Assessment of molecular structure using frame-independent orientational restraints derived from residual dipolar couplings

Nikolai R. Skrynnikov & Lewis E. Kay
Protein Engineering Network Centers of Excellence and Departments of Medical Genetics, Biochemistry and Chemistry, University of Toronto, Toronto, ON, Canada M5S 1A8

Abstract

Residual dipolar couplings measured in weakly aligning liquid-crystalline solvent contain valuable information on the structure of biomolecules in solution. Here we demonstrate that dipolar couplings (DCs) can be used to derive a comprehensive set of pairwise angular restraints that do not depend on the orientation of the alignment tensor principal axes. These restraints can be used to assess the agreement between a trial protein structure and a set of experimental dipolar couplings by means of a graphic representation termed a 'DC consistency map'. Importantly, these maps can be used to recognize structural elements consistent with the experimental DC data and to identify structural parameters that require further refinement, which could prove important for the success of DC-based structure calculations. This approach is illustrated for the 42 kDa maltodextrin-binding protein.

Abbreviations: MBP, maltodextrin-binding protein; NOE, nuclear Overhauser effect; DC, dipolar coupling.

Introduction

Since the measurements of residual dipolar couplings in weakly aligned proteins were first reported (Tolman et al., 1995; Tjandra et al., 1996; Bax and Tjandra, 1997; Tjandra and Bax, 1997a), significant effort has been directed in their optimal use in structure calculations. The interpretation of dipolar couplings (DCs) in the context of structure relies on five independent alignment tensor parameters. These parameters are typically expressed as the axial component of the alignment tensor, A_a, the rhombicity of the alignment tensor, R, and three Euler angles, α, β, γ, that specify the orientation of the alignment tensor principal axes in the molecular frame. Best-fit values of the alignment tensor parameters have been used to demonstrate an excellent correlation between the experimental DC data and dipolar couplings calculated from X-ray, NMR, or neutron diffraction structures (Tolman et al., 1995; Tjandra and Bax, 1997a; Cai et al., 1999). The utility of DC data as structural probes

has led to their use in structure refinement protocols (Tjandra et al., 1997; Clore et al., 1998a; Fischer et al., 1999; Bayer et al., 1999; Baber et al., 1999; Wu et al., 2000). In this case, alignment parameters can be estimated using a preliminary structure, such as an NOE-derived structure (Bayer et al., 1999), or treated as fitting parameters throughout the calculations. In practice, A_a and R are usually estimated from the 'powder pattern' histogram of dipolar couplings or related considerations (Clore et al., 1998a,b), assuming that internuclear vectors connecting coupled spins are of known length and that their orientations are uniformly distributed on a unit sphere. The remaining parameters, α, β, and γ, are dependent on molecular structure and are treated as fitting parameters.

The interpretation of DCs in terms of molecular structure depends on the prior availability of structural information from other sources. For example, most protocols published to date rely heavily on NOE data, chemical shifts, and scalar couplings in addition to dipolar couplings to generate protein structures (Clore

et al., 1998a; Clore and Garrett, 1999; Fischer et al., 1999; Markus et al., 1999; Baber et al., 1999; Wu et al., 2000; Tsui et al., 2000). Recently, Mueller et al. (2000) have developed a refinement protocol that utilizes the known structure of the individual peptide plane. Alternatively, several studies have used X-ray coordinates as the preliminary structure, which has been subsequently altered based on solution-state DC restraints (Fischer et al., 1999; Markus et al., 1999; Tsui et al., 2000; Skrynnikov et al., 2000). More recently, Annila et al. (1999) used a protein database to identify protein folds compatible with DC data. In the same vein, Delaglio et al. (2000) employed structural information from a protein database in concert with DC data to assemble a preliminary structure of ubiquitin which was then further refined using DC and chemical shift values.

Another important feature of DC-based structure analysis is the ability to establish the relative orientations of distant molecular fragments. For example, the relative orientations of domains in multidomain proteins (Losonczi et al., 1999; Fischer et al., 1999; Markus et al., 1999; Tsui et al., 2000; Skrynnikov et al., 2000), the orientation of ligands bound to proteins (Losonczi and Prestegard, 1998; Bolon et al., 1999), conformational changes at protein binding sites (Drohat et al., 1999), and the relative orientation of double helical stems in RNA (Bayer et al., 1999) have been investigated using DC-based approaches.

In this paper we describe a method to assess the consistency between a trial protein structure and experimental DC data. The results are visualised in the form of two-dimensional maps termed 'DC consistency maps'. As illustrated below, such maps can be used to identify the regions in a protein structure that agree well with the experimental DC data (thus providing partial structural information) and to quickly recognize any sources of discrepancy between structure and DC data originating from incorrect relative orientations of molecular fragments.

In the first part of this paper, an improved procedure for estimation of A_a and R from the 'powder pattern' histogram of dipolar couplings is presented. Next, the concept of a frame-independent orientational restraint is formulated for a pair of internuclear vectors for which DC values are available. Finally, the results are illustrated for the 370-residue maltodextrin-binding protein (MBP). In particular, we consider a number of structures ranging from a random polypeptide chain to a high-resolution structure of MBP and examine how the quality of these structures can

be assessed using the experimental DC data. Two-dimensional DC-consistency maps are presented for crystal- and solution-state structures of MBP, clearly identifying the conformational differences that have recently been described (Skrynnikov et al., 2000).

It should be noted that after this manuscript had been prepared a paper by Meiler, Blomberg, Nilges, and Griesinger was published (Meiler et al., 2000) where it was demonstrated that frame-independent pairwise DC-based restraints can be fruitfully used in structure refinement protocols.

Estimation of A_a and R from fitting the 'powder pattern' distribution of dipolar couplings

Residual dipolar couplings in solution, D_{IM}, are observed when motional averaging of dipolar interactions between spins I and M is incomplete due to weak alignment. In order to ensure that DCs involving different types of spins can be treated in the same manner, it is convenient to define normalized dipolar couplings, $\tilde{D}_{IM}$:

$$\tilde{D}_{IM} = \frac{D_{IM}}{D_0^{IM} \langle S_{IM} \rangle} = A_a \left\{ (3\cos^2\theta - 1) + \right.$$
$$\left. (3/2) R \sin^2\theta \cos 2\phi \right\} \qquad (1)$$

In Equation 1, $D_0^{IM} = -(1/2\pi)(\mu_0/4\pi)\hbar\gamma_I\gamma_M \left\langle r_{IM}^{-3} \right\rangle$ is the dipolar interaction constant, $\langle S_{IM} \rangle$ is the order parameter that reflects averaging due to fast local dynamics, and θ and ϕ are polar angles that specify the orientation of the IM internuclear vector in the molecular alignment frame. The orientation of the alignment frame is defined relative to a fixed frame, such as the X-ray coordinate frame, and is expressed by the Euler angles $\{\alpha, \beta, \gamma\}$. The analysis of DC data is particularly straightforward when the internuclear distances, r_{IM}, are known, such as for pairs of nuclei in a peptide plane, and when a single effective value of $\langle S_{IM} \rangle$ can be assigned to each type of internuclear vector (Ottiger and Bax, 1998).

In the limiting case of an infinite number of internuclear vectors randomly distributed on the surface of the unit sphere, the dipolar couplings, Equation 1, form a 'powder pattern' distribution (Bloembergen and Rowland, 1953; Slichter, 1990). It has been demonstrated that DC data measured in a single protein molecule display a distribution that resembles a 'powder pattern' (Clore et al., 1998b). The histogram of the experimental DC values can therefore be used

to extract three parameters of the 'powder pattern' distribution: the minimum, the maximum, and the most frequently occurring DC values. These three parameters can be converted into reasonably accurate estimates for A_a and R (Clore et al., 1998b).

Although it is fairly straightforward to estimate A_a and R on the basis of the three values extracted from the DC histogram, more accurate estimates can be obtained if the entire histogram is used in the fitting. The advantage of the latter approach is especially clear when one considers that the minimum and the maximum DC values determined from the histogram are likely to be significantly influenced by measurement error.

