Journal of Biomolecular NMR, **18:** 217–227, 2000.
KLUWER/ESCOM

Study of conformational rearrangement and refinement of structural homology models by the use of heteronuclear dipolar couplings

James J. Chou[a], Shipeng Li[b] & Ad Bax[a]
[a]*Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, and* [b]*Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892-0520, U.S.A.*

Received 4 July 2000; Accepted 18 August 2000

Key words: calmodulin, dipolar coupling, heteronuclear NMR, homology model, liquid crystal, refinement

Abstract

For an increasing fraction of proteins whose structures are being studied, sequence homology to known structures permits building of low resolution structural models. It is demonstrated that dipolar couplings, measured in a liquid crystalline medium, not only can validate such structural models, but also refine them. Here, experimental ^{1}H-^{15}N, $^{1}H^{\alpha}$-$^{13}C^{\alpha}$, and $^{13}C'$-$^{13}C^{\alpha}$ dipolar couplings are shown to decrease the backbone rmsd between various homology models of calmodulin (CaM) and its crystal structure. Starting from a model of the Ca^{2+}-saturated C-terminal domain of CaM, built from the structure of Ca^{2+}-free recoverin on the basis of remote sequence homology, dipolar couplings are used to decrease the rmsd between the model and the crystal structure from 5.0 to 1.25 Å. A better starting model, built from the crystal structure of Ca^{2+}-saturated parvalbumin, decreases in rmsd from 1.25 to 0.93 Å. Similarly, starting from the structure of the Ca^{2+}-ligated CaM N-terminal domain, experimental dipolar couplings measured for the Ca^{2+}-free form decrease the backbone rmsd relative to the refined solution structure of apo-CaM from 4.2 to 1.0 Å.

Introduction

With rapid advances in gene sequencing, an enormous array of proteins is becoming available for structural studies. Massive efforts are underway to study the structure of these gene products, both by crystallography and NMR (Eisenstein et al., 2000). However, with the total number of unique folds being limited to an estimated 1000 (Chothia, 1992), a structural homologue already exists in the PDB for an ever increasing fraction of these new proteins. Recently, it has been shown that if backbone dipolar couplings can be measured, it is feasible to search the PDB database for structures that are compatible with this set of dipolar couplings (Aitio et al., 1999; Annila et al., 1999; Meiler et al., 2000), making it possible to find homologous structures even if they cannot be identified on the basis of their amino acid sequences.

Due to the non-linear relationship between dipolar coupling and the orientation of the corresponding internuclear vector, it is difficult to quantitatively evaluate the degree of structural difference between a homology model (selected on the basis of dipolar coupling homology) and the structure under study, although on average a lower correlation between the experimental data and the PDB model indicates lower structural similarity. However, as we show here, considerable closer agreement between the structure under study and the PDB-derived model can be obtained if the latter is subjected to a simple simulated annealing protocol. In particular, the model can be improved by introducing experimental dipolar restraints, measured in a liquid crystalline medium for the protein's ^{1}H-^{15}N, $^{1}H^{\alpha}$-$^{13}C^{\alpha}$, and $^{13}C'$-$^{13}C^{\alpha}$ backbone pairs (Tjandra and Bax, 1997a). In a previous study, the protein backbone structure was calculated using the so-called molecular fragment replacement (MFR) approach (Delaglio et al., 2000), which relies on the availability of a nearly complete set of backbone dipolar couplings measured in two liquid crystalline media.

In contrast, the present approach is aimed at refining a model identified either on the basis of sequence homology, or from an approximate fit to a more modest number of dipolar couplings in the protein under study.

As a second application of the dipolar refinement procedure, we demonstrate that the dipolar couplings are sufficient to define large conformational rearrangements that may take place when sample conditions are altered. Conventionally, examination of large structural rearrangements of a protein by solution NMR requires one to carry out the complete structure determination for each form of the protein. Such a protocol typically involves assignment of backbone and sidechain resonances and complete analysis of NOESY spectra. Here, we demonstrate that it is possible to examine conformational changes of a protein without going through the time-consuming NOE analysis. This approach uses the structural information of one conformational state to derive, with the help of dipolar couplings, the structure of a different state.

Due to the large energetic cost associated with disruption of secondary structure, most conformational switches in nature involve reorientation of secondary structure elements, while preserving their local order. Therefore, only a relatively small fraction of the backbone dihedral angles exhibit large changes upon conformational rearrangement. Nevertheless, even with few residues changing their backbone torsion angles, quite different structures can result. For example, the conformational change in calmodulin upon Ca^{2+} ligation results in quite different packing of the hydrophobic cores of its two globular domains. For homologous proteins, the vast majority of backbone torsion angles also have roughly similar values. The essence of our approach is to employ backbone dihedral restraints derived from the starting model, and invoke the structural changes needed to obtain agreement with the dipolar couplings by means of a low temperature simulated annealing protocol.

Experimental

Experimental dipolar couplings were collected for calmodulin and a calmodulin mutant, (E31Q; E67Q), uniformly enriched with $^{13}C/^{15}N$, in a liquid crystalline medium consisting of 20 mg/ml of the filamentous phage Pf1 (Asla Labs, http://130.237.129.141//asla/asla-phage.htm), 5 mM Hepes buffer (pH 7.0), 100 mM KCl, and 5 mg protein (1 mM) in a 280-μl Shigemi microcell. Measurements were carried out both in the absence of Ca^{2+}, and in the presence of 2 mM Ca^{2+}, which is sufficient to saturate the two Ca^{2+}-binding sites in the C-terminal domain. The Ca^{2+}-free condition is achieved through three steps of buffer exchange using a centricon filter: first, CaM-bound calcium is removed with 50 mM EDTA in unbuffered water at pH 9.0. Then the sample is washed with de-ionized water. Finally, the sample buffer is exchanged into 5 mM Hepes (pH 7.0) and 100 mM KCl.

