Protein Structure Determination Using Molecular Fragment Replacement and NMR Dipolar Couplings

Frank Delaglio, Georg Kontaxis, and Ad Bax*

Laboratory of Chemical Physics
National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health
Bethesda, Maryland 20892

Received October 7, 1999
Revised Manuscript Received January 3, 2000

Determination of the three-dimensional (3D) structure of a protein in solution from NMR data has relied primarily on the measurement of a large number of interproton distances (NOEs), supplemented by torsion angle restraints derived from *J* couplings and chemical shifts.[1] Recently developed methods for measurement of dipolar couplings[2] provide additional structural information which can be used to improve the accuracy of the NMR-derived protein structure.[3] Here, we describe a novel approach for determining the backbone structure of a protein solely from dipolar couplings. The first stage of our method, which we refer to as molecular fragment replacement (MFR), is analogous to a method described by Kraulis and Jones for determining local fragment structures from NOE patterns[4] and also is similar to the commonly used database approach for fitting the main chain electron density of protein X-ray structures.[5] It also bears some similarity to a recently described approach for identifying the fold of a protein from its dipolar couplings by searching a database,[6] but this latter method requires a very similar structure to be present in the database, and therefore is not a de novo method.

The MFR method is demonstrated for the protein ubiquitin, for which a 1.8 Å X-ray crystal structure is available,[7] and which has been studied extensively by NMR.[8,9] The ordered part of its NMR structure (residues 2−72) is in excellent agreement with the X-ray structure, with a root-mean-square deviation (rmsd) of 0.35 Å.[10] Four backbone couplings ($N-H$; $C'-N$; $C'-H^N$; $C^\alpha-H^\alpha$) have previously been measured for most residues in ubiquitin, but less for Pro, residues preceding Pro, and residues with broadened or missing amide resonances.[11] Also, these couplings were measured in two different liquid crystalline phases, yielding information on the internuclear vector orientations relative to two different axis systems.[12]

In its present implementation, the MFR method uses a fragment size of 7 residues. For each fragment, the best fit between its set of measured dipolar couplings and each 7-residue fragment found in the Brookhaven Protein Data Bank (PDB) is determined by using a linear least-squares method.[13] To expedite this search, a reduced version of the PDB was created, containing only 1560 proteins, of which two-thirds are of a resolution of 2.2 Å or better. From this entire ensemble of 350 000 PDB fragments, 20 are selected on the basis of the lowest χ^2 between measured and best-fitted dipolar couplings and, to a much weaker degree, the χ^2 between experimental chemical shifts and those predicted for each PDB fragment using a structure/chemical shift database.[14] This procedure is repeated by shifting the 7-residue fragment by one residue at a time; that is, for an *N*-residue protein the search is carried out $N - 6$ times. An example of how well the dipolar data typically match those of the best fitting 7-residue PDB fragment is shown in the Supporting Information.

Ignoring the torsion angles of the first and last residue of each database fragment, the overlapping collection of 7×20 best-fitting fragments provides 5×20 pairs of ϕ and ψ angles at each residue of the query protein (less for the first and last five residues of the protein). In favorable cases, as typically found near the center of α-helices, all "hits" for a given residue cluster in the same region of the Ramachandran map. Frequently, however, there will be outliers, as the dipolar couplings may not define uniquely the conformation of each individual 7-residue stretch (see below) so that more than one type of 7-residue peptide conformation in the database matches the experimental dipolar couplings. Empirically, we find that if the largest cluster is more than twice as large as the next largest cluster, the median ϕ and ψ angles of the largest cluster provide reliable estimates for the ϕ and ψ angles of the query residue (rmsd of 11° relative to angles measured from the ubiquitin X-ray structure). When no single cluster is dominantly populated, the median angles of the most populated cluster are deemed ambiguous and should be used with caution in subsequent structure calculations.

Starting structures are built for the contiguous segments with unambiguous ϕ/ψ values, derived in the manner described above. Calculations for ubiquitin and several other proteins yield segment lengths in the 10−50 residue range. Remarkably, a higher rhombicity of the alignment tensor results in less ϕ/ψ ambiguity and longer segment lengths. This is easily understood considering that, if for any 7-residue segment one of its $N-C^\alpha$ or $C^\alpha-C'$ bonds is parallel to one of the principal axes of the alignment tensor, a database fragment which differs by 180° in the corresponding ϕ or ψ angle will yield couplings for all its residues that are identical to a segment with the correct backbone angles. For an axially symmetric tensor, this type of degeneracy occurs each time a $N-C^\alpha$ or $C^\alpha-C'$ is parallel or orthogonal to the unique axis of the alignment tensor, that is, for any orientation in the $x-y$ plane. Most of such 180°-flipped bonds are not populated in the database, because they result in steric clashes, but nevertheless this ambiguity constitutes the main barrier to building longer segment lengths.

