

# Order Matrix Analysis of Residual Dipolar Couplings Using Singular Value Decomposition

Judit A. Losonczi, Michael Andrec,<sup>1</sup> Mark W. F. Fischer, and James H. Prestegard\*

*Complex Carbohydrate Research Center, University of Georgia, 220 Riverbend Rd., Athens, Georgia 30602-4712*

Received December 30, 1998; revised March 3, 1999

**The measurement of anisotropic spin interactions, such as residual dipolar couplings, in partially ordered solutions can provide valuable information on biomolecular structure. While the information can be used to refine local structure, it can make a unique contribution in determining the relative orientation of remote parts of molecules, which are locally well structured, but poorly connected based on NOE data. Analysis of dipolar couplings in terms of Saupe order matrices provides a concise description of both orientation and motional properties of locally structured fragments in these cases. This paper demonstrates that by using singular value decomposition as a method for calculating the order matrices, principal frames and order parameters can be determined efficiently, even when a very limited set of experimental data is available. Analysis of <sup>1</sup>H–<sup>15</sup>N dipolar couplings, measured in a two-domain fragment of the barley lectin protein, is used to illustrate the computational method.** © 1999 Academic Press

**Key Words:** anisotropic spin interactions; Saupe order matrix; singular value decomposition; domain–domain orientation; barley lectin protein.

## INTRODUCTION

Traditional NMR structure determination of biomolecules is based mainly on distance constraints derived from relaxation processes such as NOEs (1). This approach is limited, however, by the short distances over which NOE interactions are effective. This is a particularly severe limitation when the relationship of remote parts of elongated or loosely connected molecules is at issue. Spin interactions such as residual dipolar couplings that have an inherent orientational dependence, i.e., anisotropic spin interactions, can, in principle, complement NOE data in these situations. However, these interactions normally average to near zero in solution and cannot easily be measured. Recently, the idea of making these spin interactions measurable by introducing a higher degree of order in the

systems under study (2) has been extended to macromolecular systems including proteins, first by using the interaction of a high static magnetic field with molecules having large anisotropic paramagnetic susceptibilities (3), then by using similar interactions with molecules having large diamagnetic susceptibilities (4, 5), and recently by using magnetically oriented liquid crystalline media as a solvent for the molecules under study (6–9).

The success of these techniques has led to a growing number of applications (10–14). Most applications to proteins have been directed toward structure refinement. It is possible to directly refine a structure against residual dipolar couplings using a simulated annealing protocol (15) if sufficient other sources of structural constraints (such as NOEs) are available, if the order parameters can be estimated by alternate methods (16), and if it can be assumed that the whole molecule is rigid enough to have a uniform alignment tensor.

This paper demonstrates a method of direct utilization of anisotropic spin interaction data that is based on the determination of the Saupe order matrix (17) for structurally well-defined fragments within a biomolecule. It can be used to establish the relative orientation of rigid units of the molecule, and it can help to determine the order parameters (axial and rhombic components of the alignment tensor) needed for other calculations even when only a small number of dipolar couplings are measured. Use of the numerical method of singular value decomposition to solve for the order matrix elements makes this approach fast, reliable, and easy to carry out. The application of the proposed method is demonstrated using <sup>15</sup>N–<sup>1</sup>H dipolar couplings measured in a <sup>15</sup>N-labeled two-domain fragment of the barley lectin protein (BLBC) that has been oriented in a dilute bicelle (6, 18, 19) solution.

## THEORY

The most commonly measured anisotropic spin interaction is residual dipolar coupling. For a pair of spin  $\frac{1}{2}$  nuclei  $n$  and  $m$ , the general expression describing the splitting due to this interaction is

<sup>1</sup> Current address: Rutgers University/CABM, 679 Hoes Lane, Piscataway, NJ 08854.

\* To whom correspondence should be addressed. Fax: 706-542-4412; E-mail: [jpresteg@ccrc.uga.edu](mailto:jpresteg@ccrc.uga.edu).

$$D^{nm} = -\frac{\gamma_n \gamma_m h}{2\pi^2 r_{nm}^3} \left\langle \frac{3 \cos^2 \Theta - 1}{2} \right\rangle$$

$$= D_{\max}^{nm} \left\langle \frac{3 \cos^2 \Theta - 1}{2} \right\rangle, \quad [1]$$

where  $\gamma_n$  and  $\gamma_m$  are the gyromagnetic ratios for the spin  $\frac{1}{2}$  nuclei,  $h$  is Plank's constant,  $r_{nm}$  is the internuclear distance, and  $\Theta$  is the angle between the internuclear vector and the external magnetic field. In the above-mentioned systems, the value of  $D^{nm}$  is radically scaled down from its maximum static value,  $D_{\max}^{nm}$ , by rapid internal motions and overall motions as restricted by the anisotropic interaction of the molecule with the bicelles. The angle brackets denote a time average over the motions with time scales short compared to the reciprocal of the splitting.

It is also possible to obtain information on the orientational properties of molecules by measuring chemical shift anisotropy effects. The chemical shift observed in an oriented system differs from the isotropic chemical shift,  $\delta_{\text{iso}} = \frac{1}{3}(\delta_{11} + \delta_{22} + \delta_{33})$ , by an anisotropic contribution that can be expressed as:

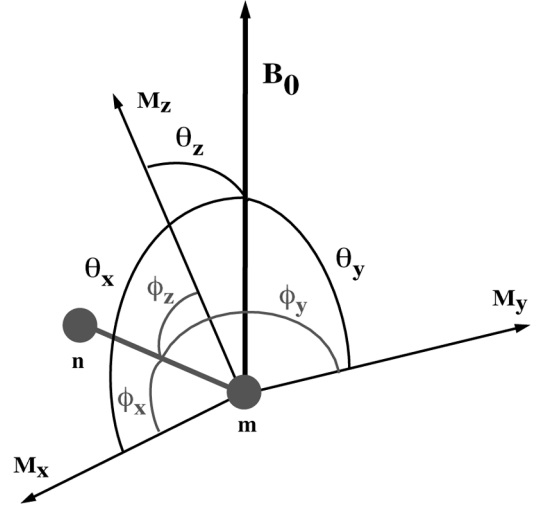
$$\delta_{an} = \frac{2}{3} \sum_{j=\{x,y,z\}} \left\langle \frac{3 \cos^2 \Theta_j - 1}{2} \right\rangle \delta_{jj}, \quad [2]$$

where  $\delta_{jj}$  are the elements of the diagonal chemical shift tensor and  $\Theta_j$  are the angles between the axis of the chemical shift principal frame and the external magnetic field. Once again, the angle brackets denote a time average over motions.