In our treatment, DC histograms are fitted using an equation which is generated from the convolution of the 'powder pattern' function (Slichter, 1990) with a Gaussian-type function representing experimental error (see Appendix). The latter construct is a weighted sum of Gaussian functions where each term represents a particular type of measured dipolar coupling. The weight and standard deviation of each Gaussian are chosen according to the number of measured DC values and the corresponding experimental uncertainty, respectively. Fitting of DC histograms is carried out using the simplex algorithm, with initial values for fitting parameters A_a and R obtained from the approach described by Clore et al. (1998b).

Results of this fitting procedure are illustrated using simulated and experimental DC data for the 370-residue maltodextrin-binding protein (MBP) with bound β-cyclodextrin. An extensive set of dipolar couplings has been measured for this protein in solution with Pf1 phage (Yang et al., 1999) and the data were subsequently used to reconstruct the solution-state conformation of the protein (Skrynnikov et al., 2000). An accurate solution-state conformation of MBP was obtained from high resolution X-ray coordinates by adjusting the relative orientation of the N- and C-terminal domains according to the dipolar coupling data. In particular, the structure 1OMP(S) derived in this manner from the set of X-ray coordinates 1OMP (Sharff et al., 1992) is used as a reference in what follows.

Figure 1 shows the results of fitting the DC histogram comprising one-bond N-H^N, CO-N, and C$^\alpha$-CO couplings measured in MBP. A total of 443 DC values from residues belonging to structured regions within the N- and C-domains of the molecule were included in the analysis (Skrynnikov et al., 2000). Estimates of experimental errors were obtained from
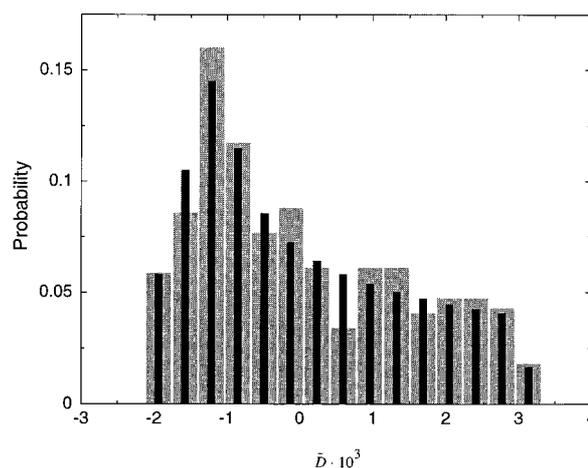
Figure 1. Experimental DC histogram (broad bars) fitted by a 'powder pattern' function (narrow bars; see text and Appendix). Experimental data comprise 443 one-bond N-H^N, CO-N, and C$^\alpha$-CO couplings from the structured regions of MBP. The fitted parameters are $A_a = 1.55 \times 10^{-3}$, $R = 0.19$ (Table 1). A fixed number of histogram bins, 15, over the entire range of normalized dipolar couplings $\bar{D}_{IM}$ have been used throughout these analyses; similar results are obtained with a larger number of bins.

repeat measurements as reported by Yang et al. (1999). It should be noted that the DC distribution can also be slightly broadened by variation of S_{IM} and r_{IM} for residues along the polypeptide chain. However, in large proteins such as MBP, the resulting uncertainty is usually small compared to the measurement error. Other potential sources of error, such as cross-correlation effects in conjunction with unresolved couplings (Tjandra and Bax, 1997b), were also found to be inconsequential.

Table 1 summarizes the results of various procedures used to estimate A_a and R in MBP. The two rows at the top of the table illustrate the effect of experimental uncertainty on the estimated values of the alignment parameters obtained from simulated DC data with added random noise. As expected, target values of A_a and R can be recovered with high accuracy when DC data are fitted using the actual structure (cf. columns 3 and 7). Reasonably good estimates for A_a and R can also be obtained from the analyses of DC histograms. The accuracy of the estimates improves when the entire DC histogram is used in the fitting procedure (columns 4 and 8) instead of the three characteristic parameters extracted from the histogram (Clore et al., 1998b) (columns 5 and 9). However, any histogram-based approach results in large uncertainties in A_a and R when the number of data points is limited, when measurement errors are significant, or

Table 1. A_a and R values estimated from N-H^N, CO-N and C$^\alpha$-CO dipolar couplings in maltodextrin-binding protein

DC data	$A_a \times 10^{-3}$				R			
	Target	Fitted with structure[c]	Fitted with histogram[d]	Calculated from histogram (Clore et al., 1998b)	Target	Fitted with structure[c]	Fitted with histogram[d]	Calculated from histogram (Clore et al., 1998b)
Simulated[a]								
485 pts	1.50	1.50±0.006	1.52±0.04	1.67±0.05	0.20	0.20±0.005	0.23±0.05	0.25±0.04
100 pts		1.50±0.014	1.51±0.12	1.46±0.10		0.20±0.011	0.24±0.12	0.26±0.14
Experimental								
443 pts	–	1.55	1.55	1.58	–	0.18	0.19	0.20
100 pts[b]		1.55±0.04	1.55±0.07	1.47±0.06		0.18±0.03	0.18±0.05	0.18±0.06

[a]The synthetic set of DC data has been produced for structure 1OMP(S) using the target values of $A_a = 1.50 \times 10^{-3}$ and $R = 0.20$ and the values of $\alpha = 6°$, $\beta = 125°$, $\gamma = 119°$ found previously for this structure (Skrynnikov et al., 2000). Prior to the analyses, normally distributed random noise with a standard deviation of 0.7 Hz (corresponding to an rmsd of 1.0 Hz between repeat measurements) was added to all data. The mean and standard deviations of A_a and R are estimated from 100 trials where random noise was applied to the complete set of synthetic DC data (first row) or, alternatively, to randomly selected subsets of DC data (second row).

[b]The mean and standard deviations of A_a and R are obtained from the analysis of 100 randomly selected subsets of DC data, where each subset consists of 100 experimental DC values.

[c]The values of A_a, R, α, β and γ are obtained by fitting Equation 1 to the set of DC values using the coordinates of structure 1OMP(S). Note that the fit is dominated by dipolar couplings with a large amplitude, i.e. N-H^N.

[d]The fitting function is described in the Appendix.

when the distribution of dipolar vectors is far from uniform. In practice we found that the gain in accuracy due to the improved histogram fitting procedure is comparable to the uncertainty in estimated values of A_a and R for the cases considered.

The two rows at the bottom of Table 1 contain the alignment parameters obtained from experimental DC data for MBP. The values of A_a and R extracted from the fitting of the DC histogram are in good agreement with the results of direct fitting using 1OMP(S) coordinates. A slight but noticeable improvement over the method of Clore et al. (1998b) is demonstrated. It is worth noting that application of the two histogram-based methods helps to quantify the uncertainty in the resulting values of A_a and R.

The present analysis indicates that sizeable errors in A_a and R arise from the histogram approaches when only ∼100 dipolar couplings measured to within a precision of ∼1 Hz are used in the calculations (second row in Table 1). In application to biomolecules where even fewer dipolar couplings are available and where the corresponding internuclear vectors show preferences for certain orientations, as is often the case for RNA and DNA fragments, any strategy relying on DC histograms is highly unreliable.

