# Supplementary Material for

*"High-Resolution Protein Structure Determination Starting with a Global Fold Calculated from Exact Solutions"*

Jianyang Zeng[1]
Jeffrey Boyles[2]
Chittaranjan Tripathy[1]
Lincong Wang[1,†]
Anthony Yan[1]
Pei Zhou[2,*]
Bruce Randall Donald[1,2,*]

July 31, 2009

**Abbreviations used:** NMR, nuclear magnetic resonance; ppm, parts per million; RMSD, root mean square deviation; HSQC, heteronuclear single quantum coherence spectroscopy; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser and exchange spectroscopy; RDC, residual dipolar coupling; PDB, Protein Data Bank; pol $\eta$ UBZ, ubiquitin-binding zinc finger domain of the human Y-family DNA polymerase Eta; CH, $C^{\alpha}$-$H^{\alpha}$; hSRI, human Set2-Rpb1 interacting domain; FF2, FF Domain 2 of human transcription elongation factor CA150 (RNA polymerase II C-terminal domain interacting protein); POF, principal order frame; SA, simulated annealing; MD, molecular dynamics; SSE, secondary structure element; C′, carbonyl carbon; WPS, well-packed satisfying; vdW, van der Waals.

The following is supplementary material which provides additional information to substantiate the claims of the paper. Section S1 presents the flow chart of RDC-PANDA. In Section S2 we give a detailed derivation of how to compute the $\phi$ and $\psi$ backbone dihedral angles from the RDC equations. Section S3 presents the details of extracting sparse NOEs between secondary structure elements using only chemical shift information. Section S4 gives the computational complexity for PACKER. Section S5 describes the details of the HANA algorithm. In Section S6 we present an analysis of the running time of HANA. Section S7 presents the details of the local minimization approach. In Section S8, additional details for the *Results* section of the main article are given. SM references are provided at the end of the SM.

## S1 The Flow Chart of RDC-PANDA

Fig. S1 shows the flow chart of RDC-PANDA.

[1]Department of Computer Science, Duke University, Durham NC 27708
[2]Department of Biochemistry, Duke University Medical Center, Durham NC 27708
[*]Corresponding authors: Bruce Randall Donald, brd+jbnmr09@cs.duke.edu, tel: 919-660-6583, Fax: 919-660-6519; Pei Zhou, peizhou@biochem.duke.edu, tel: 919-668-6409, Fax: 919-684-8885.
[†]Present address: Medicinal Chemistry, Boehringer Ingelheim Pharmaceuticals, Inc., Ridgefield CT, 06877.

## S2 Derivation of Equations for Computing Backbone Dihedral Angles from two RDCs in one Medium

In contrast to NOE restraints, which provide local distance restraints on the positions of pairs of protons, RDCs provide the global orientational restraints on internuclear vectors with respect to a global coordinate frame (Tolman et al. 1995, Tjandra & Bax 1997). RDCs have provided an alternative means for automated protein structure determination (Tolman et al. 1995, Fowler et al. 2000, Ruan et al. 2008, Prestegard et al. 2004, Delaglio et al. 2000, Hus et al. 2001, Tian et al. 2001, Wang & Donald 2004*b*, Wang et al. 2006, Rohl & Baker 2002). In contrast to RDC-PANDA, most RDC-based structure determination approaches consider RDC data as complementary restraints, and only use RDCs in the final structure refinement. In previous frameworks that incorporate both NOE and RDC restraints for structure determination, NOE assignment and RDC-based structure calculation are usually performed separatively. In these approaches, the global constraints derived from RDC data are not used for filtering ambiguous NOE assignments. Although NOE and RDC restraints can be bound together in the scoring function to compute the structure templates for pruning ambiguous NOE assignments, these methods suffer from the same error propagation problem as using the chemical shift and NOE spectra alone. In addition, previous RDC-based structure determination approaches heavily rely on stochastic techniques such as SA/MD to compute the initial structure template, and randomly sample the backbone and side-chain conformation space to satisfy the experimental restraints (Tian et al. 2001, Hus et al. 2001, Andrec et al. 2004). Since an accurate initial fold is critical to compute the correct structure, manual intervention is often required for initial NOE assignments in order to obtain a suitable initial structure template. In contrast, the backbones computed by RDC-EXACT are not abrogated by the uncertainty arising from ambiguous NOE assignments. Thus, they can be used to compute a robust and reliable initial fold for filtering ambiguous NOE assignments.

The roadmap of deriving the theoretical foundations for RDC-EXACT is given as follows. Below, we first derive a quartic equation, using basic physics (Saupe 1968) and protein backbone kinematics, that is satisfied by the $x$-component of a unit vector on which an RDC is measured. Then we show how the backbone dihedral $(\phi, \psi)$ angles can be subsequently computed from such vectors. These equations are used by the RDC-EXACT algorithm as previously described by (Wang & Donald 2004*b*,*a*, Wang et al. 2006), to compute the structure of secondary structure elements from 2 RDCs per residue in one medium. The derivation below assumes standard protein geometry, which is exploited in the kinematics. We choose to work in an orthogonal coordinate system defined at the peptide plane $i$ with $z$-axis along the bond vector $\mathrm{H}^{\mathrm{N}}(i) \rightarrow \mathrm{N}(i)$, where the symbol $\rightarrow$ means a vector from atom $\mathrm{H}^{\mathrm{N}}(i)$ to atom $\mathrm{N}(i)$. The $y$-axis is in the peptide plane $i$ and the angle between $y$-axis and the bond vector $\mathrm{N}(i) \rightarrow \mathrm{C}^{\alpha}(i)$ is $29.14^{\circ}$ as described previously in (Wang & Donald 2004*b*). The $x$-axis is defined based on the right-handedness. Let $\mathbf{R}_i$ denote the relative rotation matrix between the POF and the coordinate system defined at the peptide plane $i$. $\mathbf{R}_1$ denotes the relative rotation matrix between the coordinate system defined at the first residue of the current SSE and the POF. $\mathbf{R}_i$ is used to derive $\mathbf{R}_{i+1}$ inductively after we compute the backbone dihedral angles $\phi_i$ and $\psi_i$. $\mathbf{R}_{i+1}$, in turn, is used to compute the $(i+1)^{st}$ peptide plane.

The equations and propositions below were proven in (Wang & Donald 2004*a*). For clarity, we provide a somewhat simpler exposition here. The derivation below closely mirrors our new (open-source) software implementation, and the clearer equations are easier to interpret and build upon. A review of these techniques can be found in (Donald & Martin 2008).

## S2.1 The Computation of $\phi$ Angle

The RDC equation is given by

$$r = D_{\mathtt{max}}\mathbf{v}^T\mathbf{S}\mathbf{v}, \tag{S1}$$

where $r$ is the experimentally-observed RDC, $D_{\mathtt{max}}$ is the dipolar interaction constant, $\mathbf{S}$ is the $3 \times 3$ *Saupe order matrix* (Saupe 1968), or *alignment tensor* that specifies the ensemble-averaged anisotropic orientation of the protein in the laboratory frame, and $\mathbf{v}$ represents the internuclear bond vector. Letting $D_{\mathtt{max}} = 1$ for simplicity of exposition, and considering a global coordinate frame that diagonalizes the alignment tensor $\mathbf{S}$ (such a coordinate frame is called the *principal order frame (POF)*), Equation (S1) can be rewritten as

$$r = S_{xx}x^2 + S_{yy}y^2 + S_{zz}z^2, \tag{S2}$$

where $S_{xx}, S_{yy}$ and $S_{zz}$ are the three diagonal elements of $\mathbf{S}$, and $x, y$ and $z$ are, respectively, the $x, y$ and $z$ components of the unit vector $\mathbf{v}$ in a POF which diagonalizes $\mathbf{S}$, which is a $3 \times 3$ symmetric, traceless matrix with five independent elements (Tjandra & Bax 1997, Tolman et al. 1995, Prestegard et al. 2004, Ruan et al. 2008).

***Proposition* 1.** *If the diagonalized Saupe elements and both the* $\mathrm{NC}^\alpha$ *and* NH *vectors of residue* $i$ *in the POF are known, then the* $x$*-component of the* CH *unit internuclear vector* $\mathbf{v}$ *with RDC value* $r_{\mathrm{C}}$ *satisfies a monomial quartic equation. Additionally, the* CH *vector has at most four possible orientations.*

*Proof.* From the RDC equation (Equation (S2)) we have

$$r_{\mathrm{C}} = S_{xx}x^2 + S_{yy}y^2 + S_{zz}z^2 \tag{S3}$$

$$r_{\mathrm{N}} = S_{xx}x'^2 + S_{yy}y'^2 + S_{zz}z'^2 \tag{S4}$$

where $S_{xx}, S_{yy}$ and $S_{zz}$ are the three diagonal elements of the diagonalized Saupe matrix $\mathbf{S}$. $r_{\mathrm{C}}$ and $r_{\mathrm{N}}$ are the RDC values for CH and NH vectors, respectively. $x, y, z$ and $x', y', z'$ are the components of the CH and NH unit vectors, respectively. Alternatively, these components can be viewed as the direction cosines of those vectors. Let $\theta_{41}$ be the dihedral angle from the plane $(\mathrm{N}(i)\rightarrow\mathrm{C}^\alpha(i)\rightarrow\mathrm{C}'(i))$ to the plane $(\mathrm{N}(i)\rightarrow\mathrm{C}^\alpha(i)\rightarrow\mathrm{H}^\alpha(i))$ and $\theta_{42}$ be the angle between the two vectors $\mathrm{N}(i)\rightarrow\mathrm{C}^\alpha(i)$ and $\mathrm{C}^\alpha(i)\rightarrow\mathrm{H}^\alpha(i)$. $\mathbf{R}_\kappa(\theta)$ denotes the rotation matrix that represents a rotation by an angle $\theta$ about vector $\kappa \in \mathbb{R}^3$. From backbone kinematics we have

$$\mathbf{M}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{R}_{\mathbf{y}}(-\phi)\begin{pmatrix} C_x \\ C_y \\ C_z \end{pmatrix}. \tag{S5}$$

The matrix $\mathbf{M}$, $C_x, C_y$ and $C_z$ are known constants from standard peptide geometry, and can be computed by means of kinematics as follows:

$$\mathbf{M} = \mathbf{R}_y(\theta_8)\mathbf{R}_x(\theta_1)\mathbf{R}_i \tag{S6}$$

$$\begin{pmatrix} C_x \\ C_y \\ C_z \end{pmatrix} = \mathbf{R}_y(-\theta_{41})\mathbf{R}_x(-\theta_{42})\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{S7}$$

| Angle identity | Variable name |
|---|---|
| $\mathrm{H}^{\mathrm{N}}(i) - \mathrm{N}(i) - \mathrm{C}^{\alpha}(i)$ - $\pi/2$ | $\theta_1$ |
| $\mathrm{N}(i) - \mathrm{C}^{\alpha}(i) - \mathrm{C}'(i)$ | $\theta_3$ |
| $\mathrm{C}^{\alpha}(i) - \mathrm{C}'(i) - \mathrm{N}(i+1)$ - $\pi/2$ | $\theta_5$ |
| $\mathrm{C}^{\alpha}(i) - \mathrm{C}'(i) - \mathrm{N}(i+1) - \mathrm{H}^{\mathrm{N}}(i+1)$ | $\theta_6$ |
| $\mathrm{C}'(i) - \mathrm{N}(i+1) - \mathrm{H}^{\mathrm{N}}(i+1)$ - $\pi/2$ | $\theta_7$ |
| $\mathrm{C}'(i-1) - \mathrm{N}(i) - \mathrm{C}^{\alpha}(i) - \mathrm{H}^{\mathrm{N}}(i)$ | $\theta_8$ |

Table S1: **Six Backbone Angles.** The **Variable names** are the names assigned to the six angles in the equations.

