Improvement and analysis of computational methods for prediction of residual dipolar couplings

Konstantin Berlin [a,b], Dianne P. O'Leary [b,c], David Fushman [a,c,*]

[a] Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Organization, University of Maryland, College Park, MD 20742, USA
[b] Department of Computer Science, University of Maryland, College Park, MD 20742, USA
[c] Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA

ARTICLE INFO

ABSTRACT

We describe a new, computationally efficient method for computing the molecular alignment tensor based on the molecular shape. The increase in speed is achieved by re-expressing the problem as one of numerical integration, rather than a simple uniform sampling (as in the PALES method), and by using a convex hull rather than a detailed representation of the surface of a molecule. This method is applicable to bicelles, PEG/hexanol, and other alignment media that can be modeled by steric restrictions introduced by a planar barrier. This method is used to further explore and compare various representations of protein shape by an equivalent ellipsoid. We also examine the accuracy of the alignment tensor and residual dipolar couplings (RDC) prediction using various *ab initio* methods. We separately quantify the inaccuracy in RDC prediction caused by the inaccuracy in the orientation and in the magnitude of the alignment tensor, concluding that orientation accuracy is much more important in accurate prediction of RDCs.

1. Introduction

Knowledge of protein structure plays a critical role in our understanding of the molecular mechanisms underlying biological processes. One of the main methods for obtaining structural information at atomic-level resolution is the use of nuclear magnetic resonance (NMR) spectroscopy for determining structural constraints. The NMR-derived constraints, such as NOEs, hydrogen bonds, and torsion angles, are intrinsically local or short-range and could be insufficient for accurate structure determination of biological macromolecules and their complexes due to the scarcity of long-distance structural information. Residual dipolar couplings (RDCs), resulting from partial alignment of solute molecules relative to the magnetic field, provide valuable structural information in terms of global, long-range orientational constraints [1]. A commonly used method for aligning molecules in solution takes advantage of the anisotropy of molecular shape by imposing steric restrictions on the allowed orientations of the molecule (e.g., by means of bicelles [2], stretched gels [3,4], or PEG/hexanol-based media [5]). Such steric alignment can often be modeled as caused by planar obstacles, and we will refer to this simplified model of

molecular alignment as the *barrier model*. The alignment of a rigid molecule can be described by the so-called molecular alignment tensor. Accurate prediction of the molecular alignment tensor, and with it of the RDCs, is important for NMR-based structure determination and validation as well as applications to dynamic and disordered systems (see e.g., [6–11]). The sensitivity to molecular shape has the potential for improving structure characterization, especially in multidomain systems and macromolecular complexes (e.g., [12]), by fully integrating RDC prediction into structure refinement protocols to directly drive structure optimization. Future progress in this direction critically depends on the efficiency and accuracy of the alignment tensor prediction.

Several methods for computing the molecular alignment tensor *ab initio*, i.e., based solely on the three-dimensional shape of the molecule, have recently been proposed. In the method by Zweckstetter and Bax [13,14], implemented in a program called PALES, the alignment tensor is computed by uniformly sampling all orientations of a molecule (see e.g., [15]) at various distances away from a planar barrier, and averaging over only those orientations in which the molecule's surface does not collide with the barrier. The computational efficiency of this method is limited due to the fact that it must compute collisions between an arbitrary shape and a plane for every sample in the four-dimensional problem space.

Simpler methods based on the barrier model, but representing the shape of the molecule by an equivalent ellipsoid, have also been proposed. In Fernandes et al. [16], the alignment tensor was

* Corresponding author. Address: Department of Chemistry and Biochemistry, Center for Biomolecular Structure and Organization, University of Maryland, College Park, MD 20742-3360, USA. Fax: +1 301 314 0386.

E-mail addresses: kberlin@cs.umd.edu (K. Berlin), oleary@cs.umd.edu (D.P. O'Leary), fushman@umd.edu (D. Fushman).

computed by approximating the molecule as an axially-symmetric prolate ellipsoid and analytically solving the barrier model for the alignment tensor. In Almond and Axelsen [17] and Azurmendi and Bush [18] the barrier method is also used, but the formulae are derived empirically.

Here we describe a new, computationally efficient method for computing the molecular alignment tensor based on the barrier model. The increase in speed is achieved by re-expressing the problem as one of numerical integration, rather than one of simple uniform sampling. This formulation allowed us to simplify the problem by reducing its dimensionality from four to two. In addition to the reduction in computational complexity, numerical integration has the advantage of (i) allowing control over the size of numerical error, and (ii) allowing a more efficient sampling of the problem space [19]. Computational geometry techniques are used to increase the computational speed further. We will refer to our method as PATI (Prediction of Alignment Tensor using Integration). PATI can also be used with an equivalent ellipsoid of the molecule instead of the full surface. We will refer to this simplification as PATI-E. This simplified method is used to explore and compare various representations of protein shape by a (fully anisotropic) equivalent ellipsoid: based on the gyration tensor [16], the actual molecular surface [20], or the minimum-volume ellipsoid.

Finally, we examine the accuracy of the proposed methods (PATI and PATI-E) and the existing *ab initio* methods for RDC prediction. This analysis separately quantifies the effect of inaccuracy in the predicted RDCs caused by the inaccuracy in the orientation or in the magnitude of the alignment tensor. The results obtained for several proteins show that (i) the predicted RDCs and their agreement with experimental data are very sensitive to errors in orientation of the alignment tensor, and (ii) all *ab initio* prediction methods tested here give a rather crude estimate of the RDCs.

2. Theory

For a rigid molecule, the molecular alignment tensor **A** with respect to the magnetic field **B** is described by a 3×3 symmetric traceless matrix [13], sometimes referred to as the Saupe matrix [21], with the following elements ($i, j = 1, 2, 3$):

$$A_{ij} = \frac{1}{2}\langle F_{ij}\rangle, \qquad F_{ij} = 3\cos\theta_i \cos\theta_j - \delta_{ij}, \tag{1}$$

where θ_i is the angle between molecular axis i and the magnetic field **B**, $\langle \ldots \rangle$ is the average over all possible orientations of the molecule in solution, and δ is the Kronecker delta.

The RDC value D_{PQ} for a specific bond PQ is related to the alignment tensor and the bond's orientation relative to the molecule's coordinate frame by the following equation:

$$
\begin{aligned}
D_{PQ} &= C_{PQ}\sum_{i,j} A_{ij}\cos\phi_i \cos\phi_j, \\
C_{PQ} &= -S_{LS}\frac{\mu_0 \gamma_P \gamma_Q \hbar}{4\pi^2 r_{PQ}^3},
\end{aligned}
\tag{2}
$$

where ϕ_i is the angle between the PQ bond and the molecular axis i, S_{LS} is the Lipari–Szabo generalized order parameter, μ_0 is the permeability of free space, γ_P and γ_Q are the gyromagnetic ratios of the corresponding nuclei, $\hbar$ is the reduced Planck's constant, and r_{PQ} is the length of the bond. PQ can represent bonds such as NH, $C_\alpha H_\alpha$, $C_\alpha C'$, and $C'N$.

2.1. The model for the alignment tensor

Given the three-dimensional structure of an arbitrary molecule, we will focus on the computation of its alignment tensor **A**, defined in Eq. (1). We model the planar barrier causing steric alignment of

the molecule as a set of two infinite planes with the surface normals in the z direction, positioned at a distance $2h$ from each other. The molecule is centered around some point **m** (e.g., its center of mass), which lies somewhere inside the convex hull of the molecule's surface. The direction of the magnetic field is given by a unit vector **B**, where

$$\mathbf{B} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}, \quad b_1^2 + b_2^2 + b_3^2 = 1. \tag{3}$$

Fig. 1 shows a schematic representation of the planar barrier model. Note that due to the symmetry of the system, the possible orientations of the molecule positioned between 0 and h along the z-axis are mirror images (over the x-y plane) of the possible orientations when the molecule is between h and $2h$. Thus we can simplify the model by considering only the bottom plane and positioning the molecule's center at a height from 0 to h above this plane.