Frame-independent orientational restraints for pairs of internuclear vectors

After the alignment parameters A_a and R have been determined from the DC histogram, Equation 1 can be used to specify the possible orientations of the internuclear vector *IM* in the principal frame of the alignment tensor. Setting $z = \cos\theta$ allows Equation 1 to be cast in the following form:

$$z^2 = \frac{(\tilde{D}_{IM}/A_a) + 1 - (3/2)R + 3R\sin^2\phi}{(3 - (3/2)R) + 3R\sin^2\phi}. \quad (2a)$$

For dipolar couplings in the range $(-1 + 1.5R) \leq (\tilde{D}_{IM}/A_a) \leq 2$ the solution described by Equation 2a represents a pair of taco-shaped contours on the surface of the unit sphere, such as shown in Figure 2a (contours I(a) and I(b), corresponding to positive and negative z, respectively). Alternatively, for dipolar couplings in the range $(-1 - 1.5R) \leq (\tilde{D}_{IM}/A_a) \leq (-1 + 1.5R)$ it is convenient to use a trivial coordinate transformation $(x', y', z') = (x, -z, y)$, leading to the following parametrization:

$$z'^2 = \frac{(\tilde{D}_{IM}/A_a) + 1 - (3/2)R + (-3 + (3/2)R)\sin^2\phi'}{-3R + (-3 + (3/2)R)\sin^2\phi'} \quad (2b)$$

which gives rise to contours II(a) and II(b) in Figure 2a. Equations 2a and 2b provide a convenient
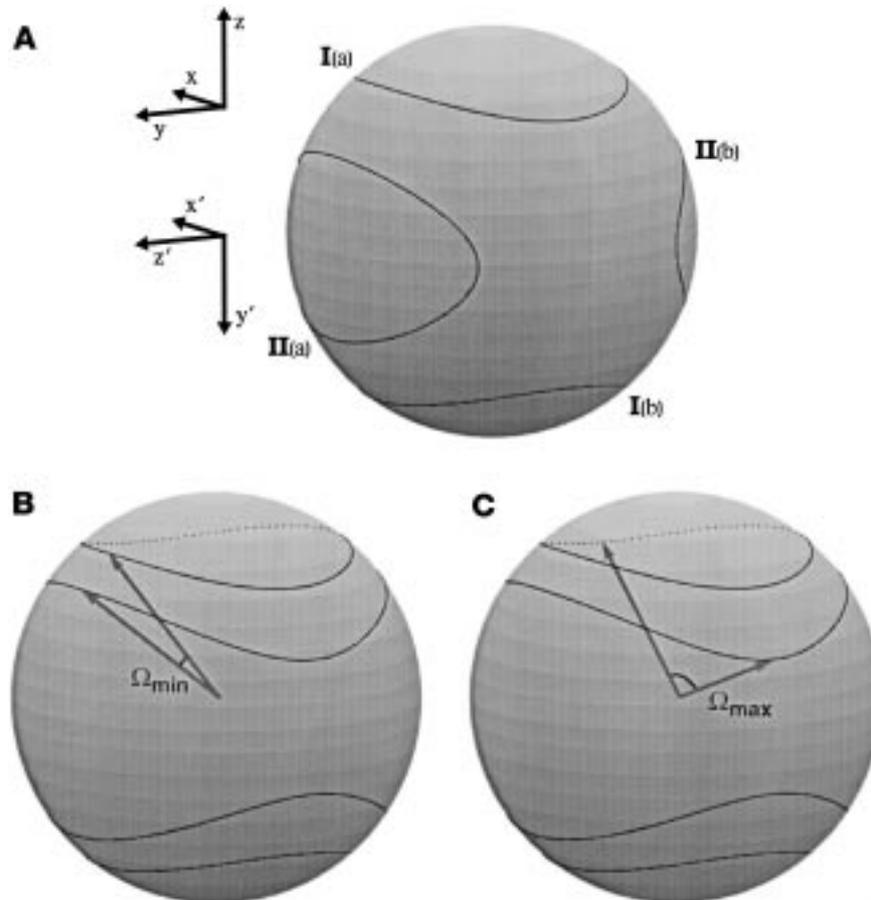
Figure 2. Possible orientations of internuclear vectors corresponding to constant DC values represented by contours on the surface of a unit sphere. The classification of the contours into types I and II is described in the text. Axes x, y, and z represent the principal axes of the alignment tensor. Ω_{min} and Ω_{max} are the minimum and maximum angles, respectively, between the two internuclear vectors tracing out the pair of contours (DC isolines).

parametrization since $z(z')$ exists for arbitrary values of $\phi(\phi')$.

Now consider two internuclear vectors, *IM* and *JN*, for which dipolar couplings, $\tilde{D}_{IM}$ and $\tilde{D}_{JN}$, have been measured. Possible orientations of these vectors with respect to the principal axes of the alignment frame (x, y, z) are specified by the curves shown in Figures 2b,c (for the sake of illustration we choose both contours to be of type I). From this illustration it is clear that the angle Ω between the vectors *IM* and *JN* is restricted to a certain interval, $[\Omega_{min}, \Omega_{max}]$.

Equation 2a can be used to calculate the dot product of the vectors *IM* and *JN* for dipolar vectors that trace out contours of type I(a). Consequently, it is straightforward to verify that Ω_{min} and Ω_{max} occur at $\phi_{IM} = \phi_{JN} = \pi/2$ and at $\phi_{IM} = 0$, $\phi_{JN} = \pi$, respectively. Another solution corresponds to the sit-

uation where one of the vectors lies on contour I(a), while the other is on contour I(b). Noting that contour I(b) is generated from I(a) by inversion about the origin, an additional interval of intervector angles can be readily determined: $[180° - \Omega_{max}, 180° - \Omega_{min}]$. These results also hold for a pair of contours of type II where orientations of the vectors are described by ϕ'_{IM}, ϕ'_{JN} according to Equation 2b. In summary, for a pair of dipolar vectors of the same type (either I or II) the allowed range of intervector angles is given by the union of the two intervals, $[\Omega_{min}, \Omega_{max}] \cup [180° - \Omega_{max}, 180° - \Omega_{min}]$, where:

$$\Omega_{\mathrm{min/max}} = \arccos\left(\frac{1}{3 \pm (3/2)R} \times\right.$$

$$\left.\left[\sqrt{\left(\frac{\tilde{D}_{IM}}{A_a} + 1 \pm \frac{3}{2}R\right)\left(\frac{\tilde{D}_{JN}}{A_a} + 1 \pm \frac{3}{2}R\right)} \pm \right.\right.$$

$$\left.\left.\sqrt{\left(2 - \frac{\tilde{D}_{IM}}{A_a}\right)\left(2 - \frac{\tilde{D}_{JN}}{A_a}\right)}\right]\right) \qquad (3)$$

with the upper sign corresponding to Ω_{min} and the lower sign to Ω_{max}.

The situation is slightly different for a pair of dipolar vectors that belong to different types, I and II. In this case, the minimum angle is obtained at $\phi_{IM} = \pi/2$, $\phi'_{JN} = 3\pi/2$, while the maximum occurs at $\phi_{IM} = 3\pi/2$, $\phi'_{JN} = \pi/2$. Straightforward calculations based on Equation 2 show that Ω_{min} is still given by Equation 3, while Ω_{max} is equal to $180° - \Omega_{\mathrm{min}}$. Therefore, the allowed range of intervector angles is reduced to a single interval, $[\Omega_{\mathrm{min}}, 180° - \Omega_{\mathrm{min}}]$.

The intervals defined above can be viewed as DC-based restraints imposed on the angle between a pair of internuclear vectors. It should be emphasized that these restraints do not depend on the orientation of the principal axes of the alignment tensor and thus can be described as frame-independent.

As an example, consider the one-bond coupled $^{15}\mathrm{N}$–$^1\mathrm{H}^N$ spin pairs from residues Glu 221 and Ser 270 in MBP, with dipolar couplings of 25.6 Hz and 20.8 Hz respectively (Yang et al., 1999). Substituting the values of $A_a = 1.55 \times 10^{-3}$, $R = 0.19$ derived in the previous section into Equation 3 we obtain that the angle between these two vectors is confined to the intervals [6.5°, 56.5°] and [123.5°, 173.5°]. In general, it is typical that each individual restraint is relatively weak: on average, one restraint eliminates only 34% of the Ω range, extending from 0° to 180° (assuming a uniform distribution of dipolar vectors and $R = 0.19$). However, the looseness of each restraint is compensated by the large number of restraints that are available for a given structure, as discussed below in more detail.

