Supporting Information for "Protein Loop Closure using Orientational Restraints from NMR Data"[¶]

Chittaranjan Tripathy[1], Jianyang Zeng[1], Pei Zhou[2], and Bruce Randall Donald[1,2,*]

[1]Department of Computer Science, Duke University, Durham, NC 27708
[2]Department of Biochemistry, Duke University Medical Center, Durham, NC 27710

Below is the supporting information (SI) for the following paper [14]:

– C. Tripathy, J. Zeng, P. Zhou, and B. R. Donald. Protein loop closure using orientational restraints from NMR data. In *Proceedings of the 15th Annual International Conference on Research in Computational Molecular Biology (RECOMB), Vancouver, BC Canada*, 2011.

Appendix

In **Appendix A**, we provide a proof of Proposition 2, and in **Appendix B** we prove Proposition 4. In **Appendix C**, we give a proof for computing all possible orientations of a peptide plane from a ϕ-defining RDC in one alignment medium and a ψ-defining RDC in a second alignment medium. **Appendix D** describes the experimental NMR data collection procedure for the protein FF2. Finally, in **Appendix E**, we describe the procedure used to simulate the RDCs for the loops studied in [4, 7, 11].

A Proof of Proposition 2

We use the notation developed in Section 2.3. For convenience, we first restate the proposition, and then provide the proof.

Proposition 2. *Given the diagonalized alignment tensor components S_{xx} and S_{yy}, the peptide plane P_i, the dihedral ϕ_i, and a ψ-defining RDC r for the corresponding internuclear vector on peptide plane P_{i+1}, there exist at most 4 possible values of the dihedral angle ψ_i that satisfy the RDC r. The possible values of ψ_i can be computed exactly and in closed form by solving a quartic equation.*

Proof. Let the unit vector $\mathbf{v}_0 = (0, 0, 1)^T$ represent the N-H^N bond vector of residue i in the local coordinate frame defined on the peptide plane P_i. Let $\mathbf{v}_1 = (x, y, z)^T$ denote the internuclear vector for the ψ-defining RDC for residue i in the principal order frame. Note that the internuclear vector for a ψ-defining RDC has at least one nucleus that belongs residue $i + 1$. The forward kinematics relation between $\mathbf{v}_0$ and $\mathbf{v}_1$ can be written as follows:

$$\mathbf{v}_1 \;=\; \mathbf{R}_{i,\mathtt{POF}}\, \mathbf{R}_l\, \mathbf{R}_z(\phi_i)\, \mathbf{R}_m\, \mathbf{R}_z(\psi_i)\, \mathbf{R}_r\, \mathbf{v}_0. \tag{A.1}$$

[¶]This work is supported by the following grant from National Institutes of Health: R01 GM-65982 to B.R.D. and R01 GM-079376 to P.Z.

[*] *Corresponding author*: Bruce Randall Donald, ✉ brd+recomb11@cs.duke.edu, ☎ 919-660-6583, 🖷 919-660-6519

Here $\mathbf{R}_l$, $\mathbf{R}_m$ and $\mathbf{R}_r$ are constant rotation matrices. $\mathbf{R}_z(\phi_i)$ is the rotation about the z-axis by ϕ_i, and is a constant rotation matrix since ϕ_i is known (already computed before computing ψ_i by using Proposition 1 in main text [14]). $\mathbf{R}_z(\psi_i)$ is the rotation about the z-axis by ψ_i.

Let c and s denote $\cos\psi_i$ and $\sin\psi_i$, respectively. Using this and expanding Eq. (A.1) we have

$$x = A_0 + A_1 c + A_2 s, \quad y = B_0 + B_1 c + B_2 s, \quad z = C_0 + C_1 c + C_2 s, \tag{A.2}$$

in which A_i, B_i, C_i for $0 \leq i \leq 2$ are constants. Using Eq. (A.2) in the reduced RDC equation (Eq. (5) in [14]) and simplifying we obtain

$$K_0 + K_1 c + K_2 s + K_3 cs + K_4 c^2 + K_5 s^2 \quad = \quad 0, \tag{A.3}$$

in which K_i, $0 \leq i \leq 5$ are constants. Using half-angle substitutions

$$u = \tan(\frac{\psi_i}{2}),\ c = \frac{1-u^2}{1+u^2},\ \text{and}\ s = \frac{2u}{1+u^2} \tag{A.4}$$

in Eq. (A.3) we have

$$L_0 + L_1 u + L_2 u^2 + L_3 u^3 + L_4 u^4 \quad = \quad 0, \tag{A.5}$$

in which L_i, $0 \leq i \leq 4$ are constants.

