

# Supporting Information for “Extracting Structural Information from Residual Chemical Shift Anisotropy: Analytic Solutions for Peptide Plane Orientations and Applications to Determine Protein Structure”<sup>†</sup>

Chittaranjan Tripathy,<sup>1</sup> Anthony K. Yan<sup>1,2</sup>, Pei Zhou,<sup>2</sup> and Bruce Randall Donald<sup>1,2,\*</sup>

<sup>1</sup>Department of Computer Science, Duke University, Durham, NC 27708

<sup>2</sup>Department of Biochemistry, Duke University Medical Center, Durham, NC 27710

Below is the supporting information (SI) for the following paper [12]:

- C. Tripathy, A. K. Yan, P. Zhou, and B. R. Donald. Extracting Structural Information from Residual Chemical Shift Anisotropy: Analytic Solutions for Peptide Plane Orientations and Applications to Determine Protein Structure. In *Proceedings of the 17th Annual International Conference on Research in Computational Molecular Biology (RECOMB), Beijing, China, 2013*.

## Appendix

In **Appendix A**, we show that solving the system of equations from N-H<sup>N</sup> RDC, and <sup>13</sup>C-RCSA or <sup>15</sup>N-RCSA to compute the peptide plane orientations leads to solving a degree 32 univariate polynomial equation. In contrast, our algorithm RDC-CSA-ANALYTIC uses a novel hybrid approach, and avoids solving such a high-degree polynomial equation. In **Appendix B**, we describe the algorithmic modules of RDC-CSA-ANALYTIC, and their inner workings. In **Appendix C**, we give a proof of Proposition 2, and describe a method to construct all possible peptide plane orientations that satisfy RCSA and/or RDC data measured for a planar structural motif such as a peptide plane. Finally, in **Appendix D**, we give the sixteen oriented peptide planes computed for the peptide plane defined by Ala28 and Lys29 of ubiquitin.

**A note on the exact solutions.** The derivations below are not short, but they must only be performed once as we have done here. After that, the simple results (the closed-form solutions) may be used without thought of the proofs of correctness given here.

### A Peptide Plane Orientations from N-H<sup>N</sup> RDC, and <sup>13</sup>C-RCSA or <sup>15</sup>N-RCSA

Here we consider the most general case when two  $\psi$ -defining RCSAs or RDCs are available for residues. From NMR experiments standpoint, this case is important, since one of <sup>13</sup>C'-RCSA or

---

<sup>†</sup>Grant sponsor: National Institutes of Health; Grant numbers: R01 GM-65982 and R01 GM-78031 to BRD, and R01 GM-079376 to PZ

\*Corresponding author: Bruce Randall Donald, ✉brd+recomb13@cs.duke.edu, ☎919-660-6583, 📠919-660-6519

$^{15}\text{N}$ -RCSA are measured along with N- $\text{H}^{\text{N}}$  RDC. Further, this includes the case when two RCSAs per residue are available. We show that given the orientation of the peptide plane  $P_i$ , and two  $\psi$ -defining RDCs and/or RCSAs for vectors on the peptide plane  $P_{i+1}$ , computing the orientations of  $P_{i+1}$  relative to  $P_i$  requires us to solve a degree 32 polynomial equation.

We assume that the peptide plane is planar. Let  $\mathbf{v}_1 = \boldsymbol{\delta}_{xx} = (x_1, y_1, z_1)^T$  and  $\mathbf{v}_2 = \boldsymbol{\delta}_{yy} = (x_2, y_2, z_2)^T$  denote the two unit vectors representing two principal components of the traceless RCSA tensor on the peptide plane in the POF. Let the unit vectors  $\mathbf{v}_3 = (x_3, y_3, z_3)^T$  and  $\mathbf{v}_4 = (x_4, y_4, z_4)^T$  denote N- $\text{H}^{\text{N}}$  and  $\text{C}^\alpha$ - $\text{C}'$  bond orientations, respectively relative to the POF.

Since  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$  and  $\mathbf{v}_4$  are coplanar, we have

$$\mathbf{v}_1 = \alpha \mathbf{v}_3 + \beta \mathbf{v}_4 \quad (\text{A.1})$$

$$\mathbf{v}_2 = \alpha \mathbf{v}_3 + \beta \mathbf{v}_4 \quad (\text{A.2})$$

where  $\alpha, \beta \in \mathbb{R}$  are constants, from which we get four independent equations

$$x_1 = \alpha_1 x_3 + \beta_1 x_4 \quad (\text{A.3})$$

$$y_1 = \alpha_1 y_3 + \beta_1 y_4 \quad (\text{A.4})$$

$$x_2 = \alpha_2 x_3 + \beta_2 x_4 \quad (\text{A.5})$$

$$y_2 = \alpha_2 y_3 + \beta_2 y_4 \quad (\text{A.6})$$

where  $\alpha_i, \beta_i$  for  $1 \leq i \leq 2$  are constants. These four equations can be used in the reduced RCSA equation (Eq. (9) of the main text [12]) to obtain

$$a_1 x_3^2 + a_2 y_3^2 + a_3 x_4^2 + a_4 y_4^2 + a_5 x_3 x_4 + a_6 y_3 y_4 = a_0 \quad (\text{A.7})$$

where  $a_i$  for  $0 \leq i \leq 6$  are constants.

The reduced RDC equation for  $\mathbf{v}_4$  can be written as

$$b_1 x_3^2 + b_2 y_3^2 = b_0 \quad (\text{A.8})$$

where  $b_i$  for  $0 \leq i \leq 2$  are constants.

Since the angle between  $\mathbf{v}_3$  and  $\mathbf{v}_4$  is fixed, we have the following equation:

$$x_3 x_4 + y_3 y_4 + z_3 z_4 = c_0 \quad (\text{A.9})$$

where the constant  $c_0$  is the cosine of the angle between  $\mathbf{v}_3$  and  $\mathbf{v}_4$ .

Since the orientation of the previous peptide plane has already been determined inductively, we let the unit vector along N- $\text{C}^\alpha$  bond of the previous peptide plane be denoted by  $(d_1, d_2, d_3)^T$ . From the fact that the angle between  $\mathbf{v}_4$  and  $(d_1, d_3, d_3)^T$  is fixed we get one more independent equation:

$$d_1 x_4 + d_2 y_4 + d_3 z_4 = d_0 \quad (\text{A.10})$$

where  $d_i$  for  $0 \leq i \leq 3$  are constants.

Since  $\mathbf{v}_3$  and  $\mathbf{v}_4$  are unit vectors we have two more independent equations:

$$x_3^2 + y_3^2 + z_3^2 = 1 \quad (\text{A.11})$$

$$x_4^2 + y_4^2 + z_4^2 = 1. \quad (\text{A.12})$$

Now all we need to do is to solve the following set of simultaneous equations:

$$a_1x_3^2 + a_2y_3^2 + a_3x_4^2 + a_4y_4^2 + a_5x_3x_4 + a_6y_3y_4 = a_0 \tag{A.13}$$

$$b_1x_3^2 + b_2y_3^2 = b_0 \tag{A.14}$$

$$x_3x_4 + y_3y_4 + z_3z_4 = c_0 \tag{A.15}$$

$$d_1x_4 + d_2y_4 + d_3z_4 = d_0 \tag{A.16}$$

$$x_3^2 + y_3^2 + z_3^2 = 1 \tag{A.17}$$

$$x_4^2 + y_4^2 + z_4^2 = 1. \tag{A.18}$$

It can be shown that solving this system (Eq. (A.13) to Eq. (A.18)) of polynomial equations is equivalent to solving a 32 degree univariate polynomial equation, which is a difficult computational problem. Our algorithm RDC-CSA-ANALYTIC uses a novel hybrid approach, and avoids solving such a high-degree polynomial equation.

## B The Four Modules of RDC-CSA-ANALYTIC

Below we give a brief description of the four core modules of RDC-CSA-ANALYTIC.