The values of Ω_{min} and Ω_{max} derived in this manner apply to the case of zero experimental error. The resulting restraints should, therefore, be relaxed somewhat in order to account for experimental uncertainty. In principle, this can be accomplished using the standard formula:

$$\Delta\Omega_{\mathrm{min/max}} =$$

$$\kappa\sqrt{\left(\frac{\partial\Omega_{\mathrm{min/max}}}{\partial\tilde{D}_{IM}}\tilde{\sigma}_{IM}\right)^2 + \left(\frac{\partial\Omega_{\mathrm{min/max}}}{\partial\tilde{D}_{JN}}\tilde{\sigma}_{JN}\right)^2}, \quad (4)$$

where $\tilde{\sigma}_{IM}$ refers to the experimental error in $\tilde{D}_{IM}$ and κ specifies the error range in units of $\tilde{\sigma}$ (in what follows $\tilde{\sigma}_{IM}$ is $1/\sqrt{2}$ of the rmsd between the repeat measurements (Yang et al., 1999) and κ is set to 2.0). However, the derivatives that enter into this expression contain singularities (see Equation 3) and we have therefore used a numeric approach for estimating $\Delta\Omega_{\mathrm{min}}$, $\Delta\Omega_{\mathrm{max}}$. Specifically, for each DC constant $\tilde{D}_{IM}$ we considered three values: $\tilde{D}_{IM} - \kappa\tilde{\sigma}_{IM}$, $\tilde{D}_{IM}$, and $\tilde{D}_{IM} + \kappa\tilde{\sigma}_{IM}$, leading to nine combinations for a given pair of dipolar couplings and subsequently to nine Ω_{min}, Ω_{max} values. The smallest of the Ω_{min} and the largest of the Ω_{max} values were used to formulate the relaxed restraints. In cases where $\tilde{D}_{IM} - \kappa\tilde{\sigma}_{IM}$, $\tilde{D}_{IM}$, or $\tilde{D}_{IM} + \kappa\tilde{\sigma}_{IM}$ fall outside the determined range of dipolar couplings, the values greater than $2 + \kappa\tilde{\sigma}_{IM}$ or less than $-1 - 1.5R - \kappa\tilde{\sigma}_{IM}$ were excluded from the analyses while the remaining outliers were rounded off to 2 and $-1 - 1.5R$, respectively.

Assessment of MBP structures using frame-independent orientational restraints

In the previous section we described how two measured dipolar couplings can be used to constrain the angle between the corresponding internuclear vectors. For a molecule where n dipolar couplings have been measured, a total of $n(n - 1)/2$ restraints can thus be formulated. Subsequently, any molecular structure can be analyzed using this set of restraints: if the angle Ω between two internuclear vectors falls outside the allowed range of angles then a violation is scored for this pair of vectors. The total number of violations calculated in this manner, N_{viol}, can be used to assess the quality of the structure and, potentially, provide some guidance in structure calculations.

In order to establish the usefulness of this parameter, N_{viol} has been calculated for a number of different MBP structures using the experimental DC data described above. The set of structures used in these computations includes X-ray and NMR-derived structures of various degrees of accuracy, as well as extended conformations of MBP (see caption of Figure 3). Among these structures, 1OMP(S) that closely approximates the solution state of MBP complexed with β-cyclodextrin (Skrynnikov et al., 2000) was
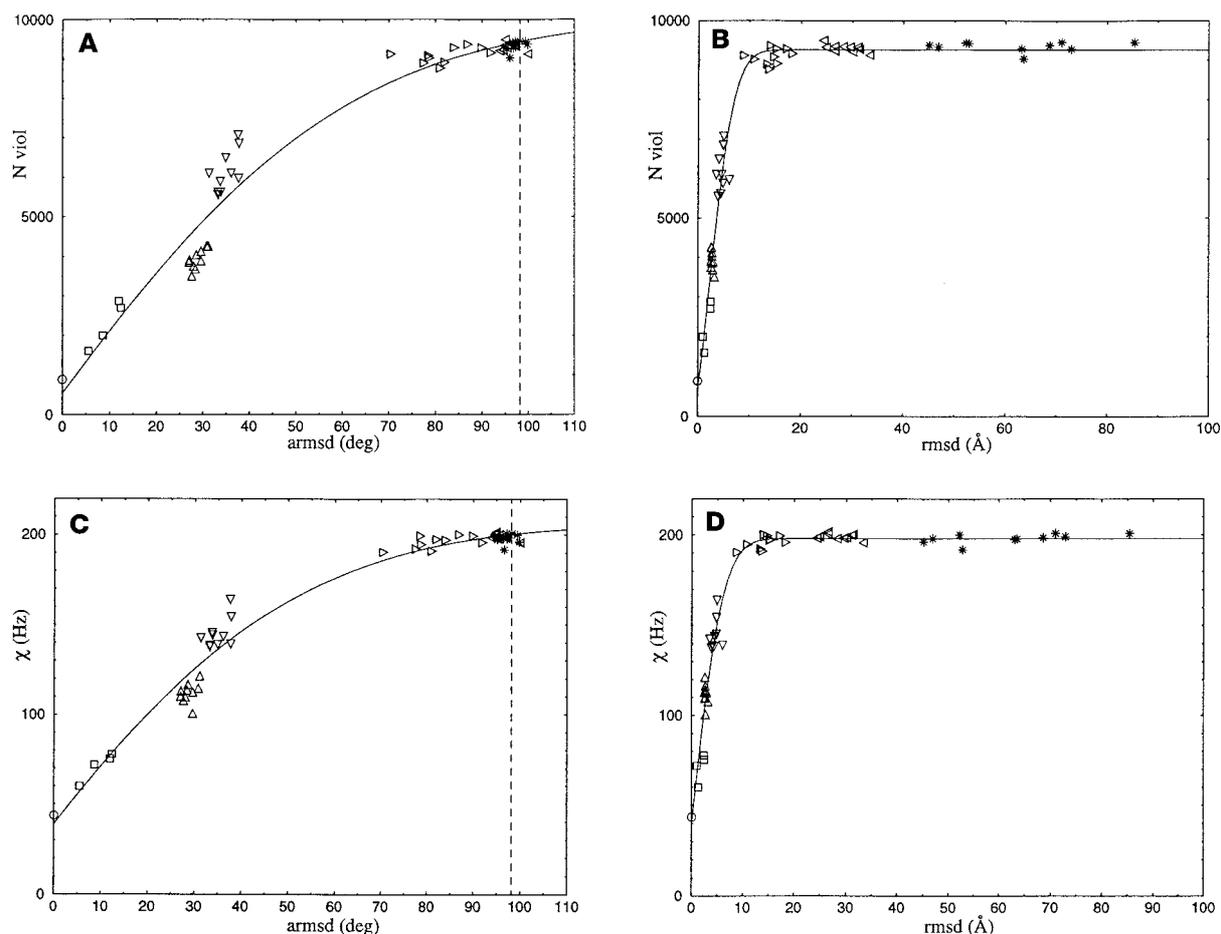
Figure 3. Two penalty functions, N_{viol} and χ, calculated on the basis of the experimental dipolar couplings for various MBP structures plotted as a function of angular root mean square deviation (armsd) and root mean square deviation (rmsd) of the structures with respect to 1OMP(S). Armsd is defined as the root mean square of the angle between the internuclear vectors from the trial structure and their counterparts from the reference structure, as computed for the set of N-H^N, CO-N, and C$^\alpha$-CO vectors used in generating Figure 1. The symbols in the plot represent: (●) structure 1OMP(S); (■) X-ray crystallography structures 1OMP (Sharff et al., 1992), 1DMB (Sharff et al., 1993), 1ANF (Quiocho et al., 1997), and 4MBP (Quiocho et al., 1997); (▲) set of NMR structures based on NOE, dihedral angle, chemical shift, and DC restraints including 1943 amide-amide, amide-methyl, and methyl-methyl NOE restraints (Mueller et al., 2000); (▼) NMR structures based on NOE, dihedral angle, and chemical shift restraints; (►) NMR structures based on partial NOE data (729 amide-amide NOE restraints); (◄) structures obtained via partial thermal unfolding of the NOE-based structures using the program CNS (Brünger et al., 1998); (∗) random conformers with the primary sequence of MBP generated using the program TRADES (Feldman and Hogue, 2000) and refined using the program CNS. The parameter N_{viol} was calculated using the set of experimental DC data consisting of one-bond N-H^N, CO-N, and C$^\alpha$-CO couplings from the structured regions in MBP. The solid lines in each of the plots help guide the eye. The dashed line indicates the limiting value of armsd, as defined in the text, for a pair of unrelated structures with uniformly distributed orientations of the internuclear vectors. The values $A_a = 1.55 \times 10^{-3}$ and $R = 0.19$ have been used in these calculations (see Table 1).

chosen as the reference structure. Accordingly, all other structures were characterized by rmsd (calculated for heavy backbone atoms) and armsd (angular rmsd computed for N-H^N, CO-N, and C$^\alpha$-CO vectors that correspond to the selected DC data) values with respect to 1OMP(S). It should be noted that armsd values are particularly useful in the analyses of dipo-

lar couplings since they are sensitive to orientations of internuclear vectors (Ottiger and Bax, 1998).