Eq. (A.5) is a quartic equation which can be solved exactly and in closed form. Let $\{u_1, u_2, u_3, u_4\}$ denote the set of (at most) four real solutions of Eq. (A.5). For each u_i, the corresponding ψ_i value can be computed by using Eq. (A.4). □

B Proof of Proposition 4

The amino acid residue glycine (Gly) (Figure 1) has two H^α atoms which we denote by H^{α_2} and H^{α_3}, respectively. The C^α-H^α RDC measured for Gly is the sum of the RDCs for these two bond vectors. Here we show that given the C^α-H^α RDC for a Gly residue we can compute all possible solutions for the dihedral ϕ.

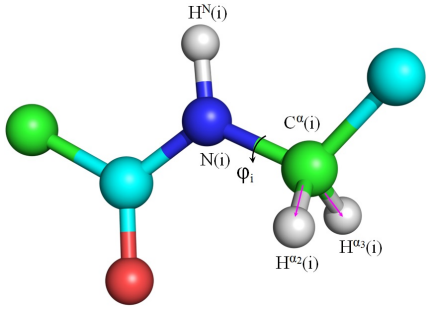

Figure 1: The amino acid residue glycine. The two H^α atoms are denote by H^{α_2} and H^{α_3}, respectively. The C^α-H^α RDC is the sum of RDCs measured on the bond vectors C^α-H^{α_2} and C^α-H^{α_3}.

Proposition 4. *Given the diagonalized alignment tensor components S_{xx} and S_{yy}, the peptide plane P_i, and the C^α-H^α RDC r for residue i which is a glycine, there exist at most 4 possible values of the dihedral angle ϕ_i that satisfy the C^α-H^α RDC r. The possible values of ϕ_i can be computed exactly and in closed form by solving a quartic equation.*

Proof. For brevity, in our derivation we will state the constants without giving explicit formulas to compute them.

Let the unit vector $\mathbf{v}_0 = (0,0,1)^T$ represent the N-H^N bond vector of residue i in the local coordinate frame defined on the peptide plane P_i. Let $\mathbf{v}_1 = (x_1, y_1, z_1)^T$ and $\mathbf{v}_2 = (x_2, y_2, z_2)^T$ be the unit vectors defined in the POF to represent C^α-H^{α_2} and C^α-H^{α_3}, respectively. We can write the forward kinematics relations between $\mathbf{v}_0$ and $\mathbf{v}_1$, and between $\mathbf{v}_0$ and $\mathbf{v}_2$ as follows:

$$\mathbf{v}_1 = \mathbf{R}_{i,\texttt{POF}}\, \mathbf{R}_l\, \mathbf{R}_z(\phi_i)\, \mathbf{R}_r\, \mathbf{v}_0 \tag{B.1}$$

$$\mathbf{v}_2 = \mathbf{R}_{i,\texttt{POF}}\, \mathbf{R}_l\, \mathbf{R}_z(\phi_i)\, \mathbf{R}'_r\, \mathbf{v}_0. \tag{B.2}$$

Here $\mathbf{R}_l$, $\mathbf{R}_r$ and $\mathbf{R}'_r$ are constant rotation matrices. $\mathbf{R}_z(\phi_i)$ is the rotation about the z-axis by ϕ_i.

Let c and s denote $\cos\phi_i$ and $\sin\phi_i$, respectively. Using this while expanding Eq. (B.1) and Eq. (B.2) we have

$$x_1 = A_{10} + A_{11}c + A_{12}s, \quad y_1 = B_{10} + B_{11}c + B_{12}s, \quad z_1 = C_{10} + C_{11}c + C_{12}s \tag{B.3}$$

$$x_2 = A_{20} + A_{21}c + A_{22}s, \quad y_2 = B_{20} + B_{21}c + B_{22}s, \quad z_2 = C_{20} + C_{21}c + C_{22}s, \tag{B.4}$$

where A_{ij}, B_{ij}, C_{ij} for $1 \le i \le 2$ and $0 \le j \le 2$ are constants.

