Journal of Biomolecular NMR, **22:** 137–151, 2002.
KLUWER/ESCOM

Exact solutions for chemical bond orientations from residual dipolar couplings

William J. Wedemeyer[a], Carol A. Rohl[b] & Harold A. Scheraga[a,*]
[a]*Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853-1301, U.S.A.;* [b]*Department of Biochemistry, Box 357350, University of Washington, Seattle, WA 98195-7350, U.S.A.*

Received 28 September 2001; Accepted 15 November 2001

Key words: orientational constraints, polynomial equations, protein structure, residual dipolar couplings, rhombicity, Saupe order matrix, solution NMR structure determination, unit sphere

Abstract

New methods for determining chemical structures from residual dipolar couplings are presented. The fundamental dipolar coupling equation is converted to an elliptical equation in the principal alignment frame. This elliptical equation is then combined with other angular or dipolar coupling constraints to form simple polynomial equations that define discrete solutions for the unit vector(s). The methods are illustrated with residual dipolar coupling data on ubiquitin taken in a single anisotropic medium. The protein backbone is divided into its rigid groups (namely, its peptide planes and C^{α} frames), which may be solved for independently. A simple procedure for recombining these independent solutions results in backbone dihedral angles ϕ and ψ that resemble those of the known native structure. Subsequent refinement of these ϕ-ψ angles by the ROSETTA program produces a structure of ubiquitin that agrees with the known native structure to 1.1 Å C^{α} rmsd.

Introduction

Orientational constraints for solution NMR structure determination may be obtained from residual dipolar couplings (Tolman et al., 1995; Tjandra and Bax, 1997; Prestegard, 1998). Such couplings may be measured by introducing a relatively weak anisotropy into the molecular tumbling, e.g., by the addition of anisotropic elements such as bicelles (Tjandra and Bax, 1997; Ottiger and Bax, 1998) or other liquid crystalline media (Rückert and Otting, 2000). Residual dipolar couplings have the advantage that they provide *long-range* constraints on the molecular structure, whereas traditional NOE and *J*-coupling measurements provide only *short-range* constraints that may not define the relative orientation of distant or poorly coupled parts of the molecule. Dipolar coupling data have been used to refine molecular structures that have already been determined by NOE and *J*-coupling data (Clore et al., 1999; Mueller et al., 2000). It has also been possible to solve NMR structures directly from dipolar coupling data (i.e., without NOE and *J*-coupling data), provided that additional constraints are imposed (Delaglio et al., 2000; Hus et al., 2000, 2001; Fowler et al., 2000). Dipolar coupling and *J*-coupling experiments can be carried out significantly faster than NOE experiments, and may be applied to larger proteins as well; hence, methods for solving protein structures from dipolar couplings and *J* couplings may be useful in high-throughput NMR structure determination (Medek et al., 2000).

The refinement of NMR structures by dipolar coupling data generally proceeds by trial-and-error methods similar to those used in the refinement of X-ray crystal structures, e.g., by a simulated annealing protocol to minimize an '*R* factor' that measures the agreement of the molecular structure with the dipolar coupling data (Clore and Garrett, 1999). Such a protocol, however, often encounters a multiple-minima problem (Chou et al., 2000), which may be overcome only by complex simulated annealing protocols.

*To whom correspondence should be addressed. E-mail: has5@cornell.edu

Moreover, this approach does not exploit the *locality* of dipolar coupling data, i.e., whereas the intensity of a single X-ray reflection depends on the electron density of the *entire* molecular structure, a dipolar coupling measurement pertains to the orientation of a *single bond* in the molecule.

In this paper, we show that dipolar coupling constraints can be expressed as a simple elliptical equation. Analytic solutions of this equation are characterized, and then combined with various additional constraints to produce polynomial equations whose real roots correspond to all possible solutions for the unit vector that are consistent with the dipolar coupling measurement(s) and the additional constraint(s). In several cases, the equation is quartic and, hence, an explicit formula can be given for the solutions. In other cases, the set of possible solutions is found by a simple one-dimensional search over a single parametric angle. These solutions may prove helpful in determining protein structures rapidly from dipolar coupling data. We illustrate these methods by applying them to the structure determination of ubiquitin from a set of dipolar couplings taken in a single anisotropic medium.

Elliptical form of the fundamental dipolar coupling equation

In this section, the fundamental dipolar coupling equation is expressed as a simple elliptical equation. Although another solution for the dipolar coupling equation has been published recently (Skrynnikov and Kay, 2000), the present solution has the advantage of simplicity and allows *exact* solutions to be determined as the roots of a quartic polynomial, as demonstrated in the next section.

The fundamental dipolar coupling equation for a chemical-bond unit vector **n** in the principal alignment frame may be written (Saupe, 1968; Losonczi et al., 1999)

$$\begin{aligned} D_n &= [-D_a + (3/2)D_r]n_x^2 \\ &\quad + [-D_a - (3/2)D_r]n_y^2 + 2D_a n_z^2, \end{aligned} \tag{1}$$

where D_a and D_r represent the axial and rhombic components of the dipolar coupling tensor. This equation may be expressed in elliptical form by employing the unit-vector condition $n_x^2 + n_y^2 + n_z^2 = 1$ to eliminate either n_y or n_z. If n_z^2 is eliminated from the fundamental Equation 1, we obtain the equation

$$\begin{aligned} 2D_a - D_n &= [3D_a - (3/2)D_r]n_x^2 \\ &\quad + [3D_a + (3/2)D_r]n_y^2. \end{aligned} \tag{2}$$

Provided that D_n/D_a is greater than a critical value $\delta_{\text{crit}} \equiv D_{xx}/D_a$, this equation is equivalent to that of an ellipse

$$\left(\frac{n_x}{\xi_n}\right)^2 + \left(\frac{n_y}{\eta_n}\right)^2 = 1, \tag{3}$$

where the semiaxes of the ellipse are given by

$$\xi_n^2 \equiv \frac{2D_a - D_n}{3D_a + (3/2)D_r}, \tag{4}$$

$$\eta_n^2 \equiv \frac{2D_a - D_n}{3D_a + (3/2)D_r}. \tag{5}$$

(The subscript on the semiaxes ξ_n and η_n indicates the unit vector to which they pertain, in this case **n**.) Both ξ_n^2 and η_n^2 are less than one provided that the condition $(D_n/D_a) > \delta_{\text{crit}}$ holds. Hence, the fundamental dipolar Equation 1 is satisfied by any vector of the form

$$n_x = \xi_n \cos\tau_n, \tag{6}$$

$$n_y = \eta_n \sin\tau_n, \tag{7}$$

$$n_z = \pm\sqrt{1 - n_x^2 - n_y^2}, \tag{8}$$

where the parametric angle τ_n may adopt any value between $-\pi$ and π. Thus, each dipolar coupling D_n defines two curves on the unit sphere, corresponding to the positive and negative roots for n_z; consistent with previous usage (Skrynnikov and Kay, 2000), we denote the solution curves for dipolar couplings $(D_n/D_a) > \delta_{\text{crit}}$ as type I solutions.

Conversely, if D_n/D_a is less than the critical value δ_{crit}, we may eliminate n_y^2 from the fundamental Equation 1 to obtain the analogous elliptical equation

$$\left(\frac{n_x}{\lambda_n}\right)^2 + \left(\frac{n_z}{\zeta_n}\right)^2 = 1, \tag{9}$$

where the semiaxes of the ellipse are now given by

$$\lambda_n^2 \equiv \frac{D_n + D_a + (3/2)D_r}{3D_r}, \tag{10}$$

$$\zeta_n^2 \equiv \frac{D_n + D_a + (3/2)D_r}{3D_a + (3/2)D_r}. \tag{11}$$

Again, both λ_n^2 and ζ_n^2 are less than one provided that the condition $(D_n/D_a) < \delta_{\text{crit}}$ holds. Hence, the fundamental dipolar Equation 1 is satisfied by any vector of the form

$$n_x = \lambda_n \cos\tau_n, \tag{12}$$

$$n_z = \zeta_n \sin \tau_n, \tag{13}$$

$$n_y = \pm\sqrt{1 - n_x^2 - n_z^2}, \tag{14}$$

where the parametric angle τ_n may again adopt any value between $-\pi$ and π. Again, each dipolar coupling defines two curves on the unit sphere, corresponding to the positive and negative roots for n_y; we denote the solution curves for dipolar couplings $(D_n/D_a) < \delta_{\text{crit}}$ as type II solutions (Skrynnikov and Kay, 2000). The minor semiaxes η_n and ζ_n of the type I and type II solutions are related by the equation $\eta_n^2 + \zeta_n^2 = 1$. As an aside, the type I and type II solutions can be used to derive the 'powder-pattern' diffraction pattern (Appendix A), which is useful in estimating D_a and R from experimental data (Skrynnikov and Kay, 2000).