Figures 3A and 3B illustrate the correlation between N_{viol} and armsd/rmsd. As expected, a strong correlation is obtained between N_{viol} and armsd (Figure 3A). Two clusters of points in the center of the plot (triangles up and triangles down) correspond to NMR structures of MBP (Mueller et al., 2000) calculated

with and without DC data. A significant drop in N_{viol} is observed when DC data are included in structure calculations. The dashed line in Figure 3A indicates the limiting value of armsd that is obtained for a pair of unrelated structures with uniformly distributed vector orientations, $\sqrt{(\pi^2/2) - 2}$. A number of structures employed in these calculations show armsd values close to this limit, indicating that these structures bear no resemblance to 1OMP(S) with regard to vector orientations.

Further insight into the relation between N_{viol} and the quality of molecular structure can be obtained by inspection of Figure 3B which illustrates the correlation between N_{viol} and rmsd. Remarkably, N_{viol} is a good indicator of structure quality in the range of rmsd from 0 to $\sim$10 Å, while beyond this range it does not show any correlation with rmsd. This can be readily understood by considering a random extended conformation of MBP with high rmsd and armsd values relative to 1OMP(S). Rotations involving several backbone dihedral angles can transform the extended conformation into a quasi-globular one, with a concomittant sharp drop in rmsd. The armsd, however, typically remains unchanged since the orientation of vectors in this new structure remains essentially random after a small number of dihedral angle rotations. Thus, in this situation a decrease in rmsd does not lead to a decrease in N_{viol}, as indicated by the plateau in Figure 3B.

The agreement between DC data and a molecular structure is often judged on the basis of the parameter $\chi = \sqrt{\sum (D_{IM}^{\text{exptl}} - D_{IM}^{\text{calc}})^2}$, where the summation extends over the set of experimental DC data and the values of D_{IM}^{calc} are calculated for a trial structure following the optimization of A_a, R, α, β, and γ (Yang et al., 1999). Note that χ is a meaningful parameter when determined for a reasonably high-quality structure; for low-quality structures calculating χ involves essentially meaningless values of A_a, R, α, β, and γ. However, the optimization with respect to A_a, R, α, β, and γ is extremely stable even for low-quality structures, and hence a unique value of χ is always obtained.

The relation between χ and armsd/rmsd is illustrated in Figures 3C and 3D for the same set of structures that were used in calculations of N_{viol} (Figures 3A and 3B). Comparison of Figures 3A and B with 3C and 3D demonstrates that the two quantities, N_{viol} and χ, provide a roughly equivalent measure for the quality of a given structure. Notably, both func-

tions level off at high rmsd values. This suggests that an intrinsic property of DC-based restraints is that they only become useful if a preliminary molecular structure can be obtained with a resolution of better than $\sim$10 Å. This feature has already been tacitly recognized in a number of studies where NOE restraints or structural database information was used to assemble a preliminary structure within this range of accuracy (see Introduction).

In the context of an iterative structure refinement algorithm, calculating both N_{viol} and χ can be time-consuming. In particular, $n(n-1)/2$ comparison operations must be performed in order to calculate N_{viol} at each iteration. However, these comparison operations can be programmed using fast matrix manipulation routines. In addition, the values of Ω_{min}, Ω_{max} can be tabulated for all combinations of $\tilde{D}_{IM}$, $\tilde{D}_{JN}$ prior to iterative computations in order to avoid repeated evaluation of Equation 3. The use of pairwise DC-based restraints in the context of simulated annealing structure determination protocols is further discussed in the recent work by Meiler et al. (2000).

The violations obtained from an analysis of DC-based angular restraints (see above) can be depicted in the form of a two-dimensional map, as shown in Figure 4. In the maps of Figure 4a and 4b each residue is represented by a unit interval along the x- and y-axes, and each interval is divided into equal parts according to the number of dipolar couplings measured for an individual residue (in this case, the three subdivisions correspond to CO($i-1$)-N(i), N(i)-H^N(i), and C$^\alpha$(i)-CO(i) couplings, in that particular order). Each node on the two-dimensional grid corresponds to a pair of specific internuclear vectors from residues i, j. The node is marked by a circle on the map if the respective angular restraint is found to be violated (i.e., if the angle between the two internuclear vectors in the given molecular structure is outside the allowed range determined on the basis of the experimental DC values). Note that it is possible to color-code the map to reflect the extent of restraint violations (results not shown).

A disadvantage of the maps shown in Figures 4a and 4b is that it is not possible to distinguish between violation-free regions and regions for which DC data are unavailable (both cases correspond to empty areas in the map). This problem can be eliminated by constructing the maps in a different fashion as illustrated in Figures 4c and 4d. Here each measured dipolar coupling, rather than each residue, is represented by a unit interval along the x- and y-axes with the order of the DCs the same as described above. Consequently, the
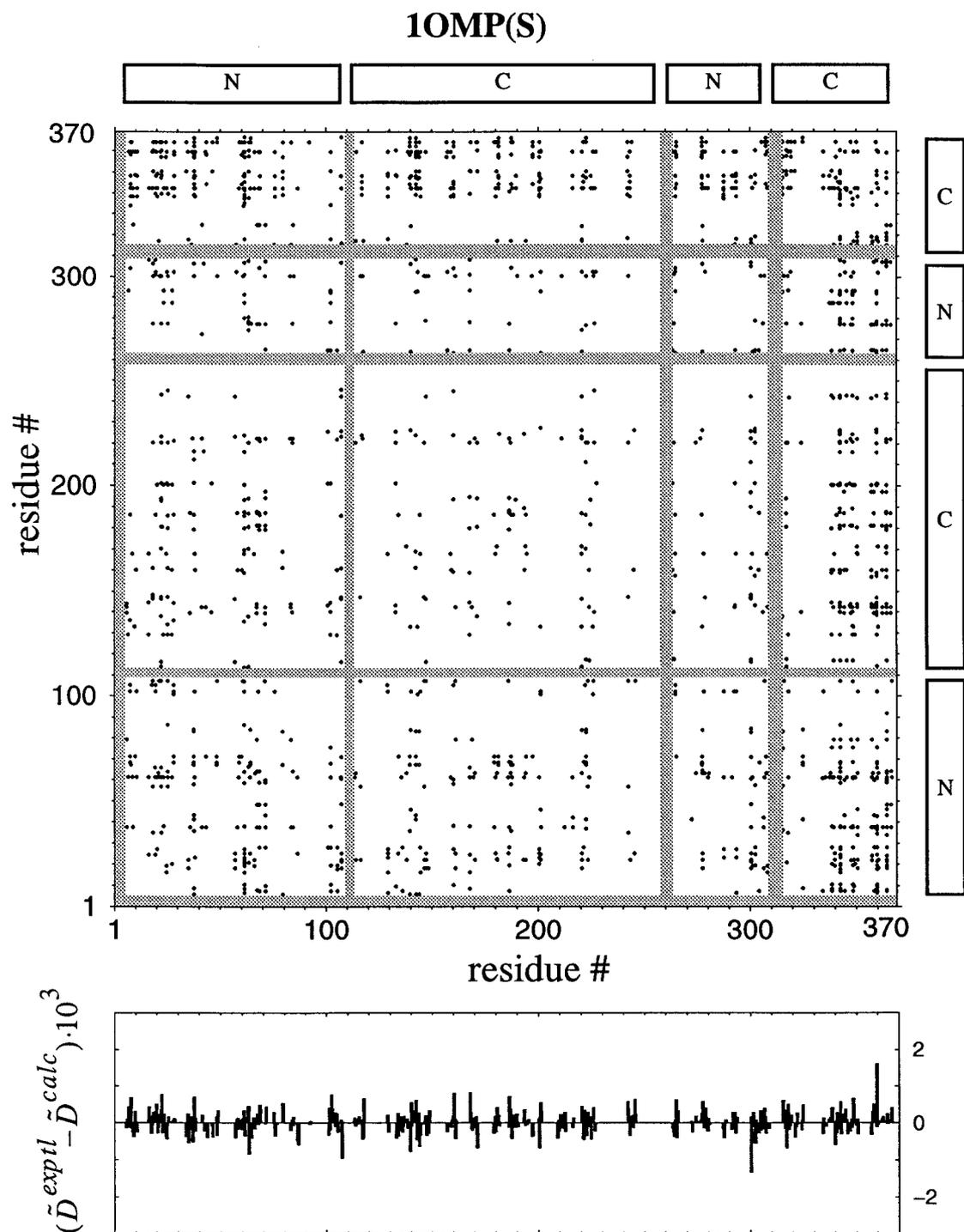
Figure 4. DC consistency maps for two structures of maltodextrin-binding protein, 1OMP(S) (a, c), and 1ANF (b, d). In maps (a, b) the grid is defined according to the residue number (sequentially from 1 to 370), while in (c, d) the grid is defined in accordance with the ordered set of measured dipolar couplings as described in the text. The DC data used to generate angular restraints are from structured regions of the polypeptide chain. The N- and C-domains consist of residues 6–109, 264–309 (N) and 114–258, 316–370 (C) as shown schematically in the plots. The regions outside N- and C-domains that have been excluded from the analyses are represented by shaded areas (a, b). Lower panels in (a, b) show the magnitude of $\tilde{D}_{IM}^{exptl} - \tilde{D}_{IM}^{calc}$, where $\tilde{D}_{IM}^{calc}$ values have been obtained via direct minimization of χ^2 with five fitting parameters, A_a, R, $\{\alpha, \beta, \gamma\}$, using the structures 1OMP(S) and 1ANF, respectively. The software used to prepare the maps is available upon request from the authors.