It is clear from Eqs. [1] and [2] that  $D^{nm}$  and  $\delta_{an}$  are sources of dynamic and structural information through their dependence upon  $\Theta$  and  $\Theta_j$  and upon the motional averaging of these quantities. Further, this information is unique in being not only distance dependent, but angle dependent, which makes it a very important tool for obtaining long-range structural information. Although the dynamic information available is highly useful, it also makes the analysis of these anisotropic interactions complicated. A simple approach to obtaining structural and motional information from such parameters is to represent their anisotropic averaging by an order matrix (17, 20, 21). For a molecule with coordinates defined in an arbitrary Cartesian system, the elements of this  $3 \times 3$  order matrix are

$$S_{ij} = \left\langle \frac{3 \cos \theta_i \cos \theta_j - k_{ij}}{2} \right\rangle, \quad [3]$$

where  $\theta_i$  denotes the instantaneous orientation of the  $i$ th molecular axis with respect to the director (which in the studied case is the direction of the external magnetic field, Fig. 1), and  $k_{ij}$  is the Kronecker delta. The order matrix is symmetric ( $S_{ij} = S_{ji}$ ) and traceless ( $S_{xx} + S_{yy} + S_{zz} = 0$ ), so it has only five



**FIG. 1.** Definition of angles with respect to the molecular frame.  $\phi_x, \phi_y, \phi_z$ : angles of an internuclear vector with respect to the molecular axis.  $\theta_x, \theta_y, \theta_z$ : angles defining the instantaneous orientation of the magnetic field ( $B_0$ ) with respect to the molecular axis.

independent elements. By diagonalizing the order matrix, it is possible to reduce the order parameter description to a principal order parameter,  $S_{z'z'}$ , and an asymmetry parameter,  $\eta = (S_{y'y'} - S_{x'x'})/S_{z'z'}$ , where  $|S_{z'z'}| > |S_{y'y'}| > |S_{x'x'}|$ . The principal order parameter and  $\eta$  are straightforwardly related to the axial and rhombic components of the alignment tensor that are used if the dipolar coupling is expressed using a polar coordinate system (16, 22). The coordinate frame in which the order matrix is diagonal is often referred to as the principal averaging frame or the principal order frame. The transformation matrix that accomplishes this diagonalization relates this principal frame to the initial molecular frame.

If the principal order frame happens to coincide with the direction of the dipolar interaction vector, or the frame of the chemical shift tensor, it is clear that the order tensor elements are identical to the angular function in the expression for  $\delta_{an}$  and  $D^{nm}$ . In the more general case where we choose an arbitrary molecular frame, the following expressions result:

$$D^{mn} = D_{\max}^{mn} \sum_{ij=\{x,y,z\}} S_{ij} \cos \phi_i^{nm} \cos \phi_j^{nm} \quad [4]$$

$$\delta_{an}^n = \frac{2}{3} \sum_{ij=\{x,y,z\}} S_{ij} \delta_{ij}^n. \quad [5]$$

Here  $\phi_i^{nm}$  is the angle of the internuclear vector connecting nuclei  $n$  and  $m$  relative to the  $i$ th molecular axis and the  $\delta_{ij}^n$  are the elements of the chemical shift tensor in an arbitrary molecular frame of nucleus  $n$ . Equation [4] suggests that if the direction cosines of the internuclear vectors in an arbitrary molecular frame are known, it is possible to determine the order parameters, and

hence the molecular orientational properties, from just five independent measurements. Equation [5] suggests that if the orientation of the chemical shift principal axis frame and the principal values are known, chemical shift anisotropy (CSA) data,  $\delta_{an}^n$ , can be substituted for some of these measurements. CSA data for certain functional groups are readily available from literature (23, 24). Knowledge of direction cosines of a sufficient number of internuclear vectors or chemical shift tensors in a particular molecular frame is only available if the local structure of a fragment is known and assumed to be rigid. Of course there is always motion, but provided that, within a fragment, all internuclear vector motion is axially symmetric about its mean position and uniform for all vectors, the fragment can be treated as rigid and this motion can be taken into account by a simple scaling factor. This is approximately the case for local N–H bond librations, for example.

If a sufficient number of interactions can be measured for each of the two fragments with well-defined structures, and if an order tensor in the molecular frame can be determined for each, then the allowed possibilities for relative orientations of these two fragments can be dramatically reduced. Because the interactions we measure are insensitive to inversion of the director, there will, in general, be four possible orientations for each fragment. However, other constraints such as molecular bonding, help to further reduce the possibilities (25, 26).

