Impact of Residual Dipolar Couplings on the Accuracy of NMR Structures Determined from a Minimal Number of NOE Restraints

G. Marius Clore,* Mary R. Starich, Carole A. Bewley, Mengli Cai, and John Kuszewski

Laboratory of Chemical Physics, Building 5
National Institute of Diabetes and Digestive and Kidney Diseases
National Institutes of Health
Bethesda, Maryland 20892-0520

Received April 12, 1999

NMR protein structure determination chiefly relies on interproton distance restraints derived from NOE measurements. As the molecular weight increases so the line widths become larger. This can be partially alleviated by perdeuteration,[1] albeit at the price of eliminating numerous observable NOEs. Thus, there has been considerable interest in attempting to determine global folds on the basis of minimal numbers of NOEs involving backbone NH, methyl, and aromatic protons.[2] These attempts have been only partially successful: the attainable accuracy is rather low, typically ranging from 2.5 to more than 7 Å for the backbone depending on topology.[1,2] It has recently been demonstrated that residual dipolar couplings (which provide direct long-range angular orientational information[3] and are typically measured in a liquid crystalline medium of bicelles[4a] or rod-shaped viruses[4b]) can provide large increases in coordinate precision and potential improvements in coordinate accuracy when employed in conjuction with virtually complete NOE data sets, as well as other NMR data derived from coupling constants and chemical shifts.[3a,5] In this paper we show that the use of dipolar couplings results in large improvements in accuracy even when minimal NOE data sets are employed. This is demonstrated by using three systems for which experimental dipolar couplings have been measured: the B1 domain of streptococcal protein G (GB1; 56 residues),[4b] the monomer of the barrier-to-autointegration factor BAF (89 residues),[5a] and cyanovirin-N (CVN; 101 residues).[5b]

All structures were calculated starting from an extended strand using conventional simulated annealing with the program XPLOR,[6] initially in torsion angle space[7a] and subsequently in Cartesian coordinate space.[7b] The terms in the target function included experimental distance (NOE-derived interproton distances and/or backbone hydrogen bonds) and dipolar coupling[7c] restraints, restraints for idealized covalent geometry, a quartic van der Waals repulsion term,[7b] a torsion angle database potential of mean force[7d] and a term for the radius of gyration.[7e] The results are summarized in Table 1 which reports the precision and accuracy of the coordinates, and the agreement with the measured dipolar couplings.

Table 1. Coordinate Precision and Accuracy and Dipolar Coupling *R*-Factors

	backbone rmsd (Å)[a]		dipolar coupling *R*-factor (%)[b]				
structure	precision	accuracy	R_{N-H}	$R_{N-C'}$	$R_{HN-C'}$	$R_{C\alpha-H}$	$R_{C\alpha-C'}$
GB1 (56 residues)[c]							
no dipolar	2.88	4.33	26/18	83/79	72/66		
dipolar (TMV)	1.37	1.12	5/6	23/25	27/22		
dipolar (bicelle)	1.45	1.19	14/5	25/30	31/21		
dipolar (TMV + bicelle)	1.03	0.93	8/7	21/24	28/22		
BAF monomer (89 residues)							
no dipolar	1.31	1.37	12	52	41		55
dipolar (bicelle)	1.23	0.91	5	25	22		34
CVN (101 residues)							
no dipolar	1.67	1.53	36	62	44	53	45
dipolar (bicelle)	1.23	1.10	3	27	18	4	35

