J|A|C|S
ARTICLES
Published on Web 11/07/2006

Geometry, Energetics, and Dynamics of Hydrogen Bonds in Proteins: Structural Information Derived from NMR Scalar Couplings

Joerg Gsponer, Harri Hopearuoho, Andrea Cavalli, Christopher M. Dobson, and Michele Vendruscolo*

Contribution from the Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

Received March 2, 2006; E-mail: mv245@cam.ac.uk

Abstract: An accurate description of hydrogen bonds is essential to identify the determinants of protein stability and function as well as folding and misfolding behavior. We describe a method of using *J* couplings through hydrogen bonds as ensemble-averaged restraints in molecular dynamics simulations. Applications to the cases of ubiquitin and protein G show that these scalar couplings provide powerful structural information that, when used through the methodology that we present here, enables the description of the geometry and energetics of hydrogen bonds with an accuracy approaching that of high-resolution X-ray structures.

Introduction

A detailed description of hydrogen bonds has been a long standing goal in structural biology, as they are crucial in the formation of α helices and β sheets[1,2] and hence are of great importance for determining protein stability,[3] enzymatic catalysis,[4] protein folding,[5,6] and the formation of amyloid aggregates.[7−9] Studies of the structures of small molecules,[10] surveys of proteins and nucleic acid structures in the Protein Data Bank (PDB),[11−14] and accurate calculations based on classical molecular dynamics simulations[15] and density functional theory (DFT)[16,17] have provided considerable insight into the geometry of hydrogen bonds. In structural studies, however, the presence of hydrogen bonds is most often inferred rather than actually detected.[18] This situation arises because in X-ray crystallography hydrogen atoms are, with a few exceptions,[19] not observed directly in electron density maps, and nuclear Overhauser effects (NOEs) exploited in NMR spectroscopy only provide approximate distance information between proton pairs.[20] Therefore, in a hydrogen bond the distances and angles between the shared proton and the donor and acceptor heavy atoms are most often not accurately defined. In recent years, residual dipolar couplings (RDCs) have been used in conjunction with NOEs in NMR structure determination to impose geometric restraints on individual bond vectors.[21−24] This use of RDCs is attractive because, besides having a global ordering character as they restrain bond orientations relative to a fixed reference frame, they also contain information about the local environment. In addition, RDCs enable to access information on time scales (up to the millisecond) that are essential to describe various important biological processes, including allosteric regulation and signal transduction. A range of methods have been proposed to exploit the orientational averaging properties of RDCs to characterize the dynamics of proteins.[25−33] Accurate information

(1) Pauling, L.; Corey, R. B.; Branson, H. R. *Proc. Natl. Acad. Sci. U.S.A.* **1951**, *37*, 205−211.
(2) Pauling, L.; Corey, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **1951**, *37*, 729−740.
(3) Shi, Z. S.; Krantz, B. A.; Kallenbach, N.; Sosnick, T. R. *Biochemistry* **2002**, *41*, 2120−2129.
(4) Frey, P. A.; Whitt, S. A.; Tobin, J. B. *Science* **1994**, *264*, 1927−1930.
(5) Krantz, B. A.; Srivastava, A.; Nauli, S.; Baker, D.; Sauer, R.; Sosnick, T. *Nat. Struct. Biol.* **2002**, *9*, 458−463.
(6) Deechongkit, S.; Nguyen, H.; Powers, E. T.; Dawson, P. E.; Gruebele, M.; Kelly, J. W. *Nature* **2004**, *430*, 101−105.
(7) Sunde, M.; Serpell, L. C.; Bartlam, M.; Fraser, P. E.; Pepys, M. B.; Blake, C. C. F. *J. Mol. Biol.* **1997**, *273*, 729−739.
(8) Dobson, C. M. *Nature* **2003**, *426*, 884−890.
(9) Tycko, R. *Curr. Opin. Struct. Biol.* **2004**, *14*, 96−103.
(10) Etter, M. C. *Acc. Chem. Res.* **1990**, *23*, 120−126.
(11) Baker, E. N.; Hubbard, R. E. *Prog. Biophys. Mol. Biol.* **2984**, *44*, 97−179.
(12) Taylor, R.; Kennard, O. *Acc. Chem. Res.* **1984**, *17*, 320−326.
(13) McDonald, I. K.; Thornton, J. M. *J. Mol. Biol.* **1994**, *238*, 777−793.
(14) Grishaev, A.; Bax, A. *J. Am. Chem. Soc.* **2004**, *126*, 7281−7292.
(15) Buck, M.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 11000−11015.
(16) Scheurer, C.; Bruschweiler, R. *J. Am. Chem. Soc.* **1999**, *121*, 8661−8662.
(17) Lipsitz, R. S.; Sharma, Y.; Brooks, B. R.; Tjandra, N. *J. Am. Chem. Soc.* **2002**, *124*, 10621−10626.
(18) Forrest, L. R.; Honig, B. *Proteins* **2005**, *61*, 296−309.
(19) Vrielink, A.; Sampson, N. *Curr. Opin. Struct. Biol.* **2003**, *13*, 709−715.
(20) Wüthrich, K. NMR of Proteins and Nucleic Acids. Wiley: New York, 1986.
(21) Bax, A.; Grishaev, A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 1−8.
(22) Schwieters, C. D.; Kuszewski, J. J.; Clore, G. M. *Prog. Nucl. Magn. Reson. Spec.* **2005**, *48*, 47−62.
(23) Blackledge, M. *Prog. Nucl. Magn. Reson. Spec.* **2005**, *46*, 23−61.
(24) Tolman, J. R.; Ruan, K. *Chem. Rev.* **2006**, *106*, 1720−1736.
(25) Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Nat. Struct. Biol.* **1997**, *4*, 292−297.
(26) Tolman, J. R.; Al-Hashimi, H. M.; Kay, L. E.; Prestegard, J. H. *J. Am. Chem. Soc.* **2001**, *123*, 1416−1424.
(27) Meiler, J.; Prompers, J. J.; Peti, W.; Griesinger, C.; Bruschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*, 6098−6107.
(28) Peti, W.; Meiler, J.; Bruschweiler, R.; Griesinger, C. *J. Am. Chem. Soc.* **2002**, *124*, 5822−5833.
(29) Hess, B.; Scheek, R. M. *J. Magn. Reson.* **2003**, *164*, 19−27.
(30) Clore, G. M.; Schwieters, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 2923−2938.
(31) Clore, G. M.; Schwieters, C. D. *Biochemistry* **2004**, *43*, 10678−10691.
(32) Bouvignies, G.; Bernado, P.; Meier, S.; Cho, K.; Grzesiek, S.; Bruschweiler, P.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13885−13890.
(33) Clore, G. M.; Schwieters, C. D. *J. Mol. Biol.* **2006**, *355*, 879−886.