Table 1 shows that the approximate backbone angles in ubiquitin can be defined unambiguously for two contiguous segments (1−52 and 54−76) when using 67 $N-H$, 69 $C'-N$, 69 $C'-H^N$, and 66 $C^\alpha-H^\alpha$ dipolar couplings, measured in two liquid

(1) Wüthrich, K. *NMR of proteins and nucleic acids*; Wiley: New York, 1986. Edison, A. S.; Abildgaard, F.; Westler, W. M.; Mooberry, E. S.; Markley, J. L. *Methods Enzymol.* **1994**, *239*, 3−79. Gronenborn, A. M.; Clore, G. M. *Crit. Rev. Biochem. Mol.* **1995**, *30*, 351−385.

(2) Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9279−9283. King, H. C.; Wang, K. Y.; Goljer, I.; Bolton, P. H. *J. Magn. Reson. Ser. B* **1995**, *109*, 323−325. Tjandra, N.; Grzesiek, S.; Bax, A. *J. Am. Chem. Soc.* **1996**, *118*, 6264−6272. Bax, A.; Tjandra, N. *J. Biomol. NMR* **1997**, *10*, 289−29. Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111−1114.

(3) Tjandra, N.; Omichinski, J. G.; Gronenborn, A. M.; Clore, G. M.; Bax, A. *Nat. Struct. Biol.* **1997**, *4*, 732−738. Ottiger, M.; Tjandra, N.; Bax, A. *J. Am. Chem. Soc.* **1997**, *119*, 9825−9830. Bewley, C. A.; Gustafson, K. R.; Boyd, M. R.; Covell, D. G.; Bax, A.; Clore, G. M.; Gronenborn, A. M. *Nature Struct. Biol.* **1998**, *5*, 571−578. Drohat, A. C.; Tjandra, N.; Baldisseri, D. M.; Weber, D. J. *Protein Sci.* **1999**, *8*, 800−809. Clore, G. M.; Starich, M. R.; Bewley, C. A.; Cai, M. L.; Kuszewski, J. *J. Am. Chem. Soc.* **1999**, *121*, 6513−6514.

(4) Kraulis, P. J.; Jones, T. A. *Proteins: Struct., Funct., Genet.* **1987**, *2*, 188−201.

(5) Jones, T. A.; Thirup, S. *EMBO J.* **1986**, *5*, 819−822.

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

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

(8) DiStefano, D. L.; Wand, A. J. *Biochemistry* **1987**, *26*, 7272−7281. Weber, P. L.; Brown, S. C.; Mueller, L. *Biochemistry* **1987**, *26*, 7282−7290. Hu, J.-S.; Bax, A. *J. Am. Chem. Soc.* **1997**, *119*, 6360−6368.

(9) Schneider, D. M.; Dellwo, M. J.; Wand, A. J. *Biochemistry* **1992**, *31*, 3645−3652. Tjandra, N.; Feller, S. E.; Pastor, R. W.; Bax, A. *J. Am. Chem. Soc.* **1995**, *117*, 12562−12566.

(10) Cornilescu, G.; Marquardt, J. L.; Ottiger, M. Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 6836−6837. PDB entry 1D3Z.

(11) Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 12334−12341.

(12) Ramirez, B. E.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 9106−9107.

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

(14) Cornilescu, G.; Delaglio, F.; Bax, A. *J. Biomol. NMR* **1998**, *13*, 289−302.

10.1021/ja993603n CCC: $19.00
Published on Web 02/16/2000

Table 1. Ubiquitin Segments Calculated from Dipolar Couplings and Chemical Shifts[a]

	initial model[b]	refined model	
segment	RMSD vs X-ray (Å)	RMSD vs X-ray (Å)	RMSD vs X-ray (Å)
1−52	5.44	0.87	0.82[c]
53−76	3.79	1.17	0.42[c]
1−76	6.95	1.21	0.88[c]

[a] Using N, C^{α}, H^{α}, C^{β}, and C′ chemical shifts and the two sets of experimental dipolar couplings measured with charged and uncharged bicelles, reported in ref 10. The amide of Gly^{53} is conformational-exchange broadened beyond detection. [b] Initial model built from MFR-derived backbone angles. [c] Excluding residues 1 and 73−76.

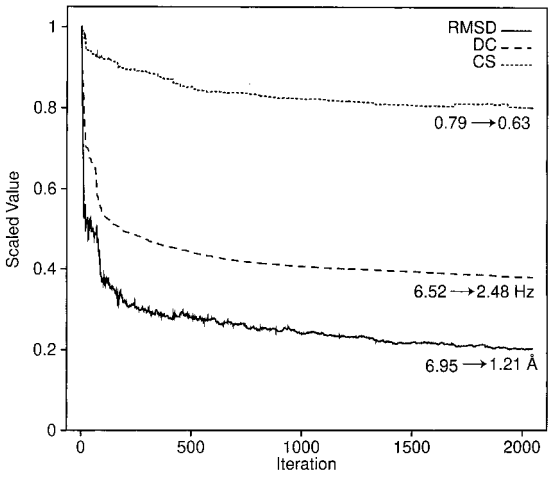

Figure 1. χ^2 between predicted and measured chemical shifts (short dashes) and rmsd between best-fitted and experimental dipolar couplings (dashed) as a function of iteration number during the minimization protocol, which aims to minimize these values. The chemical shift χ^2 drops from 0.79 to 0.63 during refinement. The dipolar coupling rmsd is calculated after normalizing all couplings relative to N−H, and decreases from 6.52 to 2.48 Hz during refinement. The solid line corresponds to the backbone rmsd between the model and the ubiquitin crystal structure and decreases from 6.95 Å for the initial model to 1.21 Å (residues 1−76). The entire refinement takes 10 min on a 400 MHz PC.

