Journal of Biomolecular NMR, **21:** 31–40, 2001.
KLUWER/ESCOM

Direct structure refinement of high molecular weight proteins against residual dipolar couplings and carbonyl chemical shift changes upon alignment: an application to maltose binding protein

Wing-Yiu Choy[a,b], Martin Tollinger[a,b], Geoffrey A. Mueller[a,b] & Lewis E. Kay[a,*]
[a]*Protein Engineering Network Center of Excellence and Departments of Medical Genetics and Microbiology, Biochemistry, and Chemistry, University of Toronto, Toronto, Ontario, Canada M5S 1A8;* [b]*Structural Biology and Biochemistry, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8*

Received 18 April 2001; Accepted 27 June 2001

Key words: chemical shift anisotropy, dipolar couplings, maltodextrin binding protein, NMR structure calculations, torsion angle molecular dynamics

Abstract

The global fold of maltose binding protein in complex with β-cyclodextrin has been determined using a CNS-based torsion angle molecular dynamics protocol involving direct refinement against dipolar couplings and carbonyl chemical shift changes that occur upon alignment. The shift changes have been included as structural restraints using a new module, CANI, that has been incorporated into CNS. Force constants and timesteps have been determined that are particularly effective in structure refinement applications involving high molecular weight proteins with small to moderate numbers of NOE restraints. Solution structures of the N- and C-domains of MBP calculated with this new protocol are within ∼ 2 Å of the X-ray conformation.

Introduction

The emergence of residual dipolar couplings as a probe of molecular structure has significantly impacted on solution NMR studies of macromolecules (Prestegard, 1998; Tjandra and Bax, 1997). Dipolar couplings have been used to refine structures determined from NOE and scalar coupling-based restraints (Tjandra et al., 1997), to determine the relative orientation of distant molecular fragments (Fischer et al., 1999; Skrynnikov et al., 2000), to derive three-dimensional folds of proteins without the use of additional restraints (Delaglio et al., 2000; Hus et al., 2001), to obtain structural information about ligands bound to receptors (Koenig et al., 2000) and to probe molecular dynamics (Tolman et al., 1997). Perhaps one of the most important applications involves the use of dipolar couplings in structure determination of high molecular weight proteins (Mueller et al., 2000). Such studies typically require proteins with high levels of deuteration, facilitating chemical shift assignment (Farmer and Venters, 1998; Gardner and Kay, 1998). However, the removal of large numbers of protons is deleterious to subsequent steps in the structure elucidation process, since the absence of large subsets of protons precludes measurement of significant numbers of distances between spins. In this regard residual dipolar couplings play an important role, since it is possible to measure large numbers of couplings even in highly deuterated molecules (Kontaxis et al., 2000; Yang et al., 1999) which can then be used as structural restraints.

In the past several years our laboratory has studied the 370 residue maltodextrin binding protein (MBP) in the apo, maltotriose and β-cyclodextrin bound states (Evenas et al., 2001; Gardner et al., 1998; Mueller et al., 2000). One of the goals of the work has

*To whom correspondence should be addressed. E-mail: kay@pound.med.utoronto.ca

been to develop a protocol which allows rapid global fold determination of large proteins. Critical to the method has been the production of Val, Leu, Ile ($C^{\delta 1}$ only) methyl protonated, highly deuterated, ^{15}N-,^{13}C-labeled proteins (Gardner and Kay, 1997; Goto et al., 1999). High quality triple resonance spectra for assignment can be obtained (Gardner et al., 1998) and NOESY data sets for measuring H^N-H^N, H^N-CH_3 and CH_3-CH_3 NOEs recorded with excellent sensitivity and resolution (Zwahlen et al., 1998a, b). In the case of MBP loaded with β-cyclodextrin 555 (φ, ψ) dihedral angle and 1943 distances restraints were obtained and the pairwise backbone rmsds between calculated structures and the X-ray structure, 1dmb (Sharff et al., 1993), were 3.1 and 3.8 Å for the N- and C-domains, respectively (Mueller et al., 2000). Subsequent direct refinement of the structures against measured one-bond $^1H^N$-^{15}N, $^{13}C^\alpha$-$^{13}C'$, ^{15}N-$^{13}C'$, two bond $^1H^N$-$^{13}C'$ and three bond $^1H^N$-$^{13}C^\alpha$ dipolar couplings using a CNS/XPLOR (Brünger, 1992; Brünger et al., 1998) molecular dynamics protocol with default parameters resulted in large numbers of violations of dipolar restraints, with distortion of peptide plane geometry (Mueller et al., 2000). With this problem in mind, a new approach was developed in which the orientation of individual peptide planes was obtained from the dipolar coupling data and these orientations used as restraints in structure calculations employing a new module that was written for CNS (Mueller et al., 2000). Structures were generated with small numbers of violations that were significantly better defined relative to their counterparts calculated exclusively from NOE, dihedral angle and hydrogen bonding restraints. However, it was clear that additional improvements were still possible. First, the plane orientation methodology is sensitive to initial NOE-based structures which are often of moderate quality. Second, restraints for a given peptide plane can only be generated if at least three dipolar couplings from the plane can be measured, and third, the approach involves several steps and includes the use of new software, and it would clearly be advantageous if structures could be generated using a single package.