**Computation of Alignment Tensor.** Our alignment tensor computation module is built upon our previous implementations in RDC-CSA-ANALYTIC that uses RDC data [17, 20, 19]. We have extended this to incorporate the constraints from RCSAs. The data fitting is done using singular value decomposition (SVD) [7, 17, 18, 6], which now uses equations for RCSAs. An ideal helix model for a helical SSE of the protein is used to bootstrap the alignment tensor computation, and using SVD the initial alignment tensor  $\mathbf{S}$  is computed. Then  $\mathbf{S}$  iteratively refined by using the computed helix structures by RDC-CSA-ANALYTIC. Once the values of the alignment tensor is estimated, other fragments of the protein are computed using these values. Usually 2 or 3 iterations (a user-defined parameter) are sufficient to obtain a very good estimate of  $\mathbf{S}$ .

**Determination of Conformations and Orientations of the SSEs.** RDC-CSA-ANALYTIC uses a divide-and-conquer approach to partition the protein backbone of size  $n$  residues into  $\Theta(n)$  fragments of bounded lengths consisting of SSEs and loops, hence it runs in  $\Theta(n)$  time. For each SSE, it inductively computes the peptide plane orientations using the method described in Section 2.3. The conformations are represented by a conformation tree grown recursively as we solve for the DOFs progressively. An internal (i.e., non-leaf) node in the tree represents the conformation of a part of a candidate SSE ( $\alpha$ -helix or  $\beta$ -Strand), and a leaf node represents a candidate SSE conformation computed from RDCs and RCSAs. As each node is visited in a depth-first traversal of the tree, a set of conformation filters [13, 14] are applied as *predicates*. If the node passes all the filters, then the subtree rooted at that node is visited; otherwise, the node is designated as a *dead-end* node, and the subtree rooted at that node is pruned. Finally, all remaining unpruned conformations (leaf nodes) are evaluated for the satisfaction of available experimental data, and deviations from standard average ( $\phi_a, \psi_a$ ) values for the SSE type. The top  $k$  (user-defined parameter) conformations are output. When  $k = 1$ , only the maximum-likelihood (ML) SSE conformation is output.

Since every conformation consistent with the data is either pruned or evaluated, RDC-CSA-ANALYTIC guarantees completeness. The experimental uncertainties in RDC and RCSA data are modeled by adding a Gaussian noise.

The RDC/RCSA RMSD between back-computed and experimental RDCs/RCSAs is computed

using the equation

$$\text{RMSD}_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (r_{x,i}^b - r_{x,i}^e)^2}, \quad (\text{B.1})$$

where  $x$  is either a  $\phi$ -defining or a  $\psi$ -defining RDC/RCSA type,  $n$  is the number of RDCs/RCSAs,  $r_{x,i}^e$  is the experimental RDC/RCSA, and  $r_{x,i}^b$  is the corresponding back-computed RDC/RCSA. The RMS deviations from standard average  $(\phi_a, \psi_a)$  is given by

$$\text{RMSD}_{\phi,\psi} = \sqrt{\frac{1}{2n} \sum_{i=1}^n ((\phi_i - \phi_a)^2 + (\psi_i - \psi_a)^2)}, \quad (\text{B.2})$$

where  $(\phi_a, \psi_a)$  are the standard averages of  $\phi$  and  $\psi$  dihedrals for the SSE type,  $\phi_i$  and  $\psi_i$  are the dihedrals computed using RDC and RCSA data, and  $n$  is the number of residues of the SSE.

**Simultaneous Structure Determination and Packing of  $\beta$ -Strands into  $\beta$ -Sheets.** Let  $E = \{E_1, \dots, E_n\}$  be a  $\beta$ -sheet, where each  $E_i$  specifies the boundaries of the  $\beta$ -strand  $E_i$ . We arbitrarily pick one of the  $\beta$ -strands in the  $\beta$ -sheet as the first strand, and compute an ML conformation using the previous module. Let the strand conformation be included in the  $\beta$ -sheet  $S$ , which is initialized to an empty set. For each strand  $E_i$  adjacent to  $S$ , all possible conformations are computed using the previous module. As each conformation is evaluated, pack (we call it the PACK procedure) it with  $S$  using a sparse set  $T_{\text{NOE}}$  of NOEs (or hydrogen bond information, if available), such that  $E_i$  forms favorable hydrogen bonds with  $S$ . The best scoring strand  $E_{i_{\text{best}}}$  is chosen to be paired with  $S$ . This process continues until all the strands in  $E$  are computed and packed to form a  $\beta$ -sheet. The pseudo-code for strand packing algorithm is given below.

COMPUTEANDPACK $\beta$ SHEET( $S, T_{\text{NOE}}$ )

1. Let  $E = \{E_1, \dots, E_n\}$  be the  $\beta$ -sheet specification, where  $E_i$ s are strand specifications
2. Initialize the partial sheet conformation  $S \leftarrow \emptyset$
3. Compute  $E_1$  using RDC-CSA-ANALYTIC
4.  $S \leftarrow S \cup \{E_1\}$
5.  $\forall E_i \in E, 2 \leq i \leq n$  in the order of strand adjacency
6.  $\forall E_{i_{\text{leaf}}}$ , a solution (leaf node) in a RDC-CSA-ANALYTIC conformation tree
7. PACK( $S, E_{i_{\text{leaf}}}, T_{\text{NOE}}$ ) and test if it scores better than the previously packed sheet
8. Let  $E_{i_{\text{best}}}$  be the strand with best packing and RDC satisfaction scores
9.  $S \leftarrow S \cup \{E_{i_{\text{best}}}\}$
10. **return**  $S$

**Assembly of  $\alpha$ -Helices and  $\beta$ -Sheets into Backbone Global Fold.** It is well known that at least three distance constraints are needed to determine the relative translation between two oriented rigid bodies. The computed SSEs from RDC and RCSA data are all oriented with a four-fold degeneracy inherent to RDCs and RCSAs. Although the orientational degeneracy can

be resolved when RDCs and RCSAs can be recorded in multiple alignment media, it cannot be resolved using only RDC and RCSA data in one alignment medium, unless other types of data such as NOEs are used. It has been shown that the orientational degeneracy can be resolved and the relative translations between the SSEs can be determined using a small number of (at least three NOEs between a pair of SSEs) inter-SSE NOEs [2, 17, 20]. Since the NOEs are interpreted as distance intervals between atoms, a small number of NOE distance restraints can bound the conformation space into a small volume which can be discretized and enumerated using a grid search of parameterized resolution. A sparse set of inter-SSE NOEs are extracted from chemical shift information alone [20]. For large and perdeuterated proteins, isoleucine-leucine-valine (ILV) selective labeling method [3, 4, 15, 11] and selective labeling of alanine residues [5, 1], in which protons and  $^{13}\text{C}$  isotopes are selectively incorporated into methyl groups of Ile $^{\delta 1}$ , Leu $^{\delta}$ , Val $^{\gamma}$  and Ala $^{\beta}$  side-chains on a deuterated background, can be used to measure NOE distances between H $^{\text{N}}$ -H $^{\text{N}}$ , methyl-H $^{\text{N}}$  and methyl-methyl protons. The inter-SSE NOEs often involve side-chain protons. Since the side-chain conformation is not known a priori, a discretized set of side-chain conformations can be used to model the possible side-chain proton positions. We used Richardson’s Rotamer Library [8] to model the discrete side-chain conformations.

Suppose  $E$  be the set of inter-SSE NOEs used to pack two SSEs. For a distance restraint  $e_i \in E$ , let  $Q_i$  be the set of possible interpretations in terms of atom positions (rotamer pairs in case of NOEs). Each element  $q_{ij} \in Q_i$ ,  $1 \leq j \leq |Q_i|$  corresponds to a pair of atom positions and the corresponding distance between them which is represented by a spherical shell with inner and outer radii that respectively represent the lower and upper bounds for the distance restraint  $e_i$ . Then the volume of translation between the two SSEs that satisfies the distance restraints in  $E$  is given by

$$\mathcal{V} = \bigcap_{i=1}^{|E|} \bigcup_{j=1}^{|Q_i|} q_{ij}. \tag{B.3}$$

Let  $S = \{s_1, \dots, s_m\}$  be the set of SSEs computed by RDC-CSA-ANALYTIC. Our algorithm PACK takes as input the set of SSEs  $S$ , the set of inter-SSE distance restraints  $E$ , and parameters to specify the grid resolution and ensemble size, and computes an ensemble of packed structures. It uses the distance restraints to resolve the four-fold degeneracy of SSE orientations due to the symmetry of dipolar operator. At each grid point during the grid search over the space  $\mathcal{V}$  of relative translations, the packed structures are checked for steric clashes. The computed ensemble of packed structures can further be clustered and evaluated [9, 10] to trim the ensemble size, in order to obtain a set of representative packed structures.