[a] Precision is defined as the average backbone rmsd of the ensemble of individual simulated annealing structures to the mean coordinate positions (each ensemble consists of 20−30 structures); accuracy is defined as the rmsd between the restrained regularized mean structure and either the X-ray structure in the case of GB1 (1PGA[10]), or the restrained regularized mean NMR structure calculated with the complete set of experimental restraints in the case of BAF (2EZX)[5a] and CVN (2EZM)[5b] (1646 and 2509, respectively, of which 800 and 1157, respectively, are NOE-derived interproton distance restraints, and 259 and 334, respectively, are dipolar couplings). For reference the precision for the backbone atoms of the latter coordinates is 0.34 Å for BAF[5a] and 0.15 Å for CVN.[5b] For BAF the rmsd's are calculated for residues 3−89; for CVN, the loop residues 23−29, 75−80, and 95−97 are excluded from the calculation of the rmsd. For GB1 there are 48 N−H, 51 N−C′ and 53 H_N−C′ dipolar couplings measured in TMV and 46 N−H, 50 N−C′ and 52 H_N−C′ dipolar couplings measured in bicelles; for BAF there are 76 N−H, 55 N−C′, 52 H_N−C′, and 76 Cα−C′ dipolar couplings measured in bicelles; and for CVN there are 84 N−H, 66 N−C′, 63 H_N−C′, 77 Cα−H, and 44 Cα−C′ dipolar couplings measured in bicelles. There are 32 backbone H-bond restraints for GB1; 245 interproton distance (107 sequential, 63 medium range with $1 < |i-j| \leq 5$, and 75 long-range with $|i-j| > 5$) and 37 H-bond restraints for BAF; and 331 interproton distance (135 sequential, 33 medium range, and 163 long-range) and 40 H-bond restraints for CVN. In the case of BAF, 59 ϕ (−60 ± 20°) and 58 ψ (−40 ± 20°) backbone torsion angle restraints were also employed for the helices (residues 5−10, 20−23, 28−36, 42−51, 56−67, 71−88), previously identified from $^{13}C\alpha$ and $^{13}C\beta$ shifts.[5a] The approximate interproton distance restraints (classified into four ranges corresponding to strong, medium, weak, and very weak NOEs[5]) used for BAF and CVN comprise the NH−NH, NH−methyl, NH−aromatic, methyl−methyl, methyl−aromatic, and aromatic−aromatic subset of NOEs taken from the complete NOE data sets (2RZXMR and 2REZMMR, respectively) used in the structure determinations reported in refs 5a and b. In all cases the agreement with the H-bond and interproton distance restraints was satisfactory (no violations greater than 0.5 Å), the deviations from idealized covalent geometry were very small, and there were no bad non-bonded contacts. [b] The *R*-factor is the ratio of the measured rms deviation between observed and calculated dipolar couplings to the expected rms deviation for a completely random set of vectors (i.e., in a random coil) given by $\{2D_a^2[4 + 3(D_r/D_a)^2]/5\}^{1/2}$, where D_a and D_r are the magnitudes of the axial and rhombic components of the molecular alignment tensor.[4b] For GB1, $^1D_a^{NH}$ and D_r/D_a have values of −5.2 Hz and 0.61, respectively, in TMV, and −9.7 Hz and 0.23, respectively, in bicelles.[4b] The values of $^1D_a^{NH}$ and D_r/D_a for BAF[5a] and CVN[5b] in bicelles are −14.9 Hz and 0.17, and −17.0 Hz and 0.17, respectively. [c] The first and second values relate to dipolar couplings measured in TMV and bicelles, respectively.

* To whom correspondence should be addressed. Telephone: (301) 496-0782. Fax: (301) 496-0825. E-mail: clore@speck.niddk.nih.gov.

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

(2) Vuister, G. W.; Kim, S.-J.; Wu, C.; Bax, A. *J. Am. Chem. Soc.* **1994**, *116*, 9206−9210; Grzesiek, S., Wingfield, P. T., Stahl, S. J.; Kaufman, J. D.; Bax, A. *J. Am. Chem. Soc.* **1995**, *117*, 9594−9595; Metzler, W. J.; Wittekind M.; Goldfarb, V.; Mueller, L.; Farmer, B. T., II. *J. Am. Chem. Soc.* **1996**, *118*, 6800−6801; Smith, B. O.; Rosen, M. K.; Gardner, K. H.; Willis, R. C.; Parris, W. E., Pawson, T.; Kay, L. E. *J. Mol. Biol.* **1996**, *263*, 627−636; Ito, Y.; Raine, A.; Teichmann, S.; Ben-Tovim, L.; Nietlispach, D.; Broadhurst, R. W.; Terada, T.; Kelly, M.; Oschkinat, H.; Shibata, T.; Yokoyama, S.; Laue, E. D. *J. Biomol. NMR* **1996**, *8*, 360−368. Gardner, K. H.; Rosen M. K.; Kay, L. E. *Biochemistry* **1997**, *36*, 1389−1401; Wang, H.; Janowick, D. A.; Schkeryantz, J. M.; Liu, X.; Fesik, S. W. *J. Am. Chem. Soc.* **1999**, *121*, 1611−1612.