Figure 4b.

1OMP(S)

| N | C | N | C |

res. 367

res. 6

res. 6   res. 367
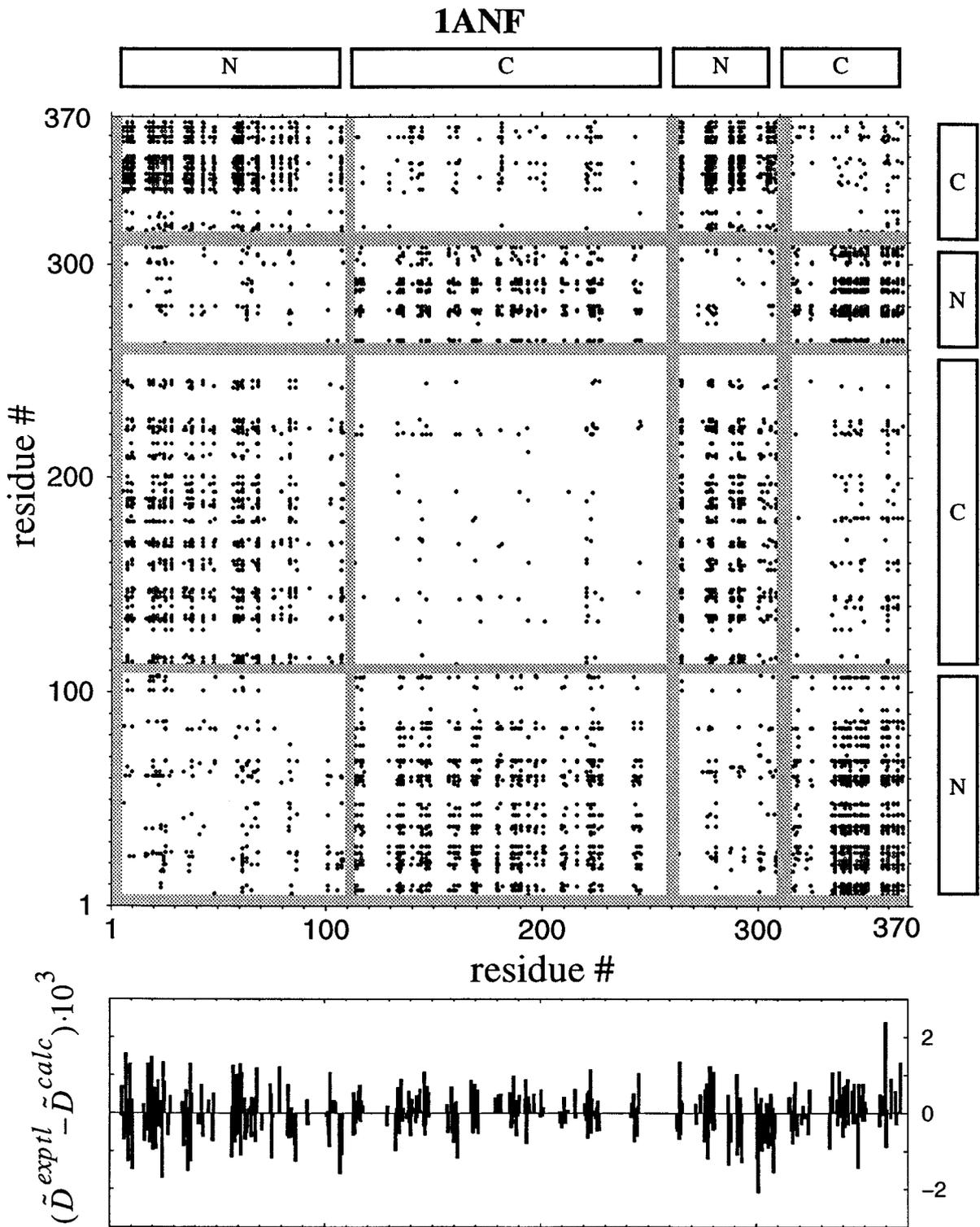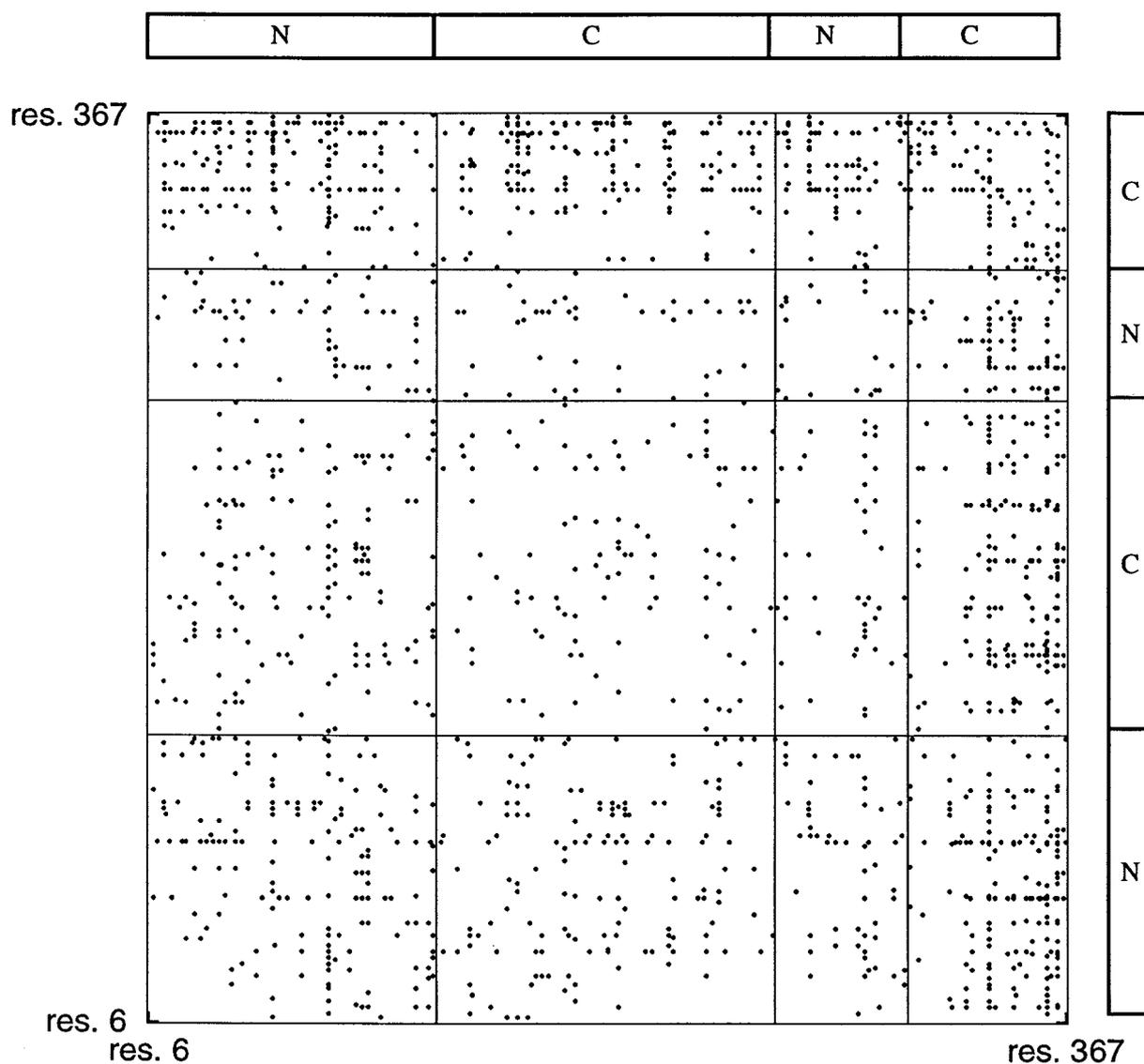
Figure 4c.