crystalline media. The "ambiguous" backbone angles of residue Gly^{53} are fortuitously correct, and building the polypeptide by using all MFR angles yields an initial fold with roughly the correct topology, differing by 6.95 Å from the X-ray structure (Table 1). For shorter segments this rmsd is correspondingly smaller. When using dipolar couplings measured in only a single liquid crystal medium, there is a larger number of residues for which the approximate backbone angles cannot be determined by using the MFR approach (Asp^{52}, Gly^{53}, and Glu^{64} for the medium with the more rhombic alignment tensor; Pro^{19}, Asp^{52}, Gly^{53}, and Thr^{55} for the less rhombic tensor). Nevertheless, reasonable initial structures can be generated for the intervening fragments (Supporting Information).

It is important to note that at this stage errors in the MFR-derived backbone angles accumulate when building the initial model because the long-range information contained in the dipolar couplings is not yet used. However, this orientational information can be reintroduced when using these rough models as starting structures in a subsequent refinement procedure. This refinement is based on a simple iterative gradient approach, which adjusts ϕ/ψ to minimize the χ^2 between measured and best-fitted dipolar couplings, and between measured chemical shifts and those

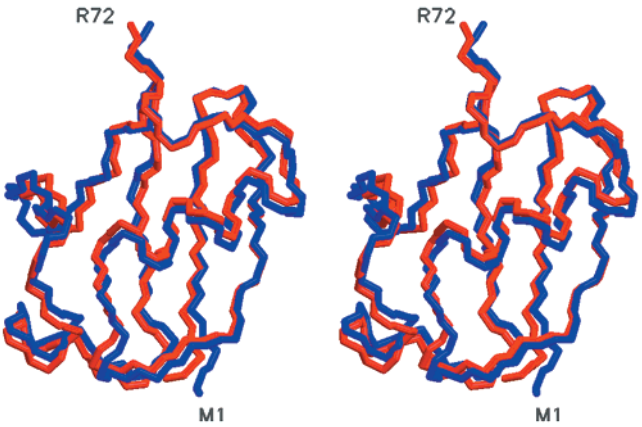

Figure 2. Stereodiagram of the backbone representations of the X-ray crystal structure (blue) of ubiquitin residues 1−72, and the refined model (red) obtained from dipolar couplings in two different liquid crystalline media, together with isotropic chemical shifts. Planar, trans peptide bonds and standard covalent geometry and bond lengths were used to calculate this model, but no vanderWaals radii or other terms were used. Further refinement may be obtained by adding such terms and additional NOE or H-bond restraints.

predicted by the model. At each iteration, a residue is selected at random, and the gradient of χ^2 with respect to ϕ and ψ is calculated numerically, and new optimal values for ϕ and ψ are predicted. If the new values improve χ^2, they are accepted and replace the prior values. Figure 1 shows how this procedure results in rapid convergence; when the model's fit to the experimental dipolar couplings and chemical shifts improves, so does the fit to the X-ray (and NMR) structure.

Figure 2 shows a superposition of the ubiquitin X-ray structure and the refined model calculated using dipolar data from both alignment tensors. The backbone rmsd is 1.21 Å; when the flexible C-terminus[9] is excluded, this rmsd drops to 0.88 Å.

This report demonstrates that it is possible to calculate the 3D structure of large protein backbone segments, and in favorable cases an entire small protein, exclusively from dipolar couplings and chemical shifts. Dipolar couplings can be rapidly measured once a protein assignment is completed, and the approach described here obviates the time-consuming NOE analysis. However, our protocol can easily integrate experimental NOE and H-bond information too, and even a handful of long-range constraints may be sufficient to correctly define the relative position of oriented fragments relative to one another. Alternatively, packing the fragments by using molecular modeling is expected to be relatively rapid and straightforward. The approach described above is just one of many possible schemes for reconstituting a protein structure from database fragments by using dipolar coupling homology, and numerous such variations are being explored.

Software is available from the authors.

Acknowledgment. G.K. is the recipient of an E. Schrödinger Fellowship (J-1706MOB). This work was supported by the AIDS Targeted Anti-Viral Program of the Office of the Director of the NIH.

Supporting Information Available: One table with structural statistics obtained from dipolar coupling data measured in a single liquid crystalline medium; one figure showing the agreement between dipolar couplings measured for ubiquitin fragment 60−66 and PDB fragment 54−60 of Hpr (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA993603N