Journal of Biomolecular NMR, **25:** 63–71, 2003.
KLUWER/ESCOM

Determination of protein global folds using backbone residual dipolar coupling and long-range NOE restraints

Alexander W. Giesen[a], Steve W. Homans[a,*] & Jonathan Miles Brown[b]
[a]*Astbury Centre for Structural Molecular Biology, School of Biochemistry and Molecular Biology University of Leeds, LS2 9JT U.K.;* [b]*ProSpect Pharma, 6480 Dobbin Road, Columbia, MD 21045, U.S.A.*

Received 25 September 2002; Accepted 31 October 2002

Key words: global fold, NMR, residual dipolar couplings, ubiquitin

Abstract

We report the determination of the global fold of human ubiquitin using protein backbone NMR residual dipolar coupling and long-range nuclear Overhauser effect (NOE) data as conformational restraints. Specifically, by use of a maximum of three backbone residual dipolar couplings per residue (N_i-H_i^N, N_i-C'_{i-1}, H_i^N - C'_{i-1}) in two tensor frames and only backbone H^N-H^N NOEs, a global fold of ubiquitin can be derived with a backbone root-mean-square deviation of 1.4 Å with respect to the crystal structure. This degree of accuracy is more than adequate for use in databases of structural motifs, and suggests a general approach for the determination of protein global folds using conformational restraints derived only from backbone atoms.

Abbreviations: NOE – nuclear Overhauser effect; rmsd – root mean square deviation; ppm – parts per million.

Introduction

Residual dipolar couplings are now well-established as valuable conformational restraints in the determination of the solution structures of proteins via high resolution multinuclear NMR (Tolman et al., 1995; Tjandra et al., 1996; Bax and Tjandra, 1997; Tjandra and Bax, 1997). The introduction of a number of lyotropic dilute liquid-crystalline solutions for weak macromolecular alignment has enabled straightforward measurement of these couplings for a variety of macromolecules (Bax and Tjandra, 1997; Clore et al., 1998c; Hansen et al., 1998; Kiddle and Homans, 1998; Losonczi and Prestegard, 1998; Prosser et al., 1998; Wang et al., 1998a; Ottiger and Bax, 1999; Fleming et al., 2000; Rückert and Otting, 2000). Recently, a great deal of interest has arisen concerning the possibilities for the rapid determination of protein global folds based on a restricted subset of experimental restraints (Mal et al., 1998; Bonvin et al., 2001). With particular reference to the application of residual dipolar couplings for this purpose, Mueller et al. have developed a methodology for orienting peptide planes using dipolar couplings which was utilised to determine the global fold of maltose binding protein in complex with β-cyclodextrin. This gave rise to pairwise rmsd values between N- and C-terminal domains of the NMR structure and the corresponding regions in the X-ray structure of 2.8 Å and 3.1 Å, respectively (Mueller et al., 2000a, 2000b). Fowler et al. (2000) have utilised N_i - H_i^N , H_i^N - H_i^α, H_i^N - $H_{i\pm1}^\alpha$, H_i^N - H_{i+1}^N residual dipolar couplings together with a small number of backbone-sidechain NOEs to determine the backbone fold of acyl carrier protein to an rmsd between backbone atoms of $\sim$3 Å. Moreover, Hus et al. (2000) have utilised long-range order restraints available from paramagnetic systems in combination with residual dipolar couplings to define the fold of cytochrome c′ in the complete absence of NOE restraints. Recently, this same group has determined the global fold of ubiquitin to 1.0 Å backbone rmsd (residues 1-71) (Hus et al., 2001) with respect to the solution structure determined by conventional methods, using restraints derived solely from N_i - H_i^N, C'_{i-1}

*To whom correspondence should be addressed. E-mail: s.w.homans@leeds.ac.uk

- N_i, C'_{i-1} - H^N_i, C^α_i - C'_i, C^α - H^α and C^α - C^β residual dipolar couplings in two independent tensor frames. Excellent results have also been obtained by use of residual dipolar couplings in concert with molecular fragment replacement (Bowers et al., 2000; Delaglio et al., 2000; Andrec et al., 2001; Rohl and Baker, 2002). Interesting recent developments include single-step determination of protein sub-structures (Zweckstetter and Bax, 2001) from residual dipolar couplings and simultaneous resonance assignment (Tian et al., 2001). The high level of accuracy obtained with these methods suggests that the global fold of a protein can be determined at a lower, though still useful level of accuracy by use of a smaller number of residual dipolar coupling restraints. In particular, global fold determination without C^α-H^α dipolar couplings is highly desirable for the structure determination of larger proteins, where perdeuteratation is usually required in order to overcome the efficient relaxation of C^α nuclei by attached H^α protons. Unfortunately, the C^α-H^α dipolar coupling is a crucial restraint in most structure determination protocols since it defines chirality at C^α. Here, we develop a protocol for the global fold determination of proteins that overcomes this limitation, while still providing good accuracy in global fold determination for the protein ubiquitin.