The orientation of the molecule's coordinate frame relative to the Cartesian coordinate system in Fig. 1 can be defined by three Euler angles α, β, and γ, which determine the rotation matrix $\mathbf{R}(\alpha, \beta, \gamma)$. (See Appendix A.)

For a specific molecule, we define S to be a finite set of sample points from its molecular surface (e.g., van der Waals surface or Richards molecular surface), and the center **m** of the molecule to be some point inside the convex hull of this molecular surface. Referring to Fig. 1, to characterize the vertical extent of the molecule under the rotation $\mathbf{R}(\alpha, \beta, \gamma)$ around its center **m**, we define $\eta(\alpha, \beta, \gamma)$, to be the difference between the z-coordinate of the center of the molecule and the minimum z-coordinate value of all the rotated points in S:

$$\eta(\alpha, \beta, \gamma) = -\min_{\mathbf{s_k} \in S}\{(\mathbf{R}(\alpha, \beta, \gamma)(\mathbf{s_k} - \mathbf{m})) \cdot [0 \quad 0 \quad 1]\}. \tag{4}$$

Note that $\eta(\alpha, \beta, \gamma)$ sets the lower limit on the height of the center of the molecule at a given orientation.

We rewrite F, from Eq. (1), in terms of the rotation matrix,

$$
\begin{aligned}
F_{ij}(\alpha, \beta, \gamma) &= 3(R_{1i}b_1 + R_{2i}b_2 + R_{3i}b_3)(R_{1j}b_1 + R_{2j}b_2 + R_{3j}b_3) \\
&\quad - \delta_{ij},
\end{aligned}
\tag{5}
$$

and average the F values at height a with the mirror cases of $2h - a$ into one equation

$$
\begin{aligned}
\overline{F}_{ij}(\alpha, \beta, \gamma) &= \frac{3}{2}(R_{1i}b_1 + R_{2i}b_2 + R_{3i}b_3)(R_{1j}b_1 + R_{2j}b_2 + R_{3j}b_3) \\
&\quad + \frac{3}{2}(R_{1i}b_1 + R_{2i}b_2 - R_{3i}b_3)(R_{1j}b_1 + R_{2j}b_2 - R_{3j}b_3) - \delta_{ij} \\
&= 3(R_{1i}b_1 + R_{2i}b_2)(R_{1j}b_1 + R_{2j}b_2) + 3R_{3i}R_{3j}b_3^2 - \delta_{ij}.
\end{aligned}
\tag{6}
$$

Due to the symmetry of the system, $\overline{\mathbf{F}}$ can be used instead of **F** to simplify our model to just one plane and a height from 0 to h.
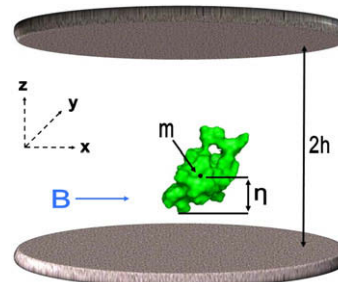

Fig. 1. Planar barrier model for molecular alignment.

For any rotation $\mathbf{R}(\alpha, \beta, \gamma)$, the center of the molecule cannot be located at a height between 0 and $\eta(\alpha, \beta, \gamma)$. Therefore, the range of interest is from $\eta(\alpha, \beta, \gamma)$ to h. The alignment tensor $\mathbf{A}$ is then computed by summing $\overline{\mathbf{F}}$ weighted by the probability of the current height and orientation, for all allowed orientations and heights from $\eta(\alpha, \beta, \gamma)$ to h.

We assume equal *a priori* probabilities of all orientations at all heights, and write the analytical expression for A_{ij} from Eq. (1). To obtain a uniform distribution of the Euler angles we multiply our integrand by the Jacobian $J = \sin \beta/(8\pi^2)$ [22] to obtain

$$A_{ij} = \frac{1}{2N} \int_{\gamma_0}^{\gamma_1} \int_{\beta_0}^{\beta_1} \int_{\alpha_0}^{\alpha_1} \int_{\eta(\alpha,\beta,\gamma)}^{h} \overline{F}_{ij} J \, dz \, d\alpha \, d\beta \, d\gamma, \tag{7}$$

where N is the normalization factor

$$N = \int_{\gamma_0}^{\gamma_1} \int_{\beta_0}^{\beta_1} \int_{\alpha_0}^{\alpha_1} \int_{\eta(\alpha,\beta,\gamma)}^{h} J \, dz \, d\alpha \, d\beta \, d\gamma, \tag{8}$$

and $[\alpha_0, \alpha_1]$, $[\beta_0, \beta_1]$, $[\gamma_0, \gamma_1]$ are the ranges in which α, β, γ are defined.

2.2. Computation of the alignment tensor: general case

We show in Appendix A, using the Euler z–y–z rotation, that the expression for the alignment tensor $\mathbf{A}$ of an arbitrarily-shaped molecule can be simplified from a quadruple integral to a double integral. Specifically,

$$A_{11} = \frac{S_c}{16N\pi} \int_0^{2\pi} \int_{-1}^{1} (-3u^2 \cos^2 \alpha + 3\cos^2 \alpha - 1)\eta(\alpha, \arccos u)\, du\, d\alpha,$$

$$A_{22} = \frac{S_c}{16N\pi} \int_0^{2\pi} \int_{-1}^{1} (3u^2 \cos^2 \alpha - 3\cos^2 \alpha - 3u^2 + 2)$$
$$\times \eta(\alpha, \arccos u)\, du\, d\alpha,$$

$$A_{33} = \frac{S_c}{16N\pi} \int_0^{2\pi} \int_{-1}^{1} (3u^2 - 1)\eta(\alpha, \arccos u)\, du\, d\alpha,$$

$$A_{21} = \frac{3S_c}{16N\pi} \int_0^{2\pi} \int_{-1}^{1} -\sin\alpha \cos\alpha (1 - u^2)\eta(\alpha, \arccos u)\, du\, d\alpha,$$

$$A_{31} = \frac{3S_c}{16N\pi} \int_0^{2\pi} \int_{-1}^{1} -u\sqrt{1 - u^2} \cos(\alpha)\eta(\alpha, \arccos u)\, du\, d\alpha,$$

$$A_{32} = \frac{3S_c}{16N\pi} \int_0^{2\pi} \int_{-1}^{1} u\sqrt{1 - u^2} \sin(\alpha)\eta(\alpha, \arccos u)\, du\, d\alpha, \tag{9}$$

where η is independent of the angle γ and

$$N = h - \frac{1}{4\pi} \int_0^{2\pi} \int_{-1}^{1} \eta(\alpha, \arccos u)\, du\, d\alpha,$$
$$S_c = 1 - 3b_3^2. \tag{10}$$

Because $\mathbf{A}$ is a traceless symmetric tensor [13], only A_{11} and A_{22}, A_{21}, A_{31}, and A_{32} need to be computed, while $A_{33} = -(A_{11} + A_{22})$, $A_{12} = A_{21}$, $A_{13} = A_{31}$, and $A_{23} = A_{32}$. One can multiply the alignment tensor by -0.8 to account for the incomplete bicelle alignment the order parameter for the liquid crystal), and to match the sign returned by PALES. The height h can be determined by the formula $d/(2V_f)$, where d is the barrier thickness (≈ 40 Å for DMPC/DHPC bicelles) and V_f ($\ll 1$) is the sample volume fraction occupied by the barriers (see [13,14]).

Thus, all one needs to know in order to compute the alignment tensor is $\eta(\alpha, \beta)$, defined in Eq. (4). Being an intrinsic geometric property of the molecule, $\eta(\alpha, \beta)$ can be computed separately, regardless of the barrier.

2.3. Computing η

In the PALES approach [13,14], $\mathbf{A}$ is estimated based on forming a mesh of the molecular surface and then rotating all the mesh triangles of this surface to check if any part of the mesh is below the barrier. Observe that the complexity of each rotation is proportional to the number of triangles in the mesh. It is possible to simplify the mesh using mesh simplification (see [23,24]); however even this is overly complex. An infinite planar barrier is not sensitive to cavities on the surface of the molecule; therefore, a convex hull of the molecule is a sufficient representation of the molecule's surface. Additional mesh simplifications could be performed on the convex hull to further reduce the number of points.