We have therefore revisited the use of protocols based on direct refinement of structures against measured dipolar couplings, recognizing that the set of parameters which is optimal for small proteins may not be satisfactory in applications involving large molecules with small numbers of NOE restraints. Stein et al. (1997) have shown that molecular dynamics in torsion-angle space (torsion angle dynamics, TAD) is more efficient than Cartesian dynamics, since the number of degrees of freedom is reduced. With this in mind we have used a TAD protocol in CNS (Brünger et al., 1998) and developed a modified set of parameters for refinement of high molecular weight proteins with a limited NOE set. An additional module has been written for CNS which uses the change in carbonyl chemical shifts upon alignment as structural restraints. The utility of the approach is demonstrated using (1) experimental data from β-cyclodextrin loaded MBP and (2) simulated restraint lists for the all-helical protein, avian farnesyl diphosphate synthase (Tarshis et al., 1994). MBP structures with low numbers of violations are obtained with a backbone rmsd of approximately 2 Å between the N- and C-domains of the average structure and the 1dmb X-ray conformation (Sharff et al., 1993). In the case of the synthase, structures showing approximately 1.8 Å decrease in backbone rmsd to the X-ray conformation were obtained by inclusion of dipolar and carbonyl chemical shift anisotropy restraints (see below) relative to structures calculated from NOEs and dihedral angles only.

Theory

The dipolar coupling (DC), D_{AB} , between two spin one-half nuclei A and B is given by (Tjandra and Bax, 1997)

$$D_{AB} = D^0_{AB} A_a \left[\left(3\cos^2\theta_{AB} - 1 \right) + \frac{3}{2} R \sin^2\theta_{AB} \cos 2\phi_{AB} \right], \quad (1)$$

where θ_{AB} and ϕ_{AB} are the polar angles that describe the orientation of the vector connecting spins A and B with respect to the principal alignment frame, A_a and R are the axial and the rhombic components of the molecular alignment tensor, respectively, and

$$D^0_{AB} = -\left(\frac{1}{2\pi}\right)\left(\frac{\mu_0}{4\pi}\right)\hbar\gamma_A\gamma_B\left\langle r^{-3}_{AB}\right\rangle \quad (2)$$

is the dipolar interaction constant. It has been shown that A_a and R can be estimated in a straightforward manner from the maximum, the minimum and the most frequently observed coupling values in the dipolar coupling histogram (Clore et al., 1998a). More accurate estimations, however, can be obtained if the entire dipolar coupling histogram is fitted (Skrynnikov and Kay, 2000). Because dipolar couplings provide long range orientational information which is often

lacking from other NMR restraints, these observables are very useful in structural studies. A direct refinement method has been developed (Clore et al., 1998b) to incorporate DCs into structure calculations and shown to improve both the accuracy and precision of structures in cases where large numbers of additional restraints are available (Clore et al., 1998b) or where small proteins with limited NOE sets are involved (Clore et al., 1999).

As demonstrated by Bax and coworkers, orientational restraints can also be derived from the changes in chemical shifts that occur upon alignment (Cornilescu and Bax, 2000; Cornilescu et al., 1998). The change in chemical shift is given by

$$\Delta_C = \frac{1}{3} \sum_{i=x,y,z} \sum_{j=x,y,z} A_{jj} \cos^2 \varphi_{ij} \delta_{ii}, \tag{3}$$

where φ_{ij} is the angle between the i principal axis of the traceless chemical shift anisotropy (CSA) tensor with principal components δ_{ii} and the j principal axis of the molecular alignment tensor and

$$\begin{aligned} A_{ZZ} &= 2A_a,\ A_{XX} = A_a\left(\tfrac{3}{2}R - 1\right), \\ A_{YY} &= -A_a\left(\tfrac{3}{2}R + 1\right). \end{aligned} \tag{4}$$

Δ_C can also be expressed as the sum of two 'virtual' dipolar couplings,

$$\begin{aligned} \Delta_C = \ & \tfrac{1}{3}\left\{(2\delta_{XX} + \delta_{YY})\, A_a\left[\left(3\cos^2\theta_1 - 1\right) + \right.\right. \\ & \left.\tfrac{3}{2}R\sin^2\theta_1 \cos 2\phi_1\right] + \\ & (2\delta_{YY} + \delta_{XX})\, A_a\left[\left(3\cos^2\theta_2 - 1\right) + \right. \\ & \left.\left.\tfrac{3}{2}R\sin^2\theta_2 \cos 2\phi_2\right]\right\}, \end{aligned} \tag{5}$$

where (θ_1, ϕ_1) and (θ_2, ϕ_2) are the polar angles that describe the orientation of the X and Y components of the CSA tensor, respectively, with respect to the principal alignment frame. The orientation of these components can, in turn, be recast in terms of the orientation of bond vectors according to

$$\begin{aligned} \vec{\delta}_{XX} &= C_1^X \overrightarrow{\mathbf{CN}} + C_2^X \overrightarrow{\mathbf{CO}}, \\ \vec{\delta}_{YY} &= C_1^Y \overrightarrow{\mathbf{CN}} + C_2^Y \overrightarrow{\mathbf{CO}}, \end{aligned} \tag{6}$$

where δ_{ii} is a unit vector along the i axis of the principle shift tensor frame, $\overrightarrow{\mathbf{CN}}$ and $\overrightarrow{\mathbf{CO}}$ are the normalized C′-N and C′-O bond vectors in the peptide plane, (C_1^X, C_2^X) are the magnitudes of $\vec{\delta}_{xx}$ bond vector components parallel to $\overrightarrow{\mathbf{CN}}$ and $\overrightarrow{\mathbf{CO}}$, respectively, with (C_1^Y, C_2^Y) the corresponding magnitudes for the $\vec{\delta}_{YY}$ vector components, and