Materials and methods

Experimental data

Experimental residual dipolar coupling data in two tensor frames were taken directly from Ottiger and Bax (1998). H^N - H^N NOEs restraints were obtained from a three-dimensional 500 MHz NOESY-HSQC spectrum of uniformly ^{13}C,^{15}N,^{2}H-enriched ubiquitin. The protein concentration was 1 mM at pH 7.0 in 50 mM phosphate buffer, at a probe temperature of 300 K. A mixing time of 1.5 s was used to emphasize NOE connectivities in the 4–6 Å range (Mal et al., 1998). Chemical shift data were taken from Wang et al. (1995).

Derivation of tensor frame orientations

The relative orientations of the tensor frames were calculated using a simulated annealing approach. Each of two idealised α-helices ($\phi = -57°$, $\varphi = -47°$) representing fragments K29-E34 (fragment 1) and N25-I30 (fragment 2) of the long ubiquitin helix were used as starting structures for molecular dynamics simulation. Experimentally obtained one-bond N^i-H^N_i and C^α_i-H^α_i dipolar couplings from two different alignment media (Ottiger and Bax, 1998) and simulated H^N-H^N NOE-data were used as restraints in XPLOR (Brünger, 1987) simulated annealing refinement protocols. The first alignment tensor was fixed in this protocol, while the second tensor and the helical fragment were allowed to reorient in the course of the calculation. By altering the structure and reorientation of the second tensor the H^N-H^N-NOEs and the dipolar coupling restraints were satisfied. A high degree of convergence of the resulting orientation of the second order tensor principal axis system relative to first system and the helical fragment was observed. In total 379 and 1568 structures and tensor frame orientations were calculated for the helical fragment 1 and fragment 2, respectively. Using straightforward geometry and linear algebra, rotation matrices describing the relative orientation of the two tensor frames were calculated. Euler-angles obtained from these rotation-matrices were grouped according to well-known inversion properties of dipolar reference frames (Fowler et al., 2000) and used for the calculation of averaged order tensor orientations. XPLOR rigid body minimization protocols and ORDERTEN_SVD (Losonczi et al., 1999) calculations were employed to check for consistency of the obtained structure ensemble and averaged tensor frames with the two sets of residual dipolar couplings. One of the averaged tensor frames was used in subsequent calculations of the ubiquitin structure. The orientation of the second tensor frame relative to the first frame is given by three Euler angles: 165°, 171° and 300°.

Determination of φ, ψ values for residue pairs

Three dimensional φ, ψ potential surfaces were calculated using XPLOR version 3.851. First, an extended structure for ubiquitin was generated by setting all φ, ψ angles to 180° (with the exception of φ for prolines). Groups comprised of all atoms of residue i, C′ and O atoms of residue i-1, and N, H^N and C^α atoms of residue i+1 were then considered stepwise from the COOH terminus. The values of φ and ψ for residue i were each varied independently through 360° in 15° increments, giving a two dimensional grid of 576 points. At each point, a rigid body minimisation was performed on the fragment, in order to minimise the difference between experimental and theoretical residual dipolar couplings (Tjandra et al., 1997) N_i-

H^{Ni}, N_i-C'_{i-1}, H^N_i - C'_{i-1}, N_{i+1}-H^N_{i+1}, N_{i+1}- C'_i, H^N_{i+1} - C'_i, C^α_i - C'_i and C^α_i - H^α_i (or N_i-H^N_i, N_i-C'_{i-1}, H^N_i - C'_{i-1}, N_{i+1}-H^N_{i+1}, N_{i+1}-C'_i, H^N_{i+1} - C'_i where a reduced set of dipolar coupling was utilised) with respect to two sets of external cartesian axes whose relative orientation is defined by the two tensor frame orientations determined above. In addition a van der Waals repulsion term was included at each grid point to account for steric clashes involving the C^β atom of residue i, together with a C^α chemical shift database potential (Kuszewski et al., 1995) with a force constant of 0.5 kcal mol^{-1} ppm^{-2}. In order to overcome local minima during the minimisation procedure, the latter was performed ten times at each grid point, starting with randomised values of the three Euler angles that describe the orientation of the fragment in the tensor frames. The axial A_a component and rhombicity R in each tensor frame was taken from Ottiger and Bax (1998). Force constants of 0.5, 0.247, 1.61,4.16 and 2.5 kcal mol^{-1} Hz^{-2} were utilised for N - H^N, C^α - H^α, H^N - C', N – C' and C^α – C' residual dipolar couplings, respectively. The resulting potential surfaces were contoured using Gnuplot 3.8. An initial coordinate set for ubiquitin was obtained first by translational rmsd fitting of global minimum energy atomic positions of overlapping atoms in neighbouring fragments , i.e. atoms C'_i, C^α_{i+1}, N_{i+1} and H^N_{i+1}. The fit for residues exhibiting an rmsd > 0.1 was then optimised by fitting the next lowest energy minimum coordinate set on the potential surface, and so on until an rmsd < 0.1 was obtained, where possible. XPLOR scripts for the above computations are available from the authors on request.