## C Peptide Plane Orientations from Three RDCs and/or RCSAs on Coplanar Internuclear Vectors

We give a proof of Proposition 2, and describe a method to construct all possible peptide plane orientations that satisfy RCSA and/or RDC data measured for a planar structural motif such as a peptide plane.

Any combination of three measurements, each of which is either an RCSA on a nucleus on the peptide plane or an RDC on an internuclear vector on the peptide plane, measured in one alignment medium, can be used in our derivation to elucidate the peptide plane orientations. In general, our derivation works for any combination of three RDCs and RCSAs measured on a planar structural motif to compute all possible discrete orientations of the planar motif that satisfy the RDCs and RCSAs. Our derivation assumes that the alignment tensor is *rhombic*. That is, magnitudes of

the two smaller principal components of the alignment tensor are not equal. In other words, if  $S_{xx}$ ,  $S_{yy}$  and  $S_{zz}$  are the three principal components of a traceless alignment tensor  $\mathbf{S}$  with  $|S_{zz}| > |S_{yy}| > |S_{xx}|$ , then  $\mathbf{S}$  is rhombic if and only if  $|S_{xx}| \neq |S_{yy}|$ . The rhombicity  $R$  of the alignment tensor is given by the expression  $R = (2/3)(S_{xx} - S_{yy})/S_{zz}$ . Clearly,  $R \in [0, 2/3]$ .

We restate Proposition 2 below and then give a proof. Then we give a procedure to construct the peptide plane orientations from the roots of the quadratic equation.

**Proposition 2.** *Given a rhombic alignment tensor, and 3 measurements, each of which is either an RCSA on a nucleus on the peptide plane  $P$  or an RDC on an internuclear vector on  $P$ , there exist at most 16 possible orientations for  $P$  that satisfy the 3 measurements, and these orientations can be written and solved in closed form by solving a quadratic equation.*

*Proof.* The proof has two parts. In the first part, we derive a quadratic equation for the peptide plane orientations from three RCSA and/or RDC equations. In the second part, we will solve the quadratic equation exactly in closed form and construct all possible peptide plane orientations.

**Derivation of a quadratic equation from three RCSA and/or RDC equations.** We assume that the peptide plane is strictly planar, and we have any combination of three RCSAs and RDCs measured in one alignment medium for three unit vectors on the plane. For example, any three data types from the set  $\{\text{N-H}^{\text{N}}, \text{C}'\text{-N}, \text{C}^{\alpha}\text{-C}', \text{C}'\text{-H}^{\text{N}}\} \cup \{^{13}\text{C}'\text{-RCSA}, ^{15}\text{N-RCSA}, ^1\text{H-RCSA}\}$  work for our derivations.

Let  $P$  denote the peptide plane. Let  $\mathbf{v}_{01} = (x_{01}, y_{01}, z_{01})^T$  and  $\mathbf{v}_{02} = (x_{02}, y_{02}, z_{02})^T$  be two unit vectors on  $P$  such that  $\mathbf{v}_{01} \perp \mathbf{v}_{02}$ . We use the literals  $x, y, z, u, v$  and  $t$ , and their subscripted versions for unknowns, and all other literals denote constants throughout the derivation. We assume that for the set of RCSAs and RDCs we know the Saupe alignment tensor  $\mathbf{S}$ , and we work in a principal order frame that diagonalizes  $\mathbf{S}$ .

Since RCSAs can be interpreted as linear combinations of two virtual RDCs (see Eq. (5) of the main text [12]), we work on the most general form of the equations, that is, we will show how to compute the peptide plane orientations when three RCSA (i.e.,  $^{13}\text{C}'\text{-RCSA}$ ,  $^{15}\text{N-RCSA}$  and  $^1\text{H-RCSA}$ ) values are known for the respective atoms on the plane  $P$ . Then we will show that any combination of three RCSAs and RDCs is a special case of our derivation.

For  $^{13}\text{C}'\text{-RCSA}$   $\Delta\delta_{\text{C}'}$ , let the unit vectors  $\mathbf{v}_1 = (x_1, y_1, z_1)^T$  and  $\mathbf{v}_2 = (x_2, y_2, z_2)^T$  denote, in the POF, the two principal components of the CSA tensor that are on the peptide plane. For  $^{15}\text{N-RCSA}$   $\Delta\delta_{\text{N}}$ , let the unit vectors  $\mathbf{v}_3 = (x_3, y_3, z_3)^T$  and  $\mathbf{v}_4 = (x_4, y_4, z_4)^T$  denote, in the POF, the two principal components of the CSA tensor that are on the peptide plane. For  $^1\text{H-RCSA}$   $\Delta\delta_{\text{H}}$ , let the unit vectors  $\mathbf{v}_5 = (x_5, y_5, z_5)^T$  and  $\mathbf{v}_6 = (x_6, y_6, z_6)^T$  denote, in the POF, the two principal components of the CSA tensor that are on the peptide plane.

Since  $\mathbf{v}_{01}, \mathbf{v}_{02}, \mathbf{v}_1, \dots, \mathbf{v}_6$  are coplanar vectors, we can write

$$\begin{aligned} \mathbf{v}_1 &= a_1\mathbf{v}_{01} + b_1\mathbf{v}_{02} & \mathbf{v}_2 &= a_2\mathbf{v}_{01} + b_2\mathbf{v}_{02} \\ \mathbf{v}_3 &= a_3\mathbf{v}_{01} + b_3\mathbf{v}_{02} & \mathbf{v}_4 &= a_4\mathbf{v}_{01} + b_4\mathbf{v}_{02} \\ \mathbf{v}_5 &= a_5\mathbf{v}_{01} + b_5\mathbf{v}_{02} & \mathbf{v}_6 &= a_6\mathbf{v}_{01} + b_6\mathbf{v}_{02}, \end{aligned}$$

where  $a_i$  and  $b_i$  for  $1 \leq i \leq 6$  are constants, determined from the fixed kinematic relationships

between these coplanar vectors. From the above equations, we can write the following equations:

$$\begin{aligned}
x_1 &= a_1x_{01} + b_1x_{02} & y_1 &= a_1y_{01} + b_1y_{02} \\
x_2 &= a_2x_{01} + b_2x_{02} & y_2 &= a_2y_{01} + b_2y_{02} \\
x_3 &= a_3x_{01} + b_3x_{02} & y_3 &= a_3y_{01} + b_3y_{02} \\
x_4 &= a_4x_{01} + b_4x_{02} & y_4 &= a_4y_{01} + b_4y_{02} \\
x_5 &= a_5x_{01} + b_5x_{02} & y_5 &= a_5y_{01} + b_5y_{02} \\
x_6 &= a_6x_{01} + b_6x_{02} & y_6 &= a_6y_{01} + b_6y_{02}.
\end{aligned}$$