The angles $\theta_1$ and $\theta_8$ are defined in Table S1. Note that

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

and

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

Since

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

we have from Equation (S5):

$$\begin{aligned} M_{11}x + M_{12}y + M_{13}z &= C_x\cos\phi + C_z\sin\phi & \text{(S8)} \\ M_{21}x + M_{22}y + M_{23}z &= C_y & \text{(S9)} \\ M_{31}x + M_{32}y + M_{33}z &= -C_x\sin\phi + C_z\cos\phi. & \text{(S10)} \end{aligned}$$

Squaring Equation (S8) and Equation (S10), and adding them together to eliminate $\phi$ we have

$$(M_{11}x + M_{12}y + M_{13}z)^2 + (M_{31}x + M_{32}y + M_{33}z)^2 = C_x^2 + C_z^2. \quad \text{(S11)}$$

Expanding Equation (S11), replacing $z$ by $\frac{C_y - M_{21}x - M_{22}y}{M_{23}}$ and letting

$$\begin{aligned}
C_0 &= C_x^2 + C_z^2 \\
C_a &= \frac{C_y M_{13}}{M_{23}} \\
C_1 &= M_{11} - \frac{M_{13}M_{21}}{M_{23}} \\
C_2 &= M_{12} - \frac{M_{13}M_{22}}{M_{23}} \\
C_b &= \frac{C_y M_{33}}{M_{23}} \\
C_3 &= M_{31} - \frac{M_{33}M_{21}}{M_{23}} \\
C_4 &= M_{32} - \frac{M_{33}M_{22}}{M_{23}},
\end{aligned}$$

we have

$$(C_1x + C_2y + C_a)^2 + (C_3x + C_4y + C_b)^2 = C_0. \tag{S12}$$

Expanding again we have

$$d_1x^2 + d_2y^2 + d_3xy + d_4x + d_5y + d_0 = 0, \tag{S13}$$

where

$$\begin{aligned}
d_0 &= C_a^2 + C_b^2 - C_0 \\
d_1 &= C_1^2 + C_3^2 \\
d_2 &= C_2^2 + C_4^2 \\
d_3 &= 2C_1C_2 + 2C_3C_4 \\
d_4 &= 2C_1C_a + 2C_3C_b \\
d_5 &= 2C_2C_a + 2C_4C_b.
\end{aligned}$$

Equation (S13) corresponds to a general conic curve, such as an ellipse.
Noting that

$$x^2 + y^2 + z^2 = 1,$$

and using this in Equation (S3) to eliminate $z$ we obtain

$$r = ax^2 + by^2, \tag{S14}$$

where

$$\begin{aligned}
a &= S_{xx} - S_{zz} \\
b &= S_{yy} - S_{zz} \\
r &= r_C - S_{zz},
\end{aligned}$$

which also defines an ellipse.

Using Equation (S14) to eliminate $y$ in Equation (S13) we obtain the following quartic equation:

$$f_4 x^4 + f_3 x^3 + f_2 x^2 + f_1 x + f_0 = 0, \tag{S15}$$

where

$$\begin{aligned}
f_4 &= e_1^2 + \frac{a e_2^2}{b} \\
f_3 &= 2e_1 e_3 + \frac{2 e_2 e_4 a}{b} \\
f_2 &= d_4^2 + 2e_1 e_0 + \frac{a e_4^2}{b} - \frac{r e_2^2}{b} \\
f_1 &= 2e_3 e_0 - \frac{2 r e_2 e_4}{b} \\
f_0 &= e_0^2 - \frac{r e_4^2}{b} \\
e_4 &= d_5 \\
e_3 &= d_4 \\
e_2 &= d_3 \\
e_1 &= d_1 - \frac{d_2 a}{b} \\
e_0 &= d_0 + \frac{d_2 r}{b}.
\end{aligned}$$

Equation (S15) can be solved in closed form to give the $x$-component of the CH unit vector. We note that there are at most four possible real roots of Equation (S15), so $x$ can have at most four real values. It remains to show that there are at most four solutions for CH unit vector, which we do next.

At most, all four solutions for $x$ are real. Let $x = \{x_1, x_2, x_3, x_4\}$ denote the set of four solutions. When we pick a root $x_i$ $(1 \leq i \leq 4)$ and substitute it in Equation (S14), we obtain at most two possible real values for $y_i$. We denote them by $+y_i$ and $-y_i$, respectively. We can discard one of the values of $y_i$ as follows. Observe the structure of Equation (S13), in which the first, second, fourth and sixth terms are independent of the sign of $y_i$, therefore they always add to the same value (denoted by $A$) given a root $x_i$ and any of the two possible $y_i$'s. The sum of the third and fifth term in Equation (S13) has the same absolute value (denoted by $B$) but $B$'s sign depends on whether $+y_i$ or $-y_i$ is chosen (call the two values $+B$ and $-B$). For Equation (S13) to hold, exactly one of $+B$ and $-B$ cancels $A$, which implies exactly one of $+y_i$ and $-y_i$ is the actual solution, and the other one is discarded. Knowing $x_i$ and its corresponding $y_i$, we can compute a unique $z_i$ using Equation (S9), which completes the proof that there are at most four solutions for the CH unit vector. □

Finally, for a given CH unit vector orientation, a unique backbone dihedral $\phi$ angle can be computed from Equation (S8) and Equation (S10), which we state formally in the following proposition:

***Proposition* 2.** *If the* CH *unit vector is known, then the backbone dihedral angle $\phi$ satisfies two simple trigonometric equations. The sine and cosine of $\phi$ can be computed exactly and in closed form.*

*Proof.* Multiplying Equation (S8) by $C_x$ and Equation (S10) by $C_z$ we have

$$\begin{aligned}
C_x(M_{11}x + M_{12}y + M_{13}z) &= C_x^2 \cos\phi + C_z C_x \sin\phi \\
C_z(M_{31}x + M_{32}y + M_{33}z) &= -C_x C_z \sin\phi + C_z^2 \cos\phi.
\end{aligned}$$

Adding together the above two equations, and then dividing both sides by $C_x^2 + C_z^2$ we have

$$\cos\phi = \frac{C_x(M_{11}x + M_{12}y + M_{13}z) + C_z(M_{31}x + M_{32}y + M_{33}z)}{C_x^2 + C_z^2}. \tag{S16}$$

Similarly, multiplying Equation (S8) by $C_z$ and Equation (S10) by $C_x$ we have

$$\begin{aligned} C_z(M_{11}x + M_{12}y + M_{13}z) &= C_xC_z\cos\phi + C_z^2\sin\phi & \text{(S17)} \\ C_x(M_{31}x + M_{32}y + M_{33}z) &= -C_x^2\sin\phi + C_xC_z\cos\phi. & \text{(S18)} \end{aligned}$$

Subtracting Equation (S18) from Equation (S17), and then dividing both sides by $C_x^2 + C_z^2$ we obtain

$$\sin\phi = \frac{C_z(M_{11}x + M_{12}y + M_{13}z) - C_x(M_{31}x + M_{32}y + M_{33}z)}{C_x^2 + C_z^2}. \tag{S19}$$

□

## S2.2 The Computation of $\psi$ Angle

The computation of the backbone dihedral $\psi$ angles proceeds very similarly with minor changes.

***Proposition* 3.** *If the backbone dihedral angle $\phi$ of residue $i$ is known, and the diagonalized Saupe elements and both the* $\mathrm{NC}^\alpha$ *and* NH *vectors of residue $i$ in the POF are known, then the $x$-component of the* NH *unit internuclear vector of residue $i+1$ with RDC value $r_{\mathrm{N}}$ satisfies a quartic monomial equation. Additionally, the* NH *vector has at most four possible orientations.*

*Proof.* Here the Equation (S5) is replaced by

$$\mathbf{M}\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \mathbf{R_z}(-\psi - \pi)\begin{pmatrix} C_x \\ C_y \\ C_z \end{pmatrix}, \tag{S20}$$

where $x', y', z'$ are the components of the NH unit vector (which we want to compute), and $\mathbf{M}$, $C_x, C_y$ and $C_z$ are known constants computed as follows using standard backbone kinematics. Here the same symbols $M$ and $C_x, C_y, C_z$ are used as in the derivation of equations for computing $\phi$. They play similar roles but are computed differently:

$$\begin{aligned} \mathbf{M} &= \mathbf{R}_x(\theta_3)\mathbf{R}_y(\phi)\mathbf{R}_y(\theta_8)\mathbf{R}_x(\theta_1)\mathbf{R}_i & \text{(S21)} \\ \begin{pmatrix} C_x \\ C_y \\ C_z \end{pmatrix} &= \mathbf{R}_x(-\theta_5)\mathbf{R}_y(-\theta_6)\mathbf{R}_x(-\theta_7)\begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}. & \text{(S22)} \end{aligned}$$

The angles $\theta_1, \theta_3, \theta_5, \theta_6$, $\theta_7$ and $\theta_8$ are defined in Table S1. Since

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

we have from Equation (S20):

$$\begin{aligned} M_{11}x' + M_{12}y' + M_{13}z' &= -C_x \cos\psi + C_y \sin\psi & \text{(S23)} \\ M_{21}x' + M_{22}y' + M_{23}z' &= -C_x \sin\psi - C_y \cos\psi & \text{(S24)} \\ M_{31}x' + M_{32}y' + M_{33}z' &= C_z. & \text{(S25)} \end{aligned}$$