Three types of dipolar couplings were measured: $^1D_{NH}$, $^1D_{C\alpha H\alpha}$, and $^1D_{C'C\alpha}$, using respectively 2D IPAP-HSQC (Ottiger et al., 1998), 3D CT-(H)CA(CO)NH without ^1H decoupling (Tjandra and Bax, 1997b), and 3D HNCO without $^{13}C^{\alpha}$ decoupling (Grzesiek and Bax, 1992) pulse schemes. The 2D IPAP-HSQC experiment was carried out at 800 MHz ^1H frequency, the 3D CT-(H)CA(CO)NH at 600 MHz, and the 3D HNCO experiment at 500 MHz. On the basis of the length of the time domain data and the signal to noise (Kontaxis et al., 2000), the accuracy of the measured dipolar couplings is estimated at ±0.3 Hz ($^1D_{NH}$), ±1.1 Hz ($^1D_{C\alpha H\alpha}$), and ±0.1 Hz ($^1D_{C'C\alpha}$).

Results and discussion

The refinement approach is demonstrated for the protein calmodulin (CaM). CaM is a ubiquitous protein that plays a central role in numerous Ca^{2+}-dependent signaling processes (Klee, 1988). Its two small, globular domains have considerable sequence homology relative to one another, and each consists of two so-called EF-hand Ca^{2+}-binding motifs (Kretsinger and Nockolds, 1973). The crystal structure for Ca^{2+}-CaM (Babu et al., 1988) shows that within each of the Ca^{2+}-binding domains, the pairs of helices that make up the 'E' and 'F' helices of each EF-hand are nearly orthogonal, with a deep hydrophobic groove between them. Targets of CaM typically interact with CaM through interaction with hydrophobic residues in this groove (Ikura et al., 1992; Meador et al., 1992, 1993; Elshorst et al., 1999). In the absence of Ca^{2+}, the 'E' and 'F' helices of each EF-hand switch to a nearly antiparallel orientation, and the four helices in each domain form an antiparallel four-helical bundle, where the hydrophobic groove is closed and no longer accessible for target binding (Finn et al., 1995; Kuboniwa et al., 1995; Zhang et al., 1995). The structure of Ca^{2+}-free, or apo-CaM was recalculated using all previously employed restraints (Kuboniwa et al., 1995),

together with the dipolar couplings measured in the present study. Except for a minor change in orientation (ca. 6°) of the third α-helix and small shifts of residues D21–D23 and D58–D60 in the first and second flexible Ca^{2+}-binding loops, the structure of the individual domains remains quite similar (the backbone rmsd for the N-terminal domain relative to PDB entry 1CFC is 0.68 Å). The structures of the N- and C-terminal domains of apo-CaM, refined with dipolar couplings added to the previously used NMR restraints, have been deposited in the PDB (access codes 1F70 and 1F71, for the N- and C-domains).

Dipolar couplings were measured for a double mutant of CaM (E31Q; E67Q) which strongly lowers Ca^{2+}-affinity of the N-terminal domain. This mutant makes it possible to study the protein in a state where only the C-terminal domain undergoes the Ca^{2+}-induced conformational change (Maune et al., 1992). The two mutated residues are the glutamates that constitute bidentate ligands to the two Ca^{2+} ions. In the Ca^{2+}-ligated protein, each Ca^{2+} is also coordinated by three Asp-carboxyl groups and a backbone carbonyl, all located in the Ca^{2+}-binding loop that immediately precedes the mutation site. In the absence of Ca^{2+}, the chemical shifts for both domains are very similar to those reported previously (Tjandra et al., 1995). In the presence of Ca^{2+}, the C-terminal domain chemical shifts are the same as those in wild-type Ca^{2+}-CaM (Ikura et al., 1990), whereas the shifts of the N-terminal domain are relatively little affected by the presence of Ca^{2+}. Assignments are confirmed by HNCA and HNCACB spectra.

Below, we show that it is possible to start from the 2.2-Å crystal structure (Babu et al., 1988; PDB code 3CLN) of the Ca^{2+}-ligated N-terminal domain of CaM and generate a reasonably accurate structure for the apo state. Next, we show that following the same protocol, homology models for the C-terminal domain of Ca^{2+}-CaM, derived from the structures of recoverin and parvalbumin, can be improved considerably.

Starting model and backbone dihedral restraints

In this study, all structure calculations are performed using XPLOR 3.84 (Brünger, 1993), with added modules for dipolar coupling refinement (Tjandra et al., 1997). A flowchart of the protocol is shown in Figure 1. If a structure already exists, but is either of low resolution or in an alternate conformation (absence or presence of ligands, mutations, etc.), this PDB structure is used as the starting model, and all of its backbone ϕ and ψ angles are used as dihedral restraints for stage 1 of the simulated annealing protocol described below. This approach is useful for studying conformational change of a protein that may occur upon ligand binding, mutation, etc., or simply for refinement of deposited structures.