Substituting the above equations for  $x_1, \dots, x_6, y_1, \dots, y_6$  in the reduced RCSA equations for  $\Delta\delta_{C'}$ ,  $\Delta\delta_N$  and  $\Delta\delta_H$ , we obtain

$$A_1x_{01}^2 + A_2y_{01}^2 + A_3x_{02}^2 + A_4y_{02}^2 + A_5x_{01}x_{02} + A_6y_{01}y_{02} + A_7 = 0 \quad (\text{C.1})$$

$$B_1x_{01}^2 + B_2y_{01}^2 + B_3x_{02}^2 + B_4y_{02}^2 + B_5x_{01}x_{02} + B_6y_{01}y_{02} + B_7 = 0 \quad (\text{C.2})$$

$$C_1x_{01}^2 + C_2y_{01}^2 + C_3x_{02}^2 + C_4y_{02}^2 + C_5x_{01}x_{02} + C_6y_{01}y_{02} + C_7 = 0, \quad (\text{C.3})$$

where

$$\begin{aligned}
A_1 &= a\lambda_2a_2^2 + a\lambda_1a_1^2, & A_2 &= b\lambda_2a_2^2 + b\lambda_1a_1^2, & A_3 &= a\lambda_2b_2^2 + a\lambda_1b_1^2, \\
A_4 &= b\lambda_1b_1^2 + b\lambda_2b_2^2, & A_5 &= 2a\lambda_1a_1b_1 + 2a\lambda_2a_2b_2, & A_6 &= 2b\lambda_1a_1b_1 + 2b\lambda_2a_2b_2, & A_7 &= -c_1, \\
B_1 &= a\lambda_4a_4^2 + a\lambda_3a_3^2, & B_2 &= b\lambda_4a_4^2 + b\lambda_3a_3^2, & B_3 &= a\lambda_4b_4^2 + a\lambda_3b_3^2, \\
B_4 &= b\lambda_3b_3^2 + b\lambda_4b_4^2, & B_5 &= 2a\lambda_3a_3b_3 + 2a\lambda_4a_4b_4, & B_6 &= 2b\lambda_3a_3b_3 + 2b\lambda_4a_4b_4, & B_7 &= -c_2, \\
C_1 &= a\lambda_6a_6^2 + a\lambda_5a_5^2, & C_2 &= b\lambda_6a_6^2 + b\lambda_5a_5^2, & C_3 &= a\lambda_6b_6^2 + a\lambda_5b_5^2, \\
C_4 &= b\lambda_5b_5^2 + b\lambda_6b_6^2, & C_5 &= 2a\lambda_5a_5b_5 + 2a\lambda_6a_6b_6, & C_6 &= 2b\lambda_5a_5b_5 + 2b\lambda_6a_6b_6, & C_7 &= -c_3.
\end{aligned}$$

In the above equations, the constants  $a, b, c_1, c_2$  and  $c_3$  are defined as

$$\begin{aligned}
a &= S_{xx} - S_{zz}, & b &= S_{yy} - S_{zz}, \\
c_1 &= \Delta\delta_{C'} - (\lambda_1 + \lambda_2)S_{zz}, & c_2 &= \Delta\delta_N - (\lambda_3 + \lambda_4)S_{zz}, & c_3 &= \Delta\delta_H - (\lambda_5 + \lambda_6)S_{zz},
\end{aligned}$$

where the constants  $\lambda_1$  and  $\lambda_2$  are defined by the magnitudes of the principal components  $^{13}\text{C}'$ -RCSA along  $\mathbf{v}_1$  and  $\mathbf{v}_2$ , the constants  $\lambda_3$  and  $\lambda_4$  are defined by the magnitudes of the principal components  $^{15}\text{N}$ -RCSA along  $\mathbf{v}_3$  and  $\mathbf{v}_4$ , and the constants  $\lambda_5$  and  $\lambda_6$  are defined by the magnitudes of the principal components  $^1\text{H}$ -RCSA along  $\mathbf{v}_5$  and  $\mathbf{v}_6$ , using Eq. (6) Eq. (7) of the main text [12].

We now make the following three observations (verified by expanding and substituting the values for  $A_i, B_i$ , and  $C_i$ , for  $1 \leq i \leq 6$ ), that will be useful in transforming Eq. (C.1), Eq. (C.2) and Eq. (C.3) into simpler forms.

$$\text{Observation 1. } A_1B_2 - A_2B_1 = A_1C_2 - A_2C_1 = B_1C_2 - B_2C_1 = 0. \quad (\text{C.4})$$

$$\text{Observation 2. } A_3B_4 - A_4B_3 = A_3C_4 - A_4C_3 = B_3C_4 - B_4C_3 = 0. \quad (\text{C.5})$$

$$\text{Observation 3. } A_5B_6 - A_6B_5 = A_5C_6 - A_6C_5 = B_5C_6 - B_6C_5 = 0. \quad (\text{C.6})$$

We use the symbol  $\times$  for scalar multiplication in our derivations unless otherwise specified. Now

$$\begin{aligned}
&\text{Eq. (C.1)} \times B_5 - \text{Eq. (C.2)} \times A_5 = 0 \\
&\Rightarrow (A_1B_5 - A_5B_1)x_{01}^2 + (A_2B_5 - A_5B_2)y_{01}^2 \\
&\quad + (A_3B_5 - A_5B_3)x_{02}^2 + (A_4B_5 - A_5B_4)y_{02}^2 + (A_7B_5 - A_5B_7) = 0, \quad (\text{C.7})
\end{aligned}$$

and

$$\begin{aligned}
& \text{Eq. (C.1)} \times C_5 - \text{Eq. (C.3)} \times A_5 = 0 \\
& \Rightarrow (A_1C_5 - A_5C_1)x_{01}^2 + (A_2C_5 - A_5C_2)y_{01}^2 \\
& \quad + (A_3C_5 - A_5C_3)x_{02}^2 + (A_4C_5 - A_5C_4)y_{04}^2 + (A_7C_5 - A_5C_7) = 0.
\end{aligned} \tag{C.8}$$

Further, we observe that

$$(A_3B_5 - A_5B_3)(A_4C_5 - A_5C_4) - (A_4B_5 - A_5B_4)(A_3C_5 - A_5C_3) = 0, \tag{C.9}$$

and

$$(A_1B_5 - A_5B_1)(A_2C_5 - A_5C_2) - (A_2B_5 - A_5B_2)(A_1C_5 - A_5C_1) = 0, \tag{C.10}$$

which can easily be shown by expanding, and then using Eq. (C.4), Eq. (C.5) and Eq. (C.6).

Now

$$\begin{aligned}
& \text{Eq. (C.7)} \times (A_4C_5 - A_5C_4) - \text{Eq. (C.8)} \times (A_4B_5 - A_5B_4) = 0 \\
& \Rightarrow E_1x_{01}^2 + E_2y_{01}^2 + E_3 = 0,
\end{aligned} \tag{C.11}$$

where

$$\begin{aligned}
E_1 &= (A_1B_5 - A_5B_1)(A_4C_5 - A_5C_4) - (A_1C_5 - A_5C_1)(A_4B_5 - A_5B_4) \\
E_2 &= (A_2B_5 - A_5B_2)(A_4C_5 - A_5C_4) - (A_2C_5 - A_5C_2)(A_4B_5 - A_5B_4) \\
E_3 &= (A_7B_5 - A_5B_7)(A_4C_5 - A_5C_4) - (A_7C_5 - A_5C_7)(A_4B_5 - A_5B_4).
\end{aligned}$$

Similarly,

$$\begin{aligned}
& \text{Eq. (C.7)} \times (A_2C_5 - A_5C_2) - \text{Eq. (C.8)} \times (A_2B_5 - A_5B_2) = 0 \\
& \Rightarrow F_1x_{02}^2 + F_2y_{02}^2 + F_3 = 0,
\end{aligned} \tag{C.12}$$

where

$$\begin{aligned}
F_1 &= (A_3B_5 - A_5B_3)(A_2C_5 - A_5C_2) - (A_3C_5 - A_5C_3)(A_2B_5 - A_5B_2) \\
F_2 &= (A_4B_5 - A_5B_4)(A_2C_5 - A_5C_2) - (A_4C_5 - A_5C_4)(A_2B_5 - A_5B_2) \\
F_3 &= (A_7B_5 - A_5B_7)(A_2C_5 - A_5C_2) - (A_7C_5 - A_5C_7)(A_2B_5 - A_5B_2).
\end{aligned}$$

Now

$$\begin{aligned}
& \text{Eq. (C.1)} \times B_1 - \text{Eq. (C.2)} \times A_1 = 0 \\
& \Rightarrow (A_3B_1 - A_1B_3)x_{02}^2 + (A_4B_1 - A_1B_4)y_{02}^2 \\
& \quad + (A_5B_1 - A_1B_5)x_{01}x_{02} + (A_6B_1 - A_1B_6)y_{01}y_{02} + (A_7B_1 - A_1B_7) = 0,
\end{aligned} \tag{C.13}$$

and

$$\begin{aligned}
& \text{Eq. (C.1)} \times C_1 - \text{Eq. (C.3)} \times A_1 = 0 \\
& \Rightarrow (A_3C_1 - A_1C_3)x_{02}^2 + (A_4C_1 - A_1C_4)y_{02}^2 \\
& \quad + (A_5C_1 - A_1C_5)x_{01}x_{02} + (A_6C_1 - A_1C_6)y_{01}y_{02} + (A_7C_1 - A_1C_7) = 0.
\end{aligned} \tag{C.14}$$



Further we observe that

$$(A_3B_1 - A_1B_3)(A_4C_1 - A_1C_4) - (A_4B_1 - A_1B_4)(A_3C_1 - A_1C_3) = 0, \quad (\text{C.15})$$

which can easily be shown by expanding, and then using Eq. (C.4), Eq. (C.5) and Eq. (C.6).

