data. For example, the backbone RMSD between the RDC-derived fragment consisting of the β-strand (Q2–T7) and helix (N25–E34), and the corresponding fragment in the X-ray structure, changed from 0.87 Å to 0.75 Å. However, the RMSDs between the RDCs back-computed from the β-strand (Q2–T7) and the experimental RDCs dropped from 2.86 Hz to 0.68 Hz for medium 1 and from 3.18 Hz to 1.00 Hz for medium 2.

Table 4. Six backbone angles. The means and variances of the *six angles* between two consecutive residues used to compute the coefficients of Equation 5. They are extracted from 23 ultra-high resolution (≤1.0Å) X-ray structures with proton coordinates. Their PDB IDs are 3AL1, 1BXO, 1CEX, 1C75, 1GDQ, 1G66, 1GDN, 1GCI, 2FDN, 1HJ9, 1IXH, 1GQV, 1IC6, 2ERL, 1HJ8, 1JFB, 1EJG, 1RB9, 3PYP, 1FY5, 1GVK, 1KQP and 1LQT

Name	Identity	Mean	Variance
θ_1	$H^N(i) - N(i) - C_\alpha(i)$	119.14°	1.20°
θ_3	$N(i) - C_\alpha(i) - CO(i)$	110.94°	2.39°
θ_5	$C_\alpha(i) - CO(i) - N(i+1)$	116.82°	1.88°
θ_6	$C_\alpha(i) - CO(i) - N(i+1) - H^N(i+1)$	−0.75°	6.07°
θ_7	$CO(i) - N(i+1) - H^N(i+1)$	119.10°	1.10°
θ_8	$CO(i-1) - N(i) - C_\alpha(i) - H^N(i)$	0.00°	0.10°

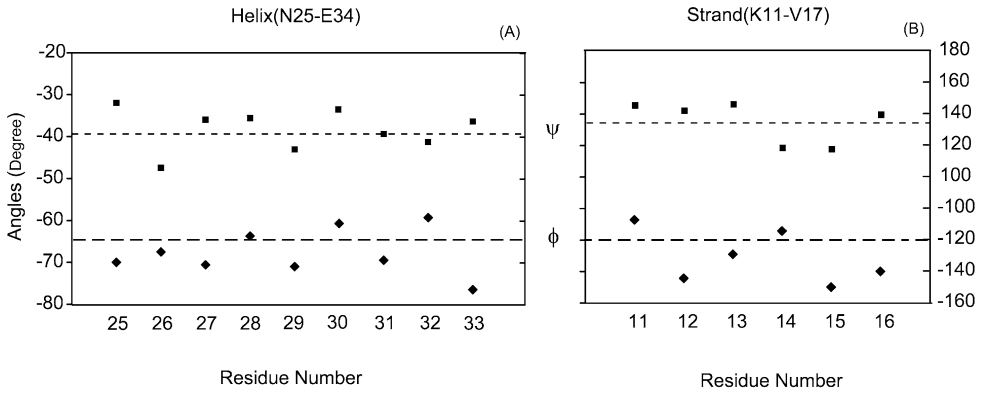

Figure 6. Distributions of backbone (ϕ, ψ) angles. The distributions of backbone (ϕ, ψ) angles of an α-helix (N25–E34) (A) and a β-strand (K11–V17) (B) both computed by our algorithm. The two dashed lines in (A) and (B) correspond, respectively, to (−63.4°, −39.4°) and (−120.0°, 138.0°), the (ϕ, ψ) averages of an α-helix and a β-strand over the PDB.

There exist four possible relative orientations between any two fragments when only one set of RDCs is used. It has been demonstrated that a second set of RDC data from a different medium can be employed to resolve the orientational degeneracy between two fragments with known structures (Al-Hashimi et al., 2000). However, we found that when the experimental errors in RDCs are taken into account it is possible to build four structures with different relative orientations between the helix (N25–E34) and the single sheet in ubiquitin. The correct orientation can be computed using four NOEs between the helix and the sheet (Table 1).

Computation of β*-sheets*

Additional translational restraints are required to pair β-strands to form a sheet since RDC data provides only orientational restraints. In our implementation, the hydrogen bonds between β-strands were used. Hydrogen bonds can be derived, for example, from backbone resonance and NOE assignment programs such as Jigsaw (Bailey-Kellogg et al., 2000) or inferred from the identification of a β-sheet structure. Among the five strands of the single β-sheet the strand consisting of residues Q2–T7 is chosen to be computed first. The other four strands are computed in the order: Q2–T7 to K11–V17, Q2–T7 to S65–V70, S65–V70 to Q41–A46 and Q41–A46 to K48–E51 using, respectively, 4, 3, 3 and 2 hydrogen bonds. The order is not important since the orientation of each strand is determined only by its RDC values, not by hydrogen bonds between pairing strands. No significant changes in orientation were observed when each strand was chosen to be refined separately without using hydrogen bonds between them. The computed sheet superimposes very well with the corresponding one in the X-ray structure with a backbone RMSD = 1.17 Å (Figure 7).