(3) (a) Tjandra, N., Omichinski, J. G.; Gronenborn, A. M.; Clore, G. M.; Bax, A. *Nat. Struct. Biol.* **1997**, *4*, 732−728. (b) Ramirez, B. E.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 9106−9107.

(4) (a) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111−1114. (b) Clore, G. M.; Starich, M. R.; Gronenborn, A. M. *J. Am. Chem. Soc.* **1998**, *120*, 10571−10572.

(5) (a) Cai, M.; Huang, Y.; Zheng, R.; Wei, S.-Q.; Ghirlando, R.; Lee, M. S.; Craigie, R.; Gronenborn, A. M.; Clore, G. M. *Nat. Struct. Biol.* **1998**, *5*, 903−909. (b) Bewley, C. A.; Gustafson, K. R.; Boyd, M. R.; Covell, D. G.; Bax, A.; Clore, G. M.; Gronenborn, A. M. *Nat. Struct. Biol.* **1998**, *5*, 571−578.

(6) Brünger, A. T. *XPLOR: A system for X-ray crystallography and NMR*; Yale University Press: New Haven, 1993.

10.1021/ja991143s
Published on Web 06/25/1999

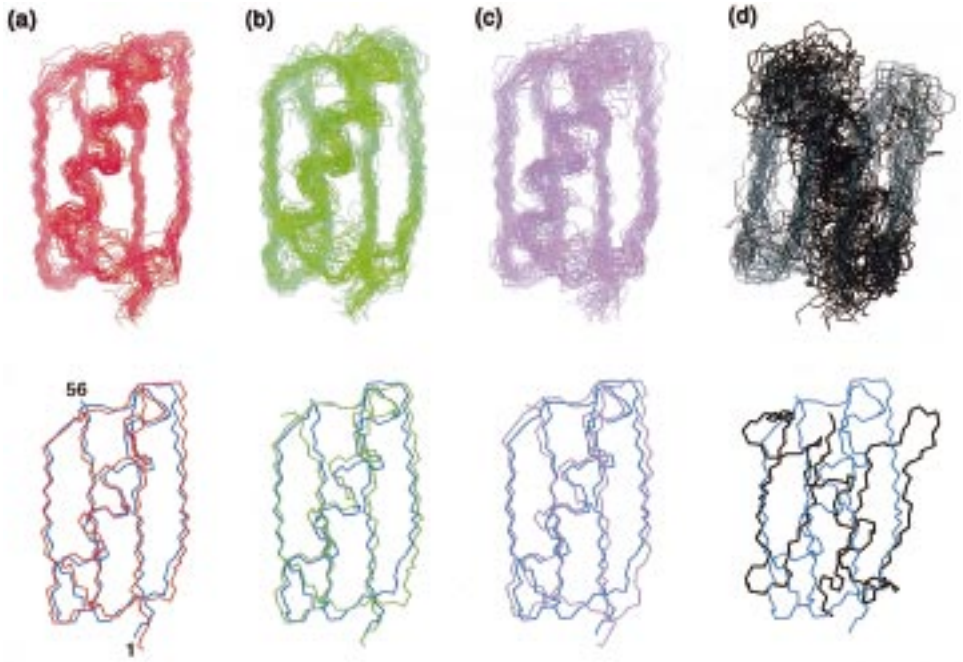

Figure 1. Effect of dipolar couplings on coordinate accuracy for GB1. The top panel shows a superposition of the ensemble of simulated annealing structures, and the bottom panel a superposition of the restrained regularized mean structure with the X-ray structure[10] (in blue). The structures shown in (a), (b), and (c) were calculated with dipolar couplings measured in TMV and bicelles (red), in TMV alone (green), and in bicelles alone (purple), respectively. There were 152 dipolar couplings measured in TMV and 150 in bicelles. The structures shown in (d) were calculated without dipolar couplings (black). All structures were calculated with 32 backbone H-bond restraints in regions of regular secondary structure.