density of angular restraint violations is well defined throughout the map; however, the residue number cannot be read off the axes directly as in Figures 4a and 4b.

The maps in Figure 4 provide comprehensive pictures of the consistency between trial molecular structures and experimental dipolar couplings and are therefore referred to as DC consistency maps. Figures 4a and 4b show DC consistency maps based on DC data from the structured regions of the polypeptide chain for 1OMP(S) (Sharff et al., 1992; Skrynnikov

et al., 2000) and 1ANF (Quiocho et al., 1997), respectively. These two structures of MBP differ by a domain closure of 23° (residues in the N- and C-domains are indicated on the sides of the maps). The DC consistency map for the structure 1OMP(S) in Figure 4a shows approximately the same density of violations in both intra- and interdomain regions (i.e., for pairs of vectors from the same domain and from the two different domains). In contrast, the DC consistency map for structure 1ANF shows a dramatically increased density of violations within the interdomain regions of

1ANF
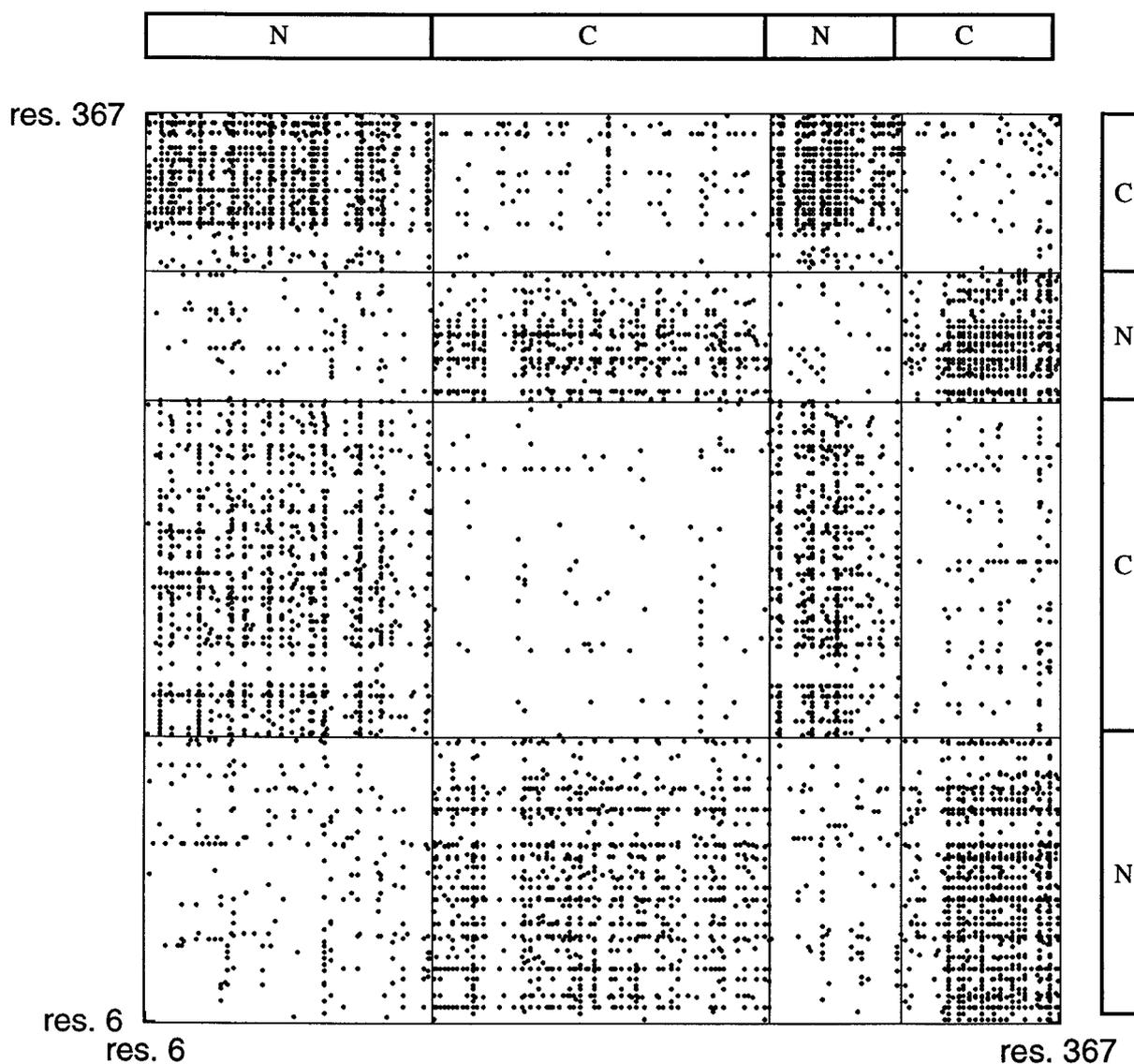
Figure 4d.

the plot, while the density of violations within each domain remains low (Figure 4b). This immediately suggests that 1OMP(S) correctly describes the relative orientation of domains in β-cyclodextrin-loaded MBP in solution, while 1ANF does not, in agreement with our previous findings (Skrynnikov et al., 2000). The same conclusion is obtained from inspection of Figures 4c and 4d.

The information contained in Figures 4b and 4d is of considerable importance since these maps

clearly indicate that the intradomain structure of β-cyclodextrin-loaded MBP in solution is reasonably well reproduced by the set of crystallographic coordinates 1ANF, whereas the relative orientation of the domains in the solution state and in the 1ANF crystal form is significantly different. Following this observation, a high-quality solution-state structure was readily obtained starting from X-ray coordinates and adjusting the relative orientation of the domains (Skrynnikov et al., 2000). In cases where structures of the indi-

vidual domains already exist, this often provides an attractive alternative to a full-scale NMR structure determination, especially since current simulated annealing protocols may not always be able to position the domains accurately despite extensive use of DC data.

More subtle details of molecular structure can also be discerned from DC consistency maps. For example, in the case of MBP, residues 340–370 show a higher than average density of both intra- and interdomain violations (Figure 4a), suggesting slight differences in local structure between solution and crystal states in this region. Moreover, DC consistency maps which include residues from the linker region (to be reported elsewhere) allow the identification of the 'hinge' residues involved in domain reorientation. The results are in agreement with X-ray crystallography analyses (Sharff et al., 1992). DC consistency maps provide an effective tool for assessment of trial protein structures. Their sensitivity is limited, however, by the uncertainty in local geometry of the trial structures (ca. 5° for orientations of dipolar vectors from high-resolution X-ray coordinates) and by the experimental errors in measured DC values.