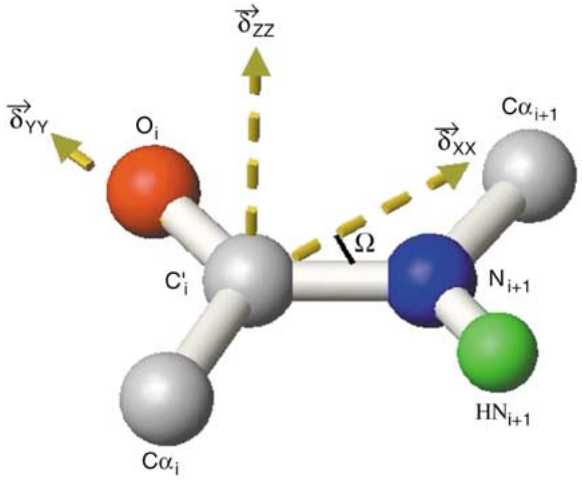

Figure 1. Orientation of the $^{13}C'$ CSA tensor components in the peptide plane. The values of δ_{xx}, δ_{yy} and δ_{zz} used in the CANI module in the present work are 73.33, 7.33 and −80.66 ppm, respectively, with the angle Ω between $\vec{\delta}_{xx}$ and the C′-N bond set to 40°.

$$\begin{aligned} C_1^X &= \frac{\sin(\angle NCO - \Omega)}{\sin(\angle NCO)}, \quad C_2^X = \frac{\sin(\Omega)}{\sin(\angle NCO)}, \\ C_1^Y &= \frac{-\cos(\angle NCO - \Omega)}{\sin(\angle NCO)} \text{ and } C_2^Y = \frac{\cos(\Omega)}{\sin(\angle NCO)} \end{aligned} \tag{7}$$

where $\angle NCO$ is the N-C′-O bond angle. Previous solid-state NMR studies on model peptides have shown that for the carbonyl carbon the δ_{ZZ} component $(-\delta_{ZZ} > -\delta_{YY} > -\delta_{XX})$ is approximately perpendicular to the peptide plane while δ_{XX} and δ_{YY} lie in the peptide plane with $\vec{\delta}_{XX}$ at an angle Ω with respect to the peptide C′-N bond (Figure 1) (Teng et al., 1992). The changes in $^{13}C'$ chemical shifts that occur upon alignment can be predicted from a protein structure so long as the magnitude and orientation of the $^{13}C'$ CSA tensor is known (Cornilescu and Bax, 2000). Cornilescu et al. (1998) have shown that there is a good agreement between the experimentally observed and the calculated changes in $^{13}C'$ chemical shifts for ubiquitin.

Carbonyl chemical shift (C-CS) restraints have been incorporated into structure calculations using a new CNS module (CANI) derived from the existing module (SANI) for including dipolar coupling restraints (Clore et al., 1998b). The C-CS restraint file format is

assign (resid 500 and name OO)
(resid 500 and name Z)
(resid 500 and name X)
(resid 500 and name Y)
(resid i and name N)
(resid $(i-1)$ and name C)
(resid $(i-1)$ and name O) δ_{XX} (in ppm)
δ_{YY} (in ppm) Ω (in degrees) Δ_C (in ppb)
$\Delta_{C,\text{error}}$ (in ppb)

where Δ_C and $\Delta_{C,\text{error}}$ pertain to the $^{13}C'$ chemical shift of residue $i-1$. Recent studies have shown that the $^{13}C'$ chemical shift tensor is sensitive to secondary structure (Cornilescu and Bax, 2000). Therefore, in the implementation of the CANI routine it is possible to include different values of δ_{ii} for each residue. In all calculations described here the same CSA tensor values ($\delta_{XX} = 73.33$ ppm and $\delta_{YY} = 7.33$ ppm, $\Omega = 40°$) have been employed (Cornilescu and Bax, 2000; Teng et al., 1992). Figure 2 shows the correlation between the observed and the predicted C-CS values based on the 1dmb X-ray structure (Sharff et al., 1993) and the CSA tensor parameters mentioned above.

Materials and methods

Structure calculations of MBP loaded with β-cyclodextrin were based on the set of NOE and hydrogen bonding restraints described previously in detail (Mueller et al., 2000). Backbone dihedral angles were also included in calculations and were determined exclusively on the basis of ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and $^{13}C'$ chemical shifts using the program TALOS (Cornilescu et al., 1999). Restraints consisting of the average ϕ, ψ value ± 2 standard deviations or at least $\pm 20°$ from the average predicted value were employed for 232 residues. These restraints were supplemented with one-bond $^{1}H^{N}$-^{15}N, $^{13}C^{\alpha}$-$^{13}C'$ and ^{15}N-$^{13}C'$ dipolar couplings measured with TROSY-based HNCO pulse sequences described by Yang et al. (1999). The $^{13}C^{\alpha}$-$^{13}C'$ and ^{15}N-$^{13}C'$ couplings were 'normalized' to the magnitude of the $^{1}H^{N}$-^{15}N couplings (with a concomitant increase in errors) so that similar deviations between each class of measured and calculated couplings will contribute equally to the dipolar energy during the structure refinement. The input coefficients for the SANI module, ($D^{0}_{^{1}H^{N}-^{15}N} \cdot A_a$) and R, were estimated from the histogram of DCs using the method proposed by Skrynnikov and Kay (2000).