To compute η for an arbitrary molecule under a rotation $\mathbf{R}$, we simply compute the convex hull of the atom positions of the molecule and consider the vertices of the convex hull as the set S in Eq. (4). We add the van der Waals radius of the atom associated with the minimum z-value to Eq. (4) to form η for the rotation $\mathbf{R}$. Fig. 2A shows the convex hull around the Cyanovirin-N molecule. The number of points used to represent the molecule drops dramatically, from 40,708 in the molecular surface representation (see [25,26]), to just 57 in the convex hull representation. For any rotation $\mathbf{R}$, the relative error in η between the two representations is less than 5% and the absolute error is less than 0.5 Å. Also the alignment tensors and the RDCs predicted by PATI (our method) and PALES are almost identical, as shown below.

2.4. Special case of an ellipsoid

A potential simplification for computing the alignment tensor is to represent a molecule by an equivalent ellipsoid. In this section we examine several methods of deriving an equivalent ellipsoidal representation of an arbitrary molecule.

An ellipsoid $\mathscr{E}$ in $\mathbb{R}^3$ is defined as

$$\mathscr{E}(\mathbf{E}, \mathbf{m}) = \left\{ \mathbf{p} \,|\, (\mathbf{p} - \mathbf{m})^T \mathbf{E}(\mathbf{p} - \mathbf{m}) = 1 \right\}, \tag{11}$$

where $\mathbf{E}$ is a 3×3 symmetric positive definite matrix that defines the shape of the ellipsoid and $\mathbf{m} \in \mathbb{R}^3$ is its center.

The ellipsoid's semi-principal axes can be derived by an eigendecomposition of $\mathbf{E}$, such that

$$\mathbf{E} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T = [\mathbf{V}_1 \mathbf{V}_2 \mathbf{V}_3] \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} [\mathbf{V}_1 \mathbf{V}_2 \mathbf{V}_3]^T, \tag{12}$$

where the lengths of the semi-principal axes are $a = 1/\sqrt{\lambda_1} \geqslant b = 1/\sqrt{\lambda_2} \geqslant c = 1/\sqrt{\lambda_3}$, and $\mathbf{V}_1$, $\mathbf{V}_2$, $\mathbf{V}_3$ are their associated directions (eigenvectors). The matrix $\mathbf{V}$ also represents the rotation matrix of the ellipsoid from the orientation where it is aligned to the principal axes of the coordinate frame to the actual orientation.

One method for finding an equivalent ellipsoid of a molecule is to find the minimum-volume ellipsoid (MVE) that encloses all of the atoms in the molecule. Several methods exist for computing MVE [27,28]. When we use the MVE in PATI, we denote the method as PATI-E.

A second method for constructing an equivalent ellipsoid of a molecule is to find an ellipsoid with the same gyration tensor as the molecule. The length of the ith axis of the equivalent ellipsoid based on the gyration tensor is computed as $\sqrt{5\lambda_i}$, where λ_i is the ith eigenvalue of the gyration tensor. We will refer to the ellipsoid with an equivalent moment of inertia as the Gyration Ellipsoid (GE) (see Fernandes et al. [16]).

A third method for finding an equivalent ellipsoid is to fit the molecular surface. This can be achieved via the Principal Component Analysis (PCA) of the coordinates of the points representing
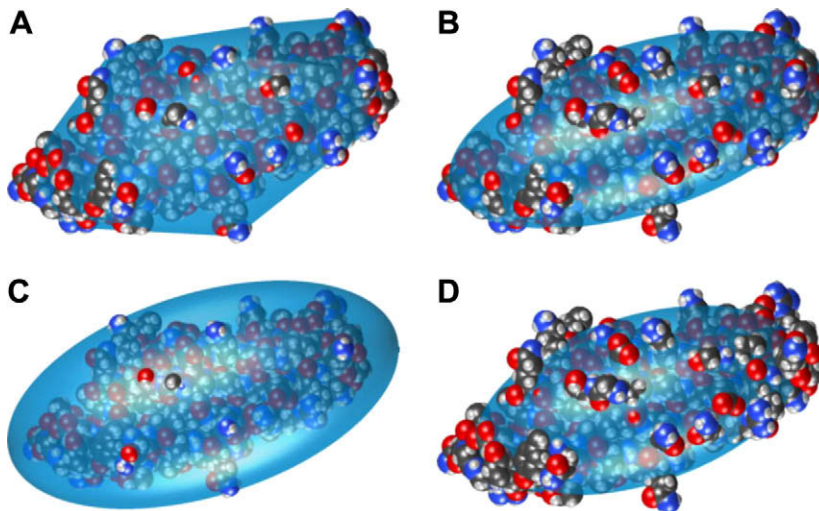
Fig. 2. Convex hull and equivalent ellipsoids for Cyanovirin-N molecule drawn on top of its van der Waals surface. (A) The convex hull around the molecule. (B) The GE ellipsoid representation. (C) The MVE ellipsoid representation. (D) The PCAE ellipsoid representation with no hydration layer included.

the surface of the molecule (see Ryabov et al. [20]). The length of the ith axis of the equivalent ellipsoid based on this method is $\sqrt{3\lambda_i}$, where λ_i is the ith eigenvalue of the corresponding covariance matrix. We refer to this method as PCAE.

Fig. 2 illustrates the three ellipsoid models of the molecular surface for the Cyanovirin-N molecule.

As shown in Appendix B, for an arbitrary ellipsoid under an Euler rotation $\mathbf{R}(\alpha, \beta, \gamma)$, $\eta(\alpha, \beta, \gamma)$ can be expressed as

$$\eta = \sqrt{R_{31}^2 a^2 + R_{32}^2 b^2 + R_{33}^2 c^2}. \tag{13}$$

3. Results

In this section we present a comprehensive comparison of several methods for computing RDCs *ab initio*. All the RDC data analyzed here are for the backbone *NH* bonds located in structurally well-defined regions of proteins, i.e., the α-helices and β-sheets. The RDC data were retrieved from the BMRB repository using the PDB code of the molecule. Only the RDC values measured using the neutral bicelle alignment medium (or, in the case of the B3 domain of protein G, the PEG/hexanol-based medium) are used. The nine proteins and their codes in the Protein Data Bank are listed in Table 1.

We assess the quality of our results by computing the *quality factor* between the vector of experimental RDCs, D_{exp}, and our predicted RDCs for those same bonds, D_{pred}, as [29]:

$$Q = \frac{\|D_{exp} - D_{pred}\|_2}{\|D_{exp}\|_2}, \tag{14}$$

where $\|\cdot\|_2$ is the l^2-norm.

Note that the predicted magnitude of the RDC values depends on the experimental conditions (they determine the barrier height h) and selection of values for equation constants, e.g., C_{PQ}. These factors affect all RDCs approximately uniformly, and hence can be represented by a scaling factor. Therefore, in order to make our analysis less sensitive to possible errors in experimental conditions and imperfect selection of values for constants, we also introduce the scaled quality factor to quantify the agreement between the experimental and predicted data with an unknown scaling factor. We define the *scaled quality factor* as

Table 1
Quality factors for the experimental data.

Protein	PDB[a]	LS[b,c]	PALES-LS[b,c,d]
Cellular factor BAF [31]	2ezx	0.03	0.03 (0.00)
B1 domain of protein G [32]	3gb1	0.05	0.05 (0.00)
B3 domain of protein G [33]	2oed	0.04	0.04 (0.00)
Rat apo-S100B [34]	1b4c	0.11	0.11 (0.00)
Cyanovirin-N [35]	2ezm	0.04	0.04 (0.00)
Gα interacting protein [36]	1cmz	0.08	0.08 (0.00)
Ubiquitin [29]	1d3z	0.04	0.04 (0.00)
Hen lysozyme [37]	1e8l	0.06	0.06 (0.00)
Oxidized putidaredoxin [38]	1yjj	0.08	0.08 (0.00)
Mean		0.06	0.06 (0.00)

[a] The RCSB Protein Data Bank code for protein coordinates. First model from the ensemble of NMR structures was used for all calculations.
[b] Values represent the quality factor Q between the predicted and experimental data.
[c] Values represent the scaled quality factor Q_s between the predicted and experimental data.
[d] Values in parentheses represent the relative error between $\widetilde{\mathbf{A}}$ and the experimental alignment tensor derived using PALES-LS.

$$Q_s = \min_\rho \frac{\|D_{exp} - \rho D_{pred}\|_2}{\|D_{exp}\|_2}, \tag{15}$$

where the scalar ρ can be computed by linear least squares. (Note that both PATI and PALES can predict the magnitude of RDC values with reasonable accuracy. See Supplementary material for the values of ρ.)