Refinement of initial structure

The initial structure obtained using the procedure described in the previous paragraph was refined by application of H^N - H^N NOE restraints, dihedral restraints and residual dipolar coupling restraints using the standard XPLOR simulated annealing script (sa.inp) (Brünger, 1987) with the following differences. Restraints corresponding to values of φ and ψ for residues with an rmsd fit < 0.1 as described above, were applied as standard biharmonic dihedral restraints. An initial force constant of 200.0 kcal mol^{-1} rad^{-2} was utilised reducing to 0.1 kcal mol^{-1} rad^{-2} at the end of each simulated annealing run. Simultaneously, residual dipolar coupling restraints were applied with an initial force constant of 0.1 kcal mol^{-1} Hz^{-2}, rising to 1.0 kcal mol^{-1} Hz^{-2} at the end of each simulated annealing run. Both dihedral angle and dipolar restraints were deleted from the database for those residues whose rmsd fit could not be optimised below 0.1 Å. A total of 12000 molecular dynamics steps of 5.0 fs were computed at a temperature of 1500 K, followed by 6000 cooling steps of 5.0 fs to a final temperature of 100 K. Twenty structures were calculated from an initial starting structure derived from the protocol described in the previous paragraph.

Results

Residual dipolar coupling data for a given dipeptide fragment in a protein, depend on five parameters, namely the backbone torsion angles φ and ψ and the three Euler angles representing the orientation of the fragment in the frame of the alignment tensor (assuming a fixed orientation of the peptide plane). Moreover, it is well-known that in the presence of two tensors, the orientation of a chiral motif is completely defined (Ramirez and Bax, 1998; Wang et al., 1998b; Al-Hashimi et al., 2000; Hus et al., 2001) and thus it should be possible to determine the global fold of a protein using a minimum of five residual dipolar couplings per dipeptide fragment. In the present work we have chosen to develop a methodology for global fold determination using N_i-H^N_i, N_i-C'_{i-1}, H^N_i - C'_{i-1}, C^α_i - C'_i and C^α_i - H^α_i residual dipolar couplings, using data reported by Ottiger and Bax (1998). Subsequently, we examine the accuracy of global fold determination that can be achieved with a reduced set of residual dipolar couplings that typically can be measured in larger proteins. Although the tensor frames for ubiquitin in two orienting media were defined in the work of Ottiger and Bax (1998), in the general case of protein fold determination these will not be known. We therefore sought to discover whether these tensors could be derived with sufficient accuracy from a limited set of residual dipolar couplings.

Tensor orientations are readily available if the structure of a molecular fragment is known. In the absence of a structure, models obtained from idealised secondary structure elements may be used instead. This approach has been recently utilised by Fowler et al. (2000). However, these authors noted that large errors for the dipolar couplings have to be allowed, mainly to account for deviation of the actual structure of the molecular fragment from the idealised structure. In unfavourable cases it is not possible to obtain

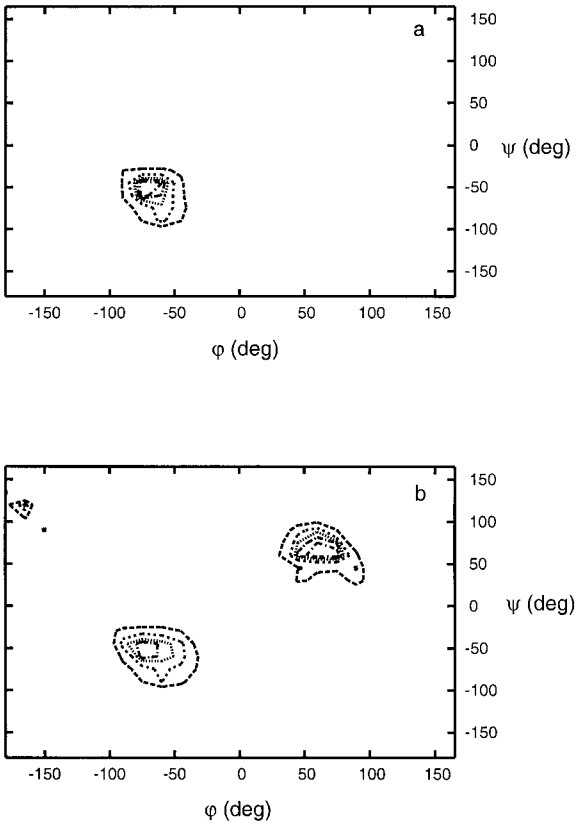

Figure 1. Typical plots of energy E ($E\ \alpha\Sigma(D_{theor}\text{-}D_{exp})^2 + E_{van\ der\ Waals} + E_{C\alpha-shift}$) as a function of the torsion angles φ and ψ of Val 26 in human ubiquitin using (a) five residual dipolar couplings $N_i\text{-}H_i^N$, $N_i\text{-}C'_{i-1}$, H_i^N - C'_{i-1}, C_i^α - C'_i and C_i^α - H_i^α for each residue in the protein; (b) three residual dipolar couplings $N_i\text{-}H_i^N$, $N_i\text{-}C'_{i-1}$, H_i^N - C'_{i-1} for each residue in the protein. Contours are plotted at 1,2,3,4, and 5 kcal/mol above the global minimum.