$^{13}C'$ chemical shifts were measured using TROSY-based HNCO experiments, as described previously (Yang and Kay, 1999), with the exception that a ^{1}H 180° pulse was inserted at the midpoint of the $^{13}C'$ chemical shift evolution period to refocus the scalar/dipolar coupling interactions between two-bond coupled $^{13}C'$ and $^{1}H^{N}$ spins. An additional ^{1}H 180° pulse was inserted prior to the $^{13}C'$ evolution period to ensure that TROSY (and not anti-TROSY) spectra were recorded. NMR experiments were performed on a Varian Inova 600 MHz spectrometer with sample conditions described previously (Yang et al., 1999). Each 3D HNCO spectrum was recorded as a complex data matrix comprised of 80 × 30 × 576 points in ($t_1[^{13}C']$, $t_2[^{15}N]$, $t_3[^{1}H^{N}]$) corresponding to acquisition times of (52.9 ms, 18.8 ms, 64.0 ms). The ^{15}N time domain was doubled using mirror image linear prediction to produce a final data set with digital resolution of (2.95 Hz/pt, 12.50 Hz/pt, 3.91 Hz/pt) in each of (F_1, F_2, F_3). Data were processed using NMRPipe software (Delaglio et al., 1995) and analyzed with the PIPP/CAPP routines (Garrett et al., 1991). Values of Δ_C were calculated from the difference in the $^{13}C'$ chemical shifts between aligned and unaligned samples. Note that in the case of the aligned sample (19 mg/ml Pf1 phage) the spectrometer is locked on one of the two ^{2}H lines (frequencies of $\omega_D \pm \Delta_D$), while in the unaligned sample the spectrometer is locked on the ^{2}H singlet resonating at ω_D. Therefore, in addition to the CSA dependent shift offset, all ^{13}C shifts are further offset by a constant amount. The offset is estimated by minimizing the difference between experimental and predicted $\Delta_C$values using the 10 lowest energy structures calculated on the basis of NOE, dihedral angle and DC restraints. An offset of 126.5 ppb was subtracted from all measured Δ_C values prior to their use in structure refinement. In principle, this offset can also be obtained from the splitting of the two ^{2}H lines using the relation,

$$|\text{offset}| = \frac{\text{splitting}}{2} x \frac{\gamma_C}{\gamma_D}.$$

In practice, however, we find slight systematic differences between $^{13}C'$ chemical shifts in samples with and without phage (∼20 ppb) that can only be accounted for using the procedure described above. A value of 10 ppb was used for $\Delta_{C,\text{error}}$ in all computations. In summary, 1943 NOE, 48 hydrogen bond, 464 dihedral angle, 815 DC and 278 C-CS restraints were used in the structure calculations.

Structure calculations of MBP were performed using CNS (Brünger et al., 1998) starting from extended structures and using a combination of TAD and Cartesian dynamics (Stein et al., 1997), as summarized in

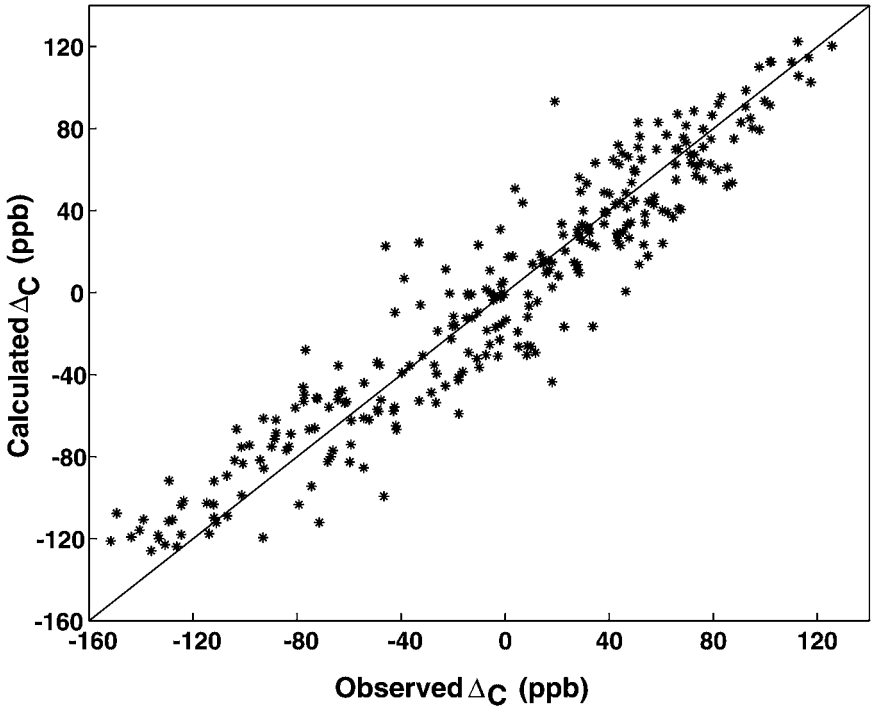

Figure 2. Correlation between measured changes in $^{13}C'$ chemical shifts of MBP complexed with β-cyclodextrin (after the offset correction, see text) with values predicted on the basis of the X-ray structure of the complex, 1dmb (Sharff et al., 1993). The correlation coefficient is 0.90 and the rmsd is 20.7 ppb.

Figure 3. Following an initial Powell minimization to regularize the geometry of the input structures an initial (hot) TAD phase was performed at a temperature of 50 000 K for 1000 molecular dynamic steps, each of 15 fs. Default CNS force constants for the energy functions were used at this stage, with the force constants for DC and C-CS restraints set to 0.003 kcal H_z^{-2} and 0.00012 kcal ppb^{-2}, respectively. Subsequently, a TAD cooling phase comprised of 60 000 steps each of 5 fs was employed with the temperature decreasing from 50 000 K to 2000 K during this interval. The force constants for DC restraints were scaled from 0.003 to 0.06 kcal Hz^{-2} while the force constants for C-CS values were increased from 0.00012 to 0.003 kcal ppb^{-2}, with other force constants ramped using default values. Subsequently a Cartesian dynamics phase was employed comprised of 10 000 steps each of 2 fs with the temperature decreasing from 2000 K to 0 K during this interval. The final values of the force constants for the DC and C-CS restraints that were used in the TAD cooling phase were employed for the duration of Cartesian dynamics, while CNS default values were used for the other parameters. As a final step standard Powell minimization of the structures was employed after the Cartesian dynamics phase.