For glycine since C^α-H^α RDC is the sum of RDCs for bond vectors C^α-H^{α_2} and C^α-H^{α_3}, we can write the RDC equation as

$$r = D_{\max}\left[\mathbf{v}_1^T \mathbf{S} \mathbf{v}_1 + \mathbf{v}_2^T \mathbf{S} \mathbf{v}_2\right]. \tag{B.5}$$

Without loss of generality, we let $D_{\max} = 1$, which is done by scaling the RDCs appropriately. Now, since $\mathbf{v}_1$ and $\mathbf{v}_2$ are unit vectors,

$$x_1^2 + y_1^2 + z_1^2 = 1 \tag{B.6}$$

$$x_2^2 + y_2^2 + z_2^2 = 1. \tag{B.7}$$

Using Eq. (B.6) and Eq. (B.7) we can expand Eq. (B.5) and rewrite in the following form:

$$a(x_1^2 + x_2^2) + b(y_1^2 + y_2^2) = c, \tag{B.8}$$

where

$$a = S_{xx} - S_{zz}, \quad b = S_{yy} - S_{zz}, \quad c = r - 2S_{zz}.$$

Using Eq. (B.3) and Eq. (B.4) in Eq. (B.8), and simplifying we obtain

$$K_0 + K_1 c + K_2 s + K_3 cs + K_4 c^2 + K_5 s^2 = 0, \tag{B.9}$$

where K_i, $0 \le i \le 5$ are constants.

Using half-angle substitutions

$$u = \tan(\frac{\phi_i}{2}),\ c = \frac{1-u^2}{1+u^2},\ \text{and}\ s = \frac{2u}{1+u^2} \tag{B.10}$$

in Eq. (B.9) we have

$$L_0 + L_1 u + L_2 u^2 + L_3 u^3 + L_4 u^4 = 0, \tag{B.11}$$

where L_i, $0 \le i \le 4$ are constants.

Eq. (B.11) is a quartic equation which can be solved exactly and in closed form. Let $\{u_1, u_2, u_3, u_4\}$ denote the set of four real solutions (at most) of Eq. (B.11). For each u_i, the corresponding ϕ_i value can be computed using Eq. (B.10). □

C Analytic Solutions for Peptide Plane Orientations from ϕ-defining RDCs in Medium 1 and ψ-defining RDCs in Medium 2

We show that it is possible to compute all possible orientations of a peptide plane from a ϕ-defining RDC in one alignment medium and a ψ-defining RDC in a second alignment medium. That is, if RDCs for the bond vectors which are missing in one alignment medium can be measured in a second medium, our algorithm POOL is able to use those to compute loop backbone conformations. The proposition below shows how to do this.

Proposition 1. *Given the diagonalized alignment tensor components S_{xx} and S_{yy} for medium 1, S'_{xx} and S'_{yy} for medium 2, a relative rotation matrix $\mathbf{R}$ between the POFs of medium 1 and 2, the peptide plane P_i, a ϕ-defining RDC in medium 1 and a ψ-defining RDC in medium 2 for ϕ_i and ψ_i, respectively, there exist at most 16 orientations of the peptide plane P_{i+1} with respect to P_i that satisfy the RDCs, which can be computed exactly and in closed form by solving two quartic equations.*

Proof. For brevity, in our derivation we will only state the constants without giving explicit formulas to compute them.

Let POF_1 and POF_2 denote the POFs for the medium 1 and 2, respectively. Without loss of generality, we choose to work in POF_1. By direct application of Proposition 1 in main text [14], we can compute ϕ_i exactly and in closed form. Now it remains to compute ψ_i. Let $\mathbf{v} = (x, y, z)^T$ be the vector in POF_1 and the same vector be $\mathbf{v}' = (x', y', z')^T$ in POF_2, for which we have a ψ-defining RDC measured in medium 2. Then

$$\begin{aligned} & \mathbf{v}' = \mathbf{R}\mathbf{v} \qquad (C.1) \\ \Rightarrow \quad & \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \end{aligned}$$

from which we have

$$\begin{aligned} x' &= R_{11}x + R_{12}y + R_{13}z \qquad (C.2) \\ y' &= R_{21}x + R_{22}y + R_{23}z \qquad (C.3) \\ z' &= R_{31}x + R_{32}y + R_{33}z. \qquad (C.4) \end{aligned}$$

The reduced RDC equation (Eq. (5) in [14]) for ψ-defining RDC can be written as

$$ax'^2 + by'^2 = c. \qquad (C.5)$$

Substituting Eq. (C.2) and Eq. (C.3) in Eq. (C.5), we obtain

$$I_0 + I_1x^2 + I_2y^2 + I_3z^2 + I_4xy + I_5yz + I_6zx = 0, \qquad (C.6)$$

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

Let the unit vector $\mathbf{v}_0 = (0, 0, 1)^T$ represent the N-$\mathrm{H^N}$ bond vector of residue i in the local coordinate frame defined on the peptide plane P_i in POF_1. Then we can write the forward kinematics relation between $\mathbf{v}_0$ and $\mathbf{v}$ as follows:

$$\mathbf{v} = \mathbf{R}_{i,\mathtt{POF}}\, \mathbf{R}_l\, \mathbf{R}_z(\phi_i)\, \mathbf{R}_m\, \mathbf{R}_z(\psi_i)\, \mathbf{R}_r\, \mathbf{v}_0 \qquad (C.7)$$