Proceeding as before and eliminating $\psi$ we have

$$(C_1x' + C_2y' + C_a)^2 + (C_3x' + C_4y' + C_b)^2 = C_0, \quad \text{(S26)}$$

where the new coefficients $C_a, C_1, C_2, C_b, C_3$ and $C_4$ are

$$\begin{aligned} C_a &= \frac{C_z M_{13}}{M_{33}} \\ C_1 &= M_{11} - \frac{M_{13}M_{31}}{M_{33}} \\ C_2 &= M_{12} - \frac{M_{13}M_{32}}{M_{33}} \\ C_b &= \frac{C_z M_{23}}{M_{33}} \\ C_3 &= M_{21} - \frac{M_{23}M_{31}}{M_{33}} \\ C_4 &= M_{22} - \frac{M_{23}M_{32}}{M_{33}}. \end{aligned}$$

From here on we can derive an analogous quartic equation as the Equation (S15) for computing the $x$-, $y$- and $z$-components of the NH unit vector, and argue *mutatis mutandis* (as in Proposition 1) that there are at most four NH unit vector orientations possible. □

Finally, for a given NH unit vector orientation, a unique backbone dihedral $\psi$ angle can be computed from Equation (S23) and Equation (S24), which we state formally in the following proposition:

***Proposition* 4.** *If the* NH *unit vector is known, then the backbone dihedral angle $\psi$ satisfies two simple trigonometric equations. The sine and cosine of $\psi$ can be computed exactly and in closed form.*

*Proof.* Multiplying Equation (S23) by $C_x$ and Equation (S24) by $C_y$ we have

$$\begin{aligned} C_x(M_{11}x' + M_{12}y' + M_{13}z') &= -C_x^2 \cos\psi + C_xC_y \sin\psi \\ C_y(M_{21}x' + M_{22}y' + M_{23}z') &= -C_xC_y \sin\psi - C_y^2 \cos\psi. \end{aligned}$$

Adding together the above two equations, and then dividing both sides by $-(C_x^2 + C_y^2)$ we have

$$\cos\psi = \frac{C_x(M_{11}x' + M_{12}y' + M_{13}z') + C_y(M_{21}x' + M_{22}y' + M_{23}z')}{-(C_x^2 + C_y^2)}. \quad \text{(S27)}$$

Similarly, multiplying Equation (S23) by $C_y$ and Equation (S24) by $C_x$ we have

$$\begin{aligned} C_y(M_{11}x' + M_{12}y' + M_{13}z') &= -C_xC_y \cos\psi + C_y^2 \sin\psi & \text{(S28)} \\ C_x(M_{21}x' + M_{22}y' + M_{23}z') &= -C_x^2 \sin\psi - C_xC_y \cos\psi. & \text{(S29)} \end{aligned}$$

Subtracting Equation (S29) from Equation (S28), and then dividing both sides by $C_x^2 + C_y^2$ we obtain

$$\sin\psi = \frac{C_y(M_{11}x' + M_{12}y' + M_{13}z') - C_x(M_{21}x' + M_{22}y' + M_{23}z')}{(C_x^2 + C_y^2)}. \tag{S30}$$

□

***Proposition* 4.** *Given the orientation of the peptide plane $i$ in the POF of RDCs, the RDC for the* CH *internuclear vector of residue $i$ and the RDC for the* NH *internuclear vector of residue $i+1$, there exist at most 16 orientations of the peptide plane $i+1$.*

*Proof.* By Proposition 1 there exist at most four possible orientations for the CH internuclear vector of residue $i$. Therefore, it follows from Proposition 2 that $\phi_i$ has at most four possible values. Given the peptide plane $i$ and $\phi_i$, by Proposition 3 there exist at most four possible orientations for the NH internuclear vector of residue $i+1$. Therefore, it follows from Proposition 4 that $\psi_i$ has at most four possible values. Hence, we conclude that there are at most sixteen $(\phi_i, \psi_i)$ pairs possible, and hence the peptide plane $i+1$ has at most sixteen orientations. □

The algorithm to compute a secondary structure element using the equations above is described in (Wang & Donald 2004*a*,*b*, Wang et al. 2006), and a detailed review of these techniques can be found in (Donald & Martin 2008).

## S3 Extraction of Sparse NOEs Using Only Chemical Shift Information

Sparse inter-SSE NOE restraints (which are usually long-range NOEs) can be obtained from unambiguous NOE assignments for NOE cross peaks using only chemical shift information combined with other auxiliary principles as described below. The following procedure describes the details of extracting sparse unambiguous inter-SSE NOE assignments from both $^{13}$C- and $^{15}$N-edited 3D NOESY spectra. First, we assign a pair of protons to an NOE cross peak if their resonances lie within certain error windows from the corresponding peak frequencies. There might exist several such pairs of protons for one NOE peak due to chemical shift degeneracy or experimental noise. We use error windows of 0.5 ppm for heavy atoms $^{13}$C and $^{15}$N, and 0.05 ppm for protons, in our NOE extraction. These error windows are slightly larger than those used in the NOE assignment module HANA, because we aim to extract more confidently unambiguous inter-SSE NOE assignments. Among all ambiguous NOE assignments within the error tolerances in chemical shift, we pick *unique* NOE assignments, in which the corresponding NOE peak can only be associated with a single NOE assignment, to be the initial set of potential long-range NOE interactions between SSEs. The following rules were applied to further prune the set of unique NOE assignments: (a) A unique NOE assignment is deleted if it is between a pair of charged and hydrophobic residues; (b) A unique NOE assignment is removed if no other NOE interaction has been observed between the pair of corresponding residues; (c) In the remaining set of unique NOE assignments, weak NOEs (i.e. those with peak intensities falling into the bottom 20% of all NOE peaks excluding diagonal NOE peaks) are removed.

We note that principles similar to (a) and (b) have been used to prune ambiguous NOE assignment in (Huang et al. 2006). The rule (c) is used to prevent a unique NOE assignment to a noise artifact. The above rules are heuristic, and might miss some correct inter-SSE NOEs. However, they are *conservative*, that is, they might prune some useful NOEs but the result is ("only") that we will have fewer NOEs during the packing phase, and this approach will be less likely to yield incorrectly assigned NOEs.

## S4 Time Complexity Analysis of PACKER

We first derive the boundaries for our grid search in PACKER. We have the following lemma that will be useful for our subsequent analysis:

***Lemma* 1.** *Suppose that we are given two sets of points, denoted by $A = \{a_1, \cdots, a_m\}$ and $B = \{b_1, \cdots, b_m\}$ respectively, where $m$ is the maximum number of points in each set. Let $a_0$ and $b_0$ be the centers of all points in $A$ and $B$ respectively. Suppose that the maximum distance for any point in $A$ to any point in $B$ is upper bounded by $u$, that is, $\max_{i,j} \|a_i - b_j\| \leq u$. Then we have $\|a_0 - b_0\| \leq u$.*

*Proof.* Since $a_0$ and $b_0$ are centers of sets $A$ and $B$ respectively, we have $a_0 = \frac{1}{m}\sum_{i=1}^{m} a_i$ and $b_0 = \frac{1}{m}\sum_{i=1}^{m} b_i$. Thus, the distance between $a_0$ and $b_0$ is

$$\|a_0 - b_0\| = \|\frac{1}{m}\sum_{i=1}^{m} a_i - \frac{1}{m}\sum_{i=1}^{m} b_i\| = \frac{1}{m}\|\sum_{i=1}^{m}(a_i - b_i)\| \leq \frac{1}{m}\sum_{i=1}^{m}\|a_i - b_i\| \leq u.$$

□

Let $\{(a_i, b_i, \ell_i, u_i)\}$ denote the set of NOE restraints for packing two SSE backbones $H_1$ and $H_2$. By Lemma 1, the distance between the center of all protons $a_i$ in $H_1$ and the center of all protons $b_i$ in $H_2$ is less than the upper bound (i.e. 6 Å) of all NOE distances. Since the center of all protons $a_i$ (or $b_i$) and the center of $H_1$ (or $H_2$) are in the same rigid body, the distance between centers of $H_1$ and $H_2$ (namely the translation between $H_1$ and $H_2$) is also upper bounded by a constant. Let $u$ denote the upper bound of the translation between $H_1$ and $H_2$. Then the grid search is bounded in a box $2u \times 2u \times 2u$.

Let $N$ be the maximum length of SSE backbones, and let $t$ be the maximum number of rotamers for each amino acid in the rotamer library. We first generate a PDB that includes backbones and all possible rotamer conformations at each residue, which takes time $O(tN)$. Let $\varepsilon$ denote the resolution in our grid search. Then the total number of grid search points is $(\frac{2u}{\varepsilon})^3$. At each grid point, we need check whether all NOE restraints are satisfied, which takes $O(t^2 \cdot q)$, where $q$ is the total number of NOE restraints. Hence, Step (1) in PACKER runs in time $O(t^2 q(\frac{2u}{\varepsilon})^3)$. In Step (2), the steric clash checking takes time $O(N^2)$, assuming that the total number of atoms in each residues is a constant. The clustering step, i.e. Step (3), takes time $O((\frac{2u}{\varepsilon})^6)$, since the maximum number of packed structures is bounded by $O((\frac{2u}{\varepsilon})^3)$. Therefore, the total running time for our packing algorithm is

$$O(tN) + O(t^2 q(\frac{2u}{\varepsilon})^3) + O(N^2) + O((\frac{2u}{\varepsilon})^6) = O(N^2 + tN + t^2 q u^3 \varepsilon^{-3} + u^6 \varepsilon^{-6}).$$

## S5 Details of the HANA Algorithm

An alternative approach for automated NOE assignment proposed by Wang and Donald (2005), based on RDC-EXACT algorithm (Wang & Donald 2004*b*,*a*, Wang et al. 2006), uses a rotamer ensemble and residual dipolar couplings, and is a provably polynomial-time algorithm for automated NOE assignment. However, this algorithm does not exploit the NOE patterns of rotamers to model the uncertainty in peak positions; therefore, assignment accuracy is reduced while processing NOE spectra with many noisy peaks.

Our NOE assignment algorithm HANA retains the paradigm of the previous algorithm by Wang and Donald (2005), and develops a novel framework that starts with a high-resolution protein backbone computed from residual dipolar couplings (Wang & Donald 2004*a*,*b*, Wang et al. 2006), and then combines this

backbone with a library of rotamers to derive critical structural information for NOE assignment. Viewing the NOE assignment problem as a pattern-recognition problem, HANA uses an extended Hausdorff distance-based probabilistic framework to model the uncertainties due to the experimental error. Unlike many other pattern-recognition algorithms, Hausdorff-based algorithms are combinatorially precise, and provide a robust method for measuring the similarity between two point sets or image patterns (Huttenlocher & Kedem 1992, Huttenlocher et al. 1993) in the presence of noise and positional uncertainties. They also provide a straight-forward method for calculating the probability of a false match (Huttenlocher & Jaquith 1995). In contrast to previous stochastic algorithms (Güntert 2003, Herrmann et al. 2002, Huang et al. 2006, Linge et al. 2003, Mumenthaler et al. 1997, Nilges et al. 1997, Kuszewski et al. 2004, 2008) for NOE assignment, HANA uses the reliable initial fold mainly solved from RDCs, and can hence effectively filter ambiguous NOE assignments.