Now

$$\begin{aligned} & \text{Eq. (C.13)} \times (A_4C_1 - A_1C_4) - \text{Eq. (C.14)} \times (A_4B_1 - A_1B_4) = 0 \\ \Rightarrow & G_1x_{01}x_{02} + G_2y_{01}y_{02} + G_3 = 0, \end{aligned} \quad (\text{C.16})$$

where

$$\begin{aligned} G_1 &= (A_5B_1 - A_1B_5)(A_4C_1 - A_1C_4) - (A_5C_1 - A_1C_5)(A_4B_1 - A_1B_4) \\ G_2 &= (A_6B_1 - A_1B_6)(A_4C_1 - A_1C_4) - (A_6C_1 - A_1C_6)(A_4B_1 - A_1B_4) \\ G_3 &= (A_7B_1 - A_1B_7)(A_4C_1 - A_1C_4) - (A_7C_1 - A_1C_7)(A_4B_1 - A_1B_4). \end{aligned}$$

We make one more observation:

$$E_1F_2 - E_2F_1 = E_1G_2 - E_2G_1 = F_1G_2 - F_2G_1 = 0, \quad (\text{C.17})$$

from which we obtain

$$\frac{E_1}{F_1} = \frac{E_2}{F_2}, \quad \frac{E_1}{G_1} = \frac{E_2}{G_2}, \quad \frac{F_1}{G_1} = \frac{F_2}{G_2}. \quad (\text{C.18})$$

Using Eq. (C.18) in Eq. (C.12) and Eq. (C.16), we obtain

$$E_1x_{02}^2 + E_2y_{02}^2 + E_4 = 0, \quad (\text{C.19})$$

and

$$E_1x_{01}x_{02} + E_2y_{01}y_{02} + E_5 = 0, \quad (\text{C.20})$$

where

$$E_4 = \frac{E_1F_3}{F_1}, \quad E_5 = \frac{E_1G_3}{G_1}.$$

We now want to solve Eq. (C.11), Eq. (C.19) and Eq. (C.20). Since there are four variables, we need one more independent equation, which we get from the fact that  $\mathbf{v}_{01} \perp \mathbf{v}_{02}$ . In fact, even when the unit vectors  $\mathbf{v}_{01}$  and  $\mathbf{v}_{02}$  are chosen not to be perpendicular to each other, our derivation works. We let  $\cos(\angle \mathbf{v}_{01}, \mathbf{v}_{02}) = \tau$ , where  $\tau$  is a constant.

$$x_{01}x_{02} + y_{01}y_{02} + z_{01}z_{02} = \cos(\angle \mathbf{v}_{01}, \mathbf{v}_{02}) = \tau \quad (\text{C.21})$$

$$\begin{aligned} \Rightarrow & z_{01}^2 z_{02}^2 = (\tau - x_{01}x_{02} - y_{01}y_{02})^2 \\ \Rightarrow & (1 - x_{01}^2 - y_{01}^2)(1 - x_{02}^2 - y_{02}^2) = (\tau - x_{01}x_{02} - y_{01}y_{02})^2 \\ \Rightarrow & x_{01}^2 + y_{01}^2 + x_{02}^2 + y_{02}^2 \\ & - x_{01}^2 y_{02}^2 - x_{02}^2 y_{01}^2 + 2x_{01}x_{02}y_{01}y_{02} - 2\tau(x_{01}x_{02} + y_{01}y_{02}) - (1 - \tau^2) = 0. \end{aligned} \quad (\text{C.22})$$

Using Eq. (C.20) in Eq. (C.22), we eliminate the terms containing factor  $x_{01}x_{02}$  as follows:

$$\begin{aligned}
\text{Eq. (C.22)} &\Rightarrow 2\tau(x_{01}x_{02} + y_{01}y_{02}) \\
&= x_{01}^2 + y_{01}^2 + x_{02}^2 + y_{02}^2 - x_{02}^2y_{01}^2 - x_{01}^2y_{02}^2 + 2x_{01}x_{02}y_{01}y_{02} - (1 - \tau^2) \\
&\Rightarrow 2\tau \left( \frac{-E_5 - E_2y_{01}y_{02}}{E_1} + y_{01}y_{02} \right) \\
&= x_{01}^2 + y_{01}^2 + x_{02}^2 + y_{02}^2 - x_{02}^2y_{01}^2 - x_{01}^2y_{02}^2 + 2 \left( \frac{-E_5 - E_2y_{01}y_{02}}{E_1} \right) y_{01}y_{02} - (1 - \tau^2) \\
&\Rightarrow \frac{-2\tau E_5}{E_1} + 2\tau \left( 1 - \frac{E_2}{E_1} \right) y_{01}y_{02} \\
&= x_{01}^2 + y_{01}^2 + x_{02}^2 + y_{02}^2 - x_{02}^2y_{01}^2 - x_{01}^2y_{02}^2 + 2 \left( \frac{-E_5 - E_2y_{01}y_{02}}{E_1} \right) y_{01}y_{02} - (1 - \tau^2) \\
&\Rightarrow \frac{-2\tau E_5}{a} + \left( 2\tau \left( 1 - \frac{E_2}{E_1} \right) + \frac{2E_5}{E_1} \right) y_{01}y_{02} \\
&= x_{01}^2 + y_{01}^2 + x_{02}^2 + y_{02}^2 - x_{02}^2y_{01}^2 - x_{01}^2y_{02}^2 - \frac{2E_2}{E_1} y_{01}^2 y_{02}^2 - (1 - \tau^2) \\
&\Rightarrow x_{01}^2 + y_{01}^2 + x_{02}^2 + y_{02}^2 - x_{02}^2y_{01}^2 - x_{01}^2y_{02}^2 \\
&\quad - \frac{2E_2}{E_1} y_{01}^2 y_{02}^2 - \left( 2\tau \left( 1 - \frac{E_2}{E_1} \right) + \frac{2E_5}{E_1} \right) y_{01}y_{02} - (1 - \tau^2) + \frac{2\tau E_5}{E_1} = 0 \\
&\Rightarrow x_{01}^2 + y_{01}^2 + x_{02}^2 + y_{02}^2 - x_{02}^2y_{01}^2 - x_{01}^2y_{02}^2 + H_1y_{01}^2y_{02}^2 + H_2y_{01}y_{02} + H_3 = 0, \quad (\text{C.23})
\end{aligned}$$

where

$$H_1 = -\frac{2E_2}{E_1}, \quad H_2 = -2\tau \left( 1 - \frac{E_2}{E_1} \right) - \frac{2E_5}{E_1}, \quad H_3 = -1 + \tau^2 + \frac{2\tau E_5}{E_1}.$$