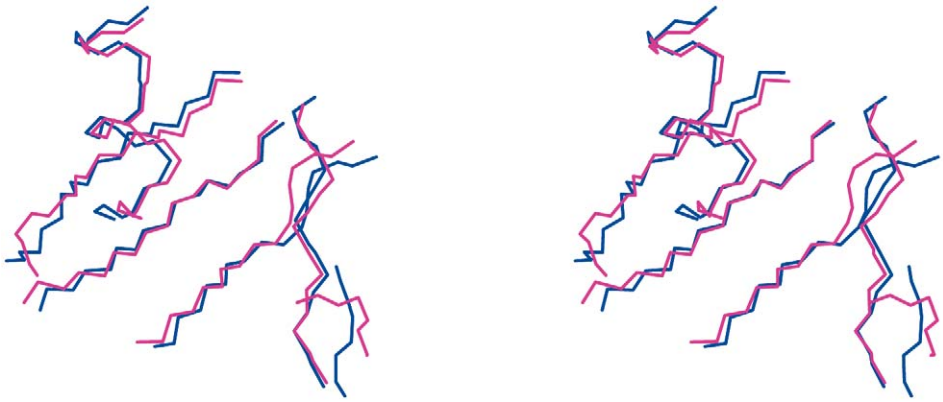

Figure 7. Stereo view of backbone superposition. The portion of RDC-derived structure (blue) consisting of the helix (N25–E34) and the single sheet can be superimposed with the corresponding portion (magenta) from the X-ray structure with a backbone RMSD = 1.21 Å. The single sheet is composed of five strands (Q2–T7, K11–V17, N41–F45, K48–E51 and S65–V70), which can be superimposed with the corresponding sheet from the X-ray structure (Vijay-Kumar et al. 1987) with a backbone RMSD of 1.17 Å.

Backbone structure determination

The last stage of our algorithm (stage III of Figure 2) is to determine a three dimensional backbone structure from oriented secondary structure elements. In principle, only three distance restraints are required to position two oriented elements. However, we found that no triple of the NOE-derived distances in Table 1 could correctly position the helix (N25–E34) relative to the single sheet in ubiquitin. Nevertheless, they could be correctly positioned using all four NOE distances. These NOE restraints between amide protons and methyl groups of regular secondary structures can be obtained, for example, from a labeling strategy introduced by Kay and coworkers (Gardner and Kay, 1997). The backbone RMSD between the RDC-derived structure consisting of the helix (N25–E34) and the single sheet, and the corresponding portion of the X-ray structure, is only 1.21 Å (Figure 7). Moreover, the computed structure satisfies the experimental RDC data very well. The RMSD between the RDCs back-computed from the RDC-derived structure and the experimental RDCs is only 0.64 Hz for medium 1 and 0.87 Hz for medium 2. The RDC-derived backbone substructure defines completely the fold of ubiquitin even though the structures of the intervening loops and turns have not been determined. The accuracy (1.2 Å) is the same as that (1.2 Å) reported for the fold of the protein GB1 (Clore et al., 1999), which shares a fold with ubiquitin, determined by XPLOR (Brünger, 1993) using *three* RDCs per residue. However, we emphasize that our results show that it is possible to build a substructure consisting of α-helices and β-sheets which can fit experimental RDCs very well *and* overlay with the X-ray structure (Vijay-Kumar et al., 1987) very well. We do not claim that our substructure is more accurate than an X-ray model with 1.8 Å resolution since it is rather tricky to compare the goodness of fit to experimental RDCs between the two substructures due to the uncertainty in the x-ray model. A reasonable estimation of a proton position from an X-ray model with 1.8 Å resolution is about 0.30 Å (Wang et al., 2001), which corresponds to a deviation of 8.6° of an NH vector from the position given in the x-ray model. Stage III of our algorithm (see Figure 2 and the *Algorithm* section) is, in principle, a rigid-body modeling method and thus is similar to several previous methods for fold prediction and recognition (Kolinski and Skolnick, 1998; Yue and Dill, 2000), to an NMR structure-determination procedure (Fowler et al., 2000) for assembling known secondary structure elements into a tertiary structure, and to a method used in the docking program HADDOCK (Dominguez et al., 2003). Our algorithm is slightly different in that (a) oriented secondary structure elements (both α-helices and β-sheets) are *refined* by RDCs before positioning, and (b) stage III can position not only oriented α-helices but also oriented β-sheets.

Extensions and limitations

The NMR data required by our algorithm can be recorded on an ^{15}N singly labeled sample, which may be significant for high-throughput fold determination. However, in practice, for a protein with more than 100 residues, it may be easier or necessary to use

^{15}N and ^{13}C double labeling to assign backbone resonances and to identify secondary structure elements. In fact, with double labeling it is rather straightforward to measure more than two RDCs per residue. Additional RDCs, if available, can be easily incorporated into our algorithm to increase the accuracy of the computed (ϕ, ψ) angles and to eliminate the requirement that the computed α-helix or β-strand have backbone (ϕ, ψ) values as close as possible to the PDB averages. The focus of the current paper is to present the derivation of low-degree polynomials used for computing, exactly, the internuclear vector and backbone (ϕ, ψ) angles from two RDCs per residue and to demonstrate the possibility of computing a backbone substructure from those RDCs using an algorithm built upon *exact solutions* and *systematic search.* In principle, these low-degree polynomials can be easily incorporated into a structure determination algorithm using restrained molecular dynamics (MD) in torsion angle space such as DYANA (Güntert et al., 1997). In fact, a module with NH orientation obtained from RDCs by numerical fitting has been implemented recently in DYANA by Luchinat and coworkers (Barbieri et al., 2002). Compared to the restrained MD with simulated annealing (SA) approaches (Brünger, 1993; Güntert et al., 1997; Clore et al., 1999; Giesen et al., 2003; Hus et al., 2001), the novelty of our algorithm lies in how the input data (RDCs) are used to limit the search space and how a global minimum is computed:

1. In our algorithm, the space of solutions is, a priori, *explicitly* restricted by the data: first, the space of NH internuclear bond vectors is restricted to a set of finite solutions. This, in turn, kinematically restricts the space of backbone dihedral angles to a finite set before a conformation is computed. In contrast, in restrained MD/SA, the space of solutions is *implicitly* restricted by an energy function that penalizes those *computed* conformations for which the back-calculated and experimental RDCs disagree.
2. In a restrained MD/SA approach, MD is used as a minimization tool to solve a multiple variable minimization problem. Since there are many local minima a heuristic search such as SA is, in general, used to search for the true global minimum. However, since SA is a non-deterministic, iterative approach that samples the search space stochastically, the computed minimum is most likely to be only a local minimum. In other words, a structure computed by such an algorithm is not guaranteed to be the true global minimum even with perfect data. In contrast, our algorithm is deterministic, non-iterative and combinatorially precise, thus, if the input data (experimental RDCs) are perfectly accurate (without any experimental error) our systematic search method is guaranteed to find the true global minimum. The advantages of systematic search vs. heuristic search have been previously described (see, for example, Rienstra et al. 2002).

The computational and modeling benefits of exact solutions have been described by Cross and coworkers (Bertram et al., 2000) for computing solid-state NMR structures and by Scheraga and co-workers (Wedemeyer et al., 2002) for computing the backbone structure of ubiquitin in solution using five RDCs in one medium. However, in the former a Monte-Carlo method is used to search over possible conformational space and in the latter final structures are refined by ROSETTA (Rohl and Baker, 2002) which uses, first, a fragment replacement method (Delaglio et al., 2000) to reduce the search space, then, a Monte-Carlo Method to search over the reduced conformation space. Systematic search has proved useful in solid-state NMR structure determination (Rienstra et al., 2002) for a tripeptide. However, Griffin and coworkers did not derive or employ exact solutions. Our algorithm is the first NMR structure determination algorithm that uses both exact solutions and systematic search.

The running time of our algorithm (see Appendix D for an analysis of the algorithmic complexity and performance), about 45 min for computing a 39-residue substructure of ubiquitin, is comparable to the time needed by a full-blown structure determination algorithm such as XPLOR (Brünger, 1993) using many NOEs and dihedral angle restraints. As is well known, a systematic search over the entire solution space needs more time than a heuristic search which samples the space stochastically. However, with exact solutions and a pruning strategy based on the Ramachandran plot our algorithm managed to compute the substructure of ubiquitin in a comparable time. Furthermore, months of time could be saved by avoiding the assignments of large number of NOEs involving sidechain protons in order to compute an accurate structure using a full-blown, traditional method. In addition, hours of spectrometer time may be saved since fewer NMR experiments are needed and single labeling is less expensive than double labeling, in general.

Conclusions

We have described a novel algorithm for determining a protein backbone structure by solution NMR spectroscopy, using almost exclusively angular restraints from RDCs measured on a single bond vector type (NH) in two media, plus very sparse distance restraints. The algorithm is built upon our newly-derived equations for computing (ϕ, ψ) angles, *exactly* and *in constant time*, from two RDCs per residue. The proposed exact solution methodology is rather general and can be applied to speed up the structure determination of both proteins and nucleic acids from RDCs in two media since similar equations can be easily derived to compute either the backbone and sidechain dihedral angles in proteins, or the backbone torsion and χ angles in nucleic acids.

We have also shown that the exact solutions make it feasible to design a novel DFS-based minimization algorithm to compute efficiently both the orientations and conformations of not only α-helices but also β-strands using only RDCs in two media. Further, we demonstrated that, after the orientations of α-helices and β-strands are calculated from NH RDCs, a β-sheet can be computed from its constituent strands using hydrogen bonds, and that the three-dimensional backbone substructure consisting of both α-helices and β-strands can be determined by adding a minimum number of NOE distances. Our success in computing such a substructure using only two RDCs per residue and very sparse distance restraints shows that solution NMR spectroscopy can play a major role in determining protein structures rapidly and inexpensively, which should be important in structural genomics.

Supporting Information

The software is available by contacting the authors, and is distributed under the Gnu Public License (Gnu, 2002).

Notes

1. A constant-time, or $O(1)$ algorithm (Cormen et al., 2001) for computing the possible (ϕ, ψ) solutions or internuclear bond vectors for a single residue always takes a fixed (i.e., constant) number of steps, and this (constant) number of steps depends neither on a grid nor upon its resolution, nor upon the size n of the protein.
2. A *depth first search* (DFS) visits all the vertices in a tree (Cormen et al., 2001). When choosing which edge to explore next, this algorithm always chooses to go 'deeper' into the tree. That is, it will pick the next adjacent unvisited vertex until reaching a vertex that has no unvisited adjacent vertices. The algorithm will then backtrack to the previous vertex and continue along any as-yet unexplored edges from that vertex. After DFS has visited all the reachable vertices from a particular source vertex, it chooses one of the remaining undiscovered vertices and continues the search.