First, we present the Q values for the experimental alignment tensor. We define the *experimental* alignment tensor, $\widetilde{\mathbf{A}}$, as the alignment tensor that optimally fits the data, i.e., gives the lowest Q value between the experimental and back-calculated RDC data. This quality factor allows us to examine whether the experimental data are well approximated by the theoretical equation for RDCs. We derive $\widetilde{\mathbf{A}}$ by solving a linear least-squares problem of the form

$$\begin{bmatrix} (v_1^1)^2 - (v_3^1)^2 & (v_2^1)^2 - (v_3^1)^2 & 2v_1^1 v_2^1 & 2v_1^1 v_3^1 & 2v_2^1 v_3^1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ (v_1^i)^2 - (v_3^i)^2 & (v_2^i)^2 - (v_3^i)^2 & 2v_1^i v_2^i & 2v_1^i v_3^i & 2v_2^i v_3^i \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ (v_1^n)^2 - (v_3^n)^2 & (v_2^n)^2 - (v_3^n)^2 & 2v_1^n v_2^n & 2v_1^n v_3^n & 2v_2^n v_3^n \end{bmatrix} \begin{bmatrix} \widetilde{A}_{11} \\ \widetilde{A}_{22} \\ \widetilde{A}_{12} \\ \widetilde{A}_{13} \\ \widetilde{A}_{23} \end{bmatrix} \approx \begin{bmatrix} D_{exp}^1 / C_{NH} \\ \vdots \\ D_{exp}^i / C_{NH} \\ \vdots \\ D_{exp}^n / C_{NH} \end{bmatrix} \tag{16}$$

where $\mathbf{v}^i = [v_1^i, v_2^i, v_3^i]$ is the normalized vector representing the orientation of the ith bond relative to the molecular coordinate frame, n is the number of bonds, and C_{NH} is the value of C_{PQ} in Eq. (2) for a NH bond. The linear least-squares problem can be solved by standard methods; see, e.g., [30].

Note that $\widetilde{\mathbf{A}}$ can be decomposed into the experimental rotation (eigenvectors) $\widetilde{\mathbf{V}}$ and experimental magnitudes (eigenvalues) $\widetilde{A}_1, \widetilde{A}_2, \widetilde{A}_3$, where

$$\widetilde{\mathbf{A}} = \widetilde{\mathbf{V}}\widetilde{\mathbf{\Lambda}}\widetilde{\mathbf{V}}^T = \begin{bmatrix} \widetilde{\mathbf{V}}_1 & \widetilde{\mathbf{V}}_2 & \widetilde{\mathbf{V}}_3 \end{bmatrix} \begin{bmatrix} \widetilde{A}_1 & 0 & 0 \\ 0 & \widetilde{A}_2 & 0 \\ 0 & 0 & \widetilde{A}_3 \end{bmatrix} \begin{bmatrix} \widetilde{\mathbf{V}}_1 & \widetilde{\mathbf{V}}_2 & \widetilde{\mathbf{V}}_3 \end{bmatrix}^T.$$

(17)

The Q values for the experimental alignment tensor derived using Eq. (16) are presented in Column 3 ('LS') in Table 1. The corresponding Q values for the best-fit alignment tensor derived from PALES are presented in Column 4, labeled PALES-LS. Naturally, $Q_s = Q$ for both methods. It is worth emphasizing here that this quality factor measures the actual quality of the experimental data (i.e., how well they fit the theoretical equation for RDCs) and therefore provides the baseline Q value for subsequent evaluation of the prediction methods. Note also that the values in the parentheses, the relative error in the alignment tensor, confirm that Eq. (16) gives the same experimental alignment tensor $\widetilde{\mathbf{A}}$ as the alignment tensor derived using PALES's best-fit algorithm.

Table 1 shows that the experimental RDC data are of high quality and consistent with the theoretical formulation of the RDC (Eqs. (1) and (2)). This is not surprising given that these RDCs were used as constraints in the calculation/refinement of the corresponding protein structures. The quality of the agreement is illustrated in Fig. 3A for Cyanovirin-N. (See also Supplementary material.)

The results of our *ab initio* calculations are presented in Table 2, for PATI, PALES, and for the ellipsoidal approximation methods

Table 2
Quality factors Q_s from RDC prediction for *ab initio* methods.

PDB[a]	PATI[b,c]	PALES[b,c,e]	PATI-E[b,c,d]	Almond[b,c,d]	PROLFIT[b,c,d]
2ezx	0.26 (0.94)	0.27 (0.94)	0.19 (0.96)	0.20 (0.96)	0.12 (0.99)
3gb1	0.14 (0.99)	0.11 (0.99)	0.27 (0.96)	0.29 (0.95)	0.20 (0.97)
2oed	0.24 (0.98)	0.19 (0.98)	0.18 (0.98)	0.17 (0.98)	0.29 (0.97)
1b4c	0.22 (0.93)	0.22 (0.93)	0.43 (0.74)	0.42 (0.75)	0.55 (0.58)
2ezm	0.46 (0.66)	0.47 (0.66)	0.53 (0.56)	0.54 (0.54)	0.49 (0.61)
1cmz	0.32 (0.90)	0.30 (0.92)	0.38 (0.86)	0.39 (0.85)	0.37 (0.88)
1d3z	0.20 (0.93)	0.23 (0.91)	0.37 (0.81)	0.41 (0.77)	0.20 (0.91)
1e8l	0.31 (0.92)	0.31 (0.91)	0.42 (0.88)	0.43 (0.87)	0.26 (0.95)
1yjj	0.52 (0.75)	0.60 (0.67)	0.56 (0.76)	0.56 (0.75)	0.86 (0.34)
Mean	0.30 (0.89)	0.30 (0.88)	0.37 (0.84)	0.38 (0.83)	0.37 (0.80)

[a] The RCSB Protein Data Bank code for protein coordinates. First model from the ensemble of NMR structures was used for the calculations. See Table 1 for the names of the proteins.
[b] Values represent the scaled quality factor Q_s between the predicted and experimental data.
[c] Values in the parentheses represent the squared Pearson's correlation coefficient, r^2 (also known as coefficient of determination).
[d] MVE ellipsoidal representation was used.
[e] All PALES prediction calculations were run with options '-bic -H -dGrid 0.5 -rA 3.1'.