The following subsections present the pseudocode of HANA.

## S5.1 Pseudocode for Computing the Similarity Score for an NOE Pattern

Let $(a_1, a_2, a_3, d)$ represent a *distance restraint* back-computed from a structure, where $a_1$ and $a_3$ are the involved protons in the structure, $a_2$ is the heavy atom covalently bound to the proton $a_1$, and $d$ is the distance between protons $a_1$ and $a_3$. Let $(p_1, p_2, p_3, I_p)$ denote an *experimental NOE peak* from a 3D NOESY spectrum, where $p_1$ and $p_3$ are frequencies of a pair of (unassigned) interacting protons, $p_2$ is the frequency of the heavy atom covalently bound to the first proton, and $I_p$ is the intensity of the cross peak. Let $(\omega(a_1), \omega(a_2), \omega(a_3), I(d))$ denote the *back-computed NOE peak* for a distance restraint $(a_1, a_2, a_3, d)$ back-computed from a structure, where $\omega(a_j)$ is the assigned chemical shift of atom $a_j$, $1 \leq j \leq 3$, and $I(d)$ is the back-computed peak intensity of distance $d$. Let $B$ be the back-computed NOE pattern, and let $Y$ be the experimental NOE spectrum. Let $\delta_j$ be the error tolerance in the NOE spectrum in the $j$th dimension, and let $\sigma_j$ be the uncertainty of the NOE peak position in the $j$th dimension, where $1 \leq j \leq 3$. The pseudocode for calculating the similarity score between $B$ and $Y$ is given in Algorithm 1. For each rotamer, the computation of its similarity score based on the Hausdorff distance using Algorithm 1 takes $O(mw)$ time, where $m$ is the number of back-computed NOE peaks, and $w$ is the total number of cross peaks in the experimental NOE spectrum.

**Algorithm 1 Similarity Score Calculation Based on the Hausdorff Distance**

Function **Hausdorff_Score** $(B, Y)$ /* $B$ is the back-computed NOE pattern, and $Y$ is the NOE spectrum. */
$x_0, x_{\max}, x, s, \theta \leftarrow 0$;
$m \leftarrow |B|$; /* $m$ is the number of back-computed NOE peaks. */
**for** each $(\omega(a_1), \omega(a_2), \omega(a_3), I(d)) \in B$ **do**
  **for** each $(p_1, p_2, p_3, I_p) \in Y$ **do**
    **if** $|p_1 - \omega(a_1)| < \delta_1$ **and** $|p_2 - \omega(a_2)| < \delta_2$ **and** $|p_3 - \omega(a_3)| < \delta_3$ **then**
      /* $\delta_j$ is the error tolerance in the NOE spectrum in the $j$th dimension, $j = 1, 2, 3$. */
      $x_0 \leftarrow \mathcal{N}(|I(d) - I_p|, \sigma_I) \prod_{j=1}^{3} \mathcal{N}(|\omega(a_j) - p_j|, \sigma_j)$;
      /* $\mathcal{N}(|x - \mu|, \sigma)$ is the probability of observing the difference $|x - \mu|$ with mean $\mu$ and deviation $\sigma$. */
      **if** $x_0 > x_{\max}$ **then**
        $x_{\max} \leftarrow x_0$;
      **end if**
    **end if**
  **end for**
  $x \leftarrow x + x_{\max}$;
**end for**
**return** $x/m$;

## S5.2 Pseudocode for NOE Assignment Algorithm HANA

The NOE assignment process is divided into three phases: initial NOE assignment (phase 1), rotamer selection (phase 2) and filtration of ambiguous NOE assignments (phase 3). In the initial NOE assignment phase, all possible ambiguous NOEs are assigned to a NOE cross peak when the resonances of corresponding atoms fall within a tolerance window around the NOE peak. In the rotamer selection phase, an extended model of the Hausdorff distance is used to measure the match between the back-computed NOE pattern and the experimental spectrum, and thus choose the ensemble of best rotamers with top match scores. In the last phase, ambiguous NOE assignments are filtered based on the structure obtained by combining the high-resolution backbone and the ensemble of computed rotamers. The final NOE assignments are fed into standard structure determination programs, such as XPLOR/CNS (Brünger 1992) for the structure calculation.

The following notations will be used in the description of our NOE assignment algorithm HANA (Algorithm 2). Let $y_i = (p_1, p_2, p_3, I_p)$ be an experimental NOE peak, where $p_1$ and $p_3$ are frequencies of a pair of (unassigned) interacting protons, $p_2$ is the frequency of the heavy atom covalently bound to the first proton, and $I_p$ is the intensity of the cross peak. Let $Y = \{y_1, \ldots, y_w\}$ denote the set of experimental NOE peaks, where $w$ is the total number of NOE peaks. Let $A_i$ denote the set of atom triples that are assigned to peak $y_i$. Let $\mathcal{A} = \{a_1, \ldots, a_q\}$ denote the set of all atoms (including all protons) in the protein, where $q$ is the total number of atoms. Let $L = \{\omega(a_1), \ldots, \omega(a_q)\}$ denote the set of chemical shifts for all atoms, where $\omega(a_i)$ is the chemical shift of atom $a_i$. Let $\delta_j$ denote the error tolerance in the $j$th dimension for the initial NOE ambiguous assignment, where $j = 1, 2, 3$. Let $n$ be the number of residues in the protein, and let $t$ be the maximum number of rotamer in a residue. Let $\mathbf{r_{ij}}$ denote the rotamer $j$ at residue $i$, where $i = 1, \ldots, n$, $j = 1, \ldots, t$. Let $u$ denote the NOE upper-limit distance bound. Let $\mathbf{P}$ denote the structure after combining the ensemble of computed rotamers with the backbone computed by RDC-EXACT, and let $d(\parallel a_1 - a_2 \parallel, \mathbf{P})$ denote the Euclidean distance between atoms $a_1$ and $a_2$ in the three-dimensional structure $\mathbf{P}$. Let $d_{min}(\parallel a_1 - a_2 \parallel, \mathbf{P})$ denote the minimum Euclidean distance between atoms $a_1$ and $a_2$ over all pairs of computed rotamers in the three-dimensional structure $\mathbf{P}$ Let $B_{ij} = \{b_1, \ldots, b_m\}$ denote the set of back-computed NOE peaks for rotamer $\mathbf{r_{ij}}$, where $m$ is the total number of back-computed NOE peaks, and $b_i = (\omega(a_1), \omega(a_2), \omega(a_3), I(d))$ denotes the back-computed NOE peak for a distance restraint $(a_1, a_2, a_3, d)$ from rotamer $\mathbf{r_{ij}}$. Let $s_{ij}$ denote the similarity score of rotamer $\mathbf{r_{ij}}$ based on the extended Hausdorff measure. Let $R_i$ denote the ensemble of top $k$ rotamers computed at residue $i$. Let $d(I_p)$ denote the distance calibrated from the peak intensity $I_p$.

The details of HANA are as follows (Algorithm 2). In Phase 1 (namely initial NOE assignment), for each cross peak $(p_1, p_2, p_3, I_p)$ in the NOE spectra, we search the resonance list and assign triple(s) of atoms $(a_1, a_2, a_3, d(I_p))$ to $(p_1, p_2, p_3, I_p)$ such that $p_1 - \delta_1 \leq \omega(a_1) \leq p_1 + \delta_1$, $p_2 - \delta_2 \leq \omega(a_2) \leq p_2 + \delta_2$, and $p_3 - \delta_3 \leq \omega(a_3) \leq p_3 + \delta_3$. In the rotamer selection phase, we first place all rotamers $\mathbf{r_{ij}}$ into backbone by rotation and translation computed based on the coordinates of $H^N$, $C^\alpha$ and N atoms. Then for each proton $a_3$ in rotamer $\mathbf{r_{ij}}$, we search the backbone structure and find all backbone protons $a_1$ that are within the NOE upper-bound limit from proton $a_3$ (an extra 2.5 Å is added as the correction of the upper-bound for every methyl group). In addition, we back compute all intra-residue NOEs for each rotamer. Next for each distance restraint $(a_1, a_2, a_3, d)$ computed from the structure, we calculate its back-computed NOE peak $(\omega(a_1), \omega(a_2), \omega(a_3), I(d))$ based on the mapping between each atom name $a$ and corresponding chemical shift $\omega(a)$ in the resonance list. Let $B_{ij} = \{(\omega(a_1), \omega(a_2), \omega(a_3), I(d))\}$ denote the set of all back-computed NOE peaks for rotamer $\mathbf{r_{ij}}$. We next call the function **Hausdorff_Score** to compute the match score between the NOE pattern $B_{ij}$ of rotamer $\mathbf{r_{ij}}$ and the experimental NOE spectrum $Y$. Finally we pick the top $k$ rotamers with highest similarity scores at each residue $i$. In Phase 3 (namely filtration of ambiguous NOE assignment), we first place the top $k$ rotamers (selected in the second phase) at each

residue into backbone, and then obtain a protein structure $\mathbf{P}$. Note that each side-chain atom in structure $\mathbf{P}$ has $k$ possible positions from the top $k$ computed rotamers. Next, for each initial NOE assignment $(a_1, a_2, a_3, d(I_p))$ obtained in the first phase, we measure the Euclidean distance between protons $a_1$ and $a_3$ in structure $\mathbf{P}$. Recall that $d_{min}(\parallel a_1 - a_2 \parallel, \mathbf{P})$ is the minimum Euclidean distance between atoms $a_1$ and $a_2$ over all pairs of computed rotamers in structure $\mathbf{P}$. In HANA, an NOE assignment $(a_1, a_2, a_3, d(I_p))$ (from the initial NOE assignment in Phase 1) is pruned, if $d_{min}(\parallel a_1 - a_2 \parallel, \mathbf{P})$ is larger than $d(I_p)$.

### S5.3 Filtering NOE Assignments Based on Low-resolution Structures

In the last phase, HANA takes as input the ambiguous NOE assignments, and uses the low-resolution structure to filter them. Each ambiguous NOE assignment is an OR of unambiguous NOE assignments. We convert each ambiguous NOE assignment into an OR over a set of unambiguous NOE assignments, and then discard the unambiguous NOE assignments that are inconsistent with the low-resolution structure computed by HANA.

When an ensemble of structures are used to filter violated NOE assignments, a *voting scheme* (Langmead & Donald 2004, Apaydin et al. 2008) is invoked to prune those NOE assignments that violate most of the structures. The voting scheme calculates a set of NOE assignments, called the *consensus NOE assignments*, that are consistent with a majority of the structures. These consensus NOE assignments obtained from the voting scheme are then input to XPLOR, in order to compute the subsequent ensemble of structures.