Acknowledgements

We would like to thank Dr Amy Anderson, Chris Langmead, Ryan Lilien, Dr Ramgopal Mettu, Elisheva Werner-Reiss, Tony Yan and all the members of Donald lab for helpful discussions and critical reading of the manuscript.

Appendix A: A quartic equation for computing an internuclear vector

In this section we prove that the direction of an internuclear vector $\mathbf{v} = (x, y, z)$ can be computed from the solutions to a quartic equation if its corresponding RDCs are measured in two different aligning media.
From the standard RDC equation (Saupe, 1968) we have

$$\begin{aligned} D_{NH} &= S_{xx}x^2 + S_{yy}y^2 + S_{zz}z^2 \\ D'_{NH} &= S'_{xx}x'^2 + S'_{yy}y'^2 + S'_{zz}z'^2 \\ \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} &= \mathbf{R_{12}} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \\ &= \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \end{aligned}$$

The symbols have the same meaning as in Equations 1, 2 of the section *Computation of vector orientations* of the main text. Here, for clarity the dipolar interaction constant D_{max} is assumed to be 1.
Eliminating x', y' and z' through squaring and substitution we obtain

$$r_2 = a_2x^2 + b_2y^2 + c_1xy + c_2xz + c_3yz \quad \text{(A1)}$$

$$r_1 = a_1x^2 + b_1y^2, \quad \text{(A2)}$$

where

$$a_2 = (S'_{xx} - S'_{zz})(R_{11}^2 - R_{13}^2)$$

$$
\begin{aligned}
&+ (S'_{yy} - S'_{zz})(R^2_{21} - R^2_{23}) \\
b_2 &= (S'_{xx} - S'_{zz})(R^2_{12} - R^2_{13}) \\
&+ (S'_{yy} - S'_{zz})(R^2_{22} - R^2_{23}) \\
c_1 &= 2(S'_{xx} - S'_{zz})R_{11}R_{12} + 2(S'_{yy} - S'_{zz})R_{21}R_{22} \\
c_2 &= 2(S'_{xx} - S'_{zz})R_{11}R_{13} + 2(S'_{yy} - S'_{zz})R_{21}R_{23} \\
c_3 &= 2(S'_{xx} - S'_{zz})R_{12}R_{13} + 2(S'_{yy} - S'_{zz})R_{22}R_{23} \\
r_2 &= D'_{NH} - S'_{zz} - (S'_{xx} - S'_{zz})R^2_{13} \\
&- (S'_{yy} - S'_{zz})R^2_{23} \\
a_1 &= S_{xx} - S_{zz} \\
b_1 &= S_{yy} - S_{zz} \\
r_1 &= D_{NH} - S_{zz}.
\end{aligned}
$$

Eliminating z from Equation A1 through squaring and substitution we obtain

$$
\begin{aligned}
&d_8x^4 + d_7x^3y + d_6x^2y^2 - d_5x^2 + d_4xy^3 - d_3xy \\
&\quad - d_2y^2 + d_1y^4 + d_0 = 0 \qquad \text{(A3)}
\end{aligned}
$$

where

$$
\begin{aligned}
d_8 &= a_2^2 + c_2^2 \\
d_7 &= 2a_2c_1 + 2c_2c_3 \\
d_6 &= c_1^2 + 2a_2b_2 + c_2^2 + c_3^2 \\
d_5 &= 2a_2r_2 + c_2^2 \\
d_4 &= 2b_2c_1 + 2c_2c_3 \\
d_3 &= 2r_2c_1 + 2c_2c_3 \\
d_2 &= 2r_2b_2 + c_3^2 \\
d_1 &= b_2^2 + c_3^2 \\
d_0 &= r_2^2
\end{aligned}
$$

Equation A3 is a degree 8 monomial in x after direct elimination of y using Equation A2. However, it can be reduced to a quartic equation by substitution since only the terms with the degrees of 0, 2, 4 and 8 appear in it.

We introduce new variables t and u:

$$
\begin{aligned}
x &= a\sin t \qquad \text{(A4)} \\
y &= b\cos t \qquad \text{(A5)} \\
u &= \cos 2t, \qquad \text{(A6)}
\end{aligned}
$$

where $a = \sqrt{\frac{r_1}{a_1}} = \sqrt{\frac{r - S_{zz}}{S_{xx} - S_{zz}}}$, $b = \sqrt{\frac{r_1}{b_1}} = \sqrt{\frac{r - S_{zz}}{S_{yy} - S_{zz}}}$ are two constants and r_1, a_1 and b_1 are all negative if we choose $S_{zz} = \max(|S_{xx}|, |S_{yy}|, |S_{zz}|)$. Through algebraic manipulation we obtain

$$
f_4u^4 + f_3u^3 + f_2u^2 + f_1u + f_0 = 0, \qquad \text{(A7)}
$$

where

$$
\begin{aligned}
f_4 &= e_1^2 + e_2^2 \\
f_3 &= 2e_1e_3 + 2e_2e_4 \\
f_2 &= 2e_1e_0 + e_3^2 + e_4^2 - e_2^2 \\
f_1 &= 2e_3e_0 - 2e_2e_4 \\
f_0 &= e_0^2 - e_4^2 \\
e_4 &= d_7a^3b + d_4ab^3 + 2d_3ab \\
e_3 &= 2(d_1b^4 + d_2b^2 - d_5a^2 - d_8a^4) \\
e_2 &= d_4ab^3 - d_7a^3b \\
e_1 &= d_8a^4 + d_1b^4 - d_6a^2b^2 \\
e_0 &= 4r_2^2 + d_8a^4 + 2d_5a^2 + 2d_2b^2 + d_1b^4 \\
&\quad + d_6a^2b^2.
\end{aligned}
$$

Finally, since $u = 1 - 2(\frac{x}{a})^2$, x^2 also satisfies a quartic equation. The y-component of the unit vector $\mathbf{v}$ can be computed from Equation A2.