Results and discussion

The CNS protocol that has been developed for direct refinement of high molecular weight proteins using a limited NOE data set in concert with DC and C-CS restraints is illustrated in Figure 3 and described in Materials and methods. The approach used is a modification of standard torsion angle molecular dynamics methodology in CNS (anneal.inp) (Stein et al., 1997). The default parameters in the anneal.inp routine have been shown to be efficient for structure refinement of proteins with less than 3000 atoms using distance and dihedral angle restraints. In the case of MBP (5700 atoms) the use of default parameters does not, however, lead to acceptable structures and the length of the cool torsion angle dynamics stage (see Figure 3) must be increased significantly. In addition, in order to ensure a stable DC energy profile it is important to obtain a proper balance between the timesteps used in the dynamics trajectory and the force constants employed.

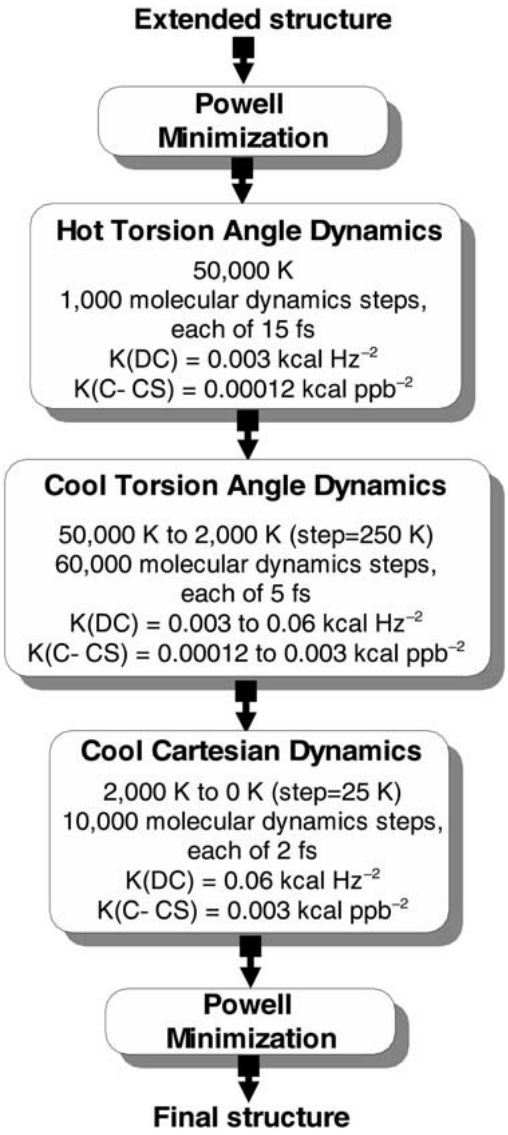

Figure 3. Flow chart illustrating the structure calculation protocol.

Figure 4 illustrates the rmsd between measured and calculated dipolar couplings during the cool phase of the TAD trajectory (extending from 50 000 K to 2000 K in this example) as a function of timestep size using the force constants for DC restraints indicated in Materials and methods above. It is clear that timesteps of 3 and 5 fs result in stable trajectories, although a value of 5 fs leads to better convergence. We have therefore used a timestep of 5 fs in all calculations. The timesteps and force constants employed ensure that there are no significant bond angle and peptide plane distortions in the resulting structures with good agreement obtained between measured and predicted DC and C-CS values.

The ten lowest energy structures of β-cyclodextrin loaded MBP calculated on the basis of different combinations of experimental restraints are shown in Figure 5 and summarized in Table 1. Inclusion of DC restraints ($^{1}H^{N}$-^{15}N, $^{13}C^{\alpha}$-$^{13}C'$ and ^{15}N-$^{13}C'$) into structure calculations significantly improves the quality of the structures (Table 1, column 2) as evaluated by the rmsd relative to the X-ray structure, 1dmb (Sharff et al., 1993). The rmsd between the N- and C-domains of the mean solution structure and 1dmb decreases by approximately 0.6 Å and 0.5 Å respectively, while the global backbone RMSD drops by 1.5 Å. Compared to the previously proposed method where the orientation of peptide planes are used as restraints (Mueller et al., 2000), the direct refinement approach described here leads to structures that show slight improvements in quality and significantly better agreement with the observed DCs.

The utility of C-CS restraints for structure refinement is illustrated in Table 1, column 3. Significant improvements in structures calculated with NOE, dihedral angle, hydrogen bond and C-CS restraints relative to their counterparts obtained without C-CS restraints are noted. In addition, the quality of structures refined with C-CS and DC restraints are improved relative to structures obtained using only DCs as orientational restraints (column 4 vs. column 2 of Table 1). It is noteworthy that structures generated using DC restraints (column 2) are more accurate than the corresponding structures obtained using C-CS restraints without DCs (column 3). This likely reflects the fact that 815 DCs ($^{1}H^{N}$-^{15}N, $^{13}C^{\alpha}$-$^{13}C'$ and ^{15}N-$^{13}C'$) are available relative to 278 C-CS values.