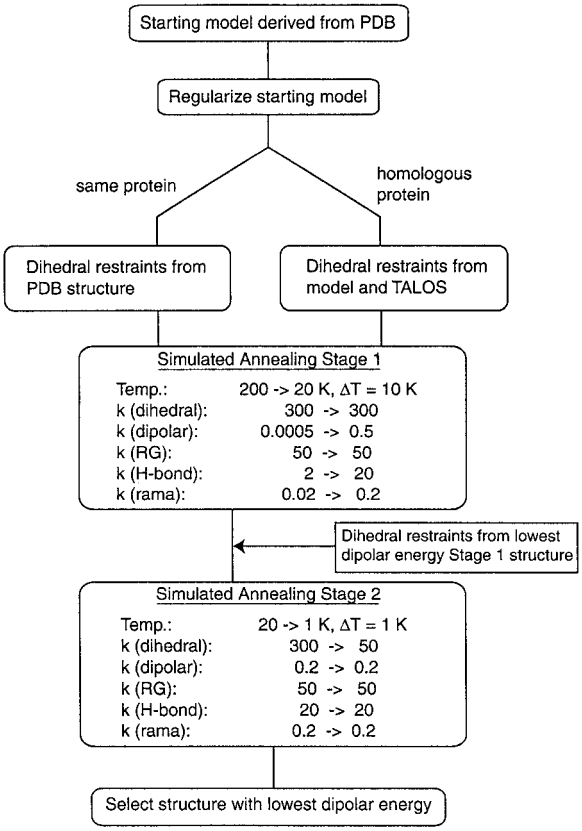

Figure 1. Flow chart representation of the structure calculation protocol used in this study. The NMR version of the crystal structure is generated by regularization using non-crystallographic symmetry (NCS) restraints, as described in the text. Force constants for the dihedral penalty term, k(dihedral) in units of kcal mol^{-1} rad^{-2}, the dipolar term, k(dipolar) in units of kcal mol^{-1} Hz^{-2} (couplings normalized to D_{NH}), the hydrogen-bond distance restraint, k(H-bond), in units of kcal mol^{-1} $Å^{-2}$, and the Ramachandran database potential, k(rama) (dimensionless), are ramped during the annealing, whereas the force constant for the radius of gyration term, k(RG) (Kuszewski et al., 1999), is kept constant at 50 kcal mol^{-1} $Å^{-2}$. The relatively small force constant for k(dipolar) reflects the larger-than-usual value for the alignment tensor. Other force constants, commonly used in NMR structure calculation, are: for Stage 1 cooling, k(vdw) = 0.002 → 4.0 kcal mol^{-1} $Å^{-4}$, k(impr) = 0.1 → 1.0 kcal mol^{-1} deg^{-2}, k(bond angle) = 0.4 → 1.0 kcal mol^{-1} deg^{-2}, and k(noe) = 2.0 (20 kcal mol^{-1} $Å^{-2}$); for Stage 2 cooling, the above k's retain the same values as they are at the end of Stage 1.

More general, the PDB database can be searched for homologous proteins, either on the basis of sequence homology or by matching dipolar couplings

(Aitio et al., 1999; Annila et al., 1999; Meiler et al., 2000). If a homologous protein is found in the PDB on the basis of sequence similarity, then a starting model is built using the program GeneMine (released by the Molecular Application Group, http://www.mag.com), which employs the segment matching approach for homology modeling (Levitt, 1992). In this case, backbone torsion restraints for stage 1 simulated annealing are imposed according to the following criteria. Since the secondary backbone structure of a protein can be identified by its characteristic chemical shifts, backbone torsion angles predicted by TALOS (Cornilescu et al., 1999) are used for validating the model angles. TALOS relies on matching Cα, C^{β}, C′, N, and H^{α} secondary chemical shifts for three consecutive residues to patterns observed in a database of previously assigned proteins of known structure. Typically, for about 65% of the residues in the protein under study, TALOS is able to predict unique φ and ψ angles on the basis of this information; residues for which predictions are not unique are marked 'ambiguous' by the TALOS program and are not used in our study. If TALOS makes a valid prediction that differs by more than 30° in φ and/or ψ from those in the homology model, the TALOS angles are used. For all other residues, the model-derived angles are used as restraints. Ideally, the TALOS angles would be incorporated in generating the original homology model, in which case only backbone angles of the model would be used during refinement. Unfortunately, modeling software which incorporates TALOS backbone angle restraints is not yet available.

Hydrogen bond restraints and side-chain potential

In order to improve convergence of the simulated annealing protocol, it is advantageous to also include hydrogen bond restraints. This is especially true for the well-defined helical regions, which can be easily identified on the basis of secondary backbone shifts. In the current study, hydrogen bond restraints are implemented as tight distance restraints (described below in the annealing protocol), and are applied to the four helices of CaM. None of the mini-antiparallel β-sheet hydrogen bonds, present in CaM and its homologues, were used because the H-bond accepting group can be difficult to identify, unless long-range NOEs or through H-bond J couplings (Cordier and Grzesiek, 1999) are available.

No information on the side chains is available from the backbone dipolar couplings. However, when using the database-derived 'Rama' potential function in XPLOR (Kuszewski et al., 1997), previously developed to exploit the well known correlation between backbone torsion angles and preferred side-chain conformers, the χ_1 angle of the vast majority of residues (especially those in helical regions) yields the correct rotameric state. Hence, a weak Rama potential is used in all simulated annealing runs.

Special care must be taken when directly using an existing X-ray structure as the starting model. Frequently, crystallographic refinement does not stringently enforce planarity of the peptide bond, or tetrahedral geometry at the C^{α} sites, and the same is true for homology models based on these structures. Therefore, when simply building an NMR starting model using the φ and ψ angles of the selected crystal structure or homology model, with planar peptide bonds and tetrahedral C^{α} geometry enforced by the NMR topology file, such an NMR starting model can differ substantially from the original starting model, despite identical φ and ψ angles. This problem can be easily solved by a brief low-temperature dynamics run, followed by energy minimization, on the NMR model, while its coordinates are kept very close to those of the homology model or X-ray structure by imposing a non-crystallographic symmetry (NCS) term in the XPLOR program. The force constant of the NCS term is adjusted such that the rmsd between the NMR model and the true starting structure is less than 0.2 Å, typically requiring a force constant of ca. 5 kcal mol^{-1} $Å^{-2}$.

Simulated annealing protocol

Refinement against dipolar couplings presents a difficult multiple minimum problem, because each dipolar coupling is compatible with two cones of bond vector orientations, pointing in opposite directions. This can lead to deep false minima, especially when parts of a structure locally get inverted during the simulated annealing trajectory. In order to avoid such false minima, a two-stage simulated annealing protocol is used (Figure 1), with both stages run at very low temperatures. In the first stage, the system is cooled from 200 to 20 K with a temperature step of 10 K, and 6.7 ps of Verlet dynamics at each temperature step, using a time step of 3 fs. Backbone torsion angle restraints for φ and ψ are enforced by harmonic quadratic potentials with strong force constants, initially fixed at 300 kcal mol^{-1} rad^{-2}. Note that this potential has no flat-well bottom, and restrains the backbone angles

quite tightly to those of the starting model. Hydrogen bond restraints for well-defined helices are applied as NOE distance restraints. O-H^N and O-N distance restraints of 2 and 3 Å are enforced with a flat-well ($\pm$ 0.25 Å) harmonic potential. The force constant is ramped exponentially from 2 to 20 kcal mol^{-1} $Å^{-2}$ during the 18 temperature steps, while the dipolar coupling restraint force constant is ramped from 0.0005 to 0.5 kcal mol^{-1} Hz^{-2} (normalized for the D_{NH} dipolar couplings).