Appendix B: Simple trigonometric equations for computing dihedral angles

In this section we prove that if the directions of any two vectors in consecutive peptide planes i and $i+1$ are known, then the intervening backbone dihedral angles (ϕ_i, ψ_i) satisfy simple trigonometric equations. The proof is given specifically for two NH bond vectors. We derive equations in a coordinate system defined on a peptide plane i with +Z-axis along the bond vector $H^N(i) \rightarrow N(i)$ (the symbol $\rightarrow$ means from atom $H^N(i)$ to atom N(i)), and +Y-axis in the peptide plane i such that the angle between +Y and the bond vector $N(i) \rightarrow C_\alpha(i)$ is $<90°$. The +X-axis is defined based on the right-handedness.

From protein backbone geometry the two NH vectors $\mathbf{v}_i$ and $\mathbf{v}_{i+1}$ of the residues i and $i+1$ can be related by 8 rotation matrices (Figure 1) between two coordinate systems defined, respectively, in the peptide planes i and $i+1$ as described in the above.

$$
\mathbf{v}_i = \mathbf{R}\mathbf{v}_{\mathbf{i+1}}, \qquad \text{(B1)}
$$

where

$$
\begin{aligned}
\mathbf{R} &= \mathbf{R}_x(\theta_7)\mathbf{R}_y(\theta_6)\mathbf{R}_x(\theta_5)\mathbf{R}_z(\psi_i + \pi)\mathbf{R}_x(\theta_3) \\
&\quad \mathbf{R}_y(\phi_i)\mathbf{R}_y(\theta_8)\mathbf{R}_x(\theta_1), \qquad \text{(B2)}
\end{aligned}
$$

where

$$
\mathbf{R}_x(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix}
$$

$$\mathbf{R}_y(\theta) = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix}$$

$$\mathbf{R}_z(\theta) = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

the six angles $\theta_1, \theta_3, \theta_5, \theta_6, \theta_7$ and θ_8 (Table 4 and Figure 1 of the main text) are given so $\mathbf{R}_l = \mathbf{R}_x(\theta_7)\mathbf{R}_y(\theta_6)\mathbf{R}_x(\theta_5)$ and $\mathbf{R}_r = \mathbf{R}_y(\theta_8)\mathbf{R}_x(\theta_1)$ are two 3×3 constant matrices. The backbone dihedral angles ϕ_i and ψ_i are defined according to the standard convention.

We define two new vectors $\mathbf{w}_1 = (x_1, y_1, z_1)$ and $\mathbf{w}_2 = (x_2, y_2, z_2)$ by

$$\mathbf{w}_1 = {\mathbf{R}_l}^{-1}\mathbf{v}_i$$
$$\mathbf{w}_2 = \mathbf{R}_r\mathbf{v}_{i+1},$$

then we have

$$\begin{aligned} x_1 = &-(\cos\phi_i\cos\psi_i + \sin\theta_3\sin\phi_i\sin\psi_i)\ x_2 \\ &- \cos\theta_3\sin\psi_i\ y_2 \\ &+ (\cos\psi_i\sin\phi_i - \cos\phi_i\sin\theta_3\sin\psi_i)\ z_2 \end{aligned} \quad \text{(B3)}$$

$$\begin{aligned} y_1 = &(\cos\phi_i\sin\psi_i - \sin\theta_3\sin\phi_i\cos\psi_i)\ x_2 \\ &- \cos\theta_3\cos\psi_i\ y_2 \\ &- (\sin\phi_i\sin\psi_i + \cos\phi_i\sin\theta_3\cos\psi_i)\ z_2 \end{aligned} \quad \text{(B4)}$$

$$\begin{aligned} z_1 = &\sin\phi_i\cos\theta_3\ x_2 - \sin\theta_3\ y_2 \\ &+ \cos\phi_i\cos\theta_3\ z_2. \end{aligned} \quad \text{(B5)}$$

By Equation (B5) and through algebraic manipulation and trigonometric identities we arrive at the following simple trigonometric equation for the ϕ_i angle:

$$\begin{aligned} \sin(\phi_i + a_1) &= \sin\phi_i\cos a_1 + \cos\phi_i\sin a_1 \\ &= b_1, \end{aligned} \quad \text{(B6)}$$

where

$$\begin{aligned} b_1 &= \frac{z_1 + y_2\sin\theta_3}{\sqrt{(x_2\cos\theta_3)^2 + (z_2\cos\theta_3)^2}} \\ \cos a_1 &= \frac{x_2\cos\theta_3}{\sqrt{(x_2\cos\theta_3)^2 + (z_2\cos\theta_3)^2}} \\ \sin a_1 &= \frac{z_2\cos\theta_3}{\sqrt{(x_2\cos\theta_3)^2 + (z_2\cos\theta_3)^2}}. \end{aligned}$$

Here, one need not to be concerned with the sign of square roots since simultaneously flipping the signs of $b_1, \cos a_1$ and $\sin a_1$ terms will not change Equation B6.