The solutions (6)–(8) and (12)–(14) are plotted for various dipolar couplings D_n in Figures 1 and 2. The solution curves for the critical dipolar coupling $D = D_{xx}$ form two great circles on the unit sphere that intersect at the $\pm\mathbf{x}$ axes of the principal alignment frame; these great circles correspond to the straight lines in Figure 1 and the cusped curves in Figure 2. Thus, the critical solution curves divide the unit sphere into four regions, which we denote as *gores*, consistent with cartographic terminology (Canters and Decleir, 1989). The solution curves for the non-critical dipolar couplings define two families of ellipses (the type I and type II solutions). The ellipses of each family are geometrically similar since the ratio of their semiaxes depends only on the rhombicity R; specificially, the semiaxis ratio S_{I} for type I solutions equals

$$S_{\text{I}} \equiv \frac{\eta_n}{\xi_n} = \sqrt{\frac{2 - R}{2 + R}} \tag{15}$$

while the semiaxis ratio S_{II} for type II solutions equals

$$S_{\text{II}} \equiv \frac{\zeta_n}{\lambda_n} = \sqrt{\frac{2R}{2 + R}}. \tag{16}$$

The semiaxis ratios S_{I} and S_{II} are related by the equation $S_{\text{I}}^2 + S_{\text{II}}^2 = 1$. For all values of the rhombicity $R < 2/3$, the semiaxis ratio S_{I} for the type I solutions is greater than the semiaxis ratio S_{II} for the type II solutions. Hence, the type II ellipses are slimmer and more eccentric (i.e., less circular) than the corresponding type I ellipses (Figure 2), except at the extreme value $R = 2/3$, where they are equal in width.

Exact solutions for unit vectors from dipolar couplings and additional constraint(s)

The elliptical solutions (6)–(8) and (12)–(14) provide orientational constraints on the direction of bond unit vectors which may be used in NMR structure determination, e.g., in restricting the search for conformations that satisfy the NOE and J-coupling constraints. However, these solutions can also be used to obtain exact solutions for the unit vectors, as described in the remainder of this article.

An unknown unit vector has two degrees of freedom, since it is confined to the surface of the unit sphere. Once its dipolar coupling D_p is specified, a unit vector **p** has only one degree of freedom, e.g., the angle τ_p in the elliptical solutions (6)–(8) and (12)–(14). If one other independent constraint equation is specified in addition to its dipolar coupling, a unit vector will have zero degrees of freedom, i.e., will have only discrete solutions. This article considers three types of additional constraints that provide discrete solutions for unit vectors. In the first case, an unknown unit vector **p** is determined from its dipolar coupling D_p and its angle θ_{np} with a *known* vector **n**. In the second case, three unknown vectors **n**, **p**, and **q** are determined from their dipolar couplings D_n, D_p and D_q and their inter-vector angles θ_{np}, θ_{nq}, and θ_{pq}. (The angular constraints required in these two cases can be obtained in several ways, as described below.) In the third case, an unknown vector **n** is determined from its dipolar coupling measured under two solution conditions of differing rhombicities. Analytic solutions are given below for each of these three cases. In all three cases, there may be multiple discrete solutions that fulfill the constraints; methods for distinguishing among these discrete solutions are considered below, although this remains an area for significant improvement.

Solving for a single unit vector from a single angular constraint

In this subsection, discrete solutions for an unknown unit vector **p** are determined from its dipolar coupling D_p and its angle θ_{np} to a *known* vector **n**. Such an angle constraint may be obtained by several means, as described below.

The angle θ_{np} between **p** and **n** is defined by the equation

$$\mathbf{p} \cdot \mathbf{n} = p_x n_x + p_y n_y + p_z n_z = \cos \theta_{np}. \tag{17}$$

For $(D_p/D_a) \geq \delta_{\text{crit}}$ (type I), the solutions for **p** may be expressed in terms of τ_p and the known

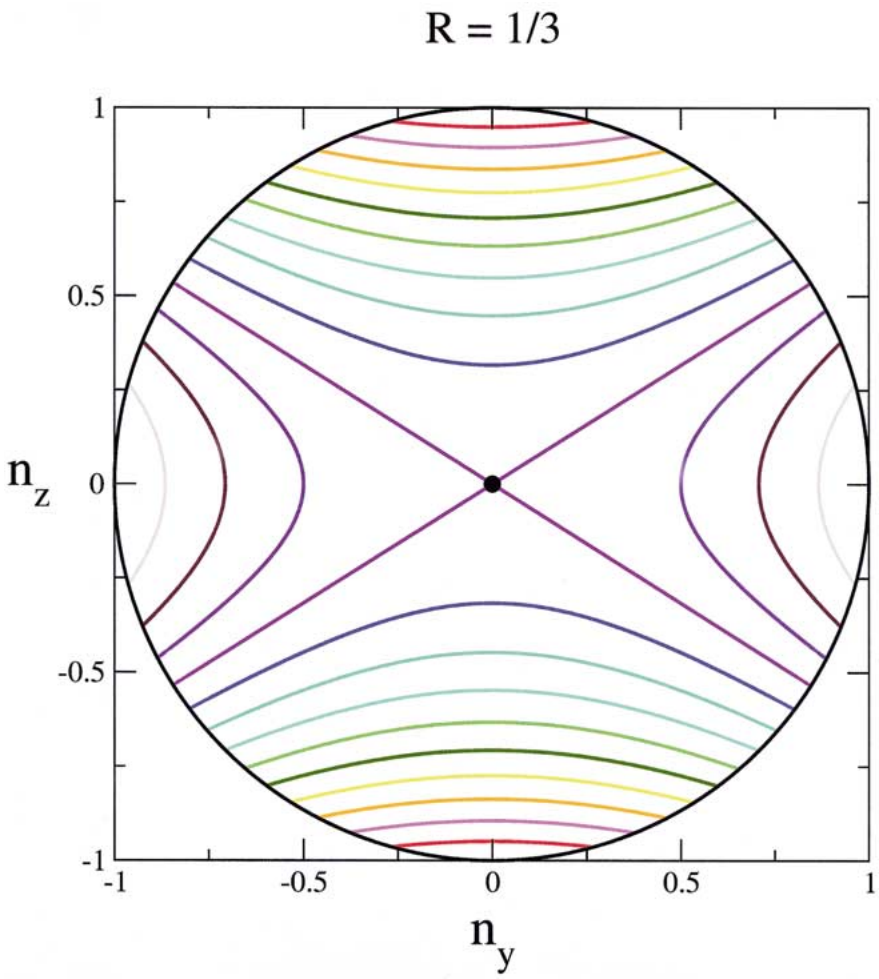

Figure 1. A family of dipolar coupling solution curves, as seen looking down the **x** axis of the principal alignment frame, for a typical value of the rhombicity ($R = 1/3$). The central black dot represents the $+\mathbf{x}$ axis, which points up out of the page, whereas the $+\mathbf{y}$ and $+\mathbf{z}$ axes point to the right and vertically, respectively, in the plane of the page. The differently colored curves represent the solution Equations 6–8 and 12–14 for values of the dipolar coupling ranging from the one extreme ($D = D_{zz} \equiv 2D_a$) to the other [$D = D_{yy} \equiv -D_a - (3/2)D_r$] in steps of 0.25 D_a. Thus, the red curve corresponds to the unit vectors with dipolar coupling $D = 1.75\ D_a$, the magenta curve to those with $D = 1.50\ D_a$, the orange curve to those with $D = 1.25\ D_a$, and so on. The solution curves for the critical dipolar coupling [$D = D_{xx} \equiv -D_a + (3/2)D_r$] form two great circles that meet at the $\pm\mathbf{x}$ axes and, hence, appear here as straight lines (in purple) radiating from the central black dot. These intersecting great circles divide the surface of the unit sphere into four sections (denoted as gores), which contain the $+\mathbf{z}$, $+\mathbf{y}$, $-\mathbf{z}$ and $-\mathbf{y}$ axes, respectively (clockwise from the top).

components of **n** (cf. Equations 6–8 and 17)

$$p_x = \xi_p \cos\tau_p, \tag{18}$$

$$p_y = \eta_p \sin\tau_p, \tag{19}$$

$$p_z = \left[\cos\theta_{np} - n_x\xi_p \cos\tau_p - n_y\eta_p \sin\tau_p\right]/n_z. \tag{20}$$

Similarly, for type II solutions, the solutions (cf. Equations 12–14 and 17) are

$$p_x = \lambda_p \cos\tau_p, \tag{21}$$

$$p_z = \zeta_p \sin\tau_p, \tag{22}$$

$$p_y = \left[\cos\theta_{np} - n_x\lambda_p \cos\tau_p - n_z\zeta_p \sin\tau_p\right]/n_y. \tag{23}$$

To avoid trivial repetition, we adopt the following notation that combines the type I and type II solutions

$$W = \cos\theta_{np}, \tag{24}$$

$$U = n_x\xi_p, \tag{25}$$

$$V = \begin{cases} n_y\eta_p & for(D_p/D_a) \geq \delta_{\text{crit}} \\ n_z\zeta_p & for(D_p/D_a) \leq \delta_{\text{crit}} \end{cases}, \tag{26}$$

$$F = \begin{cases} n_z & for(D_p/D_a) \geq \delta_{\text{crit}} \\ n_y & for(D_p/D_a) \leq \delta_{\text{crit}} \end{cases}, \tag{27}$$