As described previously, all of the structural restraints for MBP have been generated from measurements on methyl protonated, highly deuterated ^{15}N, ^{13}C and ^{15}N, perdeuterated protein samples (Mueller et al., 2000). Despite the fact that 769 H^{N}-CH_3 and 348 CH_3-CH_3 NOEs have been obtained, a substantial number of sidechain torsion angles in the calculated structures are not well defined. To reduce the possibility of backbone and sidechain torsion angles existing in physically unlikely and energetically unfavorable conformations, Kuszewski and Clore (2000) have developed a method for refining NMR structures against torsion angle potentials of mean force. We have included their approach in our calculations (Table 1, column 5) with force constants for the torsion angle potentials increased from 0.01 to 0.3 in the cool

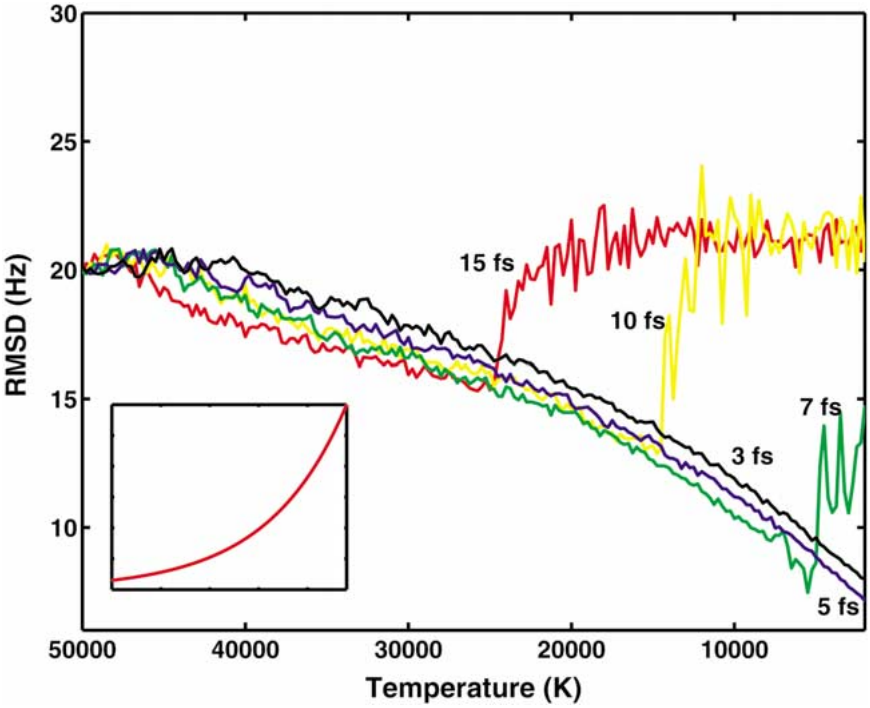

Figure 4. RMSD between observed (normalized) and calculated DC values during the cooling TAD (temperature changed from 50 000 to 2000 K) trajectory as a function of different timesteps. The RMSD values are obtained by averaging over results from 10 MBP structure calculations. The change in force constant for the DC energy terms with temperature is shown in the inset.

TAD stage and maintained at 0.3 throughout the Cartesian dynamics. Although essentially no improvements in the global and individual domain backbone rmsds were obtained, the resulting structures show better sidechain orientations as established by PROCHECK (Laskowski et al., 1998).

In previous studies we have used residual dipolar couplings to determine the relative orientation of the N- and C-domains of MBP starting from X-ray structures with the conformation of the individual domains fixed but not the linker region (Evenas et al., 2001; Skrynnikov et al., 2000). Of interest, the domain structure in the solution state of β-cyclodextrin loaded MBP is more closed (by approximately 10°) than the corresponding X-ray conformation (Evenas et al., 2001). Notably, when this procedure is repeated including both DC and C-CS restraints a closure angle of 10° relative to the X-ray structure (1dmb) is also obtained. In contrast, structures calculated entirely on the basis of NMR-derived data using methodology described herein, including DC and C-CS restraints, have closure angles of 7.7° ± 1.8° relative to the 1dmb X-ray structure. Although these values further emphasize that the solution structure is more closed than the X-ray structure, the small but quantitative difference in domain orientation obtained from the two methods (10° vs. 8°) points to the difficulty in accurately orienting structures (or domains) that are not defined to high resolution. In the present case the rmsd between each of the domains and their X-ray counterparts is approximately 2 Å; the structural accuracy of the domains is apparently not high enough to obtain the correct relative orientation. Case and coworkers have also concluded that the orientation of individual domains determined by dipolar couplings depends strongly on the accuracy of their structures (Tsui et al., 2000). Confounding the problem is that DC and C-CS restraints can be satisfied through small changes in local structure without global reorientation of individual domains.

As a further demonstration of the general applicability of our protocol we have performed structure calculations of an all-helix protein, avian farnesyl diphosphate synthase, 1fps (Tarshis et al., 1994) using simulated data derived from the X-ray structure. The simulated restraints are described in our previous work ((Mueller et al., 2000); set 2 of Table 2). In the present case C-CS restraints were simulated (70% of

