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

Appendix

In **Appendix A**, 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. In **Appendix B**, we describe the procedure used to simulate the RDCs for the loops studied in [2, 4, 6, 7], and by our algorithm POOL (see section Results and Discussion).

A 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 A.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. 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, 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

$$\mathbf{v}' = \mathbf{R}\mathbf{v} \tag{A.1}$$

$$\Rightarrow \quad \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

[¶]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+proteins11@cs.duke.edu, ☎ 919-660-6583, 🖷 919-660-6519

from which we have

$$x' = R_{11}x + R_{12}y + R_{13}z \quad (A.2)$$
$$y' = R_{21}x + R_{22}y + R_{23}z \quad (A.3)$$
$$z' = R_{31}x + R_{32}y + R_{33}z. \quad (A.4)$$

The reduced RDC equation (Eq. (5) in main text) for ψ-defining RDC can be written as

$$a'x'^2 + b'y'^2 = c', \quad (A.5)$$

where a', b' and c' are constants. Substituting Eq. (A.2) and Eq. (A.3) in Eq. (A.5), we obtain

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

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-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:

$$\mathbf{v} = \mathbf{R}_{i,\texttt{POF}}\, \mathbf{R}_l\, \mathbf{R}_z(\phi_i)\, \mathbf{R}_m\, \mathbf{R}_z(\psi_i)\, \mathbf{R}_r\, \mathbf{v}_0 \quad (A.7)$$

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. (A.7) we have

$$x = A_0 + A_1c + A_2s, \quad y = B_0 + B_1c + B_2s, \quad z = C_0 + C_1c + C_2s, \quad (A.8)$$

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

$$K_0 + K_1c + K_2s + K_3cs + K_4c^2 + K_5s^2 = 0, \quad (A.9)$$

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

Using half-angle substitutions

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

in Eq. (A.9) we have

$$L_0 + L_1u + L_2u^2 + L_3u^3 + L_4u^4 = 0, \quad (A.11)$$

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

Eq. (A.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 (at most) four real solutions of Eq. (A.11). For each u_i, we can compute the corresponding ψ_i value by using Eq. (A.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}. □

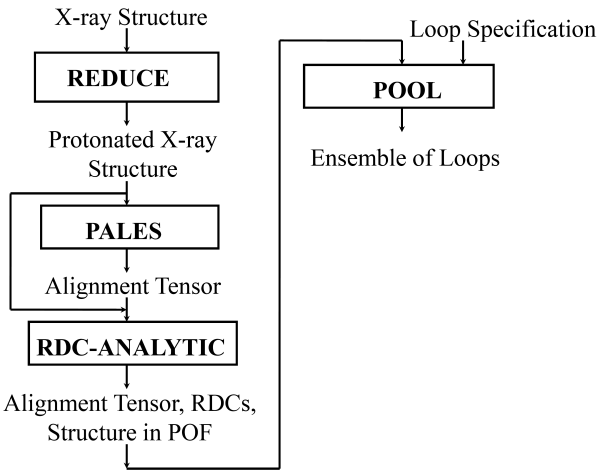

Figure S1. Block diagram of the RDC simulation procedure.

B RDC Simulations using PALES

We used the same set of loops that were previously studied by three other loop closure algorithms [2, 4, 6]. This set consists of 10 loops each with 4, 8 and 12 residues chosen from a set of nonredundant X-ray crystallographic structures obtained from PDB [1]. In addition, we also used the set of twenty 12-residue long loops published in [7]. Since no experimental RDC data was available, we simulated the RDCs for these loops. First, we used PALES [12, 11] to simulate alignment tensors. Figure S1 shows a block diagram of our RDC simulation procedure. The PDB coordinate files were obtained from the PDB [1]. Then the REDUCE [8] module of MolProbity [5, 3] was invoked to protonate the X-ray structures. The protonated structures were then input to PALES. The PALES protocol [11] 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 [10, 9] 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] 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. *Nucleic Acids Res*, 28(1):235–242, 2000.

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

[3] 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. *Acta Crystallogr*, D66:12–21, 2010.

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

[5] 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. *Nucleic Acids Res*, 35: Web Server issue:W375–W383, 2007.

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

[7] D. J. Mandell, E. A. Coutsias, and T. Kortemme. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nat Methods*, 6(8):551–552, 2009.

[8] 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. *J Mol Biol*, 285:1735–1747, 1999.

[9] 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.

[10] 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. *J Biomol NMR*, 45(3):265–281, 2009.

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

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