Carrying out an order matrix determination in practice means setting up and solving a system of linear equations of the form  $\mathbf{A} \mathbf{x} = \mathbf{b}$ . In our case,  $\mathbf{A}$  is a matrix composed of the direction cosines of the internuclear vectors and  $\delta_{ij}$ 's. The matrix  $\mathbf{A}$  and the  $\mathbf{x}$  and  $\mathbf{b}$  vectors are defined in Eq. [6]:

$$\begin{pmatrix} \cos^2 \phi_y^1 - \cos^2 \phi_x^1 & \cos^2 \phi_z^1 - \cos^2 \phi_x^1 & 2 \cos \phi_x^1 \cos \phi_y^1 & 2 \cos \phi_x^1 \cos \phi_z^1 & 2 \cos \phi_y^1 \cos \phi_z^1 \\ \cos^2 \phi_y^2 - \cos^2 \phi_x^2 & \cos^2 \phi_z^2 - \cos^2 \phi_x^2 & 2 \cos \phi_x^2 \cos \phi_y^2 & 2 \cos \phi_x^2 \cos \phi_z^2 & 2 \cos \phi_y^2 \cos \phi_z^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cos^2 \phi_y^n - \cos^2 \phi_x^n & \cos^2 \phi_z^n - \cos^2 \phi_x^n & 2 \cos \phi_x^n \cos \phi_y^n & 2 \cos \phi_x^n \cos \phi_z^n & 2 \cos \phi_y^n \cos \phi_z^n \\ -\delta_{11}^1 + \delta_{22}^1 & -\delta_{11}^1 + \delta_{33}^1 & -\delta_{12}^1 + \delta_{21}^1 & -\delta_{13}^1 + \delta_{31}^1 & -\delta_{23}^1 + \delta_{32}^1 \\ -\delta_{11}^2 + \delta_{22}^2 & -\delta_{11}^2 + \delta_{33}^2 & -\delta_{12}^2 + \delta_{21}^2 & -\delta_{13}^2 + \delta_{31}^2 & -\delta_{23}^2 + \delta_{32}^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ -\delta_{11}^n + \delta_{22}^n & -\delta_{11}^n + \delta_{33}^n & -\delta_{12}^n + \delta_{21}^n & -\delta_{12}^n + \delta_{21}^n & -\delta_{23}^n + \delta_{32}^n \end{pmatrix} \begin{pmatrix} S_{yy} \\ S_{zz} \\ S_{xy} \\ S_{xz} \\ S_{yz} \end{pmatrix} = \begin{pmatrix} D_{red}^1 \\ D_{red}^2 \\ \cdot \\ D_{red}^n \\ \delta_{an}^1 \\ \delta_{an}^2 \\ \cdot \\ \delta_{an}^n \end{pmatrix}, \quad [6]$$

In the case at hand, we will assume that the central cores of each of the two domains in the small protein, BLBC, represent an appropriate a rigid fragment. There are numerous noncollinear  $^{15}\text{N}$ – $^1\text{H}$  dipolar interactions in each of these cores. In other proteins, secondary structure elements such as helices or beta sheets may constitute appropriate fragments. In the more general case of a polypeptide chain, the peptide plane can be considered as such a rigid structural element. It is possible to measure five independent angular parameters from this unit by expanding the type of the anisotropic interactions we measure, for example,  $^1\text{D}_{\text{NH}}$ ,  $^1\text{D}_{\text{NC}}$ ,  $^1\text{D}_{\text{CC}}$ , and  $\delta_{an}$  for the carbonyl carbon and the amide nitrogen. The principal frames of the chemical shift anisotropy tensors and the internuclear vectors for these interactions each have a different orientation with respect to the averaging frame, making them nonredundant. It is useful to note that well-defined segments exist in other types of macromolecules. In the case of oligosaccharides, many sugar rings can be assumed rigid and five or more  $^1\text{D}_{\text{CH}}$ ,  $^1\text{D}_{\text{CC}}$ , or  $^n\text{D}_{\text{CH}}$  couplings can be measured.

where rows containing a reduced dipolar coupling,  $D_{red}$ , are obtained from Eq. [4], and rows with  $\delta_{an}$  are obtained from Eq. [5] after division by  $D_{\text{max}}$  and  $2/3$ , respectively. The number of equations depends on the number of measured angular parameters, but the number of unknowns is always five.

In the past, solutions to the above equation have been obtained by a simple grid search (27) or a random search (28) for allowed values of the five independent order tensor elements. It is also possible to use a fitting program that minimizes the differences between measured and observed couplings (5). Although these approaches are sufficient for simple cases, they are quite slow if the linear equation system is large, that is, if the number of measured angular parameters is large.

Singular value decomposition (SVD) is a powerful numerical technique for solving systems of linear equations, such as that given in Eq. [6], and is easy to implement for order matrix calculations. It also has the advantage that it can deal with sets of equations that are close to singular, which would happen if the measured angular data were redundant because of collinearity of vectors.

It is well known from linear algebra (29) that any matrix  $\mathbf{A}$  with  $M$  rows and  $N$  columns can be written in terms of its singular value decomposition, i.e., as the product of an  $M \times N$  column-orthogonal matrix  $\mathbf{U}$ , and an  $N \times N$  diagonal matrix  $\mathbf{W}$ , with nonnegative diagonal elements, and the transpose of an  $N \times N$  orthogonal matrix  $\mathbf{V}$ . That is,

$$(\mathbf{A}) = (\mathbf{U}) \begin{pmatrix} \omega_1 & & & 0 \\ & \omega_2 & & \\ & & \ddots & \\ 0 & & & \omega_N \end{pmatrix} (\mathbf{V}^T). \quad [7]$$

This decomposition of  $\mathbf{A}$  allows one to write the inverse of  $\mathbf{A}$  as

$$\mathbf{A}^{-1} = \mathbf{V}[\text{diag}(1/\omega_j)]\mathbf{U}^T. \quad [8]$$

The matrix  $\mathbf{A}$  is ill-conditioned with respect to inversion if any one of the  $\omega_i$ 's is zero or very small. In this case, we can set  $1/\omega_i$  to zero, which is equivalent to throwing away the part of the solution space that is most susceptible to roundoff errors.