any order tensor orientation from idealised structures. It is obvious that in these cases there is a significant deviation of the rigid structure used as a model and the actual structure and dynamics of the molecular fragment. In order to work around this problem we used a simulated annealing approach that allowed us simultaneously to refine an idealised starting structure and to obtain the relative orientation of the order tensor frames. We succeeded in obtaining the relative orientation of two tensor frames using only $N_i\text{-}H_i^N$ and $C_i^\alpha\text{-}H_i^\alpha$ residual dipolar one-bond couplings from two different alignment media and an idealised α-helix as starting structure. These particular dipolar couplings were chosen for convenience, and in general any combination of dipolar couplings could be used to determine tensor frame orientations.

Structure determination using $N_i\text{-}H_i^N$, $N_i\text{-}C'_{i-1}$, H_i^N - C'_{i-1}, C_i^α - C'_i and C_i^α - H_i^α residual dipolar couplings

Given these tensor parameters, the first stage in our approach to global fold determination involves the derivation of φ, ψ values for each dipeptide fragment that are compatible with the measured dipolar couplings in each tensor frame. By analogy with the work of Wang et al. (1998b), this can be achieved by use of a grid-search over φ and ψ, while simultaneously optimising the orientation of the dipeptide fragment relative to the two alignment tensors, that gives the best fit between predicted and measured couplings (see methods). This can be achieved by incorporating residual dipolar couplings as pseudo-energy terms (Tolman et al., 1997; Clore et al., 1998a, b, 1999) in a conventional energy minimisation protocol. In addition, we included a van der Waals repulsion term that includes all backbone and C^β atoms of each residue, together with a C^α chemical shift database potential (Spera and Bax, 1991; Kuszewski et al., 1995; Beger and Bolton, 1997). The resulting data can conveniently be analysed as three dimensional potential energy surfaces of φ vs. ψ vs. energy. Selected contour maps derived from such a procedure are shown in Figure 1. In most cases a single pair of φ, ψ values is obtained that is compatible with the measured couplings. In addition to defining the correct φ, ψ values for each dipeptide fragment, the coordinates for each dipeptide fragment at the global minimum describe the orientation of the fragment with respect to the principal axes of the alignment tensors, and hence define the correct orientation of each fragment with respect to other fragments in the molecule. However, the translational position of the dipeptide fragment is arbitrary.

In the second stage of our approach, an initial structure can thus be generated by translational (but not rotational) least-squares fit of the overlapping regions of adjacent dipeptide fragments using the global minimum energy coordinate sets for each fragment. The resulting translational rmsd vs. residue number is shown in Figure 2a. In circumstances where two or more minima with low energies are observed on the potential surface, which can arise for example from coincidences between bond vector and tensor orientations, limitations in the accuracy of experimental residual dipolar couplings, or lack of chirality in the case of glycine residues, the global minimum energy coordinate set is not necessarily the 'correct' configuration as evidenced by a high rmsd fit. In order to find the correct minimum, the coordinate set at each local