Rearranging the terms in Eq. (C.20) and squaring, we can express  $y_{01}y_{02}$  as a linear combination of terms involving  $x_{01}^2x_{02}^2$  and  $y_{01}^2y_{02}^2$  as follows:

$$\begin{aligned}
&\text{Eq. (C.20)} \\
&\Rightarrow E_1x_{01}x_{02} = -E_5 - E_2y_{01}y_{02} \\
&\Rightarrow E_1^2x_{01}^2x_{02}^2 = (-E_5 - E_2y_{01}y_{02})^2 = E_5^2 + 2E_2E_5y_{01}y_{02} + E_2^2y_{01}^2y_{02}^2 \\
&\Rightarrow y_{01}y_{02} = \frac{E_1^2x_{01}^2x_{02}^2 - E_2^2y_{01}^2y_{02}^2 - E_5^2}{2E_2E_5} = I_1x_{01}^2x_{02}^2 + I_2y_{01}^2y_{02}^2 + I_3, \quad (\text{C.24})
\end{aligned}$$

where

$$I_1 = \frac{E_1^2}{2E_2E_5}, \quad I_2 = \frac{-E_2}{2E_5}, \quad I_3 = \frac{-E_5}{2E_2}.$$

Substituting Eq. (C.24) in Eq. (C.23) we obtain

$$x_{01}^2 + y_{01}^2 + x_{02}^2 + y_{02}^2 - x_{02}^2y_{01}^2 - x_{01}^2y_{02}^2 + J_1x_{01}^2x_{02}^2 + J_2y_{01}^2y_{02}^2 + J_3 = 0, \quad (\text{C.25})$$

where

$$\begin{aligned}
J_1 = H_2I_1 &= \frac{E_1(-\tau E_1 + \tau E_2 - E_5)}{E_2E_5} \\
J_2 = H_1 + H_2I_2 &= \frac{E_2(\tau E_1 - \tau E_2 - E_5)}{E_1E_5} \\
J_3 = H_3 + H_2I_3 &= \frac{\tau^2 E_1 E_2 + \tau E_1 E_5 + \tau E_2 E_5 + E_5^2 - E_1 E_2}{E_1 E_2}.
\end{aligned}$$

We now solve the system of equations Eq. (C.11), Eq. (C.19), Eq. (C.20) and Eq. (C.25) simultaneously.

We will now transform Eq. (C.20) so that the degree of each of  $x_{01}, y_{01}, x_{02}$  and  $y_{02}$  is even.

Eq. (C.20)

$$\begin{aligned}
&\Rightarrow E_1 x_{01} x_{02} = -E_5 - E_2 y_{01} y_{02} \\
&\Rightarrow E_1^2 x_{01}^2 x_{02}^2 = E_2^2 y_{01}^2 y_{02}^2 + 2E_2 E_5 y_{01} y_{02} + E_5^2 \quad (\text{squaring both sides}) \\
&\Rightarrow 2E_2 E_5 y_{01} y_{02} = E_1^2 x_{01}^2 x_{02}^2 - E_2^2 y_{01}^2 y_{02}^2 - E_5^2 \\
&\Rightarrow 4E_2^2 E_5^2 y_{01}^2 y_{02}^2 = E_1^4 x_{01}^4 x_{02}^4 + E_2^4 y_{01}^4 y_{02}^4 \\
&\quad - 2E_1^2 E_2^2 x_{01}^2 x_{02}^2 y_{01}^2 y_{02}^2 - 2E_1^2 E_5^2 x_{01}^2 x_{02}^2 + 2E_2^2 E_5^2 y_{01}^2 y_{02}^2 + E_5^4 \quad (\text{squaring both sides}) \\
&\Rightarrow E_1^4 x_{01}^4 x_{02}^4 + E_2^4 y_{01}^4 y_{02}^4 - 2E_1^2 E_2^2 x_{01}^2 x_{02}^2 y_{01}^2 y_{02}^2 - 2E_1^2 E_5^2 x_{01}^2 x_{02}^2 - 2E_2^2 E_5^2 y_{01}^2 y_{02}^2 + E_5^4 = 0. \quad (\text{C.26})
\end{aligned}$$

Now we solve Eq. (C.11), Eq. (C.19), Eq. (C.25) and Eq. (C.26) simultaneously. Let

$$u = x_{01}^2 \quad (\text{C.27})$$

and

$$v = x_{02}^2. \quad (\text{C.28})$$

Then,

$$\text{Eq. (C.11)} \Rightarrow y_{01}^2 = \frac{-E_3 - E_1 u}{E_2} \quad (\text{C.29})$$

and

$$\text{Eq. (C.19)} \Rightarrow y_{02}^2 = \frac{-E_4 - E_1 v}{E_2}. \quad (\text{C.30})$$

Using Eq. (C.29) and Eq. (C.30) in Eq. (C.25), we obtain

$$K_0 + K_1 u + K_2 v + K_3 uv = 0, \quad (\text{C.31})$$

where

$$\begin{aligned}
K_0 &= E_3 E_4 J_2 - E_2 E_4 - E_2 E_3 + E_2^2 J_3 \\
&= \frac{E_2}{E_1 E_5} (-E_3 E_4 E_5 + E_3 E_4 \tau E_1 - E_3 E_4 E_2 \tau - E_4 E_1 E_5 - E_3 E_1 E_5 - E_2 E_1 E_5) \\
&\quad + \frac{E_2}{E_1 E_5} (\tau^2 E_1 E_2 E_5 + \tau E_2 E_5^2 + \tau E_1 E_5^2 + E_5^3) \\
K_1 &= -E_1 E_2 + E_2 E_4 + E_2^2 + E_1 E_4 J_2 = \frac{E_2(-E_1 E_5 + E_2 E_5 + \tau E_1 E_4 - \tau E_2 E_4)}{E_5} \\
K_2 &= E_2 E_3 + E_2^2 - E_1 E_2 + E_1 E_3 J_2 = \frac{E_2(E_2 E_5 - E_1 E_5 + \tau E_1 E_3 - \tau E_2 E_3)}{E_5} \\
K_3 &= E_1^2 J_2 + E_2^2 J_1 + 2E_1 E_2 = 0.
\end{aligned}$$

Noting that  $K_3 = 0$ , we can write Eq. (C.31) as

$$\begin{aligned}
&K_0 + K_1 u + K_2 v = 0 \\
\Rightarrow v &= \frac{-K_0 - K_1 u}{K_2}. \quad (\text{C.32})
\end{aligned}$$

Using Eq. (C.29) and Eq. (C.30) in Eq. (C.26), we obtain

$$L_1 u^2 + L_2 uv + L_3 v^2 + L_4 u + L_5 v + L_6 = 0, \quad (\text{C.33})$$

where

$$\begin{aligned} L_1 &= E_1^2 E_4^2 \\ L_2 &= 2E_1^2 E_3 E_4 - 4E_1^2 E_5^2 \\ L_3 &= E_1^2 E_3^2 \\ L_4 &= 2E_1 E_3 E_4^2 - 2E_1 E_4 E_5^2 \\ L_5 &= 2E_1 E_3^2 E_4 - 2E_1 E_3 E_5^2 \\ L_6 &= E_3^2 E_4^2 - 2E_3 E_4 E_5^2 + E_5^4. \end{aligned}$$

Using Eq. (C.32) in Eq. (C.33) we obtain

$$M_0 + M_1 u + M_2 u^2 = 0, \quad (\text{C.34})$$

where

$$\begin{aligned} M_0 &= -\frac{L_5 K_0}{K_2} + L_6 + \frac{L_3 K_0^2}{K_2^2} \\ M_1 &= -\frac{L_5 K_1}{K_2} - \frac{L_2 K_0}{K_2} + L_4 + \frac{2L_3 K_0 K_1}{K_2^2} \\ M_2 &= -\frac{L_2 K_1}{K_2} + L_1 + \frac{L_3 K_1^2}{K_2^2}. \end{aligned}$$

Eq. (C.34) is a quadratic equation which can be solved in closed form from which the orientations of the peptide plane can be computed as detailed below.

**Constructing the oriented peptide planes.** We give the details of how to compute the different orientations of a peptide plane from the solutions of Eq. (C.34). The construction exploits the symmetry present in the RDC equation.