For our matrix equation  $\mathbf{A} \mathbf{x} = \mathbf{b}$ ,  $\mathbf{x}$  can be determined using the above decomposition as

$$\mathbf{x} = \mathbf{V}[\text{diag}(1/\omega_j)]\mathbf{U}^T\mathbf{b}. \quad [9]$$

If the set of equations has no exact solution, as in an overdetermined linear system ( $M > N$ ), the SVD will still produce a solution that will not exactly solve the linear system, but will be the best solution in the least squares sense. Also, SVD can be used for underdetermined ( $N > M$ ) systems. In this case one can isolate the subspace about which we have no information (also called the nullspace). The columns of  $\mathbf{V}$  corresponding to zero  $\omega_i$ 's then form an orthonormal basis that spans the nullspace.

One difficulty encountered in the calculations described above is that experimental uncertainties have to be considered. One way to take this into account is to calculate  $S_{ij}$  for several sets of dipole couplings and chemical shift anisotropies that are sampled from Gaussian distributions centered at the measured values with standard deviations depending on the experimental precisions. The calculated  $S_{ij}$  values are then multiplied with the  $\mathbf{A}$  matrix and are only accepted if the predicted  $D^{nm}$  or  $\delta_{an}^n$  values are within the estimated experimental error. This step is necessary since the SVD method will produce a least squares solution for any set of input data. The width of the resulting distribution of  $S_{ij}$  values gives a representation of errors in the derived parameters. Figure 2 shows a flow chart of order matrix calculations done using SVD.

## IMPLEMENTATION

The order matrix calculations performed by both random search and SVD were coded using the C programming language. Output visualization was performed using the Xmgr software package (P.J. Turner, v. 3.01). The random number generator and the SVD algorithm were modified from published subroutines (LAPACK, v. 2.0). After each set of order parameters was determined and a symmetric tensor formed from them, the tensor was diagonalized to obtain the principal order tensor components, or order parameters, and a transformation matrix that relates the principal order frame to the initial molecular frame.

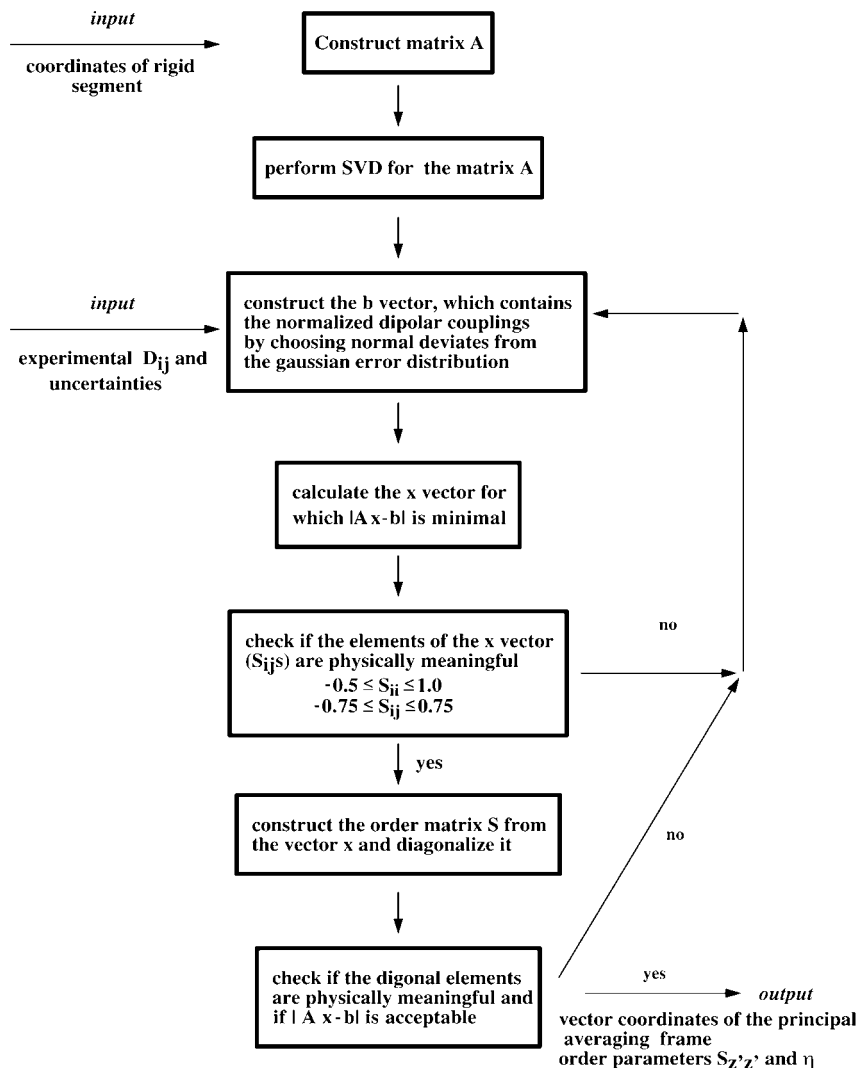
Performing the above operation yields a distribution of order parameters and a collection of vectors defining the possible directions of the principal order parameters,  $S_{x'x'}$ ,  $S_{y'y'}$ , and  $S_{z'z'}$ , all of which are consistent with the experimental data. To facilitate the visualization of these distributions of vectors, a mapping technique was employed. The equal area pseudocylindrical Sauson–Flamsted projection (30) is well suited for this purpose (28, 31).

## RESULTS

Application of the proposed method is illustrated using a  $^{15}\text{N}$ -labeled protein, BLBC, a two-domain fragment of barley lectin. The 89-residue BLBC fragment shares approximately 95% sequence homology with wheat germ agglutinin, WGA, which has been the subject of numerous crystal structure studies (32). Previous NOE-based studies of BLBC were not sufficient to determine a high-resolution structure, but they did show the domains to have folds very similar to those seen in the crystal structure of WGA, and they did indicate that the center part of each domain was well structured (33). More significantly for our purposes, the relative orientation of the B and C domains of BLBC could not be determined because of the lack of long-range restraints. Thus, BLBC is an ideal candidate for the determination of domain–domain orientation using residual dipolar couplings.