Dipolar couplings do not provide translational information, and in order to ensure global compactness of the protein, a pseudopotential for the radius of gyration (RG) is applied with a force constant of 50 kcal mol^{-1} $Å^{-2}$. In the absence of any constraints which prevent translation of parts of the structure relative to one another (such as a long-range NOE or H-bond), the RG term is essential for avoiding an overly expanded structure. This general collapsing potential was recently introduced in XPLOR and shown to improve the packing of NMR structures and their agreement with crystal structures (Kuszewski et al., 1999). All other force constants are indicated in Figure 1. During the first stage of molecular dynamics, the backbone of the molecule is kept rather rigid by the tight dihedral restraints. However, despite the high dihedral restraint force constants, the spread among the structures (resulting from different random velocities at the start of the protocol) is relatively large, ca. 1 Å when only considering the fraction of structures with the lowest dipolar energies.

Because regions of secondary structure can reorient relative to one another with relatively little change in the tightly restrained backbone angles, these first-stage structures are already reasonably close to the true conformation (typically better than 1.5–2 Å). However, most of the individual backbone dihedral angles are still rather far from their actual values, since they are too strongly restrained by the steep harmonic potential. Therefore, at the end of the first simulated annealing stage, the structures have high backbone tension and still a considerable dipolar energy term.

Subsequently, the structure with the lowest dipolar energy from the stage 1 calculation is subjected to a second round of simulated annealing, which is conducted at very low temperature: the temperature bath is cooled further from 20 to 1 K, using a 1 K temperature step. In this second stage, all backbone angle restraints are derived from the lowest dipolar energy structure obtained in stage 1. During the second stage, the dihedral force constant is ramped down from 300 to 50 kcal mol^{-1} rad^{-2}, while the dipolar force is kept constant at 0.2 kcal mol^{-1} Hz^{-2} (again normalized for D_{NH}). All other force constants are fixed at the same values as they were at the end of the first simulated annealing stage. This second round of simulated annealing lowers the dipolar energy term (after correcting for the difference in force constant), and considerably decreases the difference between the backbone angles of the resulting structure and those of the actual structure.

Ca^{2+}-Cam N-terminal domain to apo-CaM

At a Pf1 concentration of 21 mg/ml, the magnitudes of the normalized dipolar couplings in the N-terminal domain of apo-CaM range from -35 to 35 Hz. Using the histogram approach (Clore et al., 1998), the alignment tensor magnitude, D_a, is estimated at 17.5 Hz, with a rhombicity, *R*, of 0.65. The X-ray structure of the N-terminal domain of Ca^{2+}-ligated CaM ((Babu et al., 1988; PDB entry 3CLN), after regularizing the peptide bonds and tetrahedral geometry at C^{α} in the above described manner, is used as the starting model. Its backbone differs from the apo-CaM solution structure (1F70) by 4.2 Å (Figure 2A). Since this is a case of a conformational change, all backbone torsion angles of the starting model are used as initial dihedral restraints for stage 1 simulated annealing. At the end of the first annealing stage, dipolar coupling restraints have significantly reoriented all four helices. The structure with the lowest dipolar energy differs by 1.08 Å from the apo-CaM N-terminal domain solution structure (PDB entry 1F70). Relative to the original NMR solution structure (calculated without dipolar restraints) the backbone atom rmsd is slightly larger (1.4 Å relative to 1CFC; 1.7 Å relative to 1DMO), caused primarily by a slight change in the previously relatively poorly defined orientation of the third α-helix. Among the structures generated in this stage of dynamics, very few (less than 10%) of the ϕ and ψ angles differ by more than 10° from the values imposed as torsion restraints. Consequently, the dipolar restraints remain poorly satisfied and the XPLOR energy remains rather high. In the second stage, the dihedral force constant is gradually lowered, and considerably larger changes in ϕ and ψ occur in order to satisfy the dipolar restraints. However, the structure with the lowest dipolar energy changes very little relative to the structures from stage 1, and fits to the 1F70 structure with an rmsd of 1.01 Å. The differences in backbone angles between the final model calculated in the above manner and the experimental apo-CaM solution struc-

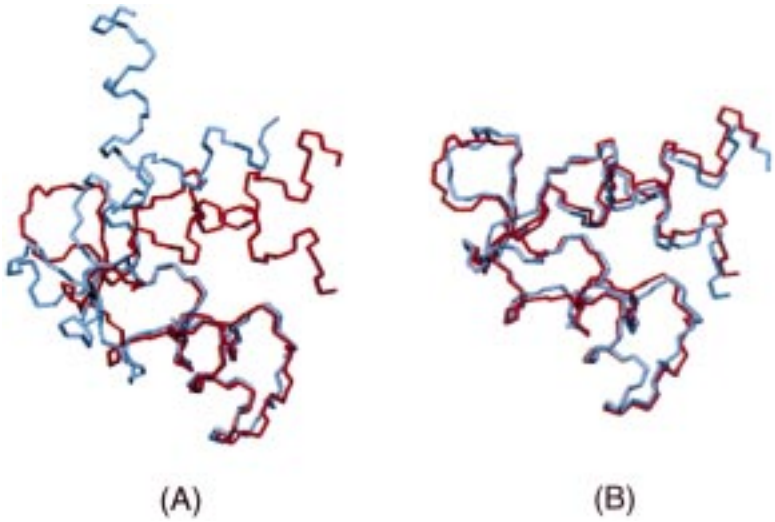

Figure 2. Structure comparison before and after dipolar coupling refinement. (A) The backbones of apo-CaM-N (red, PDB entry 1F70) and Ca^{2+}-ligated (blue) Ca^{2+}-CaM-N (N-terminal domain of CaM) (PDB entry 3CLN (Babu et al., 1988)) are best fit for residues 29–58, illustrating large structural differences between the two states. The backbone C^{α} rmsd between the two structures is 4.2 Å. (B) After refinement of 3CLN with dipolar couplings, measured for apo-CaM, using the protocol of Figure 1. The backbone of the refined structure fits the experimental apo-CaM-N structure (1F70), with a backbone C^{α} rmsd of 1.01 Å. The experimental structure of apo-CaM was obtained by recalculating the solution structure using the newly measured dipolar couplings, in addition to the previously used NOE and J-coupling derived dihedral restraints (Kuboniwa et al., 1995).

ture ('true structure') are now much smaller than those between the starting model and the true structure (Figure 3). The backbone model derived from the dipolar couplings in the above described manner is superimposed on the apo-CaM solution structure in Figure 2B, and shows remarkably good agreement.