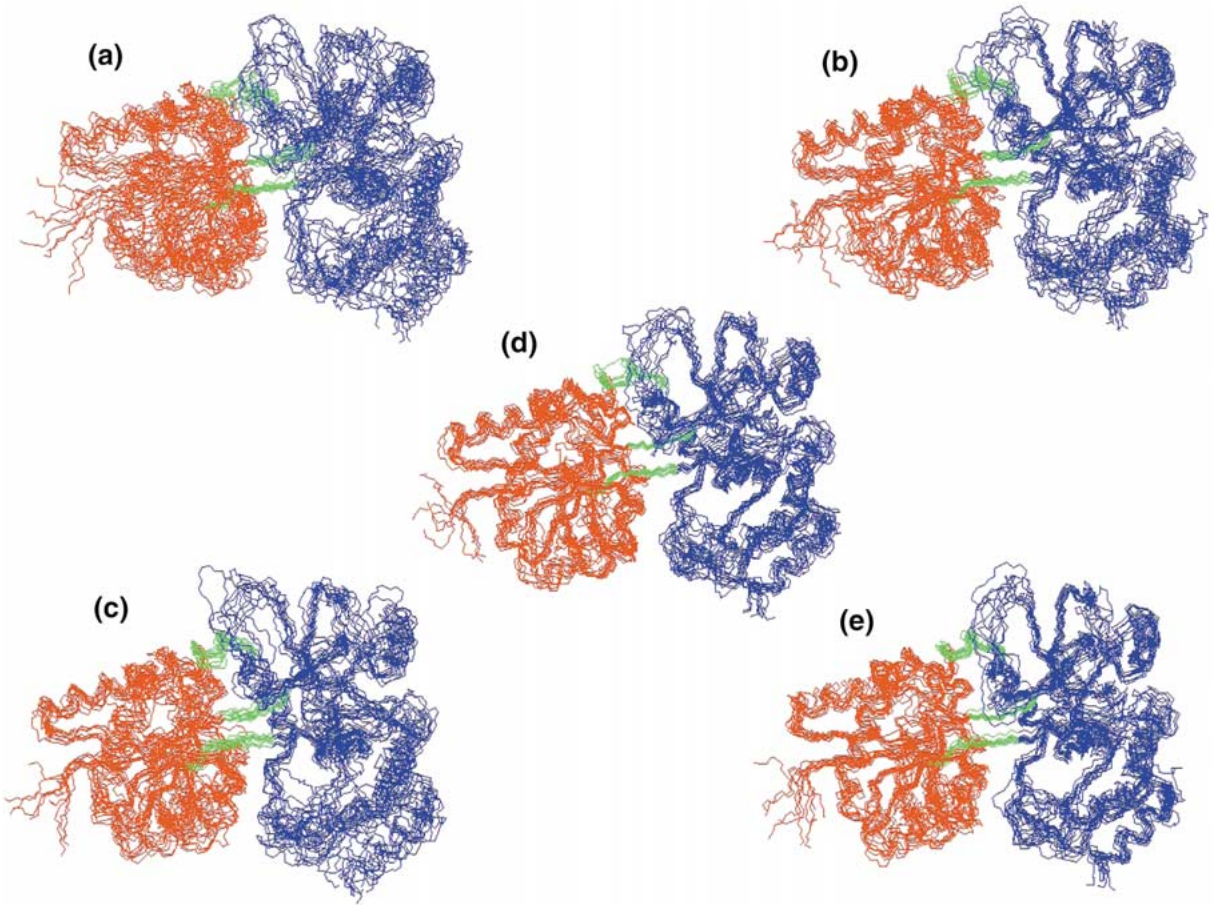

Figure 5. The ten lowest energy NMR structures calculated from different combinations of experimental restraints. (a) NOE, hydrogen bond and dihedral angle restraints, (b) NOE, hydrogen bond, dihedral angle and DC restraints, (c) NOE, hydrogen bond, dihedral angle and C-CS restraints, (d) NOE, hydrogen bond, dihedral angle, C-CS and DC restraints, (e) NOE, hydrogen bond, dihedral angle, C-CS and DC restraints along with torsion angle potentials of mean force (Kuszewski and Clore, 2000).

the residues) with the same alignment parameters used to generate DCs and a random error of 10 ppb added to Δ_C. The high helical content in this protein makes it a particularly stringent test of the methodology. For example, approximately 70% of the simulated NOEs are of the short range variety (i, $i+4$ or less). The protocol used to calculate the synthase structures is the same as described for MBP except that the number of steps executed during the cool TAD phase was increased to 80 000. The backbone rmsd of the ten lowest energy structures relative to 1fps was 3.7 $\pm$ 0.6 Å, with a 1.8 Å decrease in rmsd to the X-ray structure obtained upon including DC and C-CS restraints.

Conclusion

A TAD based approach has been developed for the calculation of structures of high molecular weight proteins from limited NOE data sets using dipolar and chemical shift anisotropy restraints. Direct refinement against orientational restraints is possible so long as the force constants and timesteps employed during the minimization trajectory are properly balanced. In the case of MBP with approximately 5 NOEs per residue and DC and C-CS data for 75% of the residues we have obtained average domain structures that are within 2 Å of the 1dmb X-ray model. Further refinement of the structures using extra information derived from protein databases is under investigation.

Table 1. Statistics for the ten lowest energy structures (out of 100) calculated on the basis of different restraints.

NOE, Dihedral, Hydrogen bond[a]	•	•	•	•	•
Dipolar coupling[b]		•		•	•
Change in carbonyl chemical shift[c]			•	•	•
Database potential					•
Average pairwise rmsd (Å)					
Global	3.79 ± 0.70	2.61 ± 0.46	3.34 ± 0.72	2.09 ± 0.21	2.24 ± 0.25
N-domain	2.28 ± 0.25	1.90 ± 0.52	2.26 ± 0.35	1.52 ± 0.24	1.74 ± 0.25
C-domain	3.34 ± 0.66	2.44 ± 0.45	2.59 ± 0.36	2.18 ± 0.30	2.26 ± 0.36
Average backbone rmsd to 1dmb (Å)					
Global	5.11 ± 0.50	3.33 ± 0.54	3.75 ± 0.47	2.78 ± 0.22	3.02 ± 0.22
N-domain	2.95 ± 0.20	2.11 ± 0.42	2.81 ± 0.24	2.01 ± 0.13	2.08 ± 0.21
C-domain	3.64 ± 0.44	2.81 ± 0.40	3.06 ± 0.32	2.34 ± 0.32	2.73 ± 0.30
[d]Backbone rmsd of mean structure to 1dmb (Å)					
Global	4.55 (4.44)	3.07 (2.87)	3.38 (3.01)	2.57 (2.40)	2.62 (2.55)
N-domain	2.66 (2.55)	2.05 (1.66)	2.73 (2.31)	1.94 (1.72)	1.82 (1.72)
C-domain	3.04 (2.83)	2.50 (2.28)	2.78 (2.49)	1.99 (1.84)	2.17 (2.01)
Experimental					
Average number of NOE violations (>0.5 Å)	0	0	0	0.4	0.3
Average number of dihedral angle violations (>5°)	0	2.6	2.5	3.2	2.9
RMSD of NH dipolar couplings (Hz)	14.2	2.7	10.7	2.9	2.9
RMSD of C-CS (ppb)	60.0	24.7	13.4	14.5	14.7