To use the order matrix approach, we need a very well defined structure within the core of each domain. The NOE derived structures are not, in themselves, adequate, since they have RMSDs of 1.8 and 1.5 Å, respectively, even for the well-structured regions. Instead, since it has been shown that BLBC and WGA have very similar backbone folds (33), we used the X-ray structure of WGA (32) as the rigid structure from which we calculated the direction cosines in matrix  $\mathbf{A}$ .

To obtain residual dipolar coupling data, an approximately 0.2 mM protein sample was oriented in a 5% 2.9:1 DMPC:DHPC (dimyristoylphosphatidylcholine/dihexanoylphosphatidylcholine) bicelle solution that has been stabilized by the addition of a positively charged lipid, CTAB (hexadecyl(cetyl)-trimethylammonium bromide) (34).  $^1\text{D}_{\text{NH}}$  scalar and dipolar



**FIG. 2.** Flow chart of the program that solves for the elements of the Saupe order matrix from experimental dipolar coupling data using the method of singular value decomposition.

couplings were measured using a quantitative  $J$ -correlation experiment ( $J_{\text{NH}}\text{-HSQC}$ ) (35) designed for the accurate measurement of one-bond amide  $^{15}\text{N}\text{-}^1\text{H}$  couplings in proteins. In this experiment the couplings are encoded in the resonance intensity, reducing the need for high resolution in the indirect dimension. Couplings were measured using this experiment in a bicelle medium at  $25^\circ\text{C}$  where the medium is isotropic and only scalar couplings contribute and at  $34^\circ\text{C}$  where the medium is oriented and both dipolar and scalar couplings contribute. The dipolar contribution was extracted by taking differences of splittings measured in these two experiments.

Normal HSQC spectra were also collected for the protein in pure water, in the bicelle medium at  $25^\circ\text{C}$  where the system is nearly isotropic, and in the bicelle medium at  $34^\circ\text{C}$  where the system has undergone the transition to a liquid crystalline state and the protein is partially aligned. Since the appearance of the

$^1\text{H}\text{-}^{15}\text{N}$  HSQC spectrum in all three cases is comparable, it is reasonable to assume that the presence of the lipid does not change the tertiary fold of the protein. For the well-structured region of domain B, 16 dipolar couplings were measured. For the well-structured region of domain C, 24 dipolar couplings were measured. These data are presented in Tables 1a and 1b. The measured dipolar couplings range from 0 to 7.5 Hz in absolute value. Experimental precision is estimated to be on the order of 0.2–0.5 Hz.

Order matrix calculations were performed on the data presented in Table 1 using the SVD method with an estimated uncertainty of 2–3 Hz. These increased uncertainty estimates partially account for the fact that in addition to experimental error, uncertainty in our rigid core structures will contribute to the precision of our calculations. Since the singular value decomposition method solves the exact system of linear equa-

**TABLE 1a**  
Residual  $^1\text{H}$ - $^{15}\text{N}$  Dipolar Couplings from the B Domain of the BLBC Protein

Secondary structure <sup>a</sup>	Residue # <sup>b</sup>	Measured dipolar coupling (Hz) <sup>c</sup>	Used in final solution <sup>d</sup>
$\beta$ -Sheet	*His59	7.3	yes
	Cys60	-2.5	yes
	Cys61	0.3	no
	Ser62	4.0	yes
	Gln63	6.7	no
	*Trp64	-1.4	yes
	Gly65	4.3	yes
	Tyr66	6.3	yes
	Cys67	3.8	yes
	Gly68	-1.0	no
Phe69	-3.4	yes	
	Gly70	N/A	N/A
$\alpha$ -Helix	Ala71	N/A	N/A
	Glu72	-2.2	yes
	Tyr73	-2.2	yes
	Cys74	-4.6	yes
	Gly75	N/A	N/A
	Ala76	-1.2	yes
	Gly77	N/A	N/A
	Cys78	5.3	yes

<sup>a</sup> Secondary structure as determined by high-resolution NMR spectroscopy. For domain B the well structured residues are 59–81 and for domain C they are 100–126.

<sup>b</sup> \* indicates residues that differ between BLBC and WGA.

<sup>c</sup> The data was collected using a quantitative  $J$  correlation experiment (see text). NMR spectroscopy was carried out on a 500-MHz Varian INOVA spectrometer. In the direct dimension, 1024 complex points were collected with a spectral width of 6000 Hz; 70  $t_1$  increments with a corresponding spectral width of 2000 Hz were collected in the indirect dimension. For each  $t_1$  increment, 32 transients were collected per FID, and four FIDs were stored per  $t_1$  increment. The constant time delay was set to 64.516 ms, which is optimized for  $J = 93$  Hz. A 2-s recycle delay and an acquisition time of 170 ms were used.

<sup>d</sup> No indicates residues from which the measured residual dipolar coupling was not consistent with rest of the data (see text).

tions, it is possible to quickly identify dipolar couplings that have uncertainties outside of these bounds. This was accomplished here by counting the number of times that particular data could not be back-calculated to within estimated uncertainty by the common solution and eliminating them if this happens for more than 50% of the cycles. Thirteen out of the 16 measured values for domain B were found to be compatible with the rigid structure, and 21 out of the 24 measured values for domain C were found to be compatible with the rigid structural model. The inconsistencies most likely result from imperfections in the assumed structure or additional motional effects at those sites. It is worth pointing out that most of the poorly fit data result from sites that are at or close to residues

that differ between BLBC and WGA (Table 1). The 10,000 cycles required to fully account for uncertainties in the acceptable order matrices (see Theory section) consumed less than 60 s of CPU time on a 150-MHz Silicon Graphics Indy R4400