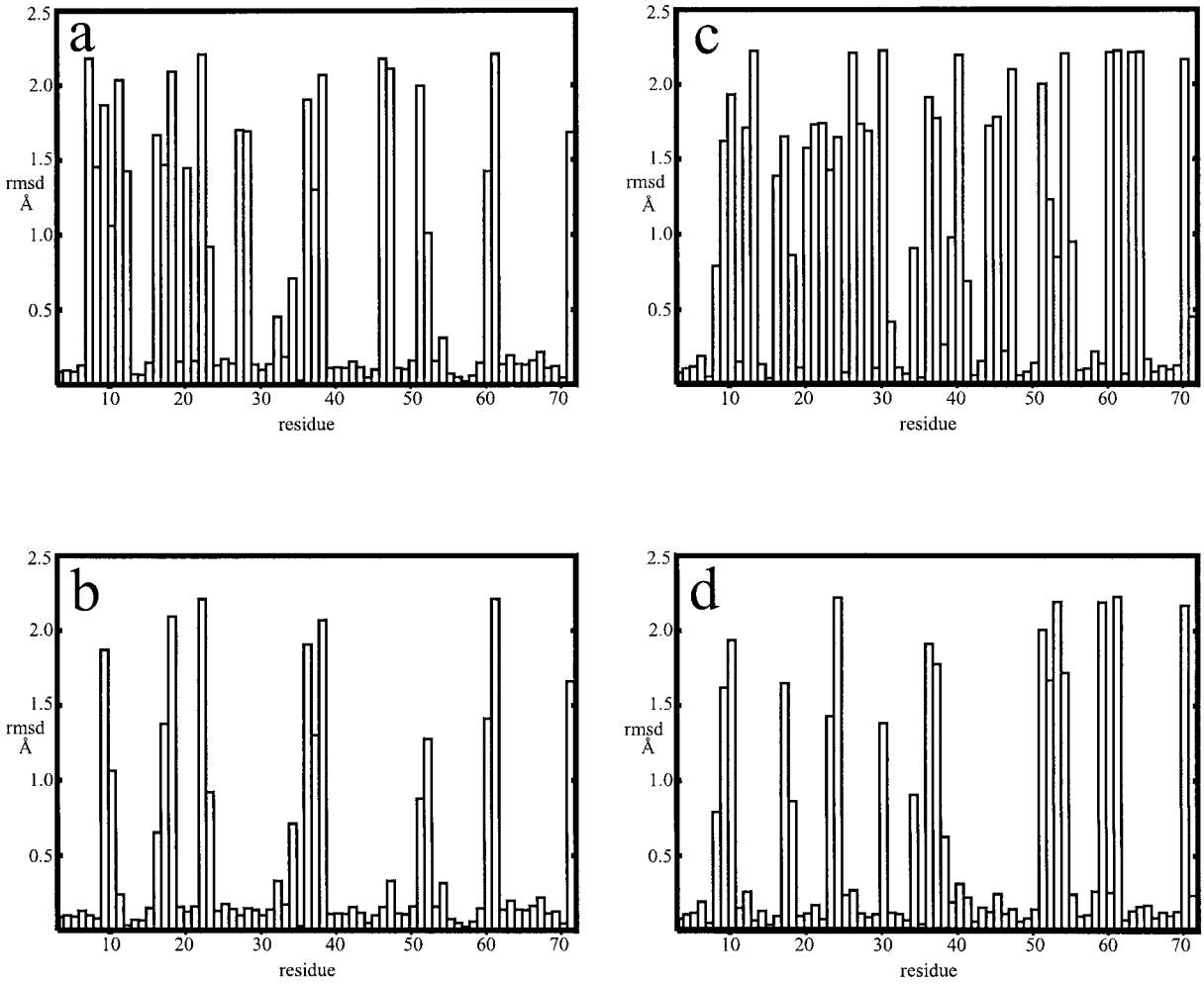

Figure 2. (a) Plots of translational rmsd fit of the C'_i, C^{α}_{i+1}, N_{i+1} and H^N_{i+1} atoms of residues i and i+1 versus residue number using (left panels) five residual dipolar couplings N_i-H^N_i, N_i-C'_{i-1}, H^N_i - C'_{i-1}, C^{α}_i - C'_i and C^{α}_i - H^{α}_i and (right panels) three residual dipolar couplings N_i-H^N_i, N_i-C'_{i-1}, H^N_i - C'_{i-1} for each residue: (a,c) Initial fit obtained from the global minimum energy coordinates for each fragment; (b,d) Optimised fit considering local energy minimum coordinates for each fragment.

minimum is in turn (lowest energy first) subjected to translational least-squares fitting with respect to the adjacent dipeptide pair in the sequence. The coordinate set with the best rmsd fit is then selected, resulting in an optimised rmsd fit over the whole protein, as shown in Figure 2b. It should be noted that a low rmsd fit could not be obtained for certain residues, particularly those in loop regions of the protein. This result derives principally from the approximation that a single order parameter can be used for all residues, which is clearly not valid for regions that exhibit significant internal motion.

In the third and final stage of the procedure, the initial structure generated in the second stage is refined by use of conventional dynamical simulated annealing, using a combination of dihedral angle, dipolar, and H^N-H^N NOE restraints. The NOE restraints applied are exclusively long-range in nature. In the present study these were derived from a three-dimensional NOESY-HSQC spectrum (not shown) of uniformly ^{13}C,^{15}N,^{2}H ubiquitin with a mixing time of 1.5 s. Thirty-seven long-range NOEs could readily be measured in this spectrum, corresponding to analogous distances in the crystal structure of up to 5.4 Å (Table 1). In the initial stages of the simulated annealing protocol, a strong dihedral angle force constant is utilised, which is slowly decreased to zero with concomitant increase in the residual dipolar coupling

Table 1. Long-range H_i^N – H_j^N NOE connectivities determined in uniformly ^{13}C,^{15}N,^{2}H-enriched ubiquitin