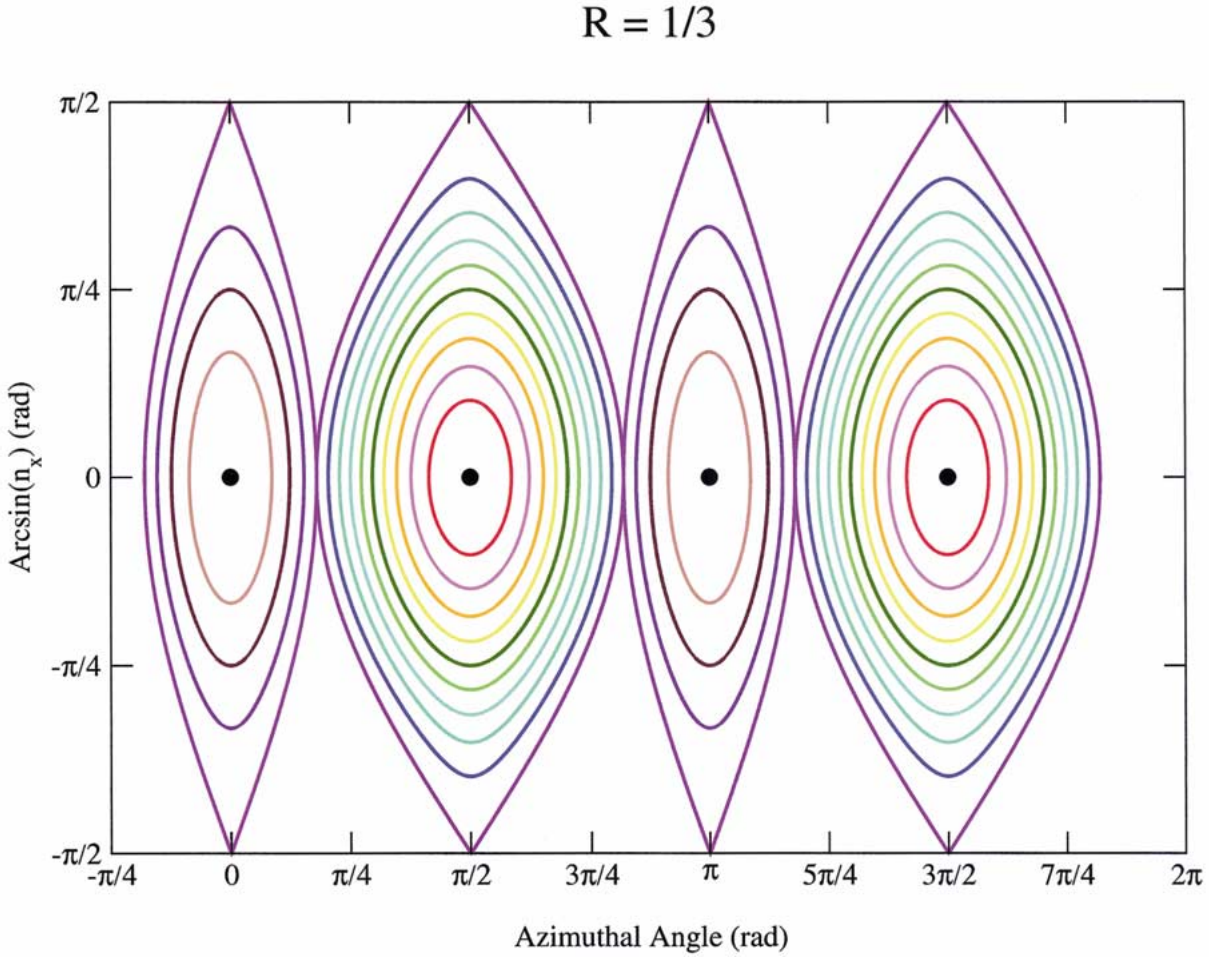

Figure 2. Flattened projection of the family of solution curves of Figure 1. The surface of the unit sphere has been interrupted along the two great circles meeting at the $\pm\mathbf{x}$ axes (the solution curves for the critical dipolar coupling, in purple) and each of the four gores has been projected onto the plane using the sinusoidal projection (Canters and Decleir, 1989). This projection (similar to peeling an orange in four sections and flattening each section) allows the whole surface of the unit-sphere to be viewed at once. The **x**-axis is vertical in this projection. From left to right, the four black central dots represent the $+\mathbf{y}$, $+\mathbf{z}$, $-\mathbf{y}$, and $-\mathbf{z}$ axes, respectively. The coloring of the solution curves is the same as that used in Figure 1, and the four cusped curves (in purple) correspond to the solution curves for the critical dipolar coupling. As is evident, the type II ellipses (those encircling the $\pm\mathbf{y}$ axes) are slimmer than the type I ellipses (those encircling the $\pm\mathbf{z}$ axes). The abscissa of this graph corresponds to the azimuthal angle (in radians) measured in the $y-z$ plane (i.e., the angle made with the $+\mathbf{y}$ axis in the projection of Figure 1) relative to the y axis, while the ordinate equals the arcsine of the x-component (in radians). Hence, the four gores have a combined width of 2π along the abscissa, while the ordinate has a maximum amplitude of $\pi/2$.

$$G = \begin{cases} \xi_p & for (D_p/D_a) \geq \delta_{\text{crit}} \\ \lambda_p & for (D_p/D_a) \leq \delta_{\text{crit}} \end{cases}, \tag{28}$$

$$H = \begin{cases} \eta_p & for (D_p/D_a) \geq \delta_{\text{crit}} \\ \zeta_p & for (D_p/D_a) \leq \delta_{\text{crit}} \end{cases}. \tag{29}$$

For the normal case when $|F| > 0$, the unit vector equation $p_x^2 + p_y^2 + p_z^2 = 1$ can be written as

$$\begin{aligned} &A_1 \cos^2 \tau_p + A_2 \sin^2 \tau_p + \\ &A_3 \cos \tau_p \sin \tau_p + A_4 \cos \tau_p + \\ &A_5 \sin \tau_p + A_6 = 0, \end{aligned} \tag{30}$$

where the six coefficients are given by

$$A_1 = U^2 + F^2 G^2, \tag{31}$$

$$A_2 = V^2 + F^2 H^2, \tag{32}$$

$$A_3 = 2UV, \tag{33}$$

$$A_4 = -2WU, \tag{34}$$

$$A_5 = -2WV, \tag{35}$$

$$A_6 = W^2 - F^2. \tag{36}$$

Using the change of variables

$$\cos\tau_p \equiv \frac{1-t_p^2}{1+t_p^2}, \tag{37}$$

$$\sin\tau_p \equiv \frac{2t_p}{1+t_p^2}, \tag{38}$$

Equation 30 can be converted into a quartic polynomial equation

$$b_4 t_p^4 + b^3 t_p^3 + b_2 t_p^2 + b_1 t_p + b_0 = 0 \tag{39}$$

with coefficients

$$b_4 = A_1 - A_4 + A_6, \tag{40}$$
$$b_3 = 2A_5 - 2A_3, \tag{41}$$
$$b_2 = 4A_2 - 2A_1 + 2A_6, \tag{42}$$
$$b_1 = 2A_3 + 2A_5, \tag{43}$$
$$b_0 = A_1 + A_4 + A_6. \tag{44}$$

Such quartic equations may be solved explicitly by either Descartes' or Ferrari's method (Korn and Korn, 1961). Each real root corresponds to a valid solution for p on the unit-sphere; hence, there are zero, two or four solutions. A simple solution can also be found for the exceptional case $F \approx 0$ (Appendix B).

The solutions of Descartes or Ferrari to the quartic equation provide explicit formulae for the possible discrete solutions of **p** given its dipolar coupling D_p and its angle θ_{np} to a known vector **n**. Additional data are required to decide between the alternative discrete solutions. For example, some solutions may be eliminated because they are inconsistent with other experimental data or involve steric overlap of atoms. As an aside, the single-vector solution presented here can also be used to derive analytical formulae for the values of θ_{np} for which there are no solutions, given two dipolar couplings D_n and D_p (Meiler et al., 2000; Skrynnikov and Kay, 2000).

The set of all solutions for **p** may also be obtained by an exhaustive one-dimensional search over the vectors that are separated by an angle θ_{np} from the given vector **n**. However, such an exhaustive-search method is much less efficient than the quartic-polynomial method described here. The CPU time required to determine the coefficients and solve Equation 39 is roughly equivalent to that required to compute and test three vectors of the exhaustive-search method. Assuming that the exhaustive search is carried out on a fine grid (say, in steps of 1 deg), the quartic-polynomial method would then be roughly one- hundred fold more efficient. This increased efficiency is all the more helpful when carrying out two-dimensional searches, e.g., in solving for triplets of vectors, as described in the next subsection.

A key question is the effect of measurement error on the orientation of the solutions. To assess this question, ten million pairs of random unit vectors **n**,**p** were generated uniformly on the unit sphere. A random measurement error (drawn from a Gaussian distribution of $\sigma = D_a/10$) was added to the ideal dipolar coupling D_p to produce a perturbed D_p^{err}. Discrete solutions $\mathbf{p}^{err}$ were then determined using the unperturbed unit vector **n** and angle θ_{np} and the perturbed dipolar coupling D_p^{err}. A histogram of the angle between the closest such solution and the corresponding ideal solution is plotted in Figure 3 (green histogram). The histogram is dominated by two peaks, one close to zero degrees deviation and a second, smaller peak at 180° that corresponds to cases in which no solution for $\mathbf{p}^{err}$ was possible. Between these two peaks, there is a low but relatively uniform distribution of deviation angles. This qualitative behavior also holds for larger measurement error ($\sigma = D_a/5$, the red curve in Figure 3) although, not surprisingly, the peak at 0° becomes lower and broader while the peak at 180° becomes commensurately higher. Thus, if any solution is possible, it is likely to lie close to the true vector.

Solving for a triplet of unit vectors from three angular constraints

A triplet of unknown unit vectors **n**, **p**, and **q** can be determined from the six constraints provided by their dipolar couplings D_n, D_p, and D_q and the angles θ_{np}, θ_{nq} and θ_{pq} between them. Methods for obtaining the three required angle constraints are discussed in the next subsection.