**TABLE 1b**  
Residual  $^1\text{H}$ - $^{15}\text{N}$  Dipolar Couplings from the C Domain of the BLBC Protein

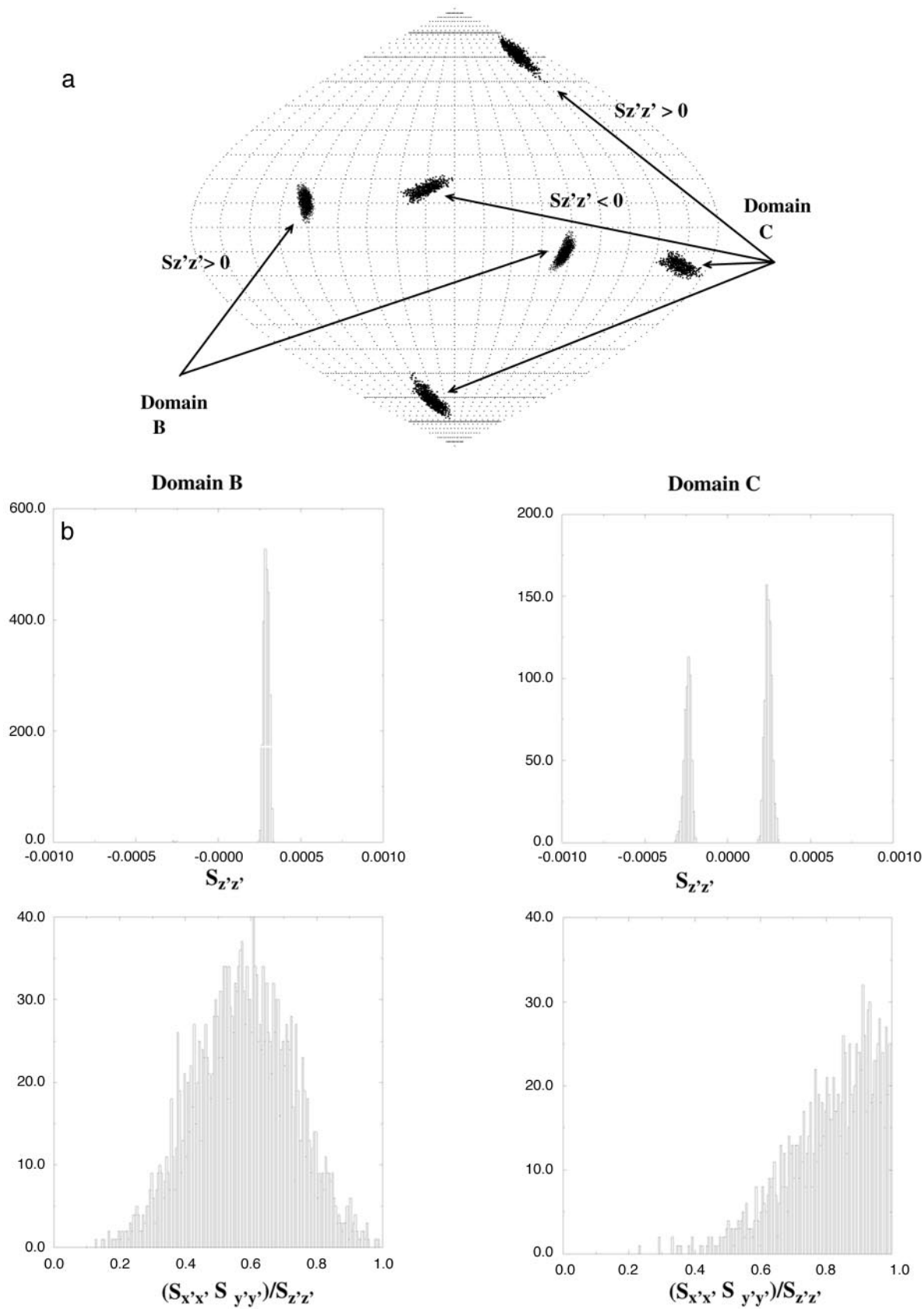
Secondary structure <sup>a</sup>	Residue # <sup>b</sup>	Measured dipolar coupling (Hz) <sup>c</sup>	Used in final solution <sup>d</sup>
$\beta$ -Sheet	Asn100	4.4	yes
	Asn101	4.8	yes
	Leu102	-0.1	yes
	Cys103	2.7	yes
	Cys104	N/A	N/A
	Ser105	-5.8	no
	Gln106	5.4	yes
	Trp107	6.7	yes
	Gly108	N/A	no
	*Tyr109	-2.7	yes
	Cys110	-4.5	yes
	Gly111	-5.1	yes
Leu112	-3.1	yes	
	Gly113	N/A	N/A
$\alpha$ -Helix	Ser114	3.0	yes
	Glu115	-3.9	yes
	Phe116	4.8	yes
	Cys117	3.4	yes
	Gly118	-2.7	yes
	*Glu119	4.5	yes
	Gly120	-2.9	yes
	Cys121	6.0	no
$\beta$ -Sheet	Gln122	-4.8	yes
	*Gly123	7.0	no
	Gly124	2.9	yes
	Ala125	-6.0	yes
	Cys126	4.2	yes

<sup>a</sup> Secondary structure as determined by high-resolution NMR spectroscopy. For domain B the well structured residues are 59–81 and for domain C they are 100–126.