Here $\mathbf{R}_l$, $\mathbf{R}_m$ and $\mathbf{R}_r$ are constant rotation matrices that describe the kinematic relationship between $\mathbf{v}_0$ and $\mathbf{v}$. $\mathbf{R}_z(\phi_i)$ is the rotation about the z-axis by ϕ_i, and is a constant rotation matrix since ϕ_i is known. $\mathbf{R}_z(\psi_i)$ is the rotation about the z-axis by ψ_i.

Let c and s denote $\cos\psi_i$ and $\sin\psi_i$, respectively. Using this while expanding Eq. (C.7) we have

$$x = A_0 + A_1 c + A_2 s, \quad y = B_0 + B_1 c + B_2 s, \quad z = C_0 + C_1 c + C_2 s, \tag{C.8}$$

where A_i, B_i, C_i for $0 \leq i \leq 2$ are constants. Substituting Eq. (C.8) in Eq. (C.6) we obtain

$$K_0 + K_1 c + K_2 s + K_3 cs + K_4 c^2 + K_5 s^2 \;\; = \;\; 0, \tag{C.9}$$

where K_i, $0 \leq i \leq 5$ are constants.

Using half-angle substitutions

$$u = \tan(\frac{\psi_i}{2}),\; c = \frac{1-u^2}{1+u^2},\; \text{and}\; s = \frac{2u}{1+u^2} \tag{C.10}$$

in Eq. (C.9) we have

$$L_0 + L_1 u + L_2 u^2 + L_3 u^3 + L_4 u^4 \;\; = \;\; 0, \tag{C.11}$$

where L_i, $0 \leq i \leq 4$ are constants.

Eq. (C.11) is a quartic equation which can be solved exactly and in closed form. Let $\{u_1, u_2, u_3, u_4\}$ denote the set of four real solutions (at most) of Eq. (C.11). For each u_i, we can compute the corresponding ψ_i value by using Eq. (C.10).

We have shown that for both ϕ_i and ψ_i there are at most four possible real solutions that satisfy the respective RDCs. Therefore, in total there are at most 16 orientations possible for the peptide plane P_{i+1}. □

D NMR Experimental Procedures for RDC Data Collection

The NMR data for FF2 was recorded and collected using Varian 600 and 800 MHz spectrometers at Duke University. NMRPipe [9] was used to process the NMR spectra. All NMR peaks were picked by the programs NMRView [10] or Xeasy/Cara [2], followed by manual editing. Backbone assignments were obtained from the set of triple resonance NMR experiments HNCA, HN(CO)CA, HN(CA)CB, HN(COCA)CB, and HNCO, combined with the HSQC spectra using the program Paces [6], followed by manual checking. The NOE cross peaks were picked from three-dimensional ^{15}N- and ^{13}C-edited NOESY-HSQC spectra. The diagonal cross peaks and water artifacts were removed from the picked NOE peak list. The C^{α}-H^{α} and N-H^{N} RDC data for FF2 was measured from a 2D ^{1}H-^{15}N IPAP experiment [12] and a modified (HACACO)NH experiment [1], respectively. The C^{α}-C' and C'-N RDCs of FF2 were measured from a set of HNCO-based experiments [13]. The RDC data for ubiquitin (PDB id: 1d3z), DinI (PDB id: 1ghh) and GB3 (PDB id: 2oed) were obtained from the BioMagResBank [15].

E RDC Simulations using PALES

We used the same set of loops that were previously studied by three other loop closure algorithms [4, 7, 11]. This set consists of 10 loops each with 4, 8 and 12 residues long chosen from a set of nonredundant X-ray crystallographic structures obtained from PDB [3]. We used PALES [20, 19]

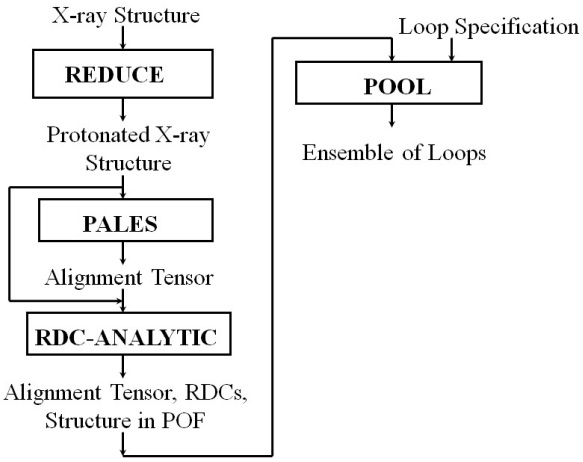

Figure 2: Block diagram of the RDC simulation procedure.