## S6 Time Complexity Analysis of HANA

In this section, we will analyze the time complexity of our NOE assignment algorithm HANA (Algorithm 2). We first state the theorem about the time complexity of HANA and then provide the proof.

***Theorem* 1.** HANA *runs in $O(tn^3 + tn \log t)$ time, where $t$ is the maximum number of rotamers at a residue and $n$ is the total number of residues in the protein sequence.*

*Proof.* To analyze the algorithmic complexity of our NOE assignment algorithm, we first recall some notations defined previously. Let $n$ be the number of residues in the protein sequence, and let $w$ denote the total number of cross peaks in the experimental NOE data. Let $t$ denote the maximum number of rotamers for every amino acid in the rotamer library. Let $\xi$ denote the maximum number of atoms per residue. Let $q$ be the total number of atoms in the protein, then $q = O(\xi n)$.

The running time of the initial NOE assignment phase is bounded by $O(wq^2)$ steps. In Phase 2, the initialization in lines $1-7$ takes $O(tn)$ time. Since the number of protons in the backbone is bounded by $O(n)$, the total number of protons in a rotamer is less than $\xi$, the loop in lines $11-19$ needs $O(\xi n)$ steps. The function **Hausdorff_Score** takes $O(mw)$ time to compute the similarity score between the back-computed NOE pattern $B_{ij}$ and the experimental NOE spectrum $Y$, where $m$ is the number of back-computed NOE peaks in $B_{ij}$. Hence, the loop in lines $9-21$ runs in $O\big(t(n\xi + mw)\big)$ time. Sorting all rotamers and selecting top $k$ rotamers in lines $22-23$ only requires $O(t \log t)$ time. Thus, the overall running time for Phase 2 is $O(tn) + n \cdot O\big(t(mw + \xi n)\big) + n \cdot O(t \log t) = O\big(tn(mw + \xi n) + tn \log t\big)$. In Phase 3 (namely the filtration of ambiguous NOE assignment), placing all rotamers into the backbone (in lines $1-5$) takes $O(kn)$ time. In worst case, $|A_i|$ is bounded by $O(q^2)$, where $q$ is the total number of atoms in the proteins. Hence the total running time for lines $6-12$ is $O(wq^2)$. Thus, Phase 3 runs in $O(kn + wq^2)$ time.

Therefore, the overall running time for HANA is $O(wq^2) + O\big(tn(mw + \xi n) + tn \log t\big) + O(kn + wq^2) = O\big(wq^2 + tn(mw + \xi n) + tn \log t\big)$.

In general, it is safe to assume the number of atoms in a residue is a constant, that is, $\xi = O(1)$. Thus, $q = O(\xi n) = O(n)$. Also, since each proton can only have NOE interactions with a constant number of other protons within 6.0 Å distance, we have $w = O(n)$ and $m = O(n)$. Therefore, the running time of HANA is $O(tn^3 + tn \log t)$ in the worst case. □

# S7 The Local Minimization Approach and NOE Assignment for Loop Regions

Since HANA only uses NOE patterns and modal rotamers to compute the side-chain conformations, some steric clashes might exist between side-chains in the low-resolution structure computed by HANA. We add the loop regions to the SSE structures, and refine the side-chain conformations previously-computed from HANA using the following *local minimization approach*: The core structure, namely previously-packed SSE backbones, is fixed as a rigid body, and only the side-chains and loops are allowed to move during the minimization (Fig. 1 D in the main article). The reason for doing this is that we are more confident in the SSE backbones that we previously determined by RDC-EXACT, while the side-chain conformations are still at low resolution. In addition, the local minimization approach can alleviate the steric clash between side-chains. The algorithm uses the empirical molecular mechanics scoring function from XPLOR, including the NOE restraints, dihedral angle restraints, plus XPLOR's empirical energy terms, such as bond angle, covalent bond, electrostatic, van der Waals, improper torsion terms, to find the full conformations (i.e. complete structures with side-chains, including both loops and SSEs) with the lowest energies.

We use an iterative process, namely NOE-assignment and structure-calculation iteration, for NOE assignment in the loop regions. Sparse unambiguous NOEs in the loop regions (viz., where at least one proton of the NOE is in a loop region) are extracted using a procedure similar to that in the long-range NOE extraction for SSE packing. The set of unambiguous NOE restraints are fed into XPLOR (with the fixed core structure) to calculate the loop structures.

# S8 Results on NOE Assignment and Structure Calculation

In this section, we give additional details supplementing the *Results* section of the main article.

## S8.1 Evaluation of SSE Backbones Determined from RDCs

For all four proteins, we used CH and NH RDCs measured in one medium to estimate the alignment tensor as previously described in (Wang & Donald 2004*b*, Wang et al. 2006). Given the alignment tensor, we applied the extended version of RDC-EXACT (as described in the Methods section) to compute the conformations and orientations of $\alpha$-helices and $\beta$-sheets for proteins pol $\eta$ UBZ, ubiquitin, FF2 and hSRI, using CH and NH RDCs. For FF2, we first used NH and CH RDCs in a systematic search (Wang & Donald 2004*b*) to compute and enumerate all possible conformations from the polynomial RDC equations, and then incorporated two other RDCs, namely NC′ and $C^{\alpha}C'$ RDCs, to prune the conformations whose back-computed NC′ or $C^{\alpha}C'$ RDCs deviate $\geq 5.0$ Hz RMSD from the experimental RDCs. The RDC RMSD for NC′ and $C^{\alpha}C'$ bond vectors is also incorporated in the RDC-EXACT scoring function. This allows the algorithm to search over all possible conformations and find the optimal solution that best fits the RDC data. The backbone conformation

computed by RDC-EXACT is the global optimum, in that the combinatorial search is performed over each entire structure fragment, and the scoring function guarantees the resulting solution best fits the experimental data.

Fig. S2 shows the plot of back-computed vs. experimental RDCs for ubiquitin hSRI and pol $\eta$ UBZ.

### S8.2 Quality of SSE Packing

To assess the quality of the packed structures computed by PACKER, we analyzed the ensemble of packed SSE structures and compared them with the SSE regions of the corresponding reference structures solved either by traditional NMR approaches or by X-ray crystallography. We focused on the ensemble of *well-packed satisfying* (WPS) structures that have both high-quality NOE satisfaction and packing scores. We computed the mean structure of the WPS ensemble and compared the mean structure with the reference structure. Note that the scheme of selecting WPS structures from a complete set of structures consistent with the experimental data has been used in the structure determination of symmetric homo-oligomers (Potluri et al. 2006, 2007). Unlike (Potluri et al. 2006, 2007) in which the side-chains were fixed before packing, here our search space also includes all possible rotamer conformations. Fig. S3 shows the evaluation of packed structures computed by PACKER for ubiquitin, hSRI and pol $\eta$ UBZ.

### S8.3 Evaluation of Rotamers Computed by HANA

Rotamers usually represent a statistical mode of side-chain conformations in torsion angle space, in a local energy well. In general, rotamers are classified into different bins according to their distribution over $\chi$ angles. Here we used the same classification rule as in (Lovell et al. 2000) for rotamer identification and comparison, in which values $\pm 30^\circ$ are used in determining most $\chi$ angle ranges, and a few specific values are used in determining several terminal $\chi$ angle boundaries. For hydrophilic residues, which are usually on surface of the protein, we used the window $\pm 40^\circ$ in determining $\chi$-angle boundaries. We did not use the RMSD measurement to compare different rotamers, because most rotamers are short, and the RMSD is not sufficient to measure the conformational dissimilarity between two rotamers.

We carefully examined and compared each individual rotamer computed by HANA vs. its corresponding side-chain conformations in the NMR reference structure (PDB ID: 1D3Z) and the X-ray structure (PDB ID: 1UBQ). We consider two rotamers *equivalent* if the difference between their corresponding $\chi$ angles are within $\pm 30^\circ$ (or $\pm 40^\circ$ for hydrophilic residues, and we choose the same ranges as in Table 1 in (Lovell et al. 2000) for terminal $\chi$ angles). In our ubiquitin test, rotamers were called *consistent* if they are equivalent with either X-ray or NMR reference side-chains. Otherwise, they are called *inconsistent*. The results on rotamers computed by HANA for ubiquitin are shown in Fig. S4 and Fig. S5, in which each line lists all $\chi$ angles in rotamers computed by HANA, side-chains in X-ray and NMR reference structures respectively. The $\chi$ angles in consistent rotamers or side-chains are shown in either green or yellow, while $\chi$ angles of inconsistent rotamers are shown in either magenta or red. HANA can select more than 70% of rotamers that are consistent with either X-ray or NMR reference structure (Fig. S4 and Fig. S5).

Next, we examined the $\chi_1/\chi_2$ angle distribution for all leucine rotamers in ubiquitin computed by HANA. All HANA-computed leucine rotamers were consistent with either X-ray or NMR reference structures (SM Fig. S6). Although the rotamer computed by HANA in residue L43 is different from their corresponding X-ray side-chains, it is consistent with the NMR reference structure. As pointed out in (Lovell et al. 2000), the side-chain conformation of L43 is possibly incorrect, and is caused by the misfitting of models to electron density maps. As suggested in (Lovell et al. 2000), the possible conformation for L43 should be in the top-right box (A), which is actually consistent with the rotamers computed by our algorithm (they are

also consistent with the side-chains in the NMR reference structure). For residue L67, although our computed rotamer is different from the NMR reference structure, it agrees with the X-ray conformation, which indicates that both rotamers might exist in different states of the protein.

HANA takes as input a backbone structure computed by RDC-EXACT. To test the sensitivity of HANA to variations in the backbone structure, we ran three independent tests on our rotamer selection algorithm using different ubiquitin backbones, that is, all input parameters were the same except for the backbone structure. We first used RDC-EXACT to generate two ubiquitin backbone structures. These structures were within RMSD 1.17 Å and 1.79 Å to the X-ray backbone structure, respectively. These two backbones and the X-ray backbone are then used as the input backbone structure in three independent tests correspondingly. SM Fig. S7 shows the fraction of consistent rotamers computed by HANA in these three tests (using different input backbone structures). As illustrated in Fig. S7, the number of consistent rotamers does not vary significantly with the backbone resolution (the variance is less than 10% of the consistent rotamers), which indicates that our rotamer selection algorithm is not sensitive to small variations in the backbone conformation.