using the MVE model. The MVE ellipsoid data are used in this table as this model provides on average a slightly more accurate estimation of the alignment tensor compared to the other two equivalent ellipsoid models considered in this study. (See Supplementary material.) Surprisingly, the scaled quality factor Q_s was rather high for all prediction methods, indicating a generally marginal agreement with experimental data, as illustrated in Fig. 3D and the Supplementary material. PATI and PALES calculations gave on average a slightly better agreement with the data compared to the other methods. It should be emphasized here that PATI gives almost identical results to PALES, as evident from Fig. 4. In order to under-
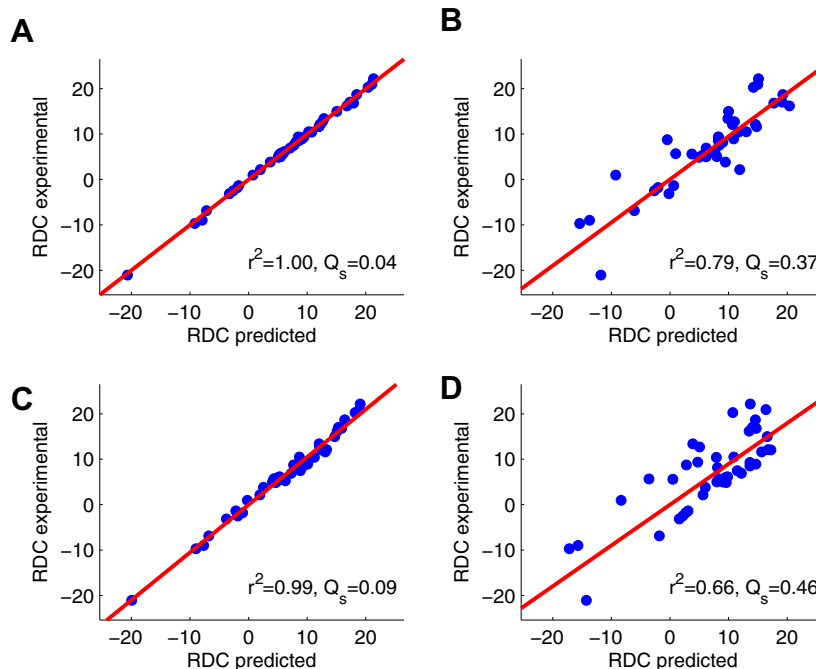
Fig. 3. Comparison of the predicted *versus* experimental $^1\text{H}^{15}\text{N}$ RDC values for the backbone amides in Cyanovirin-N, using various versions of the molecular alignment tensor derived from PATI. (A) The experimental alignment tensor was derived directly from the experimental data using least squares. (B) The alignment tensor was constructed using the magnitude (eigenvalues) of the experimental alignment tensor and the tensor orientation predicted using PATI program. (C) The alignment tensor was constructed using the orientation (eigenvectors) of the experimental alignment tensor but PATI-predicted magnitude (eigenvalues) of the tensor. (D) The alignment tensor was fully predicted from PATI simulation. The values of the squared Pearson's correlation coefficient, r^2, and the scale-insensitive quality factor, Q_s, are indicated. Similar graphs for the rest of the molecules studied here can be found in the Supplementary material.
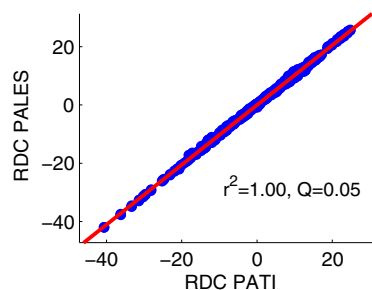
Fig. 4. The agreement between RDC values predicted using PATI and those from PALES prediction. Shown are the $^1H^{15}N$ RDCs for all backbone amides for all molecules studied here. The (unscaled) quality factor Q between the two sets of RDC values is 0.05, the RMSD is 0.6 Hz, and the maximum deviation is 1.7 Hz.

stand the reasons for the observed inaccuracy in our predictions, we now break down the contributions to the errors into those due to the principal values of the alignment tensor and those due to inaccuracy in its orientation.

In Table 3 we compare the Q_s values for 'synthetic' alignment tensors that have the same orientation as the *ab initio* calculated tensors but the correct (experimental) principal values. We constructed these tensors by combining the rotation matrix $\mathbf{V}$ determined from our five models, GE, PCAE, MVE, PATI, and PALES, with the experimental magnitudes $\tilde{A}_1, \tilde{A}_2, \tilde{A}_3$ of the alignment tensor derived from $\tilde{\mathbf{A}}$. Such a comparison is expected to rank the methods based on the accuracy of prediction of the tensor's orientation. Since the orientation of $\mathbf{V}$ for the equivalent-ellipsoid-based methods is derived directly from the orientation of the ellipsoid, this table also provides a direct comparison of the ellipsoid models. Note that there are six different combinations for $\mathbf{V}$, since it is unknown *a priori* which $\tilde{A}_i$ is associated with which $\mathbf{V}_j$. The smallest of the six Q_s values is shown. Naturally, $Q_s = Q$ in this case.

As evident from Table 3, correcting the principal values of the alignment tensor while keeping its predicted orientation did not improve the agreement with experimental data. (See also Fig. 3B.) There are large variations among the various models in the accuracy of the predicted orientation of the alignment tensor. Of the three ellipsoid models tested here, MVE gave on average a somewhat better orientation (as documented in the Supplementary material), while PATI and PALES yielded generally similar results.

We then constructed 'synthetic' alignment tensors that have the correct orientation (i.e., the $\tilde{\mathbf{V}}$ matrices derived from the experimental tensors $\tilde{\mathbf{A}}$) but the same principal values (A_1, A_2, A_3) as the *ab initio* calculated tensors. Table 4 displays the Q_s values for five prediction methods, PATI, PALES, PATI-E, Almond, and PROL-FIT. From this table, it is clear that using the correct orientation of the tensor dramatically improved the agreement with experimental data (cf. Table 2). This improvement is illustrated in Fig. 3C for Cyanovirin-N and in the Supplementary material for the other molecules.

Note that Column 3 ('PATI-E') and Column 5 ('PROLFIT') in Table 4 show that an additional degree of freedom provided by a fully anisotropic ellipsoid *versus* an axially-symmetric prolate ellipsoid approximation gives an improvement in the Q_s.

Thus, the analysis presented above demonstrates that accurate prediction of the orientation of the alignment tensor is critical for the agreement with experimental data. Accurate prediction of the principal components of the tensor is important, too. However, when experimental RDCs are available, one can make an educated guess, based on the observed histogram/distribution of the data, about the magnitude of the tensor components (e.g., as described in [39]) and scale the predicted alignment tensor appropriately, whereas there is no obvious way to predict the orientation of the tensor.

4. Conclusions

We have reformulated the planar barrier model as a numerical integration problem and implemented it in a program called PATI. Our method has accuracy similar to PALES but is computationally more efficient and allows for finer control over numerical error. In addition, the convex hull provides a simpler representation of the surface, thus further increasing the computational efficiency of the proposed method. This could allow PATI-based RDC prediction to be incorporated into the existing structure determination/refinement protocols. Because the molecular alignment tensor (and hence the RDC) is sensitive to the overall size and shape of the molecule, this would provide additional structural constraints that could potentially improve the accuracy of structure determination by NMR.

We compared several methods (old and new) for the computation of an equivalent ellipsoid of a molecule. We examined the accuracy of these equivalent ellipsoid models in predicting the alignment tensor and showed that the minimal volume ellipsoid

Table 3
Quality of prediction for the orientation of alignment tensor.

PDB[a,d]	PATI[b,c,d]	PALES[b,c,d]	GE[b,c,d]	MVE[b,c,d]	PCAE[b,c,d]
2ezx	0.27 (10°)	0.28 (11°)	0.13 (5°)	0.11 (4°)	0.15 (5°)
3gb1	0.15 (17°)	0.12 (10°)	0.21 (12°)	0.21 (28°)	0.10 (14°)
2oed	0.23 (15°)	0.19 (13°)	0.33 (20°)	0.19 (14°)	0.25 (14°)
1b4c	0.25 (11°)	0.25 (11°)	0.68 (43°)	0.32 (15°)	0.62 (31°)
2ezm	0.37 (37°)	0.37 (37°)	0.45 (26°)	0.55 (33°)	0.41 (33°)
1cmz	0.31 (25°)	0.29 (23°)	0.33 (36°)	0.33 (24°)	0.36 (37°)
1d3z	0.19 (16°)	0.21 (16°)	0.42 (23°)	0.20 (29°)	0.29 (23°)
1e8l	0.38 (42°)	0.37 (41°)	0.33 (27°)	0.18 (16°)	0.33 (26°)
1yjj	0.53 (25°)	0.61 (30°)	0.87 (48°)	0.59 (26°)	0.85 (48°)
Mean	0.30 (22°)	0.30 (22°)	0.42 (27°)	0.30 (21°)	0.37 (26°)

[a] The RCSB Protein Data Bank code for protein coordinates. First model from the ensemble of NMR structures was used for the calculations. See Table 1 for the names of the proteins.
[b] Values represent the quality factor Q between the predicted and experimental data.
[c] Values represent the scaled quality factor Q_s between the predicted and experimental data.
[d] Values in the parentheses represent the angle difference between the orientation of the experimental and predicted tensors (The angle was derived using the axis-angle representation of rotation. See Appendix C for details.).