Residue i	Residue j	NOE Intensity
Ile 3	Leu 15	medium
Ile 3	Glu 16	very weak
Ile 3	Val 17	weak
Ile 3	Glu 64	weak
Phe 4	Leu 15	very weak
Phe 4	Glu 64	very weak
Phe 4	Ser 65	weak
Phe 4	Leu 67	weak
Val 5	Ile 13	medium
Val 5	Leu 15	weak
Val 5	Leu 67	weak
Lys 6	Ile 13	weak
Lys 6	Leu 67	weak
Lys 6	Leu 69	weak
Thr 7	Ile 13	weak
Thr 7	Leu 69	weak
Glu 18	Ser 20	weak
Glu 18	Asp 21	weak
Ser 20	Leu 56	very weak
Ser 20	Ser 57	weak
Asp 21	Ser 57	weak
Ile 23	Asn 25	weak
Ile 23	Val 26	weak
Ile 23	Arg 54	weak
Ile 23	Thr 55	weak
Arg 42	Val 70	medium
Arg 42	Leu 71	weak
Leu 43	Leu 50	weak
Leu 43	Val 70	very weak
Ile 44	Leu 50	weak
Ile 44	His 68	medium
Ile 44	Leu 69	weak
Phe 45	Lys 48	medium
Phe 45	Leu 50	weak
Phe 45	His 68	weak
Glu 51	Arg 54	weak
Asp 52	Arg 54	weak

force constants (see Methods section). Empirically, we restrict the residual dipolar coupling and dihedral angle restraints database to those residues whose translational rmsd fit is less than 0.1 Å. Since the initial structure comprises dipeptide fragments whose dihedral angles and orientations are by definition consistent with the dipolar couplings restraints thus selected,

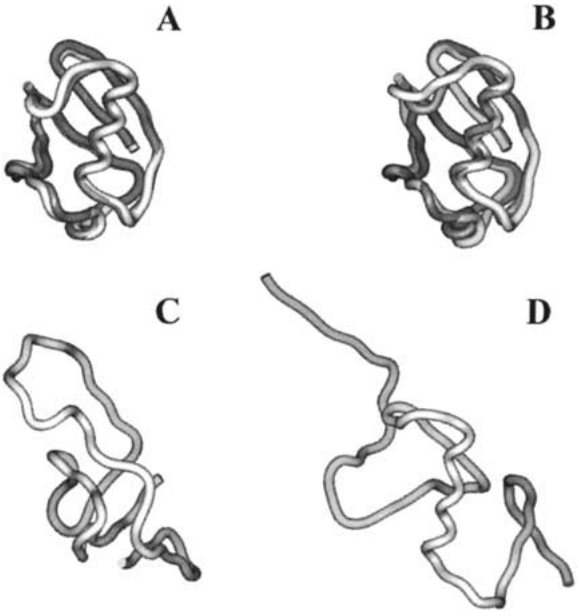

Figure 3. Backbone representations of the crystal structure (dark face) of ubiquitin, and the global fold obtained from (A) five residual dipolar coupling-derived dihedral restraints (N_i-H_i^N, N_i-C'_{i-1}, H_i^N - C'_{i-1}, C_i^α - C'_i and C_i^α - H_i^α) per residue together with 37 long-range H^N - H^N NOE restraints (Table 1); (B) three residual dipolar coupling-derived dihedral restraints (N_i-H_i^N, N_i-C'_{i-1}, H_i^N - C'_{i-1}) per residue together with H^N - H^N NOE restraints; (C) 37 long-range H^N - H^N NOE restraints (Table 1); (D) three residual dipolar coupling-derived dihedral restraints (N_i-H_i^N, N_i-C′i − 1, H_i^N - C'_{i-1}) per residue. The average backbone rmsd from the crystal structure is 0.92 Å in (A) and 1.4 Å in (B).

good convergence is obtained using the conventional dipolar coupling potential described by Clore and coworkers (Clore et al., 1998b). It should be noted that good convergence of the dipolar energy is obtained in the absence of dihedral angle restraints, but empirically we find that the backbone rmsd of the resulting structure with respect to the crystal structure is slightly improved by inclusion of a dihedral restraint term. The lowest energy structure derived from this procedure has an average backbone rmsd with respect to the crystal structure of 0.92 Å (Figure 3a and Table 2).