In principle, information of a similar nature to that obtained from DC consistency maps can be accessed using the parameter χ. For example, fitting DC data separately for the N- and C-domains of MBP using the set of coordinates 1ANF results in χ values of 2.3 and 1.9 Hz per coupling, respectively. In contrast, fitting the entire structure yields $\chi = 3.6$ Hz per coupling. The large increase in χ when the whole protein is considered indicates that the relative domain orientation in the trial structure 1ANF is at odds with DC data. Note, however, that the usefulness of a χ-based analysis is limited since it does not provide a comprehensive picture of the inconsistencies in a trial structure. For example, this approach fails to reveal the subtle structural variations that occur between residues 340–370 in the solution and X-ray conformations of MBP. This is demonstrated by the lower panels in Figures 4a and 4b where χ is plotted as a function of the residue number for the structures 1OMP(S) and 1ANF. These plots offer little structural insight. Likewise, χ-based analyses would not help in identifying the regions that are consistent with measured DC data in the case where large portions of the trial structure differ from the structure in question.

It is noteworthy that DC consistency maps bear some resemblance to C^{α} distance variance maps (Levitt, 1976) that are often used in the analyses of

protein structures. The main distinction is that C^{α} distance maps highlight differences between pairs of known structures, while DC consistency maps compare one known structure with a second hypothetical structure characterized solely by a set of measured dipolar couplings. It should also be stressed that DC consistency maps exploit the global character of dipolar couplings by fully utilizing the information from all pairs of dipolar vectors regardless of the distance that separates them in the three-dimensional structure. It is envisaged that DC consistency maps will become a useful tool in the analyses of residual dipolar couplings.

Appendix

The 'powder pattern' probability density function (Slichter, 1990) can be formulated for normalized dipolar couplings $\tilde{D}$, Equation 1, as follows:

$$p(\tilde{D}) = \begin{cases} \dfrac{4}{\sqrt{f(1-y)}} K\left(\dfrac{y(1-f)}{(1-y)f}\right) & (y < f) \\[2ex] \dfrac{4}{\sqrt{y(1-f)}} K\left(\dfrac{f(1-y)}{(1-f)y}\right) & (y > f) \end{cases}$$

$$y = \left\{\left(\dfrac{\tilde{D}}{A_a}\right) + 1 + \dfrac{3}{2}R\right\} \bigg/ \left(3 + \dfrac{3}{2}R\right)$$

$$f = (3R)/\left(3 + \dfrac{3}{2}R\right) \tag{A1}$$

where $K(x)$ is the complete elliptic integral of the first kind, and y is a dimensionless variable ranging

from zero to one (we note in passing that y and f can be used to rewrite Equations 2 and 3 in a more concise form). The relevant probability density function $p(\tilde{D})$ can be obtained by calculating the convolution of $p(\tilde{D})$, Equation A1, with a Gaussian-type function representing the measurement errors:

$$P(\tilde{D}) = \int\limits_{-(1+(3/2)R)A_a}^{2A_a} P(\tilde{D}') \sum_k \frac{w_k}{\sqrt{2\pi\tilde{\sigma}_k^2}} \times$$

$$\exp\left(\frac{-(\tilde{D}-\tilde{D}')^2}{2\tilde{\sigma}_k^2}\right) d\tilde{D}' \qquad (A2)$$

In Equation A2 the summation index k refers to different types of DCs, the weighting coefficient $w_k = n_k/\sum_l n_l$ is proportional to the number of measured dipolar couplings of a particular type, n_k, and the normalized standard deviation $\tilde{\sigma}_k$ is related to the rmsd between the repeat measurements ρ_k according to $\tilde{\sigma}_k = \rho_k/(\sqrt{2}D_0^k\langle S_k\rangle)$. The probability density described by Equation A2 is a function of A_a and R and is used to approximate the histogram of experimental DC data by fitting these two parameters (see Figure 1). The integration in Equation A2 is facilitated by use of the Taylor expansion for $K(x)$ (Byrd and Friedman, 1954):

$$K(x) = \frac{\pi}{2} \sum_{m=0}^{\infty} \left(\frac{a_m}{m!}\right)^2 x^m \qquad (A3)$$

$$a_m = \left(\frac{1}{2}\right)\left(\frac{1}{2}+1\right)\ldots\left(\frac{1}{2}+m-1\right)$$

$$a_0 = 1$$

which avoids the effect of the integrable singularity that occurs in $p(\tilde{D})$ at $y = f$.

References

Annila, A., Aitio, H., Thulin, E. and Drakenberg, T. (1999) *J. Biomol. NMR*, **14**, 223–230.

Baber, J.L., Libutti, D., Levens, D. and Tjandra, N. (1999) *J. Mol. Biol.*, **289**, 949–962.

Bax, A. and Tjandra, N. (1997) *J. Biomol. NMR*, **10**, 289–292.

Bayer, P., Varani, L. and Varani, G. (1999) *J. Biomol. NMR*, **14**, 149–155.

Bloembergen, N. and Rowland, J.A. (1953) *Acta Met.*, **1**, 731.

Bolon, P.J., Al-Hashimi, H.M. and Prestegard, J.H. (1999) *J. Mol. Biol.*, **293**, 107–115.

Brünger, A.T. (1993) X-PLOR, *A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.

Brünger, A.T., Adams, P.D., Clore, G.M., Delano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) *Acta Crystallogr.*, **D54**, 905–921.

Byrd, P.F. and Friedman, M.D. (1954) *Handbook of Elliptic Integrals for Engineers and Physicists*, Springer-Verlag, Berlin.

Cai, M.L., Wang, H., Olejniczak, E.T., Meadows, R.P., Gunasekera, A.H., Xu, N. and Fesik, S.W. (1999) *J. Magn. Reson.*, **139**, 451–453.

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

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

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

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

Drohat, A.C., Tjandra, N., Baldisseri, D.M. and Weber, D.J. (1999) *Protein Sci.*, **8**, 800–809.

Feldman, H.J. and Hogue, C.W.V. (2000) *Proteins Struct. Funct. Genet.*, **39**, 112–131.

Fischer, M.W.F., Losonczi, J.A., Weaver, J.L. and Prestegard, J.H. (1999) *Biochemistry*, **38**, 9013–9022.

Levitt, M. (1976) *J. Mol. Biol.*, **104**, 59–107.

Losonczi, J.A. and Prestegard, J.H. (1998) *Biochemistry*, **37**, 706–716.

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

Markus, M.A., Gerstner, R.B., Draper, D.E. and Torchia, D.A. (1999) *J. Mol. Biol.*, **292**, 375–387.

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

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. Am. Chem. Soc.*, **120**, 12334–12341.

Quiocho, F.A., Spurlino, J.C. and Rodseth, L.E. (1997) *Structure*, **5**, 997–1015.

Sharff, A.J., Rodseth, L.E., Spurlino, J.C. and Quiocho, F.A. (1992) *Biochemistry*, **31**, 10657–10663.

Sharff, A.J., Rodseth, L.E. and Quiocho, F.A. (1993) *Biochemistry*, **32**, 10553–10559.

Skrynnikov, N.R., Goto, N.K., Yang, D., Choy, W.-Y., Tolman, J.R., Mueller, G.A. and Kay, L.E. (2000) *J. Mol. Biol.*, **295**, 1265–1273.

Slichter, C.P. (1990) *Principles of Magnetic Resonance*, Springer-Verlag, New York, NY, pp. 605–614.

Tjandra, N., Grzesiek, S. and Bax, A. (1996) *J. Am. Chem. Soc.*, **118**, 6264–6272.

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

Tjandra, N. and Bax, A. (1997b) *J. Magn. Reson.*, **124**, 512–515.

Tjandra, N., Omichinski, J.G., Gronenborn, A.M., Clore, G.M. and Bax, A. (1997) *Nat. Struct. Biol.*, **4**, 732–738.

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

Tsui, V., Zhu, L.M., Huang, T.H., Wright, P.E. and Case, D.A. (2000) *J. Biomol. NMR*, **16**, 9–21.

Wu, B., Arumugam, S., Gao, G.H., Lee, G., Semenchenko, V., Huang, W., Brew, K. and Van Doren, S.R. (2000) *J. Mol. Biol.*, **295**, 257–268.

Yang, D., Venters, R.A., Mueller, G.A., Choy, W.-Y. and Kay, L.E. (1999) *J. Biomol. NMR*, **14**, 333–343.