Remote sequence homology: Recoverin to C-domain of Ca^{2+}-CaM

In order to test the effectiveness of our protocol in homology-based structure determination, recoverin, another member of the Ca^{2+}-binding family of proteins, is used to derive the structure of the C-terminal domain of Ca^{2+}-saturated CaM. There are many ways of building a homology model. In most cases, the model structure is built piece by piece onto the reference structure based on sequence similarity. In other cases, information on the protein function (e.g. specific interaction motifs) may be used to verify the model. In the current study, starting models are built using the program GeneMine. The solution structure of Ca^{2+}-free recoverin (Tanaka et al., 1995) is used to build the starting model. The relevant parts of the two proteins (residues 25–93 of recoverin and 82–148 of CaM) share 22% sequence identity and 57% similarity. Unlike the CaM C-terminal domain, which contains two EF-hands, the aligned region of recoverin

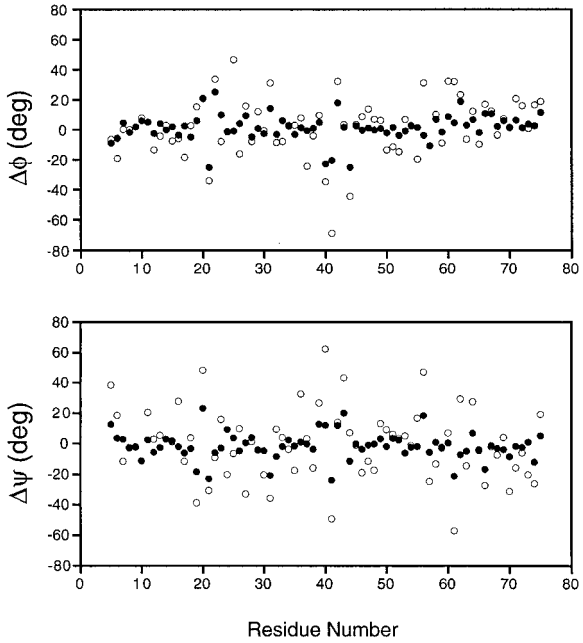

Figure 3. Differences in backbone torsion angles, $\Delta\phi$ and $\Delta\psi$, between Ca^{2+}-ligated CaM-N (starting structure; 3CLN) and apo CaM-N (target structure; 1F70) are shown as open circles. The differences between the dipolar-refined 3CLN model and the experimental 1F70 structure are indicated by solid circles.

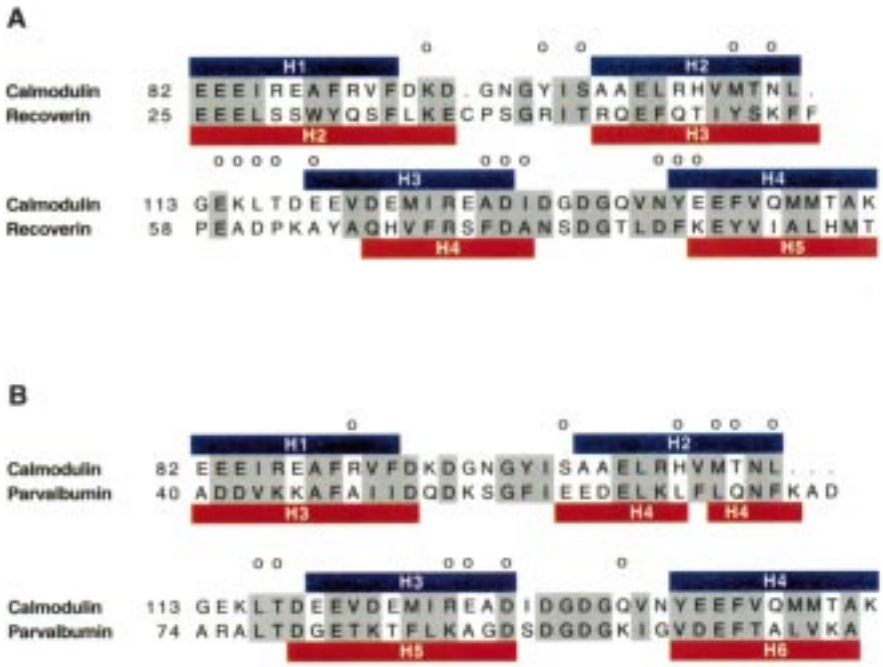

Figure 4. Sequence alignment of CaM-C (C-terminal domain of CaM) with (A) recoverin, and (B) parvalbumin. The α-helices of Ca^{2+}-ligated CaM-C (PDB access code 3CLN) are represented by red bars above the CaM sequence, whereas the α-helices of recoverin and parvalbumin (PDB access code 1CDP) are mapped by blue bars. Shaded residues are similar or identical. An open circle above a residue indicates that the TALOS prediction on the basis of the Ca^{2+}-CaM chemical shifts (using a TALOS chemical shift database from which calmodulin was removed) disagrees by more than 30° for either ϕ and/or ψ from the GeneMine-derived homology model, and the dihedral restraint is based on TALOS instead of the homology model.

encompasses only one true EF-hand, which includes its helices H4 and H5 (Figure 4A). There are also two insertions in the sequence (Figure 4A). The modeled structure differs from the 1.68-Å crystal structure of Ca^{2+}-ligated CaM C-terminal domain (Ban et al., 1994; PDB entry 1OSA) by 5 Å (Figure 5A). This large difference is primarily caused by the different relative orientations of the helices in Ca^{2+}-free recoverin relative to CaM. For the vast majority of the residues, dihedral restraints are taken from the recoverin-derived starting model, but for 16 residues where the TALOS prediction conflicts with this model (marked by open circles in Figure 4A) TALOS-derived ϕ and ψ restraints are used (see *Starting model and backbone dihedral restraints* section).