Substituting the computed ϕ_i value into Equation B4 and through similar algebraic manipulation and trigonometric identities we arrive at the following simple trigonometric equation for ψ_i:

$$\sin(\psi_i + a_2) = b_2, \quad \text{(B7)}$$

where both $b_2 \leq 1$ and a_2 are computed from $y_1, x_2, y_2, z_2, \theta_3$ and ϕ_i. These two simple trigonometric equations can be solved exactly. They are equivalent to quadratic equations as can be proved easily by the following substitution for the $\sin\phi_i$ and $\cos\phi_i$ in Equation B6:

$$\begin{aligned} w &= \tan\frac{\phi_i}{2}, \quad \sin\phi_i = \frac{2w}{1+w^2}, \\ \cos\phi_i &= \frac{1-w^2}{1+w^2}, \end{aligned} \quad \text{(B8)}$$

and by a similar substitution for the $\sin\psi_i$ and $\cos\psi_i$ in Equation B7. There are only two independent solutions for (ϕ_i, ψ_i) angles given two known NH vectors if the orientation of the first peptide plane is also known. In our implementation the first peptide plane is specified by an NH vector and an NC_α vector, where the former is solved from Equation 5 and the latter is computed by solving an optimization problem (see the section *Computation of an optimal first NC_α vector* of the main text).

Appendix C: A data structure and pseudocode for the refinement of a secondary structure element

In the following we describe a data structure and pseudocode for Stage II of our algorithm (see the section *Refinement of secondary structure elements* of the main text): the refinement of an m-residue secondary structure element based on a depth-first search (DFS). The data structure for the search over all the possible combinations of directions (cross product) of the m NH unit vectors obtained from Equation 4 is a tree constructed as follows. The height of the tree is m for an m-residue fragment and the first residue corresponds to depth 1. Each node at depth i corresponds to an NH vector from Equation 4 for residue i. The children of each node at depth i are all the NH solutions for residue $i + 1$. During the DFS backbone (ϕ, ψ) angles are computed from the jth member of the NH solution set at depth i and the kth member of the NH solution set at depth $i + 1$. A single step of

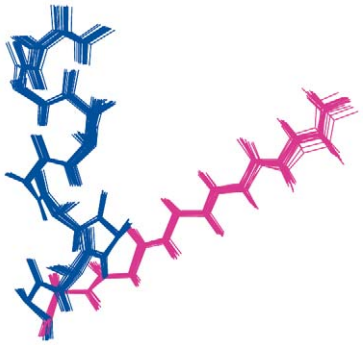

Figure 8. Superposition of backbone structures. The superposition of 20 RDC-refined fragments consisting of the helix (N25–E34) (blue) and the β-strand (Q2–T7) (magenta) with pairwise backbone RMSD of 0.31 Å ± 0.14 Å. These structures were generated with the same variance (1.0 Hz) for error distributions and the same sampling size (1024) but using 20 different seeds for a random number generator.

the DFS-based refinement is shown in Figure 3 of the main text. The pruning afforded by the Ramachandran filter occurs in three places during the refinement. At each step, if the computed (ϕ, ψ) value falls inside the favorable Ramachandran region the DFS continues using the kth member at depth $i + 1$, otherwise, the subtree rooted at the kth member of the NH solution set at depth $i + 1$ is pruned and the DFS continues immediately at the $(k + 1)$th member. If none of the (ϕ, ψ) angles computed from the cross product of the jth member of the NH solution set at depth i and all the members of NH solution set at depth $i + 1$ falls into the favorable region, the subtree rooted at the jth member at depth i is pruned and DFS continues immediately at the $(j + 1)$th member at depth i. If none of the (ϕ, ψ) angles computed from the cross product of the two NH solution sets at depth i and $i + 1$ falls into the favorable region the subtree at depth $i - 1$ rooted at the parent of the current NH solution set at depth i is pruned and the DFS continues immediately at the next member of the NH solution set at depth $i - 1$.