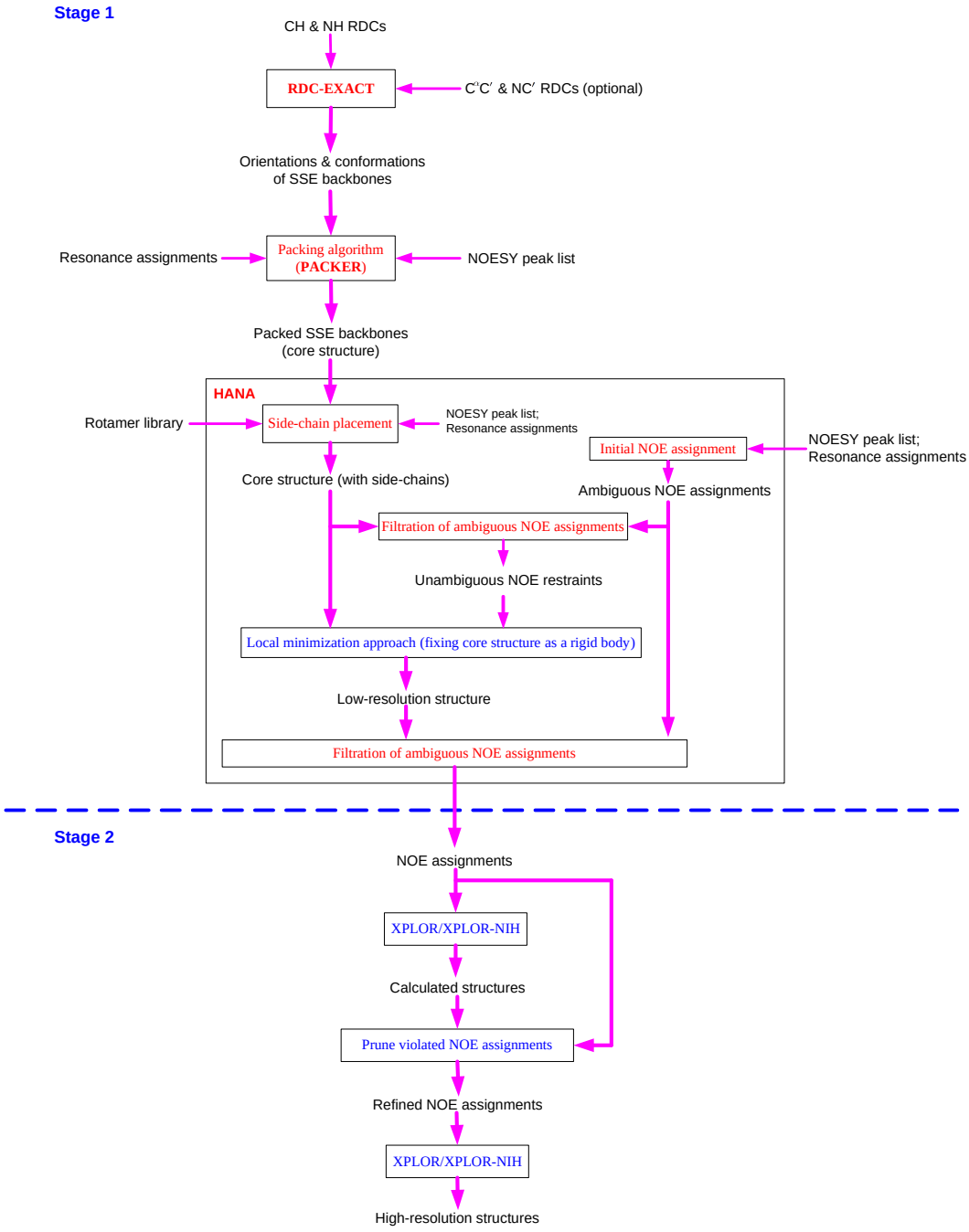


Figure S1: Flow chart of RDC-PANDA.

## Algorithm 2 Hausdorff-based NOE Assignment (HANA)

Given $L$, $Y$, backbone, rotamer library. /* $L$ is the assigned resonance list, and $Y$ is the experimental NOE spectrum. */

**Phase 1 (Initial NOE Assignment):**

**for** $i \leftarrow 1$ **to** $w$ **do** /* $w$ is the number of experimental peaks in the NOE spectrum. */
  $A_i \leftarrow \emptyset$; /* Initialization of NOE assignment for each NOE peak. */
**end for**
**for** $i \leftarrow 1$ **to** $w$ **do**
  **for** $j \leftarrow 1$ **to** $q$ **do** /* $q$ is the number of protons in the protein. */
    $a'_j \leftarrow$ heavy atom bond-connected to $a_j$;
    **for** $k \leftarrow 1$ **to** $q$ **do**
      **if** $|p_1 - \omega(a_j)| < \delta_1$ **and** $|p_2 - \omega(a'_j)| < \delta_2$ **and** $|p_3 - \omega(a_k)| < \delta_3$ **then**
        $A_i \leftarrow A_i \cup \{(a_j, a'_j, a_k, d(I_p))\}$;
      **end if**
    **end for**
  **end for**
**end for**

**Phase 2 (Rotamer Selection):**

**for** $i \leftarrow 1$ **to** $n$ **do** /* $n$ is the number of residues in the protein. */
  $R_i \leftarrow \emptyset$; /* Initialization for the set of computed rotamers at residue $i$. */
  **for** $j \leftarrow 1$ **to** $t$ **do** /* $t$ is the maximum number of rotamers per residue. */
    $B_{ij} \leftarrow \emptyset$; /* Initialization for the back-computed NOE pattern for rotamer $j$ at residue $i$. */
    $s_{ij} \leftarrow 0$; /* Initialization for the similarity score of the back-computed NOE pattern $B_{ij}$. */
  **end for**
**end for**
**for** $i \leftarrow 1$ **to** $n$ **do**
  **for** $j \leftarrow 1$ **to** $t$ **do**
    structure $\mathbf{P} \leftarrow$ rotate and translate rotamer $\mathbf{r_{ij}}$ into backbone;
    **for** each proton $a_3 \in \mathbf{r_{ij}}$ **do** /* $\mathbf{r_{ij}}$ is the rotamer $j$ at residue $i$. */
      **for** each proton $a_1 \in$ backbone or side-chain in residue $i$ **do**
        $a_2 \leftarrow$ heavy atom bond-connected to $a_1$;
        **if** $d(\| a_1 - a_3 \|,\mathbf{P}) < u$ **then**
          /*$d(\| a_1 - a_3 \|,\mathbf{P})$ is the Euclidean dist. betw. protons $a_1$ and $a_3$ in $\mathbf{P}$, and $u$ is the NOE upper-bound. */
          $B_{ij} \leftarrow B_{ij} \cup \{(\omega(a_1), \omega(a_2), \omega(a_3), I(d(\| a_1 - a_3 \|, \mathbf{P})))\}$
        **end if**
      **end for**
    **end for**
    $s_{ij} \leftarrow$ **Hausdorff_Score**$(B_{ij}, Y)$; /* Compute the similarity score between $B_{ij}$ and $Y$ (see Algorithm 1). */
  **end for**
  sort all rotamers $\{\mathbf{r_{ij}} | j = 1, \ldots, t\}$ in descending order of scores $s_{ij}$;
  $R_i \leftarrow$ top $k$ rotamers in $\{\mathbf{r_{ij}} | j = 1, \ldots, t\}$;
**end for**

**Phase 3 (Filtration of Ambiguous NOE Assignment):**

**for** $i \leftarrow 1$ **to** $n$ **do**
  **for** each rotamer $\mathbf{r} \in R_i$ **do** /* $R_i$ is the set of computed rotamers from Phase 2. */
    structure $\mathbf{P} \leftarrow$ rotate and translate $\mathbf{r}$ into backbone
  **end for**
**end for**
**for** $i \leftarrow 1$ **to** $w$ **do**
  **for** each $(a_1, a_2, a_3, d(I_p)) \in A_i$ **do** /* $A_i$ is the set of initial NOE assignments from Phase 1. */
    **if** $d_{min}(\| a_1 - a_3 \|,\mathbf{P}) > d(I_p)$ **then**
      $A_i = A_i \setminus \{(a_1, a_2, a_3, d(I_p))\}$
    **end if**
  **end for**
**end for**
**return** $A_1 \cup \ldots \cup A_w$

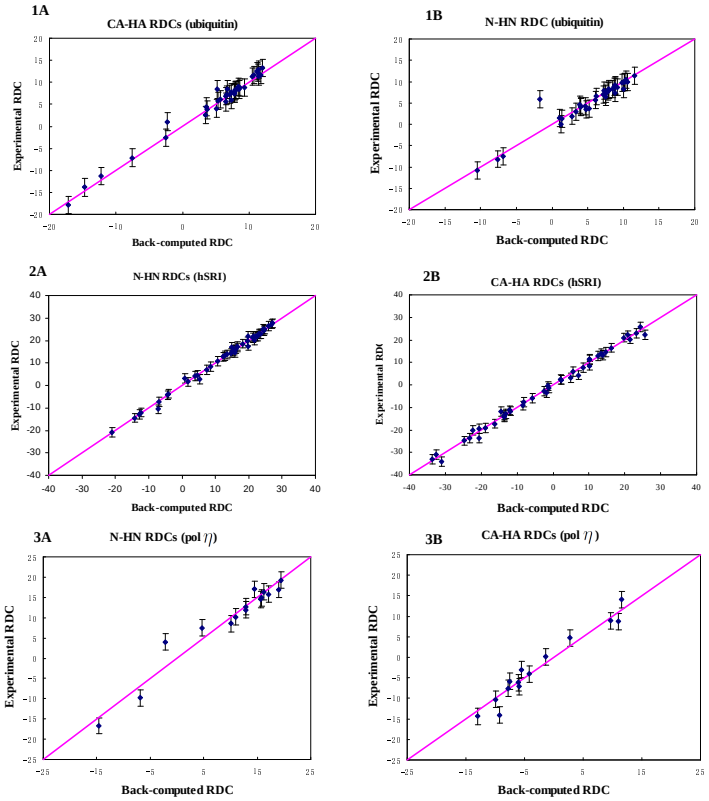


Figure S2: Back-computed vs. experimental RDCs. Panels 1A and 1B: CH and NH RDCs for ubiquitin. Panels 2A and 2B: CH and NH RDCs for hSRI. Panels 3A and 3B: CH and NH RDCs for pol $\eta$ UBZ. All RDCs are scaled to the NH RDCs; a window of 2.0 Hz is shown as the error bars for the experimental RDCs.