After the first stage of simulated annealing, the structure with the lowest dipolar energy differs from the Ca^{2+}-CaM crystal structure by 1.36 Å. The subsequent second stage decreases this difference to 1.25 Å (Figure 5B).

Close sequence homology: Parvalbumin to C-domain of Ca^{2+}-CaM

As a third example, we show that the approach can also be used to refine a reasonably accurate homology model, built from the structure of parvalbumin (Tufty and Kretsinger, 1975; PDB entry 1CDP), for the C-terminal domain of Ca^{2+}-saturated CaM. In this case, the homologous regions of the two proteins (residues 38–108 of parvalbumin, 82–148 of CaM) share 27% sequence identity and 70% sequence similarity (Figure 4B). More importantly, like the CaM C-terminal domain, parvalbumin also contains two regular EF-hands. Again, the starting homology model is built using GeneMine, and it fits the crystal structure (PDB entry 1OSA) of the Ca^{2+}-ligated CaM C-terminal domain to 1.25 Å (Figure 5C). Despite extensive homology, there is a three-residue insertion in parvalbumin relative to calmodulin between the second and third helix (Figure 4B). Since parvalbumin is highly homologous to CaM, the starting model and TALOS predictions agree better with one another than in the case of recoverin, and for only 12 residues are the dihedral restraints derived from TALOS instead of from

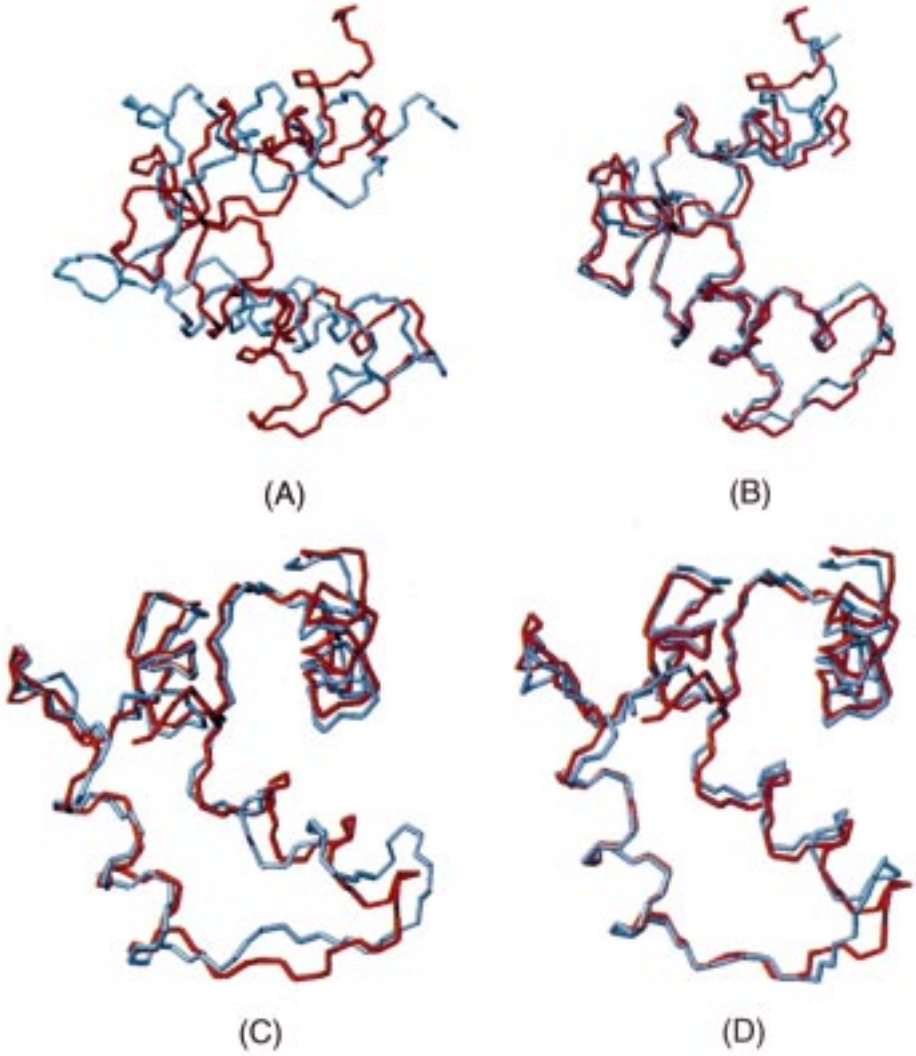

Figure 5. Structure comparisons before and after dipolar refinement of homology models. (A) The backbone of the starting structure for CaM-C (blue), constructed using GeneMine-based homology modeling starting from the structure of apo-recoverin (see text); backbone rmsd from the crystal structure of Ca^{2+}-ligated CaM-C (red) is 5 Å. (B) Superposition of X-ray and model obtained after refinement with dipolar couplings (backbone rmsd 1.25 Å). (C,D) Backbone of the parvalbumin-based homology model of Ca^{2+}-ligated CaM-C (blue) superimposed on Cam-C X-ray structure, (C) before and (D) after refinement with dipolar couplings (backbone rmsd decreases from 1.25 Å to 0.93 Å).

the starting model (Figure 4B). After the second stage of simulated annealing, the structure with the lowest dipolar energy differs from the crystal structure by 0.93 Å (Figure 5D). Although the improvement upon refinement of the parvalbumin-derived homology model, evaluated in backbone rmsd, is relatively small (1.25 to 0.93 Å), visually the improvement is more striking.