Appendix D: Algorithmic complexity and performance

The complexity analysis of the algorithm is as follows. The Saupe matrices, the coefficients of the quartic and quadratic equations and their solutions can all be computed in $O(1)$ time. An initial model with m residues can be built in $O(m)$ time. The search for an optimal first NC_α vector (see the section *Computation of an optimal first NC_α vector* of the main text) takes $O(mk_1^3)$ time on a $k_1 \times k_1 \times k_1$ grid. In practice, it takes less than one minute on a $180 \times 90 \times 180$ grid on a Pentium 4 (2.4 GHz) Linux workstation. The search for relative positions among RDC-derived structure elements using NOE distances (see the section *Backbone structure determination* of the main text) takes $O(lk_2^2)$ time on a $k_2 \times k_2$ grid with l NOEs. In practice, it takes less than one second on a 90×180 grid. The secondary structure determination step takes $O(k16^m)$ time in the worst case where k is the resolution of a grid search over a Gaussian distribution simulating the experimental errors. With a perfect input (RDCs without experimental errors) our algorithm for secondary structure determination is guaranteed to find a global minimum. However, in practice we must perturb the experimental RDC values using an error model. In summary, the total run time of the algorithm is $O(n(lk_2^2 + mk_1^3 + k16^m))$ for n fragments. In practice, the grid search over the Gaussian distribution was implemented as a random sampling. For each sampled set consisting of a RDC value for every residue of a fragment our algorithm is still guaranteed to find a global minimum. Further, our results show that the algorithm converges quickly in practice when computing both regular α-helices and β-sheets. For example, a sample size of 2048 is large enough to generate conformations with pairwise backbone RMSD $<$0.50 Å for both the helix (N25–E34) and any of the five strands (see Figure 8 for an example). The DFS step for computing all plausible conformation vectors, though having exponential running time in the worst-case, is quite efficient in practice since a vast majority of elements in the cross product of the possible NH directions were pruned very early on. For example, the average branching factor (i.e. the exponent) for a set of sampled RDCs yielding a plausible conformation vector for the helix is only 1.03, thus only a few plausible conformation vectors can be computed from such a set. Further, our algorithm (see Figure 1 and the section *Algorithm* of the main text), first divides an entire protein into secondary structure elements, then computes each of them individually. Even though n gets larger as a protein gets larger, m, the size of a strand or helix, will typically not be arbitrarily large. Therefore, we anticipate that our algorithm will have applications for the determination of backbone folds of large proteins using mainly RDC data. In summary, despite the worst-case exponential running time in m the DFS-based minimization for computing an optimal conformation vector takes, in practice, only several minutes for either a helix or any of the five strands. In

total, our algorithm takes about 45 min to compute the backbone substructure of a 39-residue portion of ubiquitin consisting of a helix (N25–E34) and the single sheet with five strands.

Algorithm 1 Refinement of secondary structure elements using RDCs.
A plausible and an optimal conformation vector is defined in the section *Refinement of secondary structure elements* of the main text.

Input:
an NC_α vector for residue 1;
/* *The loop is repeated for the size of the random sampling over Gaussian distributions* */
loop
repeat
for every residue from 1 to m /* m *is the number of residues in the fragment* */ **do**
select an RDC value for each medium based on a Gaussian distribution
compute the NH solution set for the residue i from the quartic equation Equation 4
until none of the m NH solution sets is empty /* *all the solutions exists* */
/* *Search for an optimal conformation vector over the tree by depth-first search(DFS)* */
$i \leftarrow 1$ /* *the depth of the tree* */
$j \leftarrow 1$ /* *the current member of the NH solution set at depth* i */
$k \leftarrow 1$ /* *the current member of the NH solution set at depth* $i + 1$ */
$s \leftarrow$ the size of the NH solution set at depth i
$t \leftarrow$ the size of the NH solution set at depth $i + 1$
while $j < s$ /* *DFS begins with the* j*th member of the* NH *solution set at depth* i */ **do**
while $k < t$ **do**
compute (ϕ, ψ) from Equations 23, 24 using the NC_α direction for i
if(ϕ, ψ) falls inside the Ramachandran favorable region **then**
save the (ϕ, ψ) as the ith member of a plausible conformation vector
if $(i + 1 < m)$ **then**
compute the NC_α vector for $i + 1$
DFS continues with the kth member of the NH solution set at depth $i + 1$
else
compute the score from the Equation 7 or 8 of the main text
if the new score is better **then**
update the score, the optimal conformation vector, and the first NH vector
$k \leftarrow k + 1$
/* *If the* (ϕ, ψ) *falls outside the favorable region the subtree rooted at the kth member of the solution set at depth* $i + 1$ *is pruned* */
$j \leftarrow j + 1$
/* *If none of the* (ϕ, ψ) *angles from the cross product of the* j*th member of the* NH *solution set at depth* i *and the* NH *solution set at depth* $i + 1$ *falls inside the favorable region the subtree rooted at the* j*th member is pruned* */
if none of the (ϕ, ψ) angles from depth i and $i + 1$ falls in the favorable Ramachandran region **and** $(i - 1) > 0$ **then**
DFS continues with the next member of the NH solution set at depth $i - 1$
/* *The subtree at depth* $i - 1$ *rooted at the parent of the current* NH *solution set for depth* i *is pruned* */
Build a model using the optimal conformation vector and the first NC_α and the first NH vector

References

Al-Hashimi, H.M., Valafar, H., Terrell, M., Zartler, E.R., Eidsness, M.K. and Prestegard, J.H. (2000) *J. Magn. Reson.*, **143**, 402–406.

Andrec, M., Du, P. and Levy, R.M. (2001) *J. Biomol. NMR*, **21**, 335–347.

Bailey-Kellogg, C., Widge, A., Kelley, J.J., Berardi, M.J., Bushweller, J.H. and Donald, B.R. (2000) *J. Comput. Biol.*, **7**, 537–558.

Barbieri, R., Bertini, I., Cavallaro, G., Lee, Y., Luchinat, C. and Rosato, A. (2002) *J. Am. Chem. Soc.*, **124**, 5581–5587.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) *Nucl. Acids Res.*, **28**, 235–242.