Structure determination using N_i-H_i^N, N_i-C'_{i-1}, H_i^N - C'_{i-1} residual dipolar couplings

A major difficulty with the application of the above protocol to larger proteins ($>\sim$ 20 kDa) concerns the number of measurable residual dipolar couplings. In order to realise a suitably narrow C^α linewidth in such proteins, perdeuteration becomes mandatory (Grzesiek et al., 1993). This in turn prevents the mea-

Table 2. Structure quality for the global fold of ubiquitin determined from: (structure 1); N_i-H_i^N, N_i-C'_{i-1}, H_i^N - C'_{i-1}, C_i^α - C'_i and C_i^α - H_i^α residual dipolar couplings together with 37 long range H^N - H^N NOE restraints (Table 1): (structure 2); N_i-H_i^N, N_i-C'_{i-1}, H_i^N - C'_{i-1} residual dipolar couplings together with 37 long range H^N - H^N NOE restraints (Table 1): (structure 3); 37 long range H^N - H^N NOE restraints only: (structure 4); N_i-H_i^N, N_i-C'_{i-1}, H_i^N - C'_{i-1} residual dipolar couplings only

	Structure 1	Structure 2	Structure 3	Structure 4
Backbone rmsd of lowest energy structure (Å)[a]	0.92	1.4	9.11	13.58
Coordinate precision (Å)[b]	0.69	1.67	4.97	4.46
Dipolar R factor of lowest energy structure (%)[c]	0.45	3.00	–	2.5
PROCHECK[d] quality of lowest energy structure				
most favoured regions %	80	80	28.3	66.7
additional allowed regions %	16.7	20	38.3	28.3
NOE violations of lowest energy structure[e]	0	0	0	–

[a]The values given relate to residues 3-71 with respect to the crystal structure.
[b]Defined as the average rms difference between the 20 final simulated annealing structures and the mean coordinates. The values given relate to residues 3-71.
[c]R_{dip} values (Clore and Garrett, 1999) are reported for N_i-H_i^N dipolar couplings averaged over two tensor frames.
[d]Laskowski et al. (1993).
[e]NOE distance restraints were classified into three ranges: Medium (1.8–3.3 Å), weak (1.8–5.0 Å) and very weak (1.8–6.0 Å).

surement of C^α-H^α residual dipolar couplings, with consequent loss of chirality information, i.e., the φ, ψ potential surface becomes C_2 symmetric. The consequent increase in the number of minima on the potential surface is in principle, however, not a barrier to successful derivation of protein global folds using the protocol described above. Given that it is possible to measure the three residual dipolar couplings N_i-H_i^N, N_i-C'_{i-1} and H_i^N - C'_{i-1} simultaneously in a single experiment of the HSQC type (Wang et al., 1998b), we sought to determine whether these couplings are sufficient to define the global fold of ubiquitin to reasonable accuracy. The resulting initial and optimised rmsd fits are shown in Figures 2c,d. As anticipated, a smaller number of residues can be fitted compared to calculations involving five residual dipolar couplings per residue. This arises in major part from the fact that all three residual dipolar couplings were not available for a number of residues due to incomplete experimental data, together with the previously mentioned poor fits for residues in loop regions of the molecule.

By use of the same optimisation protocol described in the previous paragraph, but using only three residual dipolar couplings per residue, the lowest energy structure has an average backbone rmsd with respect to the crystal structure of 1.4 Å (Figure 3b and Table 2).

As control experiments, structures were also calculated using three dipolar couplings in two tensor frames as above but in the absence of NOE restraints, and vice-versa (Figures 3c,d and Table 2).

Discussion

The above results demonstrate that a combination of five residual dipolar coupling measurements per residue N_i-H_i^N, N_i-C'_{i-1}, H_i^N - C'_{i-1}, C_i^α - C'_i and C_i^α - H_i^α, together with long-range H^N - H^N NOEs is sufficient to resolve the global fold of ubiquitin to better than 1 Å backbone rmsd with respect to the crystal structure. Moreover, the backbone rmsd increases to only 1.4 Å when N_i-H_i^N, N_i-C'_{i-1}, H_i^N - C'_{i-1} residual dipolar couplings are utilised. This degree of accuracy is more than adequate for use in databases of structural motifs, and suggests a general approach for the determination of protein global folds using conformational restraints derived only from backbone atoms. This is a highly desirable goal since it obviates the need to undertake the time-consuming task of sidechain assignment. However, it is clear from the results of global fold determination using only dipolar or long-range NOE restraints (Table 2 and Figure 3c,d respectively), that both types of restraint are required in the present approach in order to obtain a global fold with useable accuracy. In future applications, one issue that is likely to be prevalent in predominantly α-helical proteins concerns the availability of the long-

range H^N - H^N NOE restraints. In proteins containing a high α-helical content, it is likely that such restraints will be sparse, and backbone-sidechain or sidechain-sidechain NOEs may be essential. However, a small number of such NOEs may be sufficient. For example, Prestegard and co-workers (Fowler et al., 2000) were able to derive the global fold of *E. coli* acyl carrier protein by use of residual dipolar coupling measurements together with one backbone-backbone and four backbone-sidechain NOEs. Similarly, these same authors determined the global fold of *Rhizobium leguminosarum* NodF protein with five backbone-backbone NOEs. Clearly, the use of NOEs involving sidechain atoms requires the assignment of the latter, which could in principle dramatically increase the time required for the derivation of a global fold. However, since the number of required NOEs would appear to be small, the effort required for assignment can be reduced by careful choice of isotopic labelling strategies based upon residue type, in a manner analogous to that introduced by Kay and co-workers (Rosen et al., 1996; Gardner and Kay, 1997; Gardner et al., 1997; Goto et al., 1999; Goto and Kay, 2000). Moreover, the recent demonstration that ^{1}H-^{1}H residual dipolar couplings can provide distance restraints of up to at 7 Å in ubiquitin (Wu and Bax, 2002), suggests an alternative source of long-range restraints for global fold determination.