[a]Includes 1943 NOE, 464 (ϕ,ψ) dihedral angle and 48 hydrogen bond restraints.
[b]279 ^{1}H-^{15}N, 275 ^{13}C-^{13}C′ and 261 ^{15}N-^{13}C′ dipolar coupling restraints.
[c]278 C-CS restraints.
[d]The mean structures calculated by CNS with (and without) energy minimization.
The following residues are used in the rmsd calculation: Global: 6-235, 241–370 (most of the assignments for residues 229 to 239 are not available). N-domain: 6–109, 264–309 and C-domain: 114–235, 241–258, 316–370. A C-CS offset of 126.5 ppb is used.

Acknowledgements

We would like to thank Dr Nikolai Skrynnikov and Dr Johan Evenäs for many useful discussions. W.-Y.C. is a research fellow supported by a grant from AstraZeneca UK Limited. M.T. and G.A.M. are the recipients of a E. Schrödinger Fellowship (J-1933-GEN) and a Human Frontiers Science Program Fellowship, respectively. L.E.K. acknowledges grant support from the Medical Research Council of Canada. L.E.K. is an International Research Scholar of the Howard Hughes Medical Institute. The authors are grateful to Dr Marius Clore, NIH, for kindly making available his dihedral torsion angle potential protocol and software which was ported from XPLOR to CNS. All software developed, including the version of CNS used in the structure calculations is available from the authors upon request.

References

Brünger, A.T. (1992) *X-PLOR Version 3.1: A System for X-Ray Crystallography, and NMR*, Yale University, New Haven, CT.

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

Clore, G.M., Gronenborn, A.M. and Bax, A. (1998a) *J. Magn. Reson.*, **113**, 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. and Bax, A. (2000) *J. Am. Chem. Soc.*, **122**, 10143–10154.

Cornilescu, G., Delaglio, F. and Bax, A. (1999) *J. Biomol. NMR*, **13**, 289–302.

Cornilescu, G., Marquardt, J., Ottiger, M. and Bax, A. (1998) *J. Am. Chem. Soc.*, **120**, 6836–6837.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.

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

Evenas, J., Tugarinov, V., Skrynnikov, N.R., Goto, N.K., Muhandiram, D.R. and Kay, L.E. (2001) *J. Mol. Biol.*, **309**, 961–974.

Farmer, B.T. and Venters, R.A. (1998) In *Biological Magnetic Resonance*, Krishna, N.R. and Berliner, L.J. (Eds.), Kluwer Academic/Plenum Publishers, New York, NY, pp. 75–120.

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

Gardner, K.H. and Kay, L.E. (1997) *J. Am. Chem. Soc.*, **119**, 7599–7600.

Gardner, K.H. and Kay, L.E. (1998) *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 357–406.

Gardner, K.H., Zhang, X., Gehring, K. and Kay, L.E. (1998) *J. Am. Chem. Soc.*, **120**, 11738–11748.

Garrett, D.S., Powers, R., Gronenborn, A.M. and Clore, G.M. (1991) *J. Magn. Reson.*, **95**, 214–220.

Goto, N.K., Gardner, K.H., Mueller, G.A., Willis, R.C. and Kay, L.E. (1999) *J. Biomol. NMR*, **13**, 369–374.

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

Koenig, B.W., Mitchell, D.C., Konig, S., Grzesiek, S., Litman, B.J. and Bax, A. (2000) *J. Biomol. NMR*, **16**, 121–125.

Kontaxis, G., Clore, G.M. and Bax, A. (2000) *J. Magn. Reson.*, **143**, 184–196.

Kuszewski, J. and Clore, G.M. (2000) *J. Magn. Reson.*, **146**, 249–254.

Laskowski, R.A., Rullman, J.A.C., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1998) *J. Biomol. NMR*, **8**, 477–486.

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.

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

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

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

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.

Stein, E.G., Rice, L.M. and Brunger, A.T. (1997) *J. Magn. Reson.*, **124**, 154–164.

Tarshis, L.C., Yan, M., Poulter, C.D. and Sacchettini, J.C. (1994) *Biochemistry*, **33**, 10871–10877.

Teng, Q., Iqbal, M. and Cross, T.A. (1992) *J. Am. Chem. Soc.*, **114**, 5312–5321.

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

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

Tolman, J.R., Flanagan, J.M., Kennedy, M.A. and Prestegard, J.H. (1997) *Nat. Struct. Biol.*, **4**, 292–297.

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

Yang, D. and Kay, L.E. (1999) *J. Biomol. NMR*, **13**, 3–10.

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

Zwahlen, C., Gardner, K.H., Sarma, S.P., Horita, D.A., Byrd, R.A. and Kay, L.E. (1998a) *J. Am. Chem. Soc.*, **120**, 7617–7625.

Zwahlen, C., Vincent, S.J.F., Gardner, K.H. and Kay, L.E. (1998b) *J. Am. Chem. Soc.*, **120**, 4825–4831.