Table 4
Quality of prediction for the magnitude of alignment tensor.

PDB[a]	PATI[b,c]	PALES[b,c]	PATI-E[b,c,d]	Almond[b,c,d]	PROLFIT[b,c,d]
2ezx	0.04 (1.00)	0.04 (1.00)	0.04 (1.00)	0.03 (1.00)	0.12 (0.99)
3gb1	0.06 (1.00)	0.05 (1.00)	0.09 (0.99)	0.11 (0.99)	0.20 (0.96)
2oed	0.04 (1.00)	0.04 (1.00)	0.04 (1.00)	0.04 (1.00)	0.16 (0.99)
1b4c	0.12 (0.98)	0.12 (0.98)	0.19 (0.96)	0.19 (0.96)	0.33 (0.88)
2ezm	0.09 (0.99)	0.08 (0.99)	0.08 (0.99)	0.05 (1.00)	0.25 (0.90)
1cmz	0.08 (0.99)	0.08 (0.99)	0.12 (0.98)	0.14 (0.98)	0.16 (0.98)
1d3z	0.06 (0.99)	0.08 (0.99)	0.19 (0.93)	0.23 (0.90)	0.17 (0.93)
1e8l	0.11 (0.99)	0.10 (0.99)	0.06 (1.00)	0.06 (1.00)	0.20 (0.97)
1yjj	0.29 (0.92)	0.31 (0.90)	0.31 (0.90)	0.29 (0.91)	0.20 (0.96)
Mean	0.10 (0.98)	0.10 (0.98)	0.12 (0.97)	0.13 (0.97)	0.20 (0.95)

[a] The RCSB Protein Data Bank code for protein coordinates. First model from the ensemble of NMR structures was used for the calculations. See Table 1 for the names of the proteins.
[b] Values represent the scaled quality factor Q_s between the predicted and experimental data. The smallest of the six possible values is shown.
[c] Values in the parentheses represent the squared Pearson's correlation coefficient, r^2.
[d] MVE ellipsoidal representation was used.

gives on average a slightly better prediction of the alignment tensor orientation.

Finally, we compared all these methods against an extensive set of experimental RDC data. The analysis of the discrepancy between the experimental and predicted values emphasized the importance of the accurate prediction of the orientation of the alignment tensor. Possible sources of inaccuracy in *ab initio* alignment tensor prediction are the dynamic nature (structural flexibility) of protein molecules, not accounted for in the current prediction models, as well as the fact that the simple steric barrier model might not fully allow the correct alignment of all the molecules.

Acknowledgments

Appendix A. Derivation of Eq. (9)

In this section we will simplify the expression for **A** for an arbitrary molecule from the quadruple integral of Eq. (7) to a double integral.

Define the Euler z–y–z rotation matrix as $\mathbf{R}(\alpha, \beta, \gamma) = \mathbf{R}_z(\gamma) \mathbf{R}_y(\beta) \mathbf{R}_z(\alpha)$, where

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

$$
\mathbf{R}_y(\beta) = \begin{bmatrix} \cos \beta & 0 & \sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & \cos \beta \end{bmatrix}, \tag{A1}
$$

$$
\mathbf{R}_z(\gamma) = \begin{bmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}.
$$

Multiplying the three matrices yields the full expression

$$
\mathbf{R}(\alpha, \beta, \gamma)
$$
$$
= \begin{bmatrix} \cos \gamma \cos \beta \cos \alpha - \sin \gamma \sin \alpha & -\cos \gamma \cos \beta \sin \alpha - \sin \gamma \cos \alpha & \cos \gamma \sin \beta \\ \sin \gamma \cos \beta \cos \alpha + \cos \gamma \sin \alpha & -\sin \gamma \cos \beta \sin \alpha + \cos \gamma \cos \alpha & \sin \gamma \sin \beta \\ -\sin \beta \cos \alpha & \sin \beta \sin \alpha & \cos \beta \end{bmatrix}. \tag{A2}
$$

We now write the equations for $A_{11}, A_{22}, A_{33}, A_{21}, A_{31}, A_{32}$, and N, recalling that the Jacobian is $J = \sin \beta/(8\pi^2)$:

$$
A_{11} = \frac{1}{16N\pi^2} \int_0^{2\pi} \int_0^{\pi} \int_0^{2\pi} \int_{\eta(\alpha,\beta,\gamma)}^{h} \overline{F}_{11} \sin \beta \, dz \, d\alpha \, d\beta \, d\gamma,
$$

$$
A_{22} = \frac{1}{16N\pi^2} \int_0^{2\pi} \int_0^{\pi} \int_0^{2\pi} \int_{\eta(\alpha,\beta,\gamma)}^{h} \overline{F}_{22} \sin \beta \, dz \, d\alpha \, d\beta \, d\gamma,
$$

$$
A_{33} = \frac{1}{16N\pi^2} \int_0^{2\pi} \int_0^{\pi} \int_0^{2\pi} \int_{\eta(\alpha,\beta,\gamma)}^{h} \overline{F}_{33} \sin \beta \, dz \, d\alpha \, d\beta \, d\gamma,
$$

$$
A_{21} = \frac{1}{16N\pi^2} \int_0^{2\pi} \int_0^{\pi} \int_0^{2\pi} \int_{\eta(\alpha,\beta,\gamma)}^{h} \overline{F}_{21} \sin \beta \, dz \, d\alpha \, d\beta \, d\gamma, \tag{A3}
$$

$$
A_{31} = \frac{1}{16N\pi^2} \int_0^{2\pi} \int_0^{\pi} \int_0^{2\pi} \int_{\eta(\alpha,\beta,\gamma)}^{h} \overline{F}_{31} \sin \beta \, dz \, d\alpha \, d\beta \, d\gamma,
$$

$$
A_{32} = \frac{1}{16N\pi^2} \int_0^{2\pi} \int_0^{\pi} \int_0^{2\pi} \int_{\eta(\alpha,\beta,\gamma)}^{h} \overline{F}_{32} \sin \beta \, dz \, d\alpha \, d\beta \, d\gamma,
$$

$$
N = \frac{1}{8\pi^2} \int_0^{2\pi} \int_0^{\pi} \int_0^{2\pi} \int_{\eta(\alpha,\beta,\gamma)}^{h} \sin(\beta) \, dz \, d\alpha \, d\beta \, d\gamma.
$$

We observe that γ does not contribute to the vertical size of the molecule, and redefine $\eta(\alpha, \beta, \gamma)$ as $\eta(\alpha, \beta)$. Integrating by γ and z first gives us

$$
A_{11} = \frac{S_c}{16N\pi} \int_0^{2\pi} \int_0^{\pi} (3 \cos^2 \alpha \cos^2 \beta - 3 \cos \alpha + 1)(h - \eta(\alpha, \beta))
$$
$$
\times \sin \beta \, d\beta \, d\alpha,
$$

$$
A_{22} = \frac{S_c}{16N\pi} \int_0^{2\pi} \int_0^{\pi} -(3 \cos^2 \alpha \cos^2 \beta - 3 \cos^2 \alpha - 3 \cos^2 \beta + 2)
$$
$$
(h - \eta(\alpha, \beta)) \sin \beta \, d\beta \, d\alpha,
$$

$$
A_{33} = \frac{S_c}{16N\pi} \int_0^{2\pi} \int_0^{\pi} -(3 \cos^2 \beta - 1)(h - \eta(\alpha, \beta)) \sin \beta \, d\beta \, d\alpha,
$$

$$
A_{21} = \frac{S_c}{16N\pi} \int_0^{2\pi} \int_0^{\pi} 3 \cos \alpha \sin \alpha \sin^2 \beta (h - \eta(\alpha, \beta)) \sin \beta \, d\beta \, d\alpha,
$$

$$
A_{31} = \frac{S_c}{16N\pi} \int_0^{2\pi} \int_0^{\pi} 3 \cos \alpha \sin \beta \cos \beta (h - \eta(\alpha, \beta)) \sin \beta \, d\beta \, d\alpha,
$$

$$
A_{32} = \frac{S_c}{16N\pi} \int_0^{2\pi} \int_0^{\pi} -3 \sin \alpha \sin \beta \cos \alpha (h - \eta(\alpha, \beta)) \sin \beta \, d\beta \, d\alpha,
$$

where

$$
N = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} (h - \eta(\alpha, \beta)) \sin \beta \, d\beta \, d\alpha,
$$
$$
S_c = 1 - 3b_3^2. \tag{A4}
$$