Solving Eq. (C.34) we get at most two real solutions. Let  $\{u_1, u_2\}$  denote the set of these two solutions. For each  $u_i$ ,  $1 \leq i \leq 2$ , we can compute a unique  $v_i$  using Eq. (C.32). For a pair  $(u_i, v_i)$ ,  $1 \leq i \leq 2$ , using Eq. (C.27), Eq. (C.28), Eq. (C.29) and Eq. (C.30), we compute sixteen  $(x_{01}, y_{01}, x_{02}, y_{02})$  tuples, where any two tuples differ by the sign of at least one element, but the absolute values of the elements are the same in all of the sixteen tuples. Out of these sixteen tuples only four will satisfy Eq. (C.20). This is because,  $(x_{01}, y_{01}, x_{02}, y_{02})$  satisfies Eq. (C.20) if and only if  $(-x_{01}, y_{01}, -x_{02}, y_{02})$ ,  $(x_{01}, -y_{01}, x_{02}, -y_{02})$  and  $(-x_{01}, -y_{01}, -x_{02}, -y_{02})$  satisfy Eq. (C.20). Since  $\mathbf{v}_{01} = (x_{01}, y_{01}, z_{01})^T$  is a unit vector, we get two solutions for  $z_{01}$  which we denote by  $\{z_{01}, -z_{01}\}$ . Similarly, since  $\mathbf{v}_{02} = (x_{02}, y_{02}, z_{02})^T$  is a unit vector we get two solutions for  $z_{02}$  which we denote by  $\{z_{02}, -z_{02}\}$ . Therefore, there are four possibilities, viz.  $\{(z_{01}, z_{02}), (-z_{01}, z_{02}), (z_{01}, -z_{02}), (-z_{01}, -z_{02})\}$ . From Eq. (C.21) we observe that the product  $z_{01} z_{02}$  must have the same sign which means that only two out of the above four pairs would be considered. Without loss of generality, we let  $(z_{01}, z_{02})$  and  $(-z_{01}, -z_{02})$  be the two representative pairs. Therefore, we have eight 6-tuples which can be written as the following product using set theoretic notation:

$$\begin{aligned} &\{(x_{01}, y_{01}, x_{02}, y_{02}), (-x_{01}, y_{01}, -x_{02}, y_{02}), (x_{01}, -y_{01}, x_{02}, -y_{02}), (-x_{01}, -y_{01}, -x_{02}, -y_{02})\} \\ &\quad \times \{(z_{01}, z_{02}), (-z_{01}, -z_{02})\}, \end{aligned}$$

Table 1: Sixteen peptide plane orientations from single set of  $C^\alpha$ - $C'$ ,  $C'$ - $N$ ,  $N$ - $H^N$  RDCs, measured in one alignment medium, for the peptide plane between the residues Ala28 and Lys29 of ubiquitin.

Solution Number	Unit Vector		Index to Figure 1
	$\mathbf{v}_{N-H^N} = (x_1, y_1, z_1)$	$\mathbf{v}_{C'-N} = (x_2, y_2, z_2)$	
1	(0.19598, 0.84846, 0.49163)	(0.78319, 0.55612, -0.27810)	A
3	(0.19598, 0.84846, -0.49163)	(0.78319, 0.55612, 0.27810)	B
3	(-0.19598, 0.84846, 0.49163)	(-0.78319, 0.55612, -0.27810)	C
4	(-0.19598, 0.84846, -0.49163)	(-0.78319, 0.55612, 0.27810)	D
5	(0.19598, -0.84846, 0.49163)	(0.78319, -0.55612, -0.27810)	E
6	(0.19598, -0.84846, -0.49163)	(0.78319, -0.55612, 0.27810)	F
7	(-0.19598, -0.84846, 0.49163)	(-0.78319, -0.55612, -0.27810)	G
8	(-0.19598, -0.84846, -0.49163)	(-0.78319, -0.55612, 0.27810)	H
9	(0.31133, -0.58536, 0.74862)	(0.80635, 0.26825, 0.52711)	I
10	(0.31133, -0.58536, -0.74862)	(0.80635, 0.26825, -0.52711)	J
11	(-0.31133, -0.58536, 0.74862)	(-0.80635, 0.26825, 0.52711)	K
12	(-0.31133, -0.58536, -0.74862)	(-0.80635, 0.26825, -0.52711)	L
13	(0.31133, 0.58536, 0.74862)	(0.80635, -0.26825, 0.52711)	M
14	(0.31133, 0.58536, -0.74862)	(0.80635, -0.26825, -0.52711)	N
15	(-0.31133, 0.58536, 0.74862)	(-0.80635, -0.26825, 0.52711)	O
16	(-0.31133, 0.58536, -0.74862)	(-0.80635, -0.26825, -0.52711)	P

Experimental RDCdata for ubiquitin (PDB id: 1D3Z) is taken from the BioMagResBank (BMRB) [16]. The right-most column indicates the indexes of the corresponding sub-figure in Figure 1, that uses the two vectors in columns 2 and 3 to define the peptide plane.

where the symbol  $\times$  is used for the Cartesian product. We note that each 6-tuple in this set represents a pair of unit vectors which forms a plane. Thus, for a pair  $(u_i, v_i)$  we have at most 8 possible peptide planes. Since there are at most two  $(u_i, v_i)$  pairs when  $1 \leq i \leq 2$ , we have at most  $8 \times 2 = 16$  possible peptide plane orientations.

This completes the proof. □

## D The Sixteen peptide Plane Orientations

The algorithm 3PLANAR computed the peptide plane orientations for the peptide plane defined by the residues Ala28 and Lys29 of ubiquitin, from  $C^\alpha$ - $C'$ ,  $C'$ - $N$  and  $N$ - $H^N$  RDCs in one alignment medium. It first computed the peptide plane orientations represented by pairs of  $\mathbf{v}_{N-H^N}$  and  $\mathbf{v}_{C'-N}$  unit vectors. The 16 possible pairs of unit vectors are listed in Table 1. Then, each pair of  $\mathbf{v}_{N-H^N}$  and  $\mathbf{v}_{C'-N}$  unit vectors were used to compute the corresponding oriented peptide plane. In Figure 1, each of the peptide planes that corresponds to a pair of  $\mathbf{v}_{N-H^N}$  and  $\mathbf{v}_{C'-N}$  unit vectors (see the right-most column of Table 1) is shown.