Four calculations were carried out for GB1. In each case, the only distance information employed consisted of a set of 32 loose N−O distance restraints (2.3−3.5 Å) that define the backbone hydrogen bonds (H-bonds) in regions of regular secondary structure. These were previously identified using conventional criteria (slowly exchanging NH protons, ^{13}C chemical shifts, and a qualitative interpretation of NOEs involving backbone protons),[8] as well as from $^{3h}J_{NC'}$ couplings measured by quantitative J correlation spectroscopy.[9] Residual ^{15}N−H, ^{15}N−^{13}C′ and H$_N$−^{13}C′ dipolar couplings were measured on a sample of ^{15}N/^{13}C-labeled GB1 both in a suspension of tobacco mosaic virus (TMV) and in 3:1 DMPC/DHPC bicelles, as described previously.[4b,5] GB1 represents a special case since ca. 90% of the residues are located in secondary structure and the topology consists of a single helix sitting on top of and connected by short loops to a four stranded mixed parallel−antiparallel β-sheet (Figure 1). As a result, the secondary structure in its own right imposes severe conformational limitations on the polypeptide fold, and the H-bond restraints alone are sufficient to define an approximate topology, albeit with low accuracy (4.3 Å; Figure 1d). Inclusion of dipolar couplings results in an approximately 4-fold improvement in accuracy (<1.2 Å; Figure 1a−c). Moreover, the use of dipolar couplings measured for two different alignment tensors[3b] (obtained in bicelles and TMV) which differ in both orientation and rhombicity improves the accuracy by a further 20−30% (Figure 1a; 0.9 Å).

For both the BAF monomer[5a] and CVN,[5b] the topology is such that long-range side chain−side chain NOE data are absolutely essential to define the polypeptide fold. BAF is entirely helical, and CVN comprises two structural domains each consisting of a β-hairpin lying on top of a three-stranded β-sheet (see Supporting Information). Initial calculations with just backbone H-bond restraints in the regions of regular secondary structure, H-bond restraints and NH−NH interproton distances, or H-bond restraints and NH−NH, methyl−methyl, and NH−methyl interproton distances could not define the topology appropriately (i.e., accuracy worse than 4 Å), even in the presence of dipolar couplings. The clustering of methyl groups at the interfaces among the various structural elements in the two proteins is not sufficiently dense to define the topology. With the additional inclusion of NH−aromatic, methyl−aromatic, and aromatic−aromatic NOEs, however, the folds for both proteins are well-defined (accuracy of 1.4−1.5 Å), and the inclusion of dipolar couplings improves the accuracy by 35−50% to a level of ∼1 Å (Table 1).

The results presented here demonstrate that the inclusion of dipolar couplings in structure calculations can result in significant increases in accuracy even for data sets consisting of a minimal number of NOE restraints. In favorable cases, such as that provided by GB1, where the secondary structure alone places severe limitations on the overall topology, reasonably accurate structures can be obtained even when the only distance restraints employed consist of backbone H-bonds within elements of regular secondary structure. Thus, the use of dipolar couplings holds considerable promise for the structure determination of larger proteins where the number of NOEs that can be assigned may be limited due to resonance overlap, line broadening, or deuteration.

Acknowledgment. We thank Ad Bax for useful discussions and Angela Gronenborn for a sample of ^{15}N/^{13}C-labeled GB1.

Supporting Information Available: Table giving breakdown of NOE-derived interproton distance restraints into interaction type used in the calculations for BAF and CVN, and figures showing structures calculated for BAF and CVN (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA991143S

(7) (a) Stein, E. G.; Rice, L. M.; Brünger, A. T. *J. Magn. Reson.* **1997**, *124*, 154−164. (b) Nilges, M.; Gronenborn, A. M.; Brünger, A. T.; Clore, G. M. *Protein Eng.* **1988**, *2*, 27−38. (c) Clore, G. M.; Gronenborn, A. M.; Tjandra, N. *J. Magn. Reson.* **1998**, *131*, 159−162. (d) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. *J. Magn. Reson.* **1997**, *125*, 171−177. (e) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. *J. Am. Chem. Soc.* **1999**, *121*, 2337−2338.

(8) Gronenborn, A. M.; Filpula, D. R.; Essig, N. Z.; Achari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M. *Science* **1991**, *253*, 657−661.

(9) Cornilescu, G.; Ramirez, B.; Frank, M. K.; Clore, G. M.; Gronenborn, A. M.; Bax, A. *J. Am. Chem. Soc.*, in press.

(10) Gallagher, T.; Alexander, P.; Bryan, P., Gilliland, G. L. *Biochemistry* **1994**, *33*, 4721−4729.