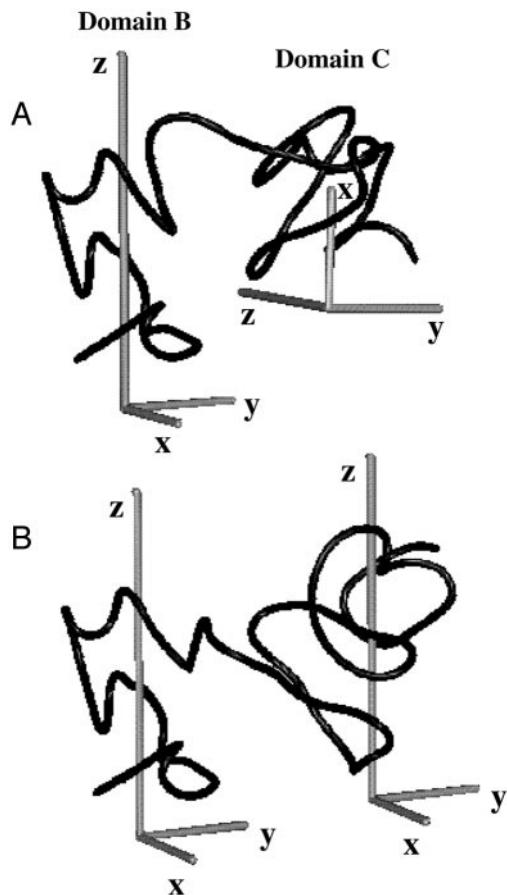
<sup>b</sup> \* indicates residues that differ between BLBC and WGA.

<sup>c</sup> The data was collected using a quantitative  $J$  correlation experiment (see text). NMR spectroscopy was carried out on a 500-MHz Varian INOVA spectrometer. In the direct dimension, 1024 complex points were collected with a spectral width of 6000 Hz; 70  $t_1$  increments with a corresponding spectral width of 2000 Hz were collected in the indirect dimension. For each  $t_1$  increment, 32 transients were collected per FID, and four FIDs were stored per  $t_1$  increment. The constant time delay was set to 64.516 ms, which is optimized for  $J = 93$  Hz. A 2-s recycle delay and an acquisition time of 170 ms were used.

<sup>d</sup> No indicates residues from which the measured residual dipolar coupling was not consistent with rest of the data (see text).



**FIG. 3.** (a) Possible orientations of the axis of the principal averaging frame that corresponds to  $S_{z'z'}$  (director). The mapping frame for both domains is the molecular frame of the X-ray structure. If the relative orientation of the two domains determined from the residual dipolar couplings was the same as in the X-ray structure, the possible director orientation for the two domains should overlap. For domain B, there are two possible orientations of the director, since it is not possible to distinguish between  $+z$  and  $-z$ . For domain C, there are four possible orientations of the director, which shows that this domain is highly rhombic. (b) Distributions of  $S_{z'z'}$  and  $\eta$  for both domains.



**FIG. 4.** (A) Domain–domain orientation of the domains B and C in the barley lectin protein fragment BLBC based on the crystal structure of WGA. Representative principal averaging frames for both domains are also shown. (B) A possible domain–domain orientation of the domains B and C, based on orientational constraints derived from residual  $^{15}\text{N}$ – $^1\text{H}$  dipolar couplings. The structure was obtained by selecting a principal averaging frame for both domains so that the corresponding  $\eta$  values were comparable ( $\eta \sim 0.7$ ). After this, the molecular frames of the two domains were reoriented in such a way that the principal averaging frames coincided.

computer. The same calculation using the random search approach (28) required several hours of CPU time.

The above calculation resulted in approximately 2000 acceptable order matrices for domain B and 1500 for domain C. Figure 3A illustrates the direction for that axis of the principal averaging frame that corresponds to  $S_{z'z'}$ , for both domains, in the molecular frame of the crystal structure. Figure 3B presents the distributions of  $S_{z'z'}$  and  $\eta$  for each domain. Note that there are two sets of solutions for domain C. This occurs because  $\eta = 1.0$ , and the choice of orientation of the  $z$  and  $y$  axes becomes arbitrary. It is clear from Fig. 3A that the clusters are, in fact,  $90^\circ$  apart.

Since the domain structures were constrained to the domain structures of WGA, and since a common coordinate frame was used for both domains, coincidence of the principal ordering

frames of the two domains provides a test for consistency with domain orientation as found in the crystal structure. In Figure 3A the representations of director orientations do not overlap, even given the degenerate representations for domain C. In other words, the director orientations appear quite different when viewed from the perspective of each domain. It is therefore obvious that the relative orientation of the two domains of BLBC must be different from that determined by X-ray crystallography of WGA. One can, in principle, generate a structure consistent with the experimental data by rotating the domains to achieve a common ordering frame (Fig. 4). This common frame should be determined in such a way that axes with order parameters of the same sign and similar magnitude coincide. An allowed structure is depicted in Fig. 4B.

The picture in Fig. 4B is, however, a significant oversimplification of the situation. In addition to determining the principal averaging frame, the above calculations return values of the principal order parameters that are consistent with the data. Figure 3B shows that while  $S_{z'z'}$  is similar for domains B and C (0.00025),  $\eta$  is broadly distributed about 0.6 in domain B and it is near 1.0 in domain C. The distributions of calculated order parameters shown in Fig. 3B do overlap to some extent, allowing for the possibility that the two domains are ordered by the same forces and behave as a single rigid entity. However, additional experiments show that even small variations in ordering conditions force a more severe divergence of principal order parameters, suggesting that the two domains of BLBC diffuse independently. This complex behavior is the subject of a forthcoming detailed study of the structure and dynamic properties of BLBC.

## CONCLUSION

In this work, we have shown that using singular value decomposition as a method for determining the Saupe order matrix is a fast and easy way to interpret anisotropic spin interactions, such as residual dipolar couplings. Although local structural information is required for the described method, a useful calculation can be performed using as few as five, and in some cases even fewer, measured couplings. When the local structure of a molecular fragment is known, it might also be the method of choice for determining the alignment parameters,  $D_a$  and  $R$ , required for further structural refinement using simulated annealing (15). Other methods that have been proposed for the determination of these order parameters (15, 16) depend on observing a large number of couplings to discern the distribution of couplings. In cases of limited data, especially with high rhombicity ( $\eta$  near 1.0), these methods can become inapplicable.