We perform a change of variable, $u = \cos\beta$, obtaining

$$
A_{11} = \frac{S_c}{16N\pi} \int_0^{2\pi} \int_{-1}^{1} (3u^2 \cos^2 \alpha - 3 \cos^2 \alpha + 1)
$$
$$
\times (h - \eta(\alpha, \arccos u)) \, du \, d\alpha,
$$

$$
A_{22} = \frac{S_c}{16N\pi} \int_0^{2\pi} \int_{-1}^{1} (3u^2 \cos^2 \alpha + 3 \cos^2 \alpha - 2)
$$
$$
\times (h - \eta(\alpha, \arccos u)) \, du \, d\alpha,
$$

$$
A_{33} = \frac{S_c}{16N\pi} \int_0^{2\pi} \int_{-1}^{1} (1 - 3u^2)(h - \eta(\alpha, \arccos u)) \, du \, d\alpha,
$$

$$
A_{21} = \frac{3S_c}{16N\pi} \int_0^{2\pi} \int_{-1}^{1} \sin \alpha \cos \alpha (1 - u^2)(h - \eta(\alpha, \arccos u)) \, du \, d\alpha,
$$

$$
A_{31} = \frac{3S_c}{16N\pi} \int_0^{2\pi} \int_{-1}^{1} u \cos \alpha \sqrt{1 - u^2}(h - \eta(\alpha, \arccos u)) \, du \, d\alpha,
$$

$$
A_{32} = \frac{3S_c}{16N\pi} \int_0^{2\pi} \int_{-1}^{1} -u \sin \alpha \sqrt{1 - u^2}(h - \eta(\alpha, \arccos u)) \, du \, d\alpha,
$$

$$
N = \frac{1}{4\pi} \int_0^{2\pi} \int_{-1}^{1} h - \eta(\alpha, \arccos u) \, du \, d\alpha,
$$
$$
S_c = 1 - 3b_3^2. \tag{A5}
$$

Integrating the terms that do not involve η gives us Eq. (9).

Appendix B. Derivation of η for an ellipsoid

In this section we derive the analytical expression for η for an arbitrary ellipsoid, following the notation of Section 2.4. Due to the symmetry of the ellipsoid we consider only one octant in our analysis, expressing all points p on the ellipsoid in that octant as

$$
p(x, y) = (x, y, z(x, y)), \tag{B1}
$$

where

$$
z(x, y) = c \sqrt{\left(1 - \frac{x^2}{a^2} - \frac{y^2}{b^2}\right)}, \tag{B2}
$$

for $x \in [0, a]$ and $y \in \left[0, b\sqrt{(1 - x^2/a^2)}\right]$.

A rotation of the ellipsoid by $\mathbf{R}(\alpha, \beta, \gamma)$ transforms the coordinates of these points into

$$x' = R_{11}(\alpha, \beta, \gamma)x + R_{12}(\alpha, \beta, \gamma)y + R_{13}(\alpha, \beta, \gamma)z(x, y),$$
$$y' = R_{21}(\alpha, \beta, \gamma)x + R_{22}(\alpha, \beta, \gamma)y + R_{23}(\alpha, \beta, \gamma)z(x, y), \quad (B3)$$
$$z' = R_{31}(\alpha, \beta, \gamma)x + R_{32}(\alpha, \beta, \gamma)y + R_{33}(\alpha, \beta, \gamma)z(x, y).$$

We observe that

$$\eta(\alpha, \beta, \gamma) = z'(x_*, y_*), \quad (B4)$$

where $x_*(\alpha, \beta, \gamma)$ and $y_*(\alpha, \beta, \gamma)$ minimize z'.

To find the minimum/maximum value of our rotated ellipsoid, we solve $\nabla z'(x, y) = 0$:

$$\frac{\partial z'(x, y)}{\partial x} = R_{31} - R_{33} \frac{cx}{a^2 \sqrt{\left(1 - \frac{x^2}{a^2} - \frac{y^2}{b^2}\right)}} = 0, \quad (B5)$$

$$\frac{\partial z'(x, y)}{\partial y} = R_{32} - R_{33} \frac{cy}{b^2 \sqrt{\left(1 - \frac{x^2}{a^2} - \frac{y^2}{b^2}\right)}} = 0. \quad (B6)$$

Let

$$x_* = \frac{R_{31}a^2}{\sqrt{R_{31}^2 a^2 + R_{32}^2 b^2 + R_{33}^2 c^2}}, \quad (B7)$$

$$y_* = \frac{R_{32}b^2}{\sqrt{R_{31}^2 a^2 + R_{32}^2 b^2 + R_{33}^2 c^2}}. \quad (B8)$$

It is easy to verify that x_* and y_* solve Eqs. (B5) and (B6):

$$\frac{\partial z'(x_*, y_*)}{\partial x} = R_{31} - R_{33} \frac{cx_*}{a^2 \sqrt{\left(1 - \frac{x_*^2}{a^2} - \frac{y_*^2}{b^2}\right)}}$$
$$= R_{31} - R_{33} \frac{\frac{cR_{31}a^2}{\sqrt{R_{31}^2 a^2 + R_{32}^2 b^2 + R_{33}^2 c^2}}}{\frac{a^2 R_{33}c}{\sqrt{R_{31}^2 a^2 + R_{32}^2 b^2 + R_{33}^2 c^2}}} = 0, \quad (B9)$$

$$\frac{\partial z'(x_*, y_*)}{\partial y} = R_{32} - R_{33} \frac{cy_*}{b^2 \sqrt{\left(1 - \frac{x_*^2}{a^2} - \frac{y_*^2}{b^2}\right)}}$$
$$= R_{32} - R_{33} \frac{\frac{cR_{32}b^2}{\sqrt{R_{31}^2 a^2 + R_{32}^2 b^2 + R_{33}^2 c^2}}}{\frac{b^2 R_{33}c}{\sqrt{R_{31}^2 a^2 + R_{32}^2 b^2 + R_{33}^2 c^2}}} = 0. \quad (B10)$$

Therefore, x_*, y_* minimizes z', and the optimal value of z is

$$z_* = \frac{R_{33}c^2}{\sqrt{R_{31}^2 a^2 + R_{32}^2 b^2 + R_{33}^2 c^2}}, \quad (B11)$$

since $R_{31}^2 + R_{32}^2 + R_{33}^2 = 1$. Therefore, from Eqs. (B4) and (B3), our solution is

$$\eta(\alpha, \beta, \gamma) = R_{31}(\alpha, \beta, \gamma)x_* + R_{32}(\alpha, \beta, \gamma)y_* + R_{33}(\alpha, \beta, \gamma)z_*$$
$$= \sqrt{R_{31}^2 a^2 + R_{32}^2 b^2 + R_{33}^2 c^2}, \quad (B12)$$

where γ is arbitrary for the Euler rotation defined in Eq. (A1).

Appendix C. Axis-angle representation

The axis-angle representation of a rotation parameterizes the rotation by two values: a unit vector indicating the orientation of the axis, $\mathbf{u}$, and an angle, θ, describing the magnitude of the rotation about that axis. The direction of rotation around the axis $\mathbf{u}$ is determined by the right-hand rule. One advantage of the axis-angle representation over Euler angles representation is that one can easily quantify the magnitude of the rotation by the size of the rotation angle θ.

Given a rotation matrix $\mathbf{R}$, θ is computed as

$$\theta = \arccos\left(\frac{1}{2}(R_{11} + R_{22} + R_{33} - 1)\right). \quad (C1)$$

Appendix D. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmr.2009.07.028.