on the position and dynamics of internuclear vectors, including the backbone NH vectors in particular, is essential to characterize accurately the conformational properties of hydrogen bonds. It has been reported that structures refined with RDCs generally have hydrogen bond geometries close to those observed in high-resolution X-ray structures.[17]

The recent discovery that scalar couplings can be observed across hydrogen bonds both in proteins[34,35] and DNA[36,37] has greatly facilitated the description of hydrogen bond donors and acceptors, as the magnitude of these scalar couplings ($^{h3}J_{NC'}$) depends strongly on hydrogen bond distances and angles.[16,38] $^{h3}J_{NC'}$ values have therefore been used to provide upper and lower limits for distance restraints in protein structure determination[39,40] and the magnitudes of these couplings have been shown to provide a very sensitive measure of hydrogen-bond geometries and energetics. Although the measurement of $^{h3}J_{NC'}$ scalar couplings is more challenging than that of NOEs and RDCs, they allow for an exquisitely accurate detection of the distance between the shared proton and the donor and acceptor heavy atoms. In addition, as for most NMR parameters measured in NMR spectroscopy, the values of $^{h3}J_{NC'}$ are motionally averaged so they also provide information about the dynamic nature of hydrogen bonds.[38,41] Thus, the dynamic character of hydrogen bonds in biomolecules is directly and individually observable by $^{h3}J_{NC'}$ couplings.

To integrate the dynamical information contained in NMR observables into the macromolecular structure determination, a number of approaches using time[41−46] and ensemble averaging[29−31,33,44,47−51] have been proposed. In particular, it has been recently shown that the agreement between predicted and experimental $^{h3}J_{NC'}$ scalar couplings is significantly improved by using RDC-derived distributions of hydrogen-bond geometries.[32] In this paper, we extend these computational approaches to $^{h3}J_{NC'}$ couplings by studying the local geometry and dynamics of backbone hydrogen bonds for two representative proteins, ubiquitin, and protein G. Structural ensembles are determined: (1) by performing unrestrained molecular dynamics simulations, (2) by using nuclear Overhauser effect (NOE) data, and (3) by using the latter in conjunction with $^{h3}J_{NC'}$ scalar couplings as ensemble-averaged restraints in simulations. The quality of the resulting ensembles are validated by predicting values of experimental NMR parameters not used in the simulations, namely RDCs and several sets of three-bond scalar backbone couplings. Distributions of structural parameters describing the geometry of hydrogen bonds and the dynamics of the NH vectors are compared between the ensembles obtained by using different types of restraints.

The analysis of the results that we present shows that hydrogen bond *J* couplings can provide powerful information to describe accurately the hydrogen-bond geometries and that they represent an