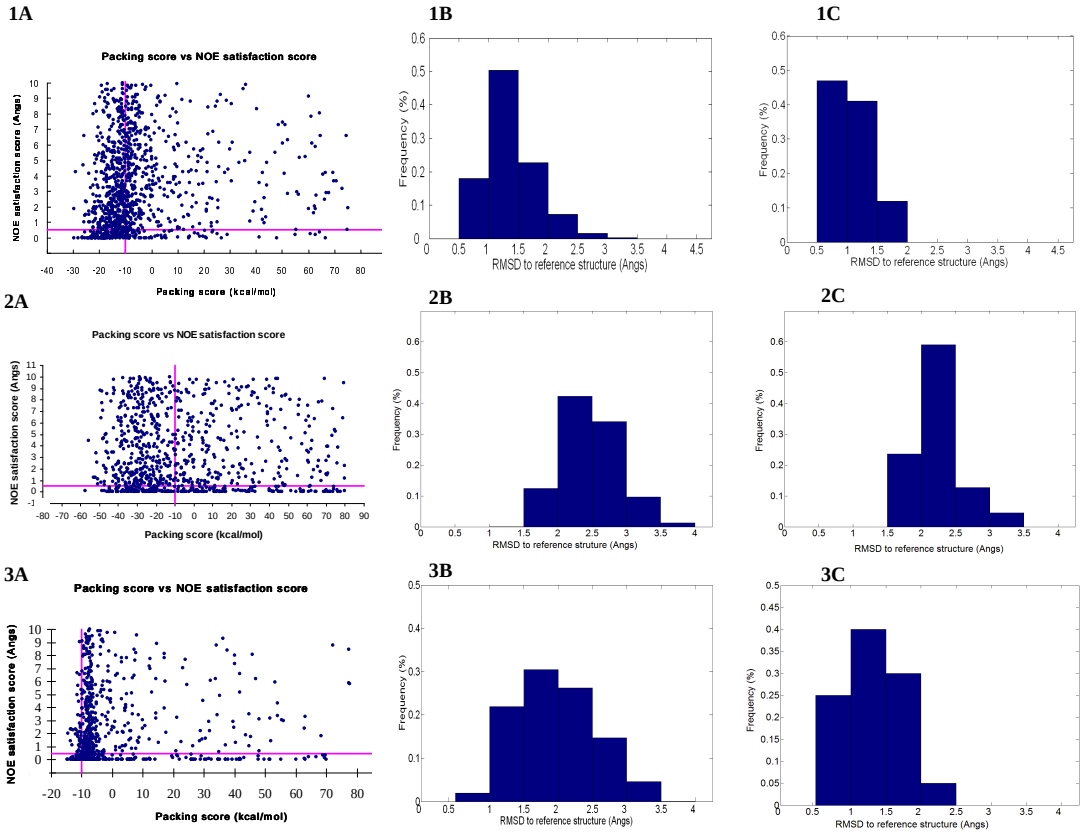


Figure S3: Evaluation of packed structures computed by PACKER. Row 1: results for ubiquitin; Row 2: results for hSRI; Row 3: results for pol $\eta$ UBZ. Column 1 (Panels 1A, 2A and 3A): NOE satisfaction score vs. packing score for all structures in the ensemble (structures with vdW energies larger than 80 and NOE score larger than 10 were truncated from the plot). Column 2 (Panels 1B, 2B and 3B): histogram of backbone RMSD to the reference structure for all packed structures. Column 3 (Panels 1C, 2C and 3C): histogram of backbone RMSD to the reference structures for WPS structures. The magenta lines show the cutoffs of NOE satisfaction score (horizontal) and packing score (vertical) for computing the WPS structures.

| Residue | χ angles of rotamers from HANA | | | | χ angles of X-ray conformations | | | | χ angles of reference NMR conformations | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\chi^1$ | $\chi^2$ | $\chi^3$ | $\chi^4$ | $\chi^1$ | $\chi^2$ | $\chi^3$ | $\chi^4$ | $\chi^1$ | $\chi^2$ | $\chi^3$ | $\chi^4$ |
| Q2 | -69.1 | 180 | -25 | | 166 | 178 | 14.4 | | -70.2 | 179 | -31.9 | |
| F4 | -66.7 | -84.9 | | | -60.6 | 97.9 | | | -62.2 | -81 | | |
| V5 | 172 | | | | -180 | | | | -176 | | | |
| T7 | 59.6 | | | | 76.7 | | | | 70.7 | | | |
| L8 | -67.3 | 174 | | | -68.2 | -166 | | | -55.6 | 167 | | |
| T9 | 59.4 | | | | 70.7 | | | | 67 | | | |
| G10 | N/A | | | | N/A | | | | N/A | | | |
| K11 | -180 | -180 | 180 | 180 | -179 | -179 | -151 | -49.4 | -170 | 179 | 175 | 168 |
| T12 | -66.9 | | | | -63.4 | | | | -60 | | | |
| I13 | -66.4 | 170 | | | 125.8 | -175 | | | -40 | -173 | | |
| T14 | -66 | | | | 66.5 | | | | -59.8 | | | |
| L15 | -66 | 174 | | | -118 | 30.2 | | | -64.6 | 167 | | |
| E16 | 178 | 180 | -60 | | -75 | -168 | 140 | | -172 | -179 | -8.5 | |
| V17 | -61 | | | | -62.6 | | | | -62.1 | | | |
| P19 | 30.6 | | | | 38.3 | | | | 29.2 | | | |
| S20 | -178 | | | | 35.7 | | | | 128 | | | |
| D21 | -65.7 | -60 | | | -80.6 | -19.8 | | | -77.8 | -23 | | |
| T22 | 59.7 | | | | 60.6 | | | | 62 | | | |
| E24 | -180 | 180 | -60 | | -170 | -175 | 72.8 | | -180 | -170 | -4.7 | |
| V26 | 172 | | | | 172.2 | | | | 168 | | | |
| K27 | -68.1 | 180 | 180 | -64.9 | -71.6 | -174 | 179 | 179 | -67.1 | -146 | -157 | -107 |
| A28 | N/A | | | | N/A | | | | N/A | | | |
| I30 | -67.6 | 170 | | | -70.9 | 167 | | | -70 | -178 | | |
| G35 | N/A | | | | N/A | | | | N/A | | | |
| I36 | -68.1 | 170 | | | -54.5 | 161 | | | -51.6 | 156 | | |
| P37 | 37 | | | | 39.7 | | | | 27.3 | | | |
| P38 | 36.9 | | | | 41.6 | | | | 24 | | | |
| D39 | -63.4 | -60 | | | 135 | -61.7 | | | -27.6 | -44 | | |
| Q40 | -66.3 | 180 | -25 | | -61.9 | -179 | -165 | | -67.1 | -150 | 7 | |
| Q41 | -68.4 | 180 | -60 | | -53.1 | -180 | 27.9 | | -70.1 | -176 | 83.2 | |
| L43 | -67 | 174 | | | -68.3 | 167 | | | -68.5 | 172 | | |
| I44 | -66 | 170 | | | -48.9 | -58.1 | | | -69 | 131 | | |
| F45 | -178 | 80 | | | 178 | 78.2 | | | 176 | 82.7 | | |
| A46 | N/A | | | | N/A | | | | N/A | | | |
| G47 | N/A | | | | N/A | | | | N/A | | | |
| L50 | -68 | 174 | | | -50.8 | -173 | | | -51.6 | 177 | | |
| E51 | -67.9 | 180 | -10 | | 173.1 | 141 | 172 | | -69.3 | -178 | -13.8 | |
| D52 | -66.1 | -60 | | | -77.9 | 178 | | | -68.8 | 13.8 | | |
| G53 | N/A | | | | N/A | | | | N/A | | | |
| T55 | 59.7 | | | | 59.9 | | | | 64 | | | |
| L56 | -52.7 | -67.6 | | | -61.3 | -60.3 | | | -60 | -62 | | |
| S57 | 61.3 | | | | 59.3 | | | | 64.4 | | | |
| D58 | -66 | -60 | | | -73.1 | 157 | | | -69.7 | -29 | | |
| Y59 | -65.8 | -85 | | | -63.3 | 104 | | | -68.6 | -83 | | |
| N60 | -66.2 | -20.9 | | | -159 | -145 | | | -59.2 | -53 | | |
| I61 | -67.2 | 170 | | | -69.7 | -179 | | | -65.5 | 172 | | |
| Q62 | -69.8 | 180 | -25 | | -79.4 | -162 | -148 | | -68.1 | -142 | -1.9 | |
| E64 | -67.8 | 180 | -10 | | -72.2 | 127 | 36.6 | | -63.3 | -156 | -12 | |
| L67 | -67.8 | 174 | | | -43.5 | -171 | | | -81 | 67.2 | | |
| H68 | -67.6 | -69 | | | -69.1 | -88.8 | | | -67.5 | 90.4 | | |
| L69 | -179 | 65 | | | 172.8 | 65.9 | | | 176 | 66.2 | | |
| V70 | 173 | | | | 177 | | | | 154 | | | |

Figure S4: All $\chi$ angles of consistent rotamers for ubiquitin computed by HANA in the low-resolution structure. The backbone has an RMSD of 1.74 Å from the X-ray structure.

| Residue | χ angles of rotamers from HANA | | | | χ angles of X-ray conformations | | | | χ angles of reference NMR conformations | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | χ1 | χ2 | χ3 | χ4 | χ1 | χ2 | χ3 | χ4 | χ1 | χ2 | χ3 | χ4 |
| I3 | -65 | 170 | | | 60.3 | 162 | | | 63.6 | 167 | | |
| K6 | -68.5 | -180 | 180 | 180 | -90.7 | -176 | -128 | 179 | -176 | 176 | -172 | -174 |
| E18 | -67.5 | 180 | -10 | | -73.8 | -55 | -18 | | -66 | -34.2 | 2.7 | |
| I23 | -67.6 | 170 | | | -61 | -55 | | | -66 | 58.7 | | |
| N25 | -67.4 | -20.9 | | | 178 | 53.8 | | | -135 | 17.7 | | |
| K29 | -68 | -180 | -180 | 180 | -60.8 | 169 | 149 | -89 | -70 | 179 | 179 | -97.1 |
| Q31 | -68.5 | 180 | -60 | | 177 | 172 | 35.2 | | 178 | 178 | 29.5 | |
| D32 | -71.9 | -15 | | | -150 | 65.5 | | | -140 | 9.6 | | |
| K33 | -177.3 | 180 | -180 | -180 | 78.3 | 138 | 62.3 | -165 | -56 | -132 | -175 | 173 |
| E34 | -67.6 | 180 | -10 | | -60.1 | -77 | -36 | | -66 | -70.2 | -20 | |
| R42 | -67.8 | -167 | -65 | -85 | 162 | 174 | 174 | -107 | -82 | 178 | -176 | -154 |
| K48 | -177.7 | 180 | -180 | -180 | -61.5 | 174 | -112 | -58.8 | -47 | 176 | -99.7 | -169 |
| Q49 | -178 | 180 | -60 | | -173 | 177 | 119 | | -129 | 173 | 149 | |
| R54 | -67.4 | -180 | 180 | -180 | -55.9 | -72 | 114 | -128 | -63 | -64.1 | -159 | -159 |
| K63 | 178.2 | 180 | 180 | -180 | 49.7 | 166 | 154 | 72.9 | 58.5 | -173 | -173 | 179 |
| S65 | 61.4 | | | | -73.3 | | | | -72 | | | |
| T66 | -177.7 | | | | -60 | | | | -61 | | | |

Figure S5: All $\chi$ angles of inconsistent rotamers for ubiquitin computed by HANA in the low-resolution structure. The backbone used by HANA is the same as in Fig. S4.