to simulate alignment tensors. Figure 2 shows a block diagram of our RDC simulation procedure. The PDB coordinate files were obtained from the PDB [3]. Then the REDUCE [16] module of MOLPROBITY [8, 5] was invoked to protonate the X-ray structures. The protonated structures were then input to PALES. The PALES protocol [19] predicts both magnitude and orientation of the steric component of the molecular alignment tensor from the molecule's three-dimensional (3D) shape. In our simulations, infinite cylinder Pf1 bacteriophage (`-pf1` flag) was used as the liquid crystalline alignment medium. The `-H` flag was enabled to include the protons. Other simulation parameters were set to their default values. The PALES-predicted alignment tensor, and the protonated crystal structure was then used by RDC-ANALYTIC [18, 17] to simulate the RDCs. RDC-ANALYTIC outputs the RDCs, the protonated structure in a principal order frame (POF) of RDCs that diagonalizes the alignment tensor by doing singular value decomposition (SVD). These, along with the loop anchor residue specifications were then input to POOL which determined the loop conformations.

References

[1] G. Ball, N. Meenan, K. Bromek, B. O. Smith, J. Bella, and D. Uhrín. Measurement of one-bond $^{13}C^{\alpha}$-$^{1}H^{\alpha}$ residual dipolar coupling constants in proteins by selective manipulation of $C^{\alpha}H^{\alpha}$ spins. *Journal of Magnetic Resonance*, 180:127–136, 2006.

[2] C. Bartels, T. Xia, M. Billeter, P. Güntert, and K. Wüthrich. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. 6:1–10, 1995.

[3] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourn. The Protein Data Bank. 28(1):235–242, 2000.

[4] A. A. Canutescu and R. L. Dunbrack, Jr. Cyclic coordinate descent: a robotics algorithm for protein loop closure. 12(5):963–972, 2003.

[5] V. B. Chen, W. B. Arendall, III, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, J. S. R. Laura W. Murray, and D. C. Richardson. MolProbity: all-atom structure validation for macromolecular crystallography. D66:12–21, 2010.

[6] B. E. Coggins and P. Zhou. PACES: Protein sequential assignment by computer-assisted exhaustive search. 26:93–111, 2003.

[7] E. A. Coutsias, C. Seok, M. P. Jacobson, and K. A. Dill. A kinematic view of loop closure. 25:510–528, 2004.

[8] I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, III, J. Snoeyink, J. S. Richardson, and D. C. Richardson. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. 35: Web Server issue:W375–W383, 2007.

[9] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. 6:277–293, 1995.

[10] B. A. Johnson and R. A. Blevins. NMRView: a computer program for the visualization and analysis of NMR data. 4:603–614, 1994.

[11] P. Liu, F. Zhu, D. N. Rassokhin, and D. K. Agrafiotis. A self-organizing algorithm for modeling protein loops. 5(8):e1000478, 08 2009.

[12] M. Ottiger, F. Delaglio, and A. Bax. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *Journal of Magnetic Resonance*, 131:373–378, 1998.

[13] P. Permi, P. R. Rosevear, and A. Annila. A set of HNCO-based experiments for measurement of residual dipolar couplings in ^{15}N, ^{13}C, (^{2}H)-labeled proteins. 17:43–54, 2000.

[14] C. Tripathy, J. Zeng, P. Zhou, and B. R. Donald. Protein loop closure using orientational restraints from NMR data. In *Proceedings of the 15th Annual International Conference on Research in Computational Molecular Biology (RECOMB), Vancouver, BC Canada*, 2011.

[15] 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. 36(Database issue):D402–D408, 2008.

[16] J. M. Word, S. C. Lovell, J. S. Richardson, and D. C. Richardson. Asparagine and Glutamine: Using Hydrogen Atom Contacts in the Choice of Side-chain Amide Orientation. 285:1735–1747, 1999.

[17] 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.

[18] 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. 45(3):265–281, 2009.

[19] M. Zweckstetter. NMR: prediction of molecular alignment from structure using the PALES software. 3:679–690, 2008.

[20] M. Zweckstetter and A. Bax. Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR. 122(15):3791–3792, 2000.