Bertram, R., Quine, J.R., Chapman, M.S. and Cross, T.A. (2000) *J. Magn. Reson.*, **147**, 9–16.

Brünger, A.T. (1993) *XPLOR: A System for X-Ray Crystallography and NMR*, Yale University Press, New Haven.

Chou, J.J., Gaemers, S., Howder, B., Louis, J.M. and Bax, A. (2001) *J. Biomol. NMR*, **21**, 377–382.

Clore, G.M., Gronenborn, A.M. and Bax, A. (1998) *J. Magn. Reson.*, **133**, 216–221.

Clore, G.M., Starich, M.R., Bewley, C.A., Cai, M.L. and Kuszewski, J. (1999) *J. Am. Chem. Soc.*, **121**, 6513–6514.

Cormen, T.H., Leiserson, C.E., Rivest, R.L. and Stein, C. (2001) *Introduction to Algorithms*, The MIT Press.

Cornilescu, G., Marquardt, J.L., Ottiger, M. and Bax, A. (1998) *J. Am. Chem. Soc.*, **120**, 6836–6837.

Delaglio, F., Kontaxis, G. and Bax, A. (2000) *J. Am. Chem. Soc.*, **122**, 2142–2143.

Dominguez, C., Boelens, R. and Bonvin, A.M.J.J. (2003) *J. Am. Chem. Soc.*, **125**, 1731–1737.

Engh, R.A. and Huber, R. (1991) *Acta Cryst.*, **A47**, 392–400.

Fowler, A.C., Tian, F., Al-Hashimi, H.M. and Prestegard, J.H. (2000) *J. Mol. Biol.*, **304**, 447–460.

Gardner, K.H. and Kay, L.E. (1997) *J. Am. Chem. Soc.*, **119**, 7599–7600.

Giesen, A.W., Homans, S.W. and Brown, J.M. (2003) *J. Biomol. NMR* **25**, 63–71.

Gnu (2002) The gnu general public license, http://www.gnu.org/licenses/licenses.html

Güntert, P., Mumenthaler, C. and Wüthrich, K. (1997) *J. Mol. Biol.*, **273**, 283–298.

Hansen, M.R., Hanson, P. and Pardi, A. (2000) *Meth. Enzymol.*, **317**, 220–240.

Hus, J.C., Marion, D. and Blackledge, M. (2001) *J. Am. Chem. Soc.*, **123**, 1541–1542.

Kac, M. (1948) *Proc. London Math. Soc.*, **50**, 390–408.

Kemple, M.D., D., R.B., Lipkowitz, K.B., Prendergast, F.G. and Rao, B.D. (1988) *J. Am. Chem. Soc.*, **110**, 8275– 8287.

Kolinski, A. and Skolnick, J. (1998) *Proteins*, **32**, 475–494.

Losonczi, J.A., Andrec, M., Fischer, M.W. and Prestegard, J.H. (1999) *J. Magn. Reson.*, **138**, 334–342.

Meiler, J., Blomberg, N., Nilges, M. and Griesinger, C. (2000) *J. Biomol. NMR*, **16**, 245–252.

Nomura, K. and Kainosho, M. (2002) *J. Magn. Reson.*, **154**, 146–153.

Ottiger, M. and Bax, A. (1998) *J. Am. Chem. Soc.*, **120**, 12334–12341.

Quine, J.R., Brenneman, M. and Cross, T. (1997) *Biophys. J.*, **72**, 2342–2348.

Ramirez, B.E. and Bax, A. (1998) *J. Am. Chem. Soc.*, **120**, 9106–9107.

Rienstra, C.M., Tucker-Kellogg, L., Jaroniec, C.P., Hohwy, M., Reif, B., McMahon, M.T., Tidor, B., Lozano-Pèrez, T. and Griffin, R.G. (2002) *Proc. Natl. Acad. Sci. USA*, **99**, 10260–10265.

Rohl, C.A. and Baker, D. (2002) *J. Am. Chem. Soc.*, **124**, 2723–2729.

Saupe, A. (1968) *Angew. Chem.*, **7**, 97–112.

Stryer, L. (1994) *Biochemistry*, W.H. Freeman and Company.

Tian, F., Valafar, H. and Prestegard, J.H. (2001) *J. Am. Chem. Soc.*, **123**, 11791–11796.

Tjandra, N. and Bax, A. (1997) *Science*, **278**, 1111–1114.

Tolman, J.R., Flanagan, J.M., Kennedy, M.A. and Prestegard, J.H. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 9279–9283.

Vijay-Kumar, S., Bugg, C.E. and Cook, W.J. (1987) *J. Mol. Biol.*, **194**, 531–544.

Wang, L., Pang, Y., Holder, T., Brender, J., Kurochkin, A.V. and Zuiderweg, E.R.P. (2001) *Proc. Natl. Acad. Sci. USA*, **98**, 7684–7689.

Wang, Y.X., Marquardt, J.L., Wingfield, P., Stahl, S.J., Lee-Huang, S., Torchia, D. and Bax, A. (1998) *J. Am. Chem. Soc.*, **120**, 7385–7386.

Wedemeyer, W.J., Rohl, C.A. and Scheraga, H.A. (2002) *J. Biomol. NMR*, **22**, 137–151.

Yue, K. and Dill, K.A. (2000) *Protein Sci.*, **9**, 1935–1946.