References

[1] A. Bax, Weak alignment offers new NMR opportunities to study protein structure and dynamics, Protein Sci. 12 (1) (2003) 1–16.
[2] N. Tjandra, A. Bax, Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium, Science 278 (5340) (1997) 1111–1114.
[3] R. Tycko, F.J. Blanco, Y. Ishii, Alignment of biopolymers in strained gels: a new way to create detectable dipole–dipole couplings in high-resolution biomolecular NMR, J. Am. Chem. Soc. 122 (38) (2000) 9340–9341.
[4] H. Sass, G. Musco, S. Stahl, P. Wingfield, S. Grzesiek, Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes, J. Biomol. NMR 18 (4) (2000) 303–309.
[5] M. Ruckert, G. Otting, Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments, J. Am. Chem. Soc. 122 (32) (2000) 7793–7797.
[6] C. Bewley, G. Clore, Determination of the relative orientation of the two halves of the domain-swapped dimer of Cyanovirin-N in solution using dipolar couplings and rigid body minimization, J. Am. Chem. Soc. 122 (25) (2000) 6009–6016.
[7] J. Warren, P. Moore, Application of dipolar coupling data to the refinement of the solution structure of the Sarcin–Ricin loop RNA, J. Biomol. NMR 20 (4) (2001) 311–323.
[8] C. Bewley, Rapid validation of the overall structure of an internal domain-swapped mutant of the anti-HIV protein Cyanovirin-N using residual dipolar couplings, J. Am. Chem. Soc. 123 (5) (2001) 1014–1015.
[9] H. Azurmendi, M. Martin-Pastor, C. Bush, Conformational studies of Lewis X and Lewis A trisaccharides using NMR residual dipolar couplings, Biopolymers 63 (2) (2002) 89–98.
[10] P. Bernadó, L. Blanchard, P. Timmins, D. Marion, R. Ruigrok, M. Blackledge, A structural model for unfolded proteins from residual dipolar couplings and small-angle X-ray scattering, Proc. Natl. Acad. Sci. 102 (47) (2005) 17002–17007.
[11] M. Mukrasch, P. Markwick, J. Biernat, M. von Bergen, P. Bernado, C. Griesinger, E. Mandelkow, M. Zweckstetter, M. Blackledge, Highly populated turn conformations in natively unfolded tau protein identified from residual dipolar couplings and molecular simulation, J. Am. Chem. Soc. 129 (16) (2007) 5235–5243.
[12] Y. Ryabov, D. Fushman, Structural assembly of multidomain proteins and protein complexes guided by the overall rotational diffusion tensor, J. Am. Chem. Soc. 129 (25) (2007) 7894–7902.
[13] M. Zweckstetter, 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) (2000) 3791–3792.
[14] M. Zweckstetter, NMR: prediction of molecular alignment from structure using the PALES software, Nat. Protocols 3 (4) (2008) 679–690.
[15] F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander, M. Scharf, The double cubic lattice method: efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies, J. Comput. Chem. 16 (3) (1995) 273–284.
[16] M.X. Fernandes, P. Bernado, M. Pons, J. Garcia de la Torre, An analytical solution to the problem of the orientation of rigid particles by planar obstacles. Application to membrane systems and to the calculation of dipolar couplings in protein NMR spectroscopy, J. Am. Chem. Soc. 123 (48) (2001) 12037–12047.
[17] A. Almond, J. Axelsen, Physical interpretation of residual dipolar couplings in neutral aligned media, J. Am. Chem. Soc. 124 (34) (2002) 9986–9987.
[18] H. Azurmendi, C. Bush, Conformational studies of blood group A and blood group B oligosaccharides using NMR residual dipolar couplings, Carbohydr. Res. 337 (10) (2002) 905–915.
[19] C.F. Van Loan, Introduction to Scientific Computing: A Matrix–Vector Approach Using MATLAB, Prentice-Hall Inc., Upper Saddle River, NJ, USA, 1997.
[20] Y. Ryabov, C. Geraghty, A. Varshney, D. Fushman, An efficient computational method for predicting rotational diffusion tensors of globular proteins using an ellipsoid representation, J. Am. Chem. Soc. 128 (48) (2006) 15432–15444.
[21] A. Saupe, G. Englert, High-resolution nuclear magnetic resonance spectra of orientated molecules, Phys. Rev. Lett. 11 (10) (1963) 462–464.

[22] F. Murnaghan, The element of volume of the rotation group, Proc. Natl. Acad. Sci. 36 (11) (1950) 670–672.
[23] P. Lindstrom, G. Turk, Fast and memory efficient polygonal simplification, in: VIS '98: Proceedings of the Conference on Visualization '98, IEEE Computer Society Press, Los Alamitos, CA, USA, 1998, pp. 279–286.
[24] P. Lindstrom, G. Turk, Evaluation of memoryless simplification, IEEE Trans. Vis. Comput. Graph. 5 (2) (1999) 98–115.
[25] A. Varshney, F. Brooks Jr., W. Wright, Linearly scalable computation of smooth molecular surfaces, IEEE Comput. Graph. Appl. 14 (5) (1994) 19–25.
[26] A. Varshney, F. Brooks Jr., Fast analytical computation of Richard's smooth molecular surface, in: IEEE Visualization'93 Proceedings, 1993, pp. 300–307.
[27] E. Welzl, Smallest enclosing disks (balls and ellipsoids), New Results and New Trends in Computer Science, Vol. 555, Springer-Verlag, 1991, pp. 359–370.
[28] M.J. Todd, E.A. Yildirim, On Khachiyan's algorithm for the computation of minimum volume enclosing ellipsoids, Discrete Appl. Math. 155 (13) (2007) 1731–1744.
[29] G. Cornilescu, J.L. Marquardt, M. Ottiger, A. Bax, Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase, J. Am. Chem. Soc. 120 (27) (1998) 6836–6837.
[30] J.A. Losonczi, M. Andrec, M.W.F. Fischer, J.H. Prestegard, Order matrix analysis of residual dipolar couplings using singular value decomposition, J. Magn. Res. 138 (2) (1999) 334–342.
[31] M. Cai, Y. Huang, R. Zheng, S. Wei, R. Ghirlando, M. Lee, R. Craigie, A. Gronenborn, G. Clore, Solution structure of the cellular factor BAF responsible for protecting retroviral DNA from autointegration, Nat. Struct. Biol. 5 (10) (1998) 903–909.
[32] J. Kuszewski, A.M. Gronenborn, G.M. Clore, Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration, J. Am. Chem. Soc. 121 (10) (1999) 2337–2338.
[33] T.S. Ulmer, B.E. Ramirez, F. Delaglio, A. Bax, Evaluation of backbone proton positions and dynamics in a small protein by liquid crystal NMR spectroscopy, J. Am. Chem. Soc. 125 (30) (2003) 9179–9191.
[34] A. Drohat, N. Tjandra, D. Baldisseri, D. Weber, The use of dipolar couplings for determining the solution structure of rat apo-S100B ($\beta\beta$), Protein Sci. 8 (4) (1999) 800–809.
[35] C. Bewley, K. Gustafson, M. Boyd, D. Covell, A. Bax, G. Clore, A. Gronenborn, Solution structure of Cyanovirin-N, a potent HIV-inactivating protein, Nat. Struct. Biol. 5 (7) (1998) 571–578.
[36] E. de Alba, L. De Vries, M. Farquhar, N. Tjandra, Solution structure of human GAIP (Gα interacting protein): a regulator of G protein signaling, J. Mol. Biol. 291 (4) (1999) 927–939.
[37] H. Schwalbe, S. Grimshaw, A. Spencer, M. Buck, J. Boyd, C. Dobson, C. Redfield, L. Smith, A refined solution structure of hen lysozyme determined using residual dipolar coupling data, Protein Sci. 10 (4) (2001) 677–688.
[38] N.U. Jain, E. Tjioe, A. Savidor, J. Boulie, Redox-dependent structural differences in putidaredoxin derived from homologous structure refinement via residual dipolar couplings, Biochemistry 44 (25) (2005) 9067–9078.
[39] A. Bax, G. Kontaxis, N. Tjandra, Dipolar couplings in macromolecular structure determination, in: V.D. Thomas, L. James, U. Schmitz (Eds.), Part B: Nuclear Magnetic Resonance of Biological Macromolecules. Vol. 339 of Methods in Enzymology, Academic Press, 2001, pp. 127–174.