Supporting Information for

“Protein Loop Closure Using Orientational Restraints from NMR Data”*

Chittaranjan Tripathy1, Jianyang Zeng1, Pei Zhou2, and Bruce Randall Donald1,2,*

1Department of Computer Science, Duke University, Durham, NC 27708
2Department 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 \( \phi \)-defining RDC in one alignment medium and a \( \psi \)-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 \( \phi \)-defining RDCs in Medium 1 and \( \psi \)-defining RDCs in Medium 2

We show that it is possible to compute all possible orientations of a peptide plane from a \( \phi \)-defining RDC in one alignment medium and a \( \psi \)-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 \( R \) between the POFs of medium 1 and 2, the peptide plane \( P_i \), a \( \phi \)-defining RDC in medium 1 and a \( \psi \)-defining RDC in medium 2 for \( \phi_i \) and \( \psi_i \), 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 \( \phi_i \) exactly and in closed form. Now it remains to compute \( \psi_i \). Let \( v = (x, y, z)^T \) be the vector in POF\(_1\) and the same vector be \( v' = (x', y', z')^T \) in POF\(_2\), for which we have a \( \psi \)-defining RDC measured in medium 2. Then

\[
\begin{align*}
    v' &= Rv \\
    \Rightarrow \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} &= \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}
\end{align*}
\]

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

\[
x' = R_{11}x + R_{12}y + R_{13}z \tag{A.2}
\]

\[
y' = R_{21}x + R_{22}y + R_{23}z \tag{A.3}
\]

\[
z' = R_{31}x + R_{32}y + R_{33}z \tag{A.4}
\]

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

\[
a'x'^2 + b'y'^2 = c', \tag{A.5}
\]

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

\[
I_0 + I_1 x^2 + I_2 y^2 + I_3 z^2 + I_4 x y + I_5 y z + I_6 x z = 0, \tag{A.6}
\]

where \( I_i \) for \( 0 \leq i \leq 6 \) are constants.

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

\[
\mathbf{v} = R_{i, POF} R_{l} R_{z}(\phi_i) R_{m} R_{z}(\psi_i) R_{r} \mathbf{v}_0 \tag{A.7}
\]

Here \( R_l, R_m \) and \( R_r \) are constant rotation matrices that describe the kinematic relationship between \( \mathbf{v}_0 \) and \( \mathbf{v} \). \( R_z(\phi_i) \) is the rotation about the \( z \)-axis by \( \phi_i \), and is a constant rotation matrix since \( \phi_i \) is known. \( R_z(\psi_i) \) is the rotation about the \( z \)-axis by \( \psi_i \).

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

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

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

\[
K_0 + K_1 c + K_2 s + K_3 c s + K_4 c^2 + K_5 s^2 = 0, \tag{A.9}
\]

where \( K_i \) for \( 0 \leq i \leq 5 \) are constants.

Using half-angle substitutions

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

in Eq. (A.9) we have

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

where \( L_i \) for \( 0 \leq i \leq 4 \) 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 \( \psi_i \) value by using Eq. (A.10).

We have shown that for both \( \phi_i \) and \( \psi_i \) 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 \( P_{i+1} \).
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 $\text{–H}$ 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