### Availability

Information on obtaining the software described in this paper can be found on our Web page accessed through <http://www.ccrcc.uga.edu>.

### ACKNOWLEDGMENTS

We thank Mr. Hashim M. Al-Hashimi for stimulating discussions on order matrix calculations. We also thank Dr. Jeanne Lim Weaver for providing us with the  $^{15}\text{N}$ -labeled BLBC sample. This work was supported by grant GM33225 from the National Institutes of Health and grant MCB-9726341 from the National Science Foundation.

### REFERENCES

1. K. Wüthrich, "NMR of Proteins and Nucleic Acids." Wiley, New York (1986).
2. E. W. Bastian, C. Maclean, P. C. M. V. Zijl, and A. A. Bothner-By, *Ann. Rep. NMR Spect.* **19**, 35–77 (1987).
3. J. R. Tolman, J. M. Flanagan, M. A. Kenedy, and J. H. Prestegard, *Proc. Natl. Acad. Sci. USA* **92**, 9297–9283 (1995).
4. H. C. Kung, K. Y. Wang, I. Goljer, and P. H. Bolton, *J. Magn. Reson. Ser. B* **109**, 323–325 (1995).
5. N. Tjandra, S. Grzesiek, and A. Bax, *J. Am. Chem. Soc.* **118**, 6264–6272 (1996).
6. A. Bax and N. Tjandra, *J. Biomol. NMR* **10**, 289–293 (1997).
7. G. M. Clore, M. R. Starich, and A. M. Gronenborn, *J. Am. Chem. Soc.* **120**, 10571–10572 (1998).
8. M. R. Hansen, L. Mueller, and A. Pardi, *Nature Struct. Biol.* **5**, 1065–1074 (1998).
9. R. S. Prosser, J. A. Losonczi, and I. V. Shiyonovskaya, *J. Am. Chem. Soc.* **120**, 11010–11011 (1998).
10. X. Liu, H. Wang, M. Eberstadt, A. Schnuchel, E. T. Olejniczak, R. P. Meadows, J. M. Schkeryantz, D. A. Janowick, J. E. Harlan, D. E. S. Harris, D. E. Staunton, and S. W. Fesik, *Cell* **95**, 269–277 (1998).
11. C. A. Bewley, K. R. Gustafson, M. R. Boyd, D. G. Covell, A. Bax, G. M. Clore, and A. M. Gronenborn, *Nature Struct. Biol.* **5**, 571–578 (1998).
12. P. J. Bolon and J. H. Prestegard, *J. Am. Chem. Soc.* **120**, 9366–9367 (1998).
13. G. R. Kiddle and S. W. Homans, *FEBS Lett.* **436**, 128–130 (1998).
14. R. D. Beger, V. M. Marathias, B. F. Volkman, and P. H. Bolton, *J. Magn. Reson.* **135**, 256–259 (1998).
15. G. M. Clore, A. M. Gronenborn, and N. Tjandra, *J. Magn. Reson.* **131**, 159–162 (1998).
16. G. M. Clore, A. M. Gronenborn, and A. Bax, *J. Magn. Reson.* **133**, 216–221 (1998).
17. A. Saupe, *Angew. Chem., Int. Ed. Engl.* **7**, 97 (1968).
18. C. R. Sanders, B. J. Hare, K. P. Howard, and J. H. Prestegard, *Prog. NMR Spectrosc.* **26**, 421–444 (1994).
19. C. R. Sanders and G. C. Landis, *Biochemistry* **34**, 4030–4040 (1995).
20. J. W. Emsly and J. C. Lindon, "NMR Spectroscopy Using Liquid Crystal Solvents." Pergamon Press, Oxford, UK (1975).
21. P. Diehl and C. L. Khetrapal, "NMR: Basic Principles and Progress." Springer-Verlag, New York (1969), Vol. 1.
22. A. A. Bothner-By, Magnetic Field Induced Alignment of Molecules, in "Encyclopedia of Nuclear Magnetic Resonance" (D. M. Grant and R. K. Harris, Eds.), pp. 2932–2938. Wiley, Chichester (1995).
23. D. Fushman, N. Tjandra, and D. Cowburn, *J. Am. Chem. Soc.* **120**, 10947–10952 (1998).
24. T. G. Oas, C. J. Hartzell, T. J. McMahon, G. P. Drobny, and F. W. Dahlquist, *J. Am. Chem. Soc.* **109**, 5956–5962 (1987).
25. R. R. Ketchum, K. C. Lee, S. Huo, and T. A. Cross, *J. Biomol. NMR* **8**, 1–14 (1996).
26. R. R. Ketchum, B. Roux, and T. A. Cross, *Structure* **5**, 1655–1669 (1997).
27. C. R. Sanders and J. H. Prestegard, *J. Am. Chem. Soc.* **114**, 7096–7107 (1992).
28. J. A. Losonczi and J. H. Prestegard, *Biochemistry* **37**, 706–716 (1998).
29. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical Recipes in C." Cambridge University Press, Cambridge (1992).
30. L. M. Bugayevskiy and J. P. Snyder, "Map Projections." Taylor & Francis, London (1995).
31. J. H. Prestegard, J. R. Tolman, H. M. Al-Hashimi, and M. Andreac, Protein Structure and Dynamics from Field Induced Residual Dipolar Couplings, in "Modern Techniques in Protein NMR" (N. R. Krishna and L. J. Berliner, Eds.). Plenum, New York (1998).
32. C. S. Wright, *J. Mol. Biol.* **215**, 635–651 (1990).
33. J. L. Weaver and J. H. Prestegard, *Biochemistry* **37**, 116–128 (1998).
34. J. A. Losonczi and J. H. Prestegard, *J. Biomol. NMR* **12**, 447–451 (1998).
35. J. R. Tolman and J. H. Prestegard, *J. Magn. Reson. B* **112**, 245–252 (1996).