In real applications of the above described protocol, the true structure is not known, but the agreement between the experimental dipolar couplings and the final structure provides an indicator for the local quality of the structure. Figure 6 plots the rmsd (averaged over the normalized D_{NH}, $D_{C'C\alpha}$, and $D_{C\alpha H\alpha}$ couplings) between experimental and structure-predicted dipolar couplings. For example, for the structure of the N-terminal domain of apo-CaM, dipolar couplings for residues T28, T29, Q41 and V55 agree somewhat less with experimental data than the rest. This suggests that either the protocol has not managed to find the true minimum in conformational space, or previously identified extensive local backbone mobility for these residues makes it impossible to fit these dipolar couplings by a single structure. For the

recoverin- and parvalbumin-derived CaM C-domain structures, agreement is slightly higher, but there also remain individual residues with poor fits to the dipolar data. For several of these (H107, G113, T117), the rmsd between the dipolar couplings and the 1.68-Å X-ray structure of *Paramecium-Tetraurelia* CaM (Ban et al., 1994; PDB entry1OSA) is also rather large (>4 Hz), suggesting that the backbone structure of human calmodulin differs somewhat from that of *Paramecium-Tetraurelia* CaM (84% sequence identity) in the crystalline state. Remarkably, however, the fit between dipolar couplings and the X-ray structure of mammalian CaM (Babu et al., 1988) is much poorer, presumably due to its lower resolution (2.2 Å).

Concluding remarks

In recent years, the size of the Protein Data Bank has grown at a rapid pace. As a result, for an increasing fraction of proteins under study there are already homologous structures in the PDB. In other cases, work may focus on domain reorientation in a protein whose structure has already been solved under different conditions. Our study shows that, by using dipolar couplings, it is possible to obtain a reasonably accurate structure if the starting and target structures have similar folds, even if they differ significantly in their atomic coordinates, as in the case of apo- and Ca^{2+}-ligated CaM.

There are numerous sophisticated methods to search a sequence database for potentially homologous structures. In addition, comparison of the chemical shifts for such a 'hit' with the experimental shifts in the protein under study can be used to evaluate the degree of structural similarity. If a significant fraction of secondary structural elements adopt sufficiently similar relative orientations in the PDB protein and the one under study, the goodness of the fit between experimental dipolar couplings and the PDB protein may be used as a criterion for identifying structural homology. We have shown here that these structures subsequently can be refined by dipolar coupling restraints.

It is important to note that our refinement procedure only will be successful if the fold of the starting model is correct. Since any rigid segment for which dipolar couplings are measured in only a single alignment medium has a fourfold orientation degeneracy, refinement of an incorrect starting model could result in a wrong structure. Note, however, that if the set of measured dipolar couplings is relatively complete,

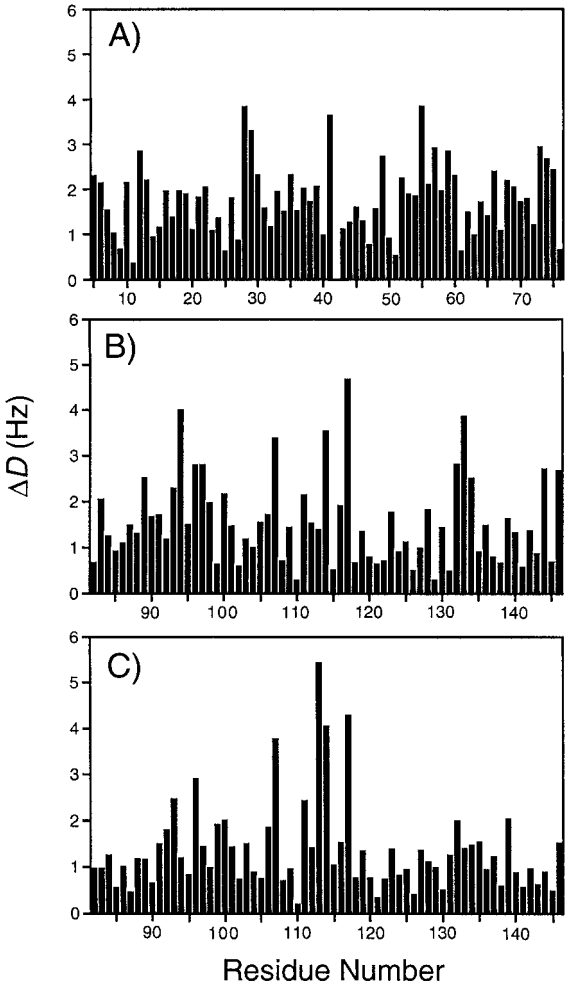

Figure 6. Residue-specific average difference, ΔD, between the measured dipolar couplings and those predicted by the structure, after optimizing the magnitude and orientation of the alignment tensor (Losonczi et al., 1999) for (A) the final structure of the N-terminal domain of apo-CaM, derived by starting from the Ca^{2+}-ligated N-terminal domain, (B) the structure of the Ca^{2+}-ligated C-terminal domain of CaM derived from apo-recoverin, and (C) the structure of the Ca^{2+}-ligated C-terminal domain of CaM obtained when using the homology model derived from Ca^{2+}-saturated parvalbumin as a starting structure. $\Delta D = \{[(\Delta D_{NH})^2+(\Delta D_{CAHA})^2+(\Delta D_{C'CA})^2]/3\}^{1/2}$ represents the normalized root-mean-square difference between measured and best-fitted couplings, where D_{CAHA} and $D_{C'CA}$ have been normalized relative to D_{NH} by multiplying the experimental coupling by 0.50 and 5.0, respectively (Ottiger and Bax, 1998).

with at least three dipolar couplings for the vast majority of residues, incorrect structures are expected to show local regions where the fit between the dipolar coupling and the refined model is poor. Incorrectly refined models typically can also be recognized by an unusually large fraction of buried charged residues. On average, sequence similarity greater than 50% invariably indicates similar global folds, and therefore should make use of the dipolar coupling refinement method relatively straightforward. Of course, in many cases, proteins with unrelated sequences can also have very similar folds, but the only way to identify the possible structural similarity is a match between the structure in the PDB and measured dipolar couplings and chemical shifts (Aitio et al., 1999; Annila et al., 1999; Meiler et al., 2000).

In our study, we have used three dipolar couplings per residue. This appears to be about the minimum number needed for successful refinement using the above approach; with an average of less than two parameters per pair of backbone ϕ/ψ angles, the local structure becomes under-determined. However, when entire secondary structure elements are kept fixed during refinement, it may be feasible to conduct calculations analogous to those used in the present study, but with only a single type of backbone dipolar coupling. In this respect, it is interesting to note that N-H bond vectors in a regular, straight α-helix are not quite parallel to the helix axis, but point away from the axis by about 15° (Marassi and Opella, 2000; Wang et al., 2000). This strongly reduces the degeneracy in the orientation of the helix relative to the alignment tensor.