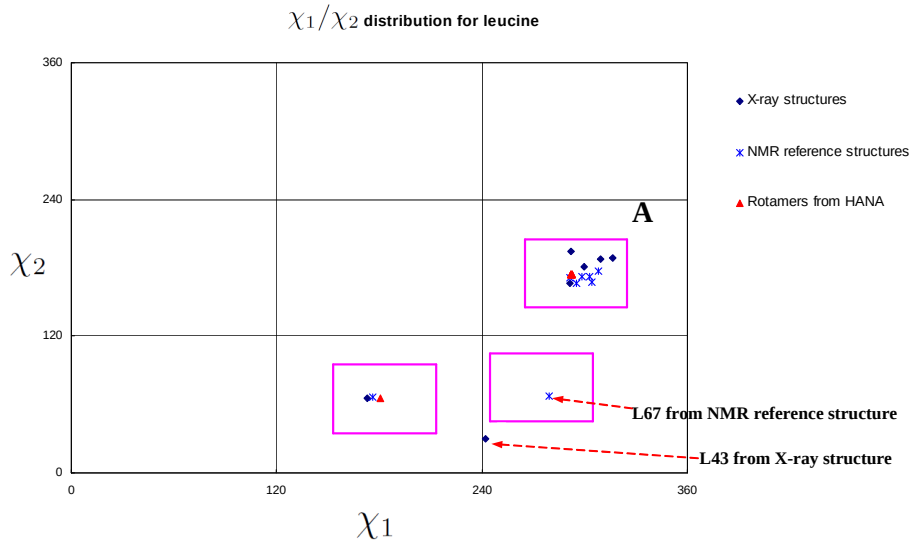


Figure S6: The $\chi_1/\chi_2$ angle distribution for leucine residues in ubiquitin's rotamers computed by HANA. Boxes in magenta represent the dominant leucine rotamers in the high-resolution structure database (Lovell et al. 2000). Residues shown in the figure include L8, L15, L43, L50, L56, L67, L69.

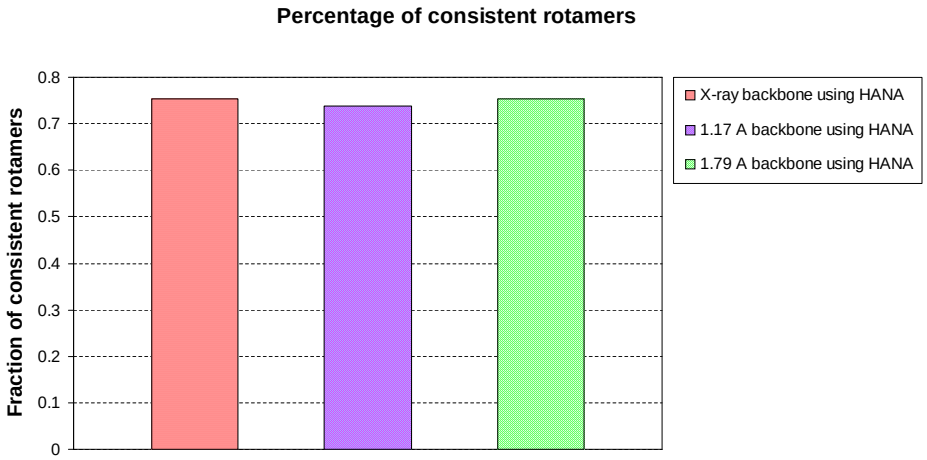


Figure S7: Percentage of consistent rotamers computed by HANA using different backbone structures as input.

## Suplementary References


Andrec, M., Du, P. & Levy, R. M. (2004), 'Protein backbone structure determination using only residual dipolar couplings from one ordering medium', *Journal of Biomolecular NMR* **21**, 335–347.

Apaydin, S., Conitzer, V. & Donald, B. R. (2008), 'Structure-Based Protein NMR Assignments using Native Structural Ensembles', *Journal of Biomolecular NMR* **40**, 263–276.

Brünger, A. T. (1992), 'X-PLOR, Version 3.1: a system for X-ray crystallography and NMR', *Journal of the American Chemical Society* .

Delaglio, F., Kontaxis, G. & Bax, A. (2000), 'Protein structure determination using molecular fragment replacement and NMR dipolar couplings', *J. Am. Chem. Soc.* **122**, 2142–2143.

Donald, B. R. & Martin, J. (2008), 'Automated NMR assignment and protein structure determination using sparse dipolar coupling constraints', *Progress in NMR Spectroscopy.* [Epub ahead of Print] doi:10.1016/j.pnmrs.2008.12.001.

Fowler, C. A., Tian, F., Al-Hashimi, H. M. & Prestegard, J. H. (2000), 'Rapid determination of protein folds using residual dipolar couplings', *Journal of Molecular Biology* **304**, 447–460.

Güntert, P. (2003), 'Automated NMR Protein Structure Determination', *Progress in Nuclear Magnetic Resonance Spectroscopy* **43**, 105–125.

Herrmann, T., Güntert, P. & Wüthrich, K. (2002), 'Protein NMR Structure Determination with Automated NOE Assignment Using the New Software CANDID and the Torsion Angle Dynamics Algorithm DYANA', *Journal of Molecular Biology* **319**(1), 209–227.

Huang, Y. J., Tejero, R., Powers, R. & Montelione, G. T. (2006), 'A topology-constrained distance network algorithm for protein structure determination from NOESY data', *Proteins: Structure Function and Bioinformatics* **62**(3), 587–603.

Hus, J. C., Marion, D. & Blackledge, M. (2001), 'Determination of protein backbone structure using only residual dipolar couplings', *J. Am. Chem. Soc.* **123**, 1541–1542.

Huttenlocher, D. P. & Jaquith, E. W. (1995), Computing visual correspondence: Incorporating the probability of a false match, *in* 'Procedings of the Fifth International Conference on Computer Vision (ICCV 95)', pp. 515–522.

Huttenlocher, D. P. & Kedem, K. (1992), *Distance Metrics for Comparing Shapes in the Plane*, In B. R. Donald and D. Kapur and J. Mundy, editors, Symbolic and Numerical Computation for Artificial Intelligence, pages 201-219, Academic press.

Huttenlocher, D. P., Klanderman, G. A. & Rucklidge, W. (1993), 'Comparing Images Using the Hausdorff Distance', *IEEE Trans. Pattern Anal. Mach. Intell.* **15**(9), 850–863.

Kuszewski, J., Schwieters, C. D., Garrett, D. S., Byrd, R. A., Tjandra, N. & Clore, G. M. (2004), 'Completely automated, highly error-tolerant macromolecular structure determination from multidimensional nuclear overhauser enhancement spectra and chemical shift assignments', *J. Am. Chem. Soc.* **126**(20), 6258–6273.

Kuszewski, J., Thottungal, R., Clore, G. & Schwieters, C. (2008), 'Automated error-tolerant macromolecular structure determination from multidimensional nuclear Overhauser enhancement spectra and chemical shift assignments: improved robustness and performance of the PASD algorithm', *J. Biomol. NMR* **41**(4), 221–239.

Langmead, C. & Donald, B. (2004), 'An expectation/maximization nuclear vector replacement algorithm for automated NMR resonance assignments', *J. Biomol. NMR* **29**(2), 111–138.

Linge, J. P., Habeck, M., Rieping, W. & Nilges, M. (2003), 'ARIA: Automated NOE assignment and NMR structure calculation', *Bioinformatics* **19**(2), 315–316.

Lovell, S. C., Word, J. M., Richardson, J. S. & Richardson, D. C. (2000), 'The Penultimate Rotamer Library', *Proteins: Structure Function and Genetics* **40**, 389–408.

Mumenthaler, C., Güntert, P., Braun, W. & Wüthrich, K. (1997), 'Automated combined assignment of NOESY spectra and three-dimensional protein structure determination', *Journal of Biomolecular NMR* **10**(4), 351–362.

Nilges, M., Macias, M. J., O'Donoghue, S. I. & Oschkinat, H. (1997), 'Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from $\beta$-spectrin', *Journal of Molecular Biology* **269**(3), 408–422.

Potluri, S., Yan, A. K., Chou, J. J., Donald, B. R. & Bailey-Kellogg, C. (2006), 'Structure Determination of Symmetric Homo-oligomers by a Complete Search of Symmetry Configuration Space using NMR Restraints and van der Waals Packing', *Proteins* **65**, 203–219.

Potluri, S., Yan, A. K., Donald, B. R. & Bailey-Kellogg, C. (2007), 'A complete algorithm to resolve ambiguity for intersubunit NOE assignment in structure determination of symmetric homo-oligomers', *Protein Science* **16**, 69–81.

Prestegard, J. H., Bougault, C. M. & Kishore, A. I. (2004), 'Residual Dipolar Couplings in Structure Determination of Biomolecules', *Chemical Reviews* **104**, 3519–3540.

Rohl, C. A. & Baker, D. (2002), 'De Novo Determination of Protein Backbone Structure from Residual Dipolar Couplings Using Rosetta', *J. Am. Chem. Soc.* **124**, 2723 –2729.

Ruan, K., Briggman, K. B. & Tolman, J. R. (2008), 'De novo determination of internuclear vector orientations from residual dipolar couplings measured in three independent alignment media', *Journal of Biomolecular NMR* **41**, 61–76.

Saupe, A. (1968), 'Recent results in the field of liquid crystals', *Angew. Chem.* **7**, 97–112.

Tian, F., Valafar, H. & Prestegard, J. H. (2001), 'A dipolar coupling based strategy for simultaneous resonance assignment and structure determination of protein backbones', *J Am Chem Soc.* **123**, 11791–11796.

Tjandra, N. & Bax, A. (1997), 'Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium', *Science* **278**, 1111–1114.

Tolman, J. R., Flanagan, J. M., Kennedy, M. A. & Prestegard, J. H. (1995), 'Nuclear magnetic dipole interactions in field-oriented proteins: Information for structure determination in solution', *Proc. Natl. Acad. Sci. USA* **92**, 9279–9283.

Wang, L. & Donald, B. R. (2004*a*), Analysis of a Systematic Search-Based Algorithm for Determining Protein Backbone Structure from a Minimal Number of Residual Dipolar Couplings, *in* 'Proceedings of The IEEE Computational Systems Bioinformatics Conference (CSB), Stanford CA (August, 2004)', pp. 319–330. PMID: 16448025.

Wang, L. & Donald, B. R. (2004*b*), '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', *Jour. Biomolecular NMR* **29**(3), 223–242.

Wang, L., Mettu, R. & Donald, B. R. (2006), 'A Polynomial-Time Algorithm for De Novo Protein Backbone Structure Determination from NMR Data', *Journal of Computational Biology* **13**(7), 1276–1288.