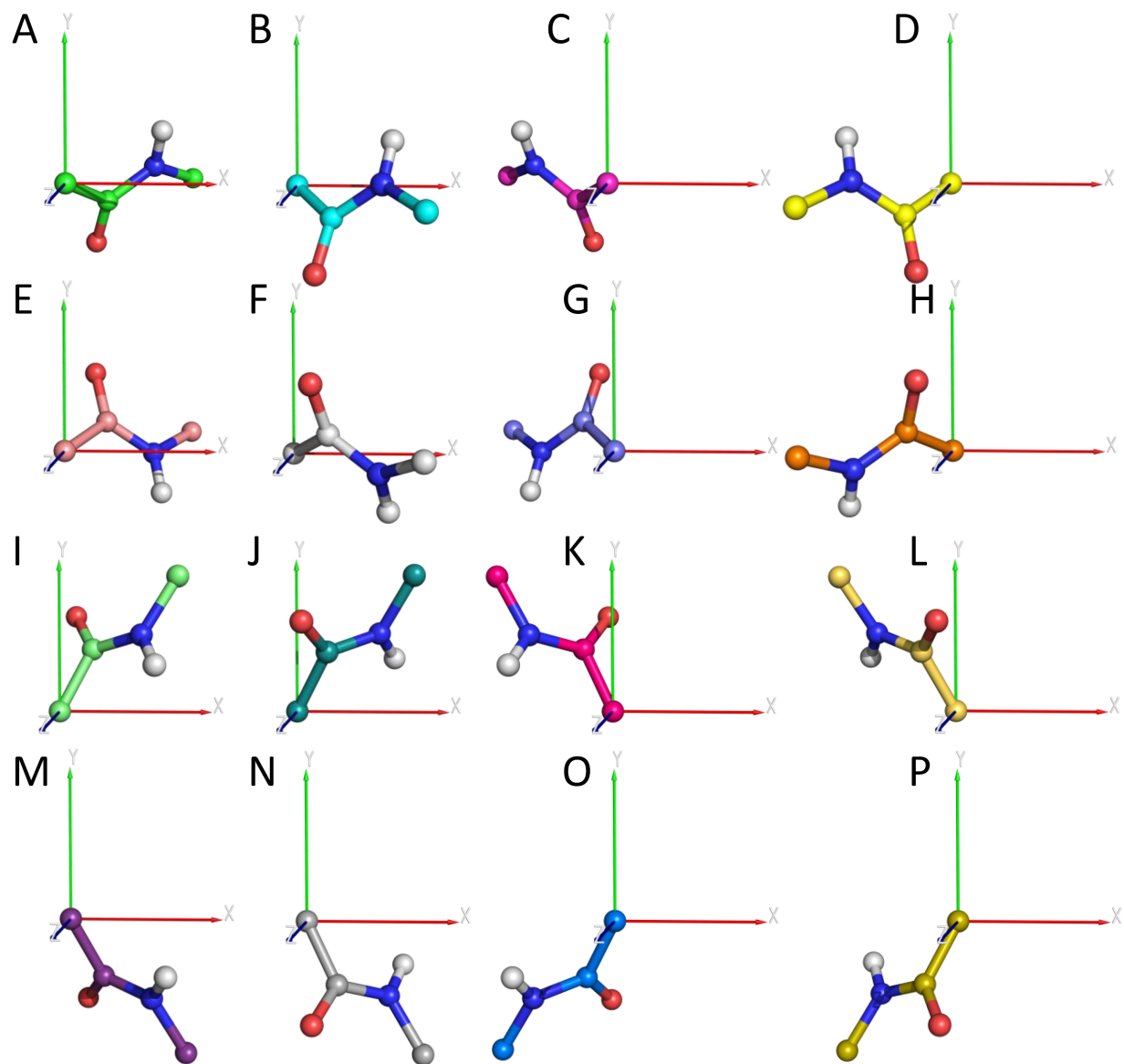


Figure 1: Sixteen peptide plane orientations from single set of  $C^\alpha$ - $C'$ ,  $C'$ -N, N- $H^N$  RDCs, measured in one alignment medium, for the peptide plane between the residues Ala28 and Lys29 of ubiquitin. The peptide planes were constructed using the unit vectors given in the Table 1.

## References

- [1] I. Ayala, R. Sounier, N. Usé, P. Gans, and J. Boisbouvier. An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. *Journal of Biomolecular NMR*, 43(2):111–119, 2009.
- [2] C. A. Fowler, F. Tian, H. M. Al-Hashimi, and J. H. Prestegard. Rapid determination of protein folds using residual dipolar couplings. *Journal of Molecular Biology*, 304(3):447–460, 2000.
- [3] K. H. Gardner and L. E. Kay. Production and incorporation of  $^{15}\text{N}$ ,  $^{13}\text{C}$ ,  $^2\text{H}$  ( $^1\text{H}$ - $\delta 1$  methyl) isoleucine into proteins for multidimensional NMR studies. *Journal of the American Chemical Society*, 119(32):7599–7600, 1997.
- [4] N. K. Goto, K. H. Gardner, G. A. Mueller, R. C. Willis, and L. E. Kay. A robust and cost-effective method for the production of Val, Leu, Ile ( $\delta 1$ ) methyl-protonated  $^{15}\text{N}$ -,  $^{13}\text{C}$ -,  $^2\text{H}$ -labeled proteins. *Journal of Biomolecular NMR*, 13(4):369–374, 1999.
- [5] R. L. Isaacson, P. J. Simpson, M. Liu, E. Cota, X. Zhang, P. Freemont, and S. Matthews. A new labeling method for methyl transverse relaxation-optimized spectroscopy NMR spectra of alanine residues. *Journal of the American Chemical Society*, 129(50):15428–15429, 2007.
- [6] Y. Liu and J. H. Prestegard. A device for the measurement of residual chemical shift anisotropy and residual dipolar coupling in soluble and membrane-associated proteins. *Journal of Biomolecular NMR*, 47(4):249–258, 2010.
- [7] J. A. Losonczi, M. Andrec, M. W. F. Fischer, and J. H. Prestegard. Order matrix analysis of residual dipolar couplings using singular value decomposition. *Journal of Magnetic Resonance*, 138:334–342, 1999.
- [8] S. C. Lovell, J. M. Word, J. S. Richardson, and D. C. Richardson. The Penultimate Rotamer Library. *Proteins: Structure, Function, and Genetics*, 40:389–408, 2000.
- [9] S. Potluri, A. K. Yan, J. J. Chou, B. R. Donald, and C. Bailey-Kellogg. Structure Determination of Symmetric Homo-oligomers by a Complete Search of Symmetry Configuration Space using NMR Restraints and van der Waals Packing. *Proteins: Structure, Function and Bioinformatics*, 65(1):203–219, 2006.
- [10] S. Potluri, A. K. Yan, B. R. Donald, and C. Bailey-Kellogg. A complete algorithm to resolve ambiguity for intersubunit NOE assignment in structure determination of symmetric homo-oligomers. *Protein Science*, 16(1):69–81, 2007.
- [11] R. Sprangers, A. Velyvis, and L. E. Kay. Solution NMR of supramolecular complexes: providing new insights into function. *Nat Methods*, 4(9):697–703, 2007.
- [12] C. Tripathy, A. K. Yan, P. Zhou, and B. R. Donald. Extracting Structural Information from Residual Chemical Shift Anisotropy: Analytic Solutions for Peptide Plane Orientations and Applications to Determine Protein Structure. In *Proceedings of the 17th Annual International Conference on Research in Computational Molecular Biology (RECOMB)*, Beijing, China, 2013.
- [13] C. Tripathy, J. Zeng, P. Zhou, and B. R. Donald. Protein loop closure using orientational restraints from NMR data. In V. Bafna and S. Sahinalp, editors, *Proceedings of the 15th Annual International Conference on Research in Computational Molecular Biology (RECOMB)*,

- Vancouver, BC Canada, volume 6577 of *Lecture Notes in Computer Science*, pages 483–498. Springer Berlin / Heidelberg, 2011.
- [14] C. Tripathy, J. Zeng, P. Zhou, and B. R. Donald. Protein loop closure using orientational restraints from NMR data. *Proteins: Structure, Function, and Bioinformatics*, 80(2):433–453, 2012.
  - [15] V. Tugarinov, V. Kanelis, and L. E. Kay. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nature Protocols*, 1(2):749–754, 2006.
  - [16] E. L. Ulrich, H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C. F. Schulte, D. E. Tolmie, R. K. Wenger, H. Yao, and J. L. Markley. BioMagResBank. *Nucleic Acids Research*, 36(Database issue):D402–D408, 2008.
  - [17] L. Wang and B. R. Donald. Exact solutions for internuclear vectors and backbone dihedral angles from NH residual dipolar couplings in two media, and their application in a systematic search algorithm for determining protein backbone structure. *Journal of Biomolecular NMR*, 29(3):223–242, 2004.
  - [18] L. Yao, A. Grishaev, G. Cornilescu, and A. Bax. Site-Specific Backbone Amide  $^{15}\text{N}$  Chemical Shift Anisotropy Tensors in a Small Protein from Liquid Crystal and Cross-Correlated Relaxation Measurements. *Journal of the American Chemical Society*, 132(12):4295–4309, 2010.
  - [19] A. Yershova, C. Tripathy, P. Zhou, and B. R. Donald. Algorithms and Analytic Solutions using Sparse Residual Dipolar Couplings for High-Resolution Automated Protein Backbone Structure Determination by NMR. *The Ninth International Workshop on the Algorithmic Foundations of Robotics (WAFR)*, 68:355–372, 2010.
  - [20] J. Zeng, J. Boyles, C. Tripathy, L. Wang, A. Yan, P. Zhou, and B. R. Donald. High-resolution protein structure determination starting with a global fold calculated from exact solutions to the RDC equations. *Journal of Biomolecular NMR*, 45(3):265–281, 2009.