Finally, as an independent method for evaluating the correctness of the calculated structure, its experimentally determined alignment tensor may be compared to the tensor predicted on the basis of the calculated structure. If dipolar couplings are measured in nearly neutral bicelles, the shape of the protein structure is sufficient for defining its alignment tensor (Zweckstetter and Bax, 2000). For measurements in charged liquid crystalline media such as filamentous phage, similar predictions can be made, provided electrostatic interactions are taken into account (M. Zweckstetter, unpublished). If the structure is accurate, the measured magnitude, rhombicity and orientation of the alignment tensor must agree with those predicted by the structure.

Acknowledgements

We thank Claude Klee for continuous encouragement and the labeled double mutant calmodulin sample, and Marius Clore, Gabriel Cornilescu, and Nico Tjandra for useful discussions.

References

Aitio, H., Annila, A., Heikkinen, S., Thulin, E., Drakenberg, T. and Kilpelainen, I. (1999) *Protein Sci.*, **8**, 2580–2588.

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

Babu, Y.S., Bugg, C.E. and Cook, W.J. (1988) *J. Mol. Biol.*, **204**, 191–204.

Ban, C., Ramakrishnan, B., Ling, K.Y., Kung, C. and Sundaralingam, M. (1994) *Acta Crystallogr.*, **D50**, 50–63.

Brünger, A.T. (1993) *XPLOR: A System for X-ray Crystallography and NMR*, v. 3.1, Yale University Press, New Haven, CT.

Chothia, C. (1992) *Nature*, **357**, 543–544.

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

Cordier, F. and Grzesiek, S. (1999) *J. Am. Chem. Soc.*, **121**, 1601–1602.

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

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

Eisenstein, E., Gilliland, G.L., Herzberg, O., Moult, J., Orban, J., Poljak, R.J., Banerjei, L., Richardson, D. and Howard, A.J. (2000) *Curr. Opin. Biotechnol.*, **11**, 25–30.

Elshorst, B., Hennig, M., Forsterling, H., Diener, A., Maurer, M., Schulte, P., Schwalbe, H., Griesinger, C., Krebs, J., Schmid, H., Vorherr, T. and Carafoli, E. (1999) *Biochemistry*, **38**, 12320–12332.

Finn, B.E., Evenas, J., Drakenberg, T., Waltho, J.P., Thulin, E. and Forsen, S. (1995) *Nat. Struct. Biol.*, **2**, 777–783.

Grzesiek, S. and Bax, A. (1992) *J. Magn. Reson.*, **96**, 432–440.

Ikura, M., Clore, G.M., Gronenborn, A.M., Zhu, G., Klee, C.B. and Bax, A. (1992) *Science*, **256**, 632–638.

Ikura, M., Kay, L.E. and Bax, A. (1990) *Biochemistry*, **29**, 4659–4667.

Klee, C.B. (1988) In *Molecular Aspects of Cellular Regulation* (Cohen, P. and Klee, C.B., Eds.), Elsevier, Amsterdam, pp. 35–56.

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

Kretsinger, R.H. and Nockolds, C.E. (1973) *J. Biol. Chem.*, **248**, 3313–3326.

Kuboniwa, H., Tjandra, N., Grzesiek, S., Ren, H., Klee, C.B. and Bax, A. (1995) *Nat. Struct. Biol.*, **2**, 768–776.

Kuszewski, J., Gronenborn, A.M. and Clore, G.M. (1997) *J. Magn. Reson.*, **125**, 171–177.

Kuszewski, J., Gronenborn, A.M. and Clore, G.M. (1999) *J. Am. Chem. Soc.*, **121**, 2337–2338.

Levitt, M. (1992) *J. Mol. Biol.*, **226**, 507–533.

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

Marassi, F.M. and Opella, S.J. (2000) *J. Magn. Reson.*, **144**, 150–155.

Maune, J.F., Beckingham, K., Martin, S.R. and Bayley, P.M. (1992) *Biochemistry*, **31**, 7779–7786.

Meador, W.E., Means, A.R. and Quiocho, F.A. (1992) *Science*, **257**, 1251–1255.
Meador, W.E., Means, A.R. and Quiocho, F.A. (1993) *Science*, **262**, 1718–1721.
Meiler, J., Peti, W. and Griesinger, C. (2000) *J. Biomol. NMR*, **17**, 283–294.
Ottiger, M. and Bax, A. (1998) *J. Am. Chem. Soc.*, **120**, 12334–12341.
Ottiger, M., Delaglio, F. and Bax, A. (1998) *J. Magn. Reson.*, **131**, 373–378.
Tanaka, T., Ames, J.B., Harvey, T.S., Stryer, L. and Ikura, M. (1995) *Nature*, **376**, 444–447.
Tjandra, N. and Bax, A. (1997a) *Science*, **278**, 1111–1114.
Tjandra, N. and Bax, A. (1997b) *J. Am. Chem. Soc.*, **119**, 9576–9577.
Tjandra, N., Garrett, D.S., Gronenborn, A.M., Bax, A. and Clore, G.M. (1997) *Nat. Struct. Biol.*, **4**, 443–449.
Tjandra, N., Kuboniwa, H., Ren, H. and Bax, A. (1995) *Eur. J. Biochem.*, **230**, 1014–1024.
Tufty, R.M. and Kretsinger, R.H. (1975) *Science*, **187**, 167–169.
Wang, J., Denny, J., Tian, C., Kim, S., Mo, Y., Kovacs, F., Song, Z., Nishimura, K., Gan, Z., Fu, R., Quine, J.R. and Cross, T.A. (2000) *J. Magn. Reson.*, **144**, 162–167.
Zhang, M., Tanaka, T. and Ikura, M. (1995) *Nat. Struct. Biol.*, **2**, 758–767.
Zweckstetter, M. and Bax, A. (2000) *J. Am. Chem. Soc.*, **122**, 3791–3792.