If the three unknown unit vectors are denoted as **n**, **p**, and **q**, the relevant six constraint equations can be written as

$$D_n = \mathbf{n}\cdot\mathbf{D}\cdot\mathbf{n}, \tag{45}$$
$$D_p = \mathbf{p}\cdot\mathbf{D}\cdot\mathbf{p}, \tag{46}$$
$$D_q = \mathbf{q}\cdot\mathbf{D}\cdot\mathbf{q}, \tag{47}$$
$$\cos\theta_{np} = \mathbf{n}\cdot\mathbf{p}, \tag{48}$$
$$\cos\theta_{nq} = \mathbf{n}\cdot\mathbf{q}, \tag{49}$$
$$\cos\theta_{pq} = \mathbf{p}\cdot\mathbf{q}, \tag{50}$$

where the unit vector condition is assumed

$$\mathbf{n}\cdot\mathbf{n} = \mathbf{p}\cdot\mathbf{p} = \mathbf{q}\cdot\mathbf{q} = 1. \tag{51}$$

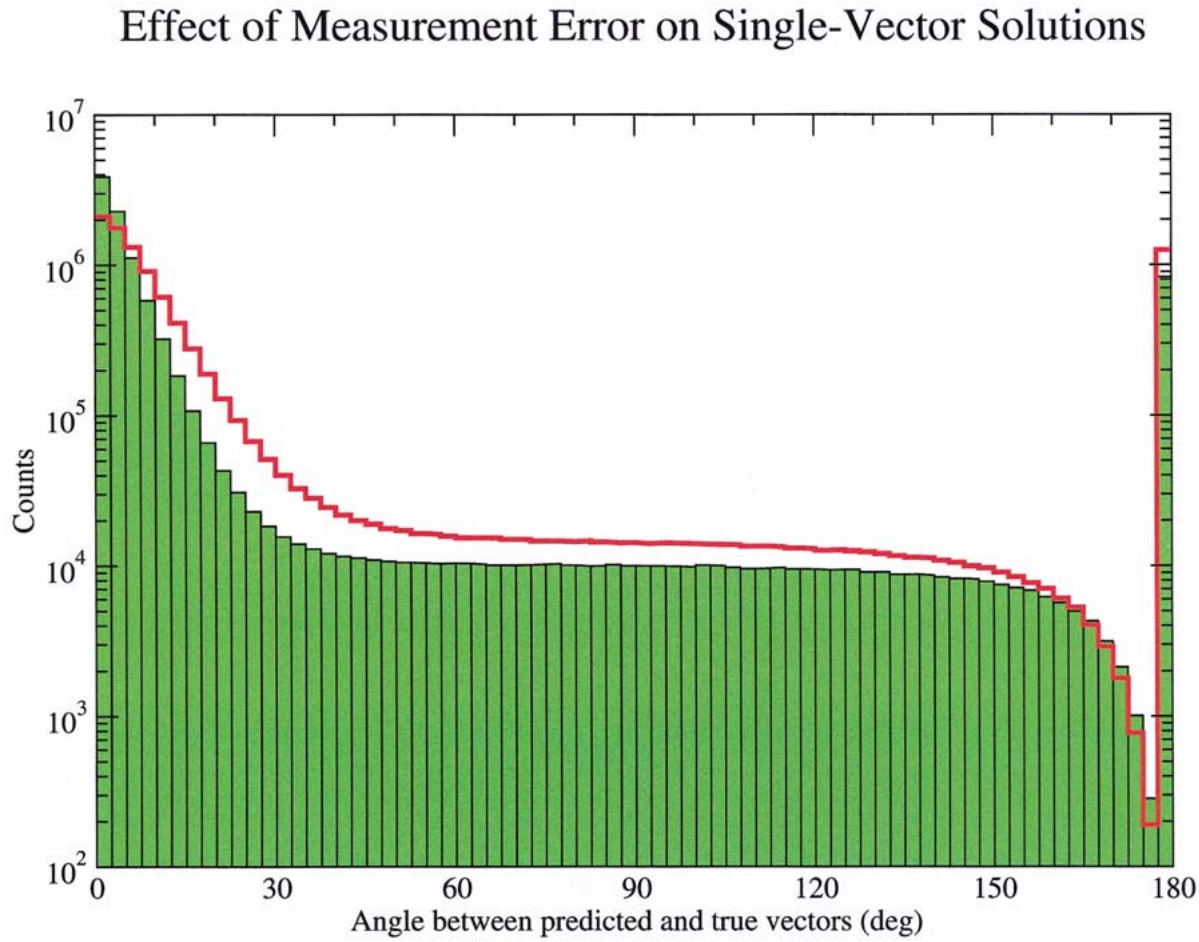

Figure 3. Error histogram for single-vector solutions. Ten million pairs of random unit vectors **n** and **p** were generated uniformly on the unit sphere, and the corresponding dipolar couplings D_n and D_p determined with a rhombicity of 1/3. The dipolar coupling D_p was then perturbed by a Gaussian random variable of zero mean and standard deviation $\sigma = D_a/10$, where D_a is the axial component of the dipolar coupling tensor. Solutions were then determined from the known vector **n**, the inter-vector angle θ_{np} and the perturbed dipolar coupling D_p. The deviation angle on the abscissa is defined as the minimum angle between the original vector **p** and the resulting solutions. The histogram of the deviation angle is dominated by two peaks, near 0° and 180°; the latter peak corresponds to cases in which no solution was feasible for **p**. This procedure was repeated for larger perturbations ($\sigma = D_a/5$); the resulting histogram (red curve) is qualitatively similar, although the peak at 0° is lower and broader and the peak at 180° is higher.

As shown above, the solutions of the dipolar coupling equations can be expressed in terms of three variables τ_n, τ_p, and τ_q for the vectors **n**, **p**, and **q**, respectively [cf. Equations 6–8 and 12–14]. Therefore, the present problem is equivalent to finding appropriate values of the three τ variables that satisfy the three angular constraints.

A simple method for solving these equations is as follows. An exhaustive search of one variable (e.g., τ_n) is carried out. For each value of τ_n, the corresponding **n** unit vector is calculated (the positive root is always adopted, for reasons described in the next paragraph). By treating **n** as a *known* vector, the solution of the previous subsection may be used to obtain all possible unit vectors **p** and **q** that satisfy the angular constraints

$$\cos\theta_{np} = \mathbf{n}\cdot\mathbf{p}, \tag{52}$$

$$\cos\theta_{nq} = \mathbf{n}\cdot\mathbf{q}. \tag{53}$$

The angles between these possible solutions for **p** and **q** are then calculated to determine whether the final angular constraint

$$\cos\theta_{pq} = \mathbf{p}\cdot\mathbf{q} \tag{54}$$

can be satisfied to an acceptable tolerance for that value of τ_n. This computation is very fast, so that even a very fine exhaustive search of τ_n (say, in steps of 0.1 deg) requires negligible computer time.

Only a quarter of the whole range of τ_n (e.g., from 0 to $\pi/2$) need be searched, because of the eightfold symmetry among the solutions. For every solution (n_x, n_y, n_z), (p_x, p_y, p_z), and (q_x, q_y, q_z)of the constraint equations, the eight vectors ($e_x n_x$, $e_y n_y$, $e_z n_z$), ($e_x p_x$, $e_y p_y$, $e_z p_z$), and ($e_x q_x$, $e_y q_y$, $e_z q_z$) (where e_x, e_y, and e_z may equal plus or minus one) also satisfy the constraint equations. (The eightfold multiplicity arises from the three binary choices of e_x, e_y and e_z.) By judicious choice of e_x, e_y, and e_z, it is always possible to find a solution for **n** in the first octant of the unit sphere, in which all the components of **n** are positive. Thus, only the values of τ_n from 0 to $\pi/2$ need be searched and the positive root may always be adopted in determining **n** (as described in the previous paragraph). If the chirality of the triplet of vectors is also specified, four of the eight solutions will have an incorrect chirality and may be eliminated from further consideration.

The exhaustive search of τ_n between 0 and $\pi/2$ may uncover several feasible solutions for **n** in the first octant of the unit-sphere, and every such solution corresponds to *eight* solutions on the whole unit- sphere (four if the chirality is specified). As in the previous single-vector method, experimental or theoretical data, such as other dipolar coupling data, other angle restraints or steric overlaps of atoms, are required to decide between the alternative discrete solutions.

The sensitivity of such vector-triplet solutions to random measurement errors was investigated. One million random vector triplets **n**, **p**, and **q** were generated uniformly on the unit sphere, and the corresponding dipolar couplings (D_n, D_p, and D_q) and inter-vector angles (θ_{np}, θ_{nq}, and θ_{pq}) were computed. All three dipolar couplings were then perturbed by a random Gaussian error of standard deviation $\sigma = D_a/10$, and the corresponding discrete solutions for the vector triplet $\mathbf{n}^{err}$, $\mathbf{p}^{err}$, and $\mathbf{q}^{err}$ determined. The error was taken as the maximum of the three pairwise deviation angles **n**-$\mathbf{n}^{err}$, **p**-$\mathbf{p}^{err}$ and **q**-$\mathbf{q}^{err}$. The resulting histogram of this maximum deviation angle (the green shaded curve in Figure 4) shows that the errors are again dominated by two peaks, a low-deviation peak close to zero degrees and a second peak at 180°, corresponding to those cases in which no solution whatsoever was possible. Thus, if any solution is possible, it is likely to lie close to the true vector triplet. Qualitatively similar results are obtained for the special cases of tetrahedral and coplanar arrangements of **n**, **p**, and **q**, corresponding to the covalent bonds at the C^α atom and in the peptide plane of proteins (the red and black curves in Figure 4, respectively).