Acknowledgements

This work was supported by BBSRC grant ref. 24/B12993, and is a contribution from the North of England Structural Biology Centre.

References

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

Andrec, M., Du, P. and Levy, R.M. (2001) *J. Biomol. NMR*, **21**, 335–347.

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

Beger, R.D. and Bolton, P.H. (1997) *J. Biomol. NMR*, **10**, 129–142.

Bonvin, A., Houben, K., Guenneugues, M., Kaptein, R. and Boelens, R. (2001) *J. Biomol. NMR*, **21**, 221–233.

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

Brünger, A.T. (1987) *XPLOR Version 3.1: A System for X-Ray Crystallography and NMR*, Yale University Press, New Haven, CT.

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

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

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

Clore, G.M., Starich, M.R. and Gronenborn, A.M. (1998c) *J. Am. Chem. Soc.*, **120**, 10571–10572.

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

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

Fleming, K., Gray, D., Prasannan, S. and Matthews, S. (2000) *J. Am. Chem. Soc.*, **122**, 5224–5225.

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

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

Gardner, K.H., Rosen, M.K. and Kay, L.E. (1997) *Biochemistry*, **36**, 1389–1401.

Goto, N.K. and Kay, L.E. (2000) *Curr. Opin. Struct. Biol.*, **10**, 585–592.

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

Grzesiek, S., Anglister, J., Ren, H. and Bax, A. (1993) *J. Am. Chem. Soc.*, **115**, 4369–4370.

Hansen, M.R., Mueller, L. and Pardi, A. (1998) *Nat. Struct. Biol.*, **5**, 1065–1074.

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

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

Kiddle, G.R. and Homans, S.W. (1998) *FEBS Lett.*, **436**, 128–130.

Kuszewski, J., Qin, J., Gronenborn, A.M. and Clore, G.M. (1995) *J. Magn Reson. B*, **106**, 92–96.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) *J. Appl. Crystallogr.*, **26**, 283–291.

Losonczi, J.A. and Prestegard, J.H. (1998) *J. Biomol. NMR*, **12**, 447–451.

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

Mal, T.K., Matthews, S.J., Kovacs, H., Campbell, I.D. and Boyd, J. (1998) *J. Biomol. NMR*, **12**, 259–276.

Mueller, G.A., Choy, W.Y., Skrynnikov, N.R. and Kay, L.E. (2000a) *J. Biomol. NMR*, **18**, 183–188.

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

Ottiger, M. and Bax, A. (1998) *J. Am. Chem. Soc.*, **120**, 12334–12341.

Ottiger, M. and Bax, A. (1999) *J. Biomol. NMR*, **13**, 187–191.

Prosser, R.S., Losonczi, J.A. and Shiyanovskaya, I.V. (1998) *J. Am. Chem. Soc.*, **120**, 11010–11011.

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

Rohl, C.A. and Baker, D. (2002) *J. Am. Chem. Soc.*, **124**, 2723–2729.

Rosen, M.K., Gardner, K.H., Willis, R.C., Parris, W.E., Pawson, T. and Kay, L.E. (1996) *J. Mol. Biol.*, **263**, 627–636.

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

Spera, S. and Bax, A. (1991) *J. Am. Chem. Soc.*, **113**, 5490–5492.

Tian, F., Valafar, H. and Prestegard, J.H. (2001) *J. Am. Chem. Soc.*, **123**, 11791–11796.

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

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

Tjandra, N., Omichinski, J.G., Gronenborn, A.M., Clore, G.M. and Bax, A. (1997) *Nat. Struct. Biol.*, **4**, 732–737.

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

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

Wang, A.C., Grzesiek, S., Tschudin, R., Lodi, P.J. and Bax, A. (1995) *J. Biomol. NMR*, **5**, 376–382.

Wang, H., Eberstadt, M., Olejniczak, E.T., Meadows, R.P. and Fesik, S.W. (1998a) *J. Biomol. NMR*, **12**, 443–446.

Wang, Y.X., Marquardt, J.L., Wingfield, P., Stahl, S.J., LeeHuang, S., Torchia, D. and Bax, A. (1998b) *J. Am. Chem. Soc.*, **120**, 7385–7386.

Wu, Z.R. and Bax, A. (2002) *J. Am. Chem. Soc.*, **124**, 9672–9673.

Zweckstetter, M. and Bax, A. (2001) *J. Am. Chem. Soc.*, **123**, 9490–9491.