This method can be extended to rigid groups in which more than three dipolar couplings are known. For example, let there be four unit vectors **n**, **p**, **q**, and **r** in a rigid group, for which the dipolar couplings (D_n, D_p, D_q, and D_r) and inter-vector angles (θ_{np}, θ_{nq}, θ_{nr}, θ_{pq}, θ_{pr} and θ_{qr}) are known. An exhaustive search of one variable (e.g., τ_n) can be carried out, corresponding to an exhaustive search over all orientations of the **n** vector consistent with its dipolar coupling D_n. For each such orientation of **n**, discrete solutions for the remaining three vectors **p**, **q**, and **r** can be obtained from their dipolar couplings and the angle constraints with **n** (θ_{np}, θ_{nq} and θ_{nr}) by the single-vector method of the previous subsection. These discrete solutions can then be checked to see whether the three remaining bond angle constraints (θ_{pq}, θ_{pr}, and θ_{qr}) are satisfied to an acceptable tolerance. As in the three-vector case, some of the resulting solutions may be eliminated if known constraints on the chirality of the vectors are violated.

Obtaining angular constraints

The solution methods of the two previous subsections require angular constraints, either the angle between a known vector and an unknown vector or the angles between three unknown vectors. Such angular constraints may be obtained in three general ways. First, the angle between some vectors is well defined by covalent bonding. In particular, there are two rigid groups per residue in the backbone of proteins, the peptide plane and the C^α frame defined by the orientations of the four covalent bonds made by the C^α atom. Since the inter-vector angles within such rigid groups are well-defined by the covalent bond geometry, the orientations of these rigid groups can be determined provided that at least three dipolar couplings are measured (Mueller et al., 2000; Hus et al., 2001). Second, the angle between some vectors may be determined experimentally, e.g., by the method of cross-correlated NMR relaxations (Reif et al., 1997). Third, the angle between some vectors may be defined (albeit with much less confidence than in the first two cases) by sampling the low-energy conformations of the molecule. For example, rigid fragments consistent with the amino-acid sequence may be used to define angular constraints between the NH bonds of successive residues in a protein (Delaglio et al., 2000; Bowers et al., 2000). Alternatively, a more complete search of

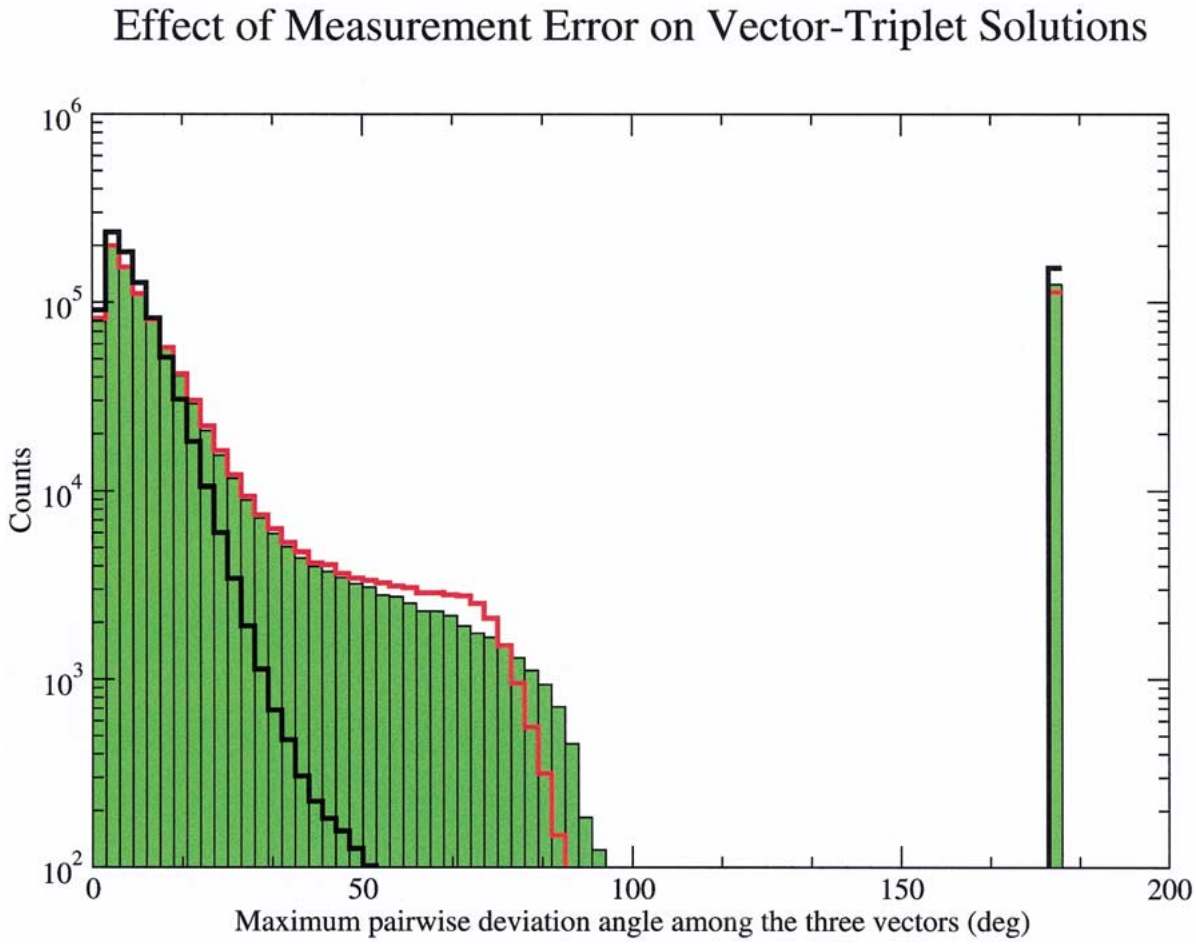

Figure 4. Error histogram for three-vector solutions. One million triplets of random unit vectors **n**, **p** and **q** were generated uniformly on the unit sphere, and the corresponding dipolar couplings D_n, D_p and D_q were determined with a rhombicity of 1/3. All three dipolar couplings were then perturbed by a Gaussian random variable of zero mean and standard deviation $\sigma = D_a/10$, where D_a is the axial component of the dipolar coupling tensor. Solutions for the three vectors $\mathbf{n}^{err}$, $\mathbf{p}^{err}$ and $\mathbf{q}^{err}$ were then determined from the known inter-vector angles θ_{np}, θ_{nq} and θ_{pq} and the perturbed dipolar couplings. The error was taken as the maximum of the three deviation angles between corresponding true and predicted vectors **n**-$\mathbf{n}^{err}$, **p**-$\mathbf{p}^{err}$ and **q**-$\mathbf{q}^{err}$. The histogram of the maximum deviation angle is dominated by two peaks, near 0° and 180°; the latter peak corresponds to cases in which no solution was feasible. The absence of a maximum deviation angle greater than 90° results from the eightfold degeneracy of the solutions. The black and red curves correspond to similar error plots for simulated data for peptide planes and for C^{α} frames, respectively.

the conformational space may indicate angles that are consistent in low-energy conformations, e.g., in segments with a high helical propensity. Such angles may then be constrained to their mean value for the purpose of solving for the unit vectors. Several methods for sampling low-energy protein conformations to aid in NMR structure determination have already been developed (Skolnick et al., 1997; Kolinski and Skolnick, 1998; Debe et al., 1999; Standley et al., 1999; Bowers et al., 2000).

Solving for a single unit vector from two dipolar coupling measurements

The direction of a unit vector may also be determined *without* angular constraints, provided that dipolar coupling measurements are carried out in two media in which the molecule has different molecular orientation tensors (Ramirez and Bax, 1998; Al-Hashimi et al., 2000). Since each independent dipolar coupling measurement eliminates one degree of freedom, two such measurements constrain the unit vector to only discrete solutions (zero degrees of freedom). How-

ever, this method requires that the rhombicities of both media and the relative orientation of the principal alignment frames of the two media be determined.

In the simplest case, the principal alignment frames of the two media are identical and the rhombicities alone are different. In this case, a unit vector **n** may be solved for explicitly, as follows. Let the semiaxes of the ellipses in the two media be denoted as (ξ_{n1}, η_{n1}) and (ξ_{n2}, η_{n2}), respectively. (For simplicity, we treat only the type I/type I case; the other cases proceed exactly analogously.) The solution of the first elliptical equation may be written

$$n_x = \xi_{n1} \cos \tau_{n1}, \tag{55}$$

$$n_y = \eta_{n1} \sin \tau_{n1}, \tag{56}$$

$$n_z = \pm\sqrt{1 - n_x^2 - n_y^2}. \tag{57}$$

This solution must also satisfy the second elliptical equation

$$\left(\frac{\xi_{n1}}{\xi_{n2}}\right)^2 \cos^2 \tau_{n1} + \left(\frac{\eta_{n1}}{\eta_{n2}}\right)^2 \sin^2 \tau_{n1} = 1. \tag{58}$$

Using the change of variables

$$\cos \tau_{n1} \equiv \frac{1 - t_{n1}^2}{1 + t_{n1}^2}, \tag{59}$$

$$\sin \tau_{n1} \equiv \frac{2t_{n1}}{1 + t_{n1}^2}, \tag{60}$$

the second elliptical equation can be converted into a quartic equation

$$t_{n1}^4 - 2Bt_{n1}^2 + 1 = 0, \tag{61}$$

where the coefficient

$$B = \frac{\xi_{n1}^2 + \xi_{n2}^2 - 2(\eta_{n1}^2/\eta_{n2}^2)\xi_{n2}^2}{\xi_{n1}^2 - \xi_{n2}^2}. \tag{62}$$

This quartic equation is actually a quadratic equation in t_{n1}^2 and, thus, may be solved by the quadratic formula, yielding zero or two real solutions for t_{n1}^2. Each such solution (if positive) corresponds to two solutions for t_{n1} (since the sign of t_{n1} may be chosen freely) and four solutions for the unit vector **n** (since the sign of n_z may be chosen freely). As in the previous two methods, these discrete solutions must be distinguished by other experimental data or by other considerations such as atomic clashes.

In the more general case, when the two principal alignment frames do not agree, it appears necessary to solve the equations by an exhaustive search over a single variable. In this approach, the unit vector in the first principal alignment frame is determined from Equations 55–57 for a large number of closely spaced values of τ_{n1}. The corresponding components $\mathbf{n}'$ in the second principal alignment frame may then be determined by multiplying the unit vector **n** by the appropriate rotation matrix **R** (which must be determined, e.g., by minimizing the deviations in the resulting covalent bond angles from known peptide-group or C^α bond geometries)

$$\mathbf{n}' = \mathbf{R} \cdot \mathbf{n}. \tag{63}$$

The resulting components of $\mathbf{n}'$ may then be substituted to determine whether the second elliptical equation

$$\left(\frac{n_x'}{\xi_{n2}}\right)^2 + \left(\frac{n_y'}{\eta_{n2}}\right)^2 = 1 \tag{64}$$

is satisfied to an acceptable tolerance. By exhaustively searching τ_{n1} from $-\pi$ to π, all solutions for **n** that are consistent with both elliptical equations may be found rapidly.

This method requires that the relative orientation of the two principal alignment frames (e.g., their relative Euler angles or the **R** matrix) be known. It may be possible to determine this relative orientation as part of the structure determination. For example, the unit vectors can be determined for a densely sampled set of Euler angles (**R** matrices). The quality of each relative orientation can then be assessed by the consistency of the unit vector solutions with each other, with well-known covalent bond angles, and with other experimental data, analogous to the manner in which the discrete solutions are distinguished. This method also requires that the rhombicities of the two media be known, which may be determined by the usual methods (Clore et al., 1998a, b; Losonczi et al., 1999; Skrynnikov and Kay, 2000).

Application to protein structure determination

The analytic solutions of the previous section may be applied to determine protein structure from residual dipolar coupling data. The essence of the strategy is to decompose the protein into rigid groups such as peptide planes (Mueller et al., 2000) and C^α frames (Hus et al., 2001) (i.e., the tetrahedral arrangement of covalent bonds at each C^α atom). Discrete solutions can be found for the orientation of these groups by the methods described above, provided that at least three dipolar couplings have been measured for each

rigid group, e.g., the C^{α}-C', C'-N and N-H dipolar couplings for the peptide groups or the C^{α}-C', C^{α}-C^{β}, and C^{α}-H^{α} dipolar couplings for the C^{α} frames.

A key difficulty of this strategy lies in recombining the independent solutions for the peptide-plane and C^{α}-frame problems into a valid protein structure, i.e., in determining which of the discrete solutions for each rigid body is most representative of the protein structure. A simple technique is to exploit the fact that the peptide-plane and C^{α}-frame problems often overlap, providing multiple independent solutions for the same vector. For example, the C^{α}-C' vector of residue i belongs to three such problems: the C^{α}-frame problem of residue i, the peptide-plane problem spanning residues i and $i + 1$ and the C^{α}-frame problem of residue $i + 1$ (since it is nearly parallel to the N-C^{α} vector of residue $i + 1$). In the absence of measurement error, these multiple solutions would agree exactly, indicating which discrete solutions should be adopted in reconstructing the protein. Once the orientation of a rigid group has been determined reliably, the single-vector method can be used to determine the orientation of unknown bond vectors that are adjacent to the known orientations of bonds in the rigid body.

We tested our methods on the well-known protein ubiquitin, for which several X-ray and NMR solutions already exist (Vijay-Kumar et al., 1987; Cornilescu et al., 1999). The initial data included up to five dipolar couplings per residue, namely, the N-H, C^{α}-C', C'-N, C'-H and C^{α}-H^{α} dipolar couplings, measured in a single anisotropic medium (Ottiger and Bax, 1998). (For this preliminary study, the two-media solutions described above were not employed.) The bond lengths and bond angles were taken from the literature (Engh and Huber, 1991). Unfortunately, the simple technique proposed in the previous paragraph (involving overlapping vector solutions) does not determine the absolute orientations of the backbone bond vectors unambiguously, since there were gaps in the data and the experimental data had significant measurement errors, possibly arising (in part) from the dynamic internal motions of the protein (Meiler et al., 2001).

Therefore, we adopted the following protocol for estimating the ϕ-ψ backbone dihedral angles at each residue. Discrete solutions were determined for the peptide groups in which at least three dipolar couplings had been measured. Since the orientations of two successive peptide planes determine the $\phi-\psi$ angles of the central residue, the m discrete solutions of the ith peptide plane can be combined with the n discrete solutions of the $(i + 1)$th peptide plane to produce nm estimates of (ϕ, ψ) of the central residue (Quine et al., 1997). These estimates are then filtered to eliminate those that lie in forbidden regions of the Ramachandran map. A second filtering criterion is the angle between the N-C^{α} vector of the ith peptide group and the C^{α}-C' vector of the $(i + 1)$th peptide group, which is constrained to lie within a few degrees of 70° by the N-C^{α}-C' bond-angle constraint at the C^{α} atom of residue i (Engh and Huber, 1991). For each such pair of peptide-plane solutions in which the deviation in this angle was less than 10°, the predicted orientation of the C^{α}-H^{α} bond vector was determined, from which the corresponding C^{α}-H^{α} dipolar coupling was predicted. The optimal pair of successive peptide-plane solutions was then determined by the minimum deviation between the predicted and experimental C^{α}-H^{α} dipolar couplings. If no solution-pair was found within a 10° angular deviation, the procedure was repeated with an angle deviation limit of 20°, and so on, up to 40° deviation. The entire solution of ubiquitin required 34 seconds on a 1 GHz Pentium III processor.

The dihedral angles resulting from this procedure are reasonably close to the experimental values (Figure 5), although only 70% (106/152) of the dihedral angles could be determined, due to the gaps in the experimental data. The majority of backbone dihedral angles (88/106) deviate from their native values by a relatively small amount ($\sigma = 11°$). However, large deviations are observed for some residues; for example, residues 31, 39 and 58 are α-helical in the native structure but were predicted to adopt β-strand structure. In general, these large deviations correlated with large violations of the N-C^{α}-C' angle constraint and/or poor predictions of the C^{α}-H^{α} dipolar coupling. However, large deviations from experiment were also observed in two dihedral angles despite good agreement with the bond angle and measured dipolar coupling.

These initial ϕ-ψ values were used to generate a centroid-based reduced representation of ubiquitin, which was then refined using the ROSETTA package (Simons et al., 1997, 1999), as modified to include dipolar coupling data (Rohl and Baker, 2001). A library of 150 fragments for each three- and nine-residue segment of the protein sequence was generated from a non-redundant database of protein structures; these fragments were selected based on their sequence similarity to ubiquitin and the agreement between the fragment dihedral angles and those predicted from the dipolar couplings. To cover the segments for which no dihedral angles could be predicted, 25 additional fragments were selected for every three- and nine-

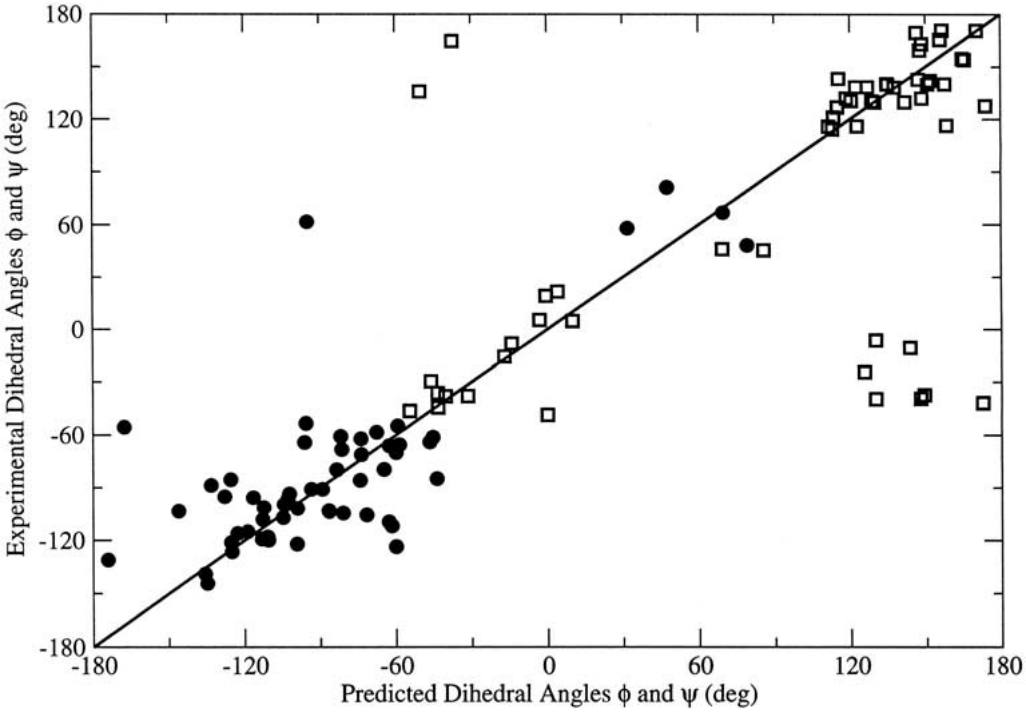

Figure 5. Plot comparing the predicted and experimental dihedral angles (Cornilescu et al., 1999). The solid circles indicate ϕ angles, and the open squares indicate ψ angles. The predicted dihedral angles generally lie within 10° of the experimental values, although there are some significant outliers.

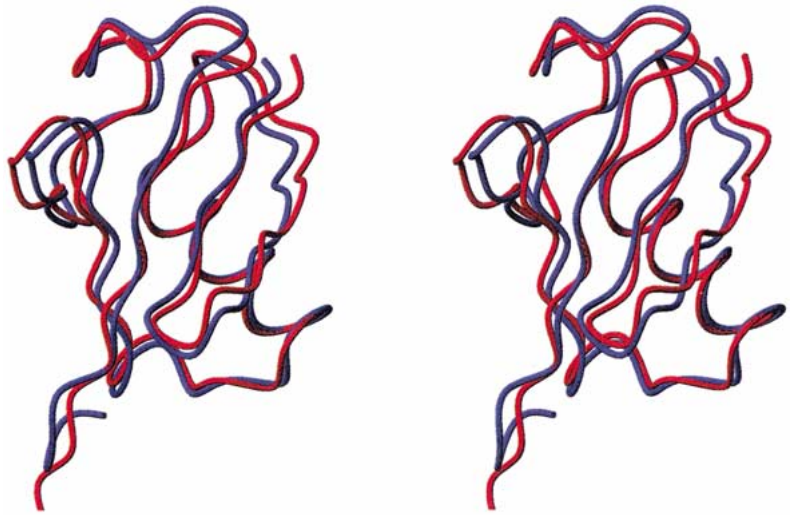

Figure 6. Comparison of the experimental structure of ubiquitin (1d3z, in blue) with the lowest-energy predicted structure (in red), which agree with a C^{α} rmsd of 1.1 Å. Data were not available for the last five (disordered) residues, which were not counted in the rmsd score.

residue segment, based on their sequence similarity to ubiquitin and the agreement of the fragment dihedral angles with the predicted secondary structure. A Monte Carlo-minimization procedure (Li and Scheraga, 1987) was used to generate low-energy protein structures from these fragments. One thousand structures were generated initially and the ten structures showing the best agreement with the dipolar coupling data were refined further. The final ten lowest-energy structures agree with the native structure of ubiqui-

tin (Cornilescu et al., 1999) very well (0.9–1.5 Å C^{α} rmsd); the lowest-energy structure (1.1 Å C^{α} rmsd) is depicted in Figure 6. The dipolar data are necessary for such good structural agreement; the corresponding ten lowest-energy structures produced by ROSETTA in the absence of dipolar coupling data range from 3.7–10.6 Å C^{α} rmsd. The initial generation of one thousand structures and the final refinement of the ten structures required roughly one hour and five hours, respectively, of CPU time on a 1 GHz Pentium III processor.

A drawback of the present method for finding ϕ and ψ angles is that it does not exploit the long-range quality of the dipolar coupling data. However, the subsequent ROSETTA refinement does consider all dipolar coupling data simultaneously and, hence, does exploit the long- range quality of the data. Moreover, the short-range solutions provide excellent 'seeds' of local structure for the ROSETTA refinement, which relies on the insertion of fragments of correct local structure (Simons et al., 1997, 1999).

The treatment of uncertainties

The most straightforward approach to the error analysis of protein structures derived from dipolar coupling data would seem to be a Monte Carlo method (Press et al., 1992). In this approach, an ensemble of mock data-sets is generated using the uncertainty in the dipolar coupling measurements and angle constraints. Specifically, each mock data-set is generated from the original dipolar coupling and angular constraint data by adding a Gaussian random variable scaled by the uncertainty in that quantity (Press et al., 1992). By assumption, each mock data-set determines a unique molecular structure; hence, the ensemble of mock data-sets corresponds to an ensemble of molecular structures. The ensemble of molecular structures may then be used to assess the statistical uncertainties in the molecular structure resulting from the uncertainties in the dipolar coupling and angular constraint data.

Acknowledgements

We thank Dr J. Meiler for helpful comments on the manuscript. This work was supported by NIH grant GM-14312. WJW was an NIH post- doctoral fellow (1998-2001) (GM-19399). CAR is a fellow of Interdisciplinary Training in Genomic Sciences (T32 HG00035-06). Software that solves the various dipolar coupling problems of this article is provided in the Supplemental Material to this article, comprising roughly seven thousand lines of ANSI-compliant portable C.

Appendix A. The powder-pattern probability distribution

The 'powder pattern' probability distribution corresponds to the distribution of dipolar couplings for a statistical ensemble of unit vectors that uniformly cover the unit sphere. This 'powder pattern' distribution is useful in estimating the axial and rhombic components D_a and D_r (Skrynnikov and Kay, 2000) and may be derived from Equations 6–8 and 12–14, as follows. An infinitesimal area on the unit sphere can be written as

$$\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi = J(\theta, \phi;\, D, \tau)\,\mathrm{d}D\,\mathrm{d}\tau, \tag{A1}$$

where $J(\theta, \phi;\, D, \tau)$ is the Jacobian for the change of variables from the polar angles θ and ϕ to the dipolar coupling D and elliptical angle τ. For type I solutions (cf. Equations 6–8), the four variables are related by the equations

$$\tan\phi \equiv \frac{n_y}{n_x} = \left(\frac{\eta}{\xi}\right)\tan\tau, \tag{A2}$$

$$\cos\theta \equiv n_z = \sqrt{1 - \xi^2\cos^2\tau - \eta^2\sin^2\tau}. \tag{A3}$$

Taking the partial derivatives of equations (A2) and (A3), the corresponding Jacobian equals

$$J(\theta, \phi;\, D, \tau) = \left(\frac{1}{3D_a}\right)\sqrt{\frac{1}{4 - R^2}} \times \frac{1}{\sqrt{1 - \xi^2\cos^2\tau - \eta^2\sin^2\tau}}. \tag{A4}$$

The normalized probability density of unit vectors distributed uniformly on the unit sphere equals $1/(4\pi)$; the corresponding probability density for D and τ for type I solutions is given by

$$\begin{aligned} f_{\mathrm{I}}(D, \tau) &= \left(\frac{1}{4\pi}\right)\frac{\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi}{\mathrm{d}D\,\mathrm{d}\tau} \\ &= \left(\frac{1}{4\pi}\right) J(\theta, \phi;\, D, \tau) \\ &= \left(\frac{1}{12\pi D_a}\right)\sqrt{\frac{1}{4 - R^2}} \\ &\quad \times \frac{1}{\sqrt{1 - \xi^2\cos^2\tau - \eta^2\sin^2\tau}}. \end{aligned} \tag{A5}$$

The 'powder-pattern' probability distribution may be obtained by integrating over τ

$$\begin{aligned} P_I(D) &\equiv 8\int_0^{\pi/2} f_I(D,\tau)\,d\tau \\ &= \left(\frac{2}{3\pi D_a}\right)\sqrt{\frac{1}{4-R^2}} \\ &\times \int_0^{\pi/2} \frac{d\tau}{\sqrt{1-\xi^2\cos^2\tau-\eta^2\sin^2\tau}} \\ &= \left(\frac{2}{3\pi D_a\zeta}\right)\sqrt{\frac{1}{4-R^2}}K(\xi,\lambda), \end{aligned} \quad (A6)$$

where $K(\xi/\lambda)$ is the complete elliptic integal of the first kind (Beyer, 1991). The factor of eight (instead of four) in front of the integral results from the fact that a specified D and τ corresponds to *two* positions on the unit sphere, corresponding to the positive and negative roots for n_z (cf. Equation 8).

For type II solutions (cf. Equations 12–14), the four variables are related by the equations

$$\tan\phi' \equiv \frac{n_z}{n_x} = \left(\frac{\zeta}{\lambda}\right)\tan\tau, \quad (A7)$$

$$\cos\theta' \equiv n_y = \sqrt{1-\lambda^2\cos^2\tau-\zeta^2\sin^2\tau}. \quad (A8)$$

The polar angles ϕ' and θ' defined by these equations differ from the normal polar angles in that they are measured relative to the **y** axis; however, the infinitesimal area is still given by $\sin\theta'\,d\theta'\,d\phi'$. Proceeding as in the type I solutions, the Jacobian equals

$$\begin{aligned} J(\theta',\phi';D,\tau) &= \left(\frac{1}{6D_a}\right)\sqrt{\frac{2}{2R+R^2}} \\ &\times \frac{1}{\sqrt{1-\lambda^2\cos^2\tau-\zeta^2\sin^2\tau}}. \end{aligned} \quad (A9)$$

The corresponding type II probability distribution for D and τ is given by

$$\begin{aligned} f_{II}(D,\tau) &= \left(\frac{1}{24\pi D_a}\right)\sqrt{\frac{2}{2R+R^2}} \\ &\times \frac{1}{\sqrt{1-\lambda^2\cos^2\tau-\zeta^2\sin^2\tau}} \end{aligned} \quad (A10)$$

The 'powder-pattern' probability distribution is again obtained by integrating over τ

$$\begin{aligned} P_{II}(D) &\equiv 8\int_0^{\pi/2} f_{II}(D,\tau)\,d\tau \\ &= \left(\frac{1}{3\pi D_a}\right)\sqrt{\frac{2}{2R+R^2}} \\ &\times \int_0^{\pi/2} \frac{d\tau}{\sqrt{1-\lambda^2\cos^2\tau-\zeta^2\sin^2\tau}} \\ &= \left(\frac{1}{3\pi D_a\eta}\right)\sqrt{\frac{2}{2R+R^2}}K(\lambda,\xi). \end{aligned} \quad (A11)$$

The formulae (A6) and (A11) for the powder-pattern probability distributions agree with earlier solutions (Skrynnikov and Kay, 2000), except that the present distributions are normalized to unity.

Appendix B. The single-vector solution for $F \approx 0$

In this Appendix, we again consider the problem of determining the solutions for an unknown vector **p** from its dipolar coupling D_p and its angle θ_{np} to a known vector **n**

$$\mathbf{p}\cdot\mathbf{n} = p_x n_x + p_y n_y + p_z n_z = \cos\theta_{np} \quad (B1)$$

Under normal conditions, for $(D_p/D_a) \geq \delta_{crit}$ (type I), the solutions for **p** may be expressed in terms of τ_p and the known components of **n** (cf. Equations 6–8 and 17)

$$p_x = \xi_p\cos\tau_p, \quad (B2)$$

$$p_y = \eta_p\sin\tau_p, \quad (B3)$$

$$\begin{aligned} p_z &= [cos\theta_{np} - n_x\xi_p\cos\tau_p \\ &\quad - n_y\eta_p\sin\tau_p]/n_z. \end{aligned} \quad (B4)$$

Similarly, for type II dipolar couplings, the solutions (cf. Equations 12–14 and 17) are normally

$$p_x = \lambda_p\cos\tau_p, \quad (B5)$$

$$p_z = \zeta_p\sin\tau_p, \quad (B6)$$

$$\begin{aligned} p_y &= [cos\theta_{np} - n_x\lambda_p\cos\tau_p \\ &\quad - n_z\zeta_p\sin\tau_p]/n_y. \end{aligned} \quad (B7)$$

However, it can occur that either $n_z \approx 0$ for a type I solution or $n_y \approx 0$ for a type II solution, leading to division by zero in the p_z and p_y components, respectively.

It is straightforward to overcome this difficulty, however. We consider here only the type I case, since the type II case is exactly analogous. In this case, the angle Equation 17 can be written as

$$p_x n_x + p_y n_y \approx \cos\theta_{np} \quad (B8)$$

corresponding to a linear equation in $\sin\tau_p$ and $\cos\tau_p$

$$n_x\xi_p\cos\tau_p + n_y\eta_p\sin\tau_p = \cos\theta_{np}. \quad (B9)$$

Using the change of variables (37)–(38), this equation may be converted into a quadratic equation

$$C_2 t_p^2 + C_1 t + C_0 = 0, \tag{B10}$$

where

$$C2 \equiv \cos\theta_{np} + n_x \xi_p, \tag{B11}$$

$$C1 \equiv -2 n_y \eta_p, \tag{B12}$$

$$C0 \equiv \cos\theta_{np} - n_x \xi_p. \tag{B13}$$

The real roots of this quadratic equation for t_p determine the τ_p corresponding to valid solutions for **p**. The p_x and p_y components can be determined directly from these τ_p values, while two solutions for the p_z component can be determined from the unit-vector condition $p_z = \pm\sqrt{1 - p_x^2 - p_y^2}$.

References

Al-Hashimi, H.M., Valafar, H., Terrell, M., Zartler, E.R., Eidsness, M.K. and Prestegard, J.H. (2000) *J. Magn. Reson.*, **143**, 402–406.

Beyer, W.H., Ed. (1991) *CRC Standard Mathematical Tables and Formulae*, 29th edn., CRC Press, Boston, pp. 360–361.

Bowers, P.M., Strauss, C.E.M. and Baker, D. (2000) *J. Biomol. NMR*, **18**, 311–318.

Canters, F. and Decleir, H. (1989) *The World in Perspective: A Directory of World Map Projections*. John Wiley & Sons, New York, pp. 80–81.

Chou, J.J., Li, S. and Bax, A. (2000) *J. Biomol. NMR*, **18**, 217–227.

Clore, G.M. and Garrett, D.S. (1999) *J. Am. Chem. Soc.*, **121**, 9008–9012.

Clore, G.M., Gronenborn, A.M. and Bax, A. (1998a) *J. Magn. Reson.*, **133**, 216–221.

Clore, G.M., Gronenborn, A.M. and Tjandra, N. (1998b) *J. Magn. Reson.*, **131**, 159–162.

Clore, G.M., Starich, M.R., Bewley, C.A., Cai, M. and Kuszewski, J. (1999) *J. Am. Chem. Soc.*, **121**, 6513–6514.

Cornilescu, G., Marquardt, J.L., Ottiger, M. and Bax, A. (1999) *J. Am. Chem. Soc.*, **120**, 6836–6837.

Debe, D.A., Carlson, M.J., Sadanobu, J., Chan, S.I. and Goddard III, W.A. (1999) *J. Phys. Chem.*, **B103**, 3001–3008.

Delaglio, F., Kontaxis, G. and Bax, A. (2000) *J. Am. Chem. Soc.*, **122**, 2142–2143.

Engh, R.A. and Huber, R. (1991) *Acta Cryst. A*, **47**, 392–400.

Fowler, C.A., Tian, F., Al-Hashimi, H.M. and Prestegard, J.H. (2000) *J. Mol. Biol.*, **304**, 447–460.

Hus, J.-C., Marion, D. and Blackledge, M. (2000) *J. Mol. Biol.*, **298**, 927–936.

Hus, J.-C., Marion, D. and Blackledge, M. (2001) *J. Am. Chem. Soc.*, **123**, 1541–1542.

Kolinski, A. and Skolnick, J. (1998) *Proteins*, **32**, 475–494.

Korn, G.A. and Korn, T.M. (1961) *Mathematical Handbook for Scientists and Engineers*, McGraw-Hill, New York, pp. 22–24.

Li, Z. and Scheraga, H.A. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 6611–6615.

Losonczi, J.A., Andrec, M., Fischer, M.W.F. and Prestegard, J.H. (1999) *J. Magn. Reson.*, **138**, 334–342.

Medek, A., Olejniczak, E.T., Meadows, R.P. and Fesik, S.W. (2000) *J. Biomol. NMR*, **18**, 229–238.

Meiler, J., Blomberg, N., Nilges, M. and Griesinger, C. (2000) *J. Biomol. NMR*, **16**, 245–252.

Meiler, J., Prompers, J.J., Peti, W., Griesinger, C. and Brüschweiler, R. (2001) *J. Am. Chem. Soc.*, **123**, 6098–6107.

Mueller, G.A., Choy, W.Y., Yang, D., Forman-Kay, J.D., Venters, R.A. and Kay, L.E. (2000) *J. Mol. Biol.*, **300**, 197–212.

Ottiger, M. and Bax, A. (1998) *J. Biomol. NMR*, **12**, 361–372.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1992) *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, New York, pp. 689–699.

Prestegard, J.H. (1998) *Nat. Struct. Biol.*, **5 (Suppl.)**, 517–522.

Quine, J.R., Brenneman, M.T. and Cross, T.A. (1997) *Biophys. J.*, **72**, 2342–2348.

Ramirez, B.E. and Bax, A. (1998) *J. Am. Chem. Soc.*, **120**, 9106–9107.

Reif, B., Hennig, M. and Griesinger, C. (1997) *Science*, **276**, 1230–1233.

Rohl, C.A. and Baker, D.J. *Am. Chem. Soc.*, submitted.

Rückert, M. and Otting, G. (2000) *J. Am. Chem. Soc.*, **122**, 7793–7797.

Saupe, A. (1968) *Angew. Chem., Int. Ed. Engl.*, **7**, 97–112.

Simons, K.T., Kooperberg, C., Huang, E. and Baker, D. (1997) *J. Mol. Biol.*, **268**, 209–225.

Simons, K.T., Ruczinski, I., Kooperberg, C., Fox, B.A., Bystroff, C. and Baker, D. (1999) *Proteins: Struct. Funct. Genet.*, **34**, 82–95.

Skolnick, J., Kolinski, A. and Ortiz, A.R. (1997) *J. Mol. Biol.*, **265**, 217–241.

Skrynnikov, N.R. and Kay, L.E. (2000) *J. Biomol. NMR*, **18**, 239–252.

Standley, D.M., Eyrich, V.A., Felts, A.K., Friesner, R.A. and McDermott, A.E. (1999) *J. Mol. Biol.*, **285**, 1691–1710.

Tjandra, N. and Bax, A. (1997) *Science*, **278**, 1111–1114.

Tolman, J.R., Flanagan, J.M., Kennedy, M.A. and Prestegard, J.H. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 9279–9283.

Vijay-Kumar, S., Bugg, C.E. and Cook, W.J. (1987) *J. Mol. Biol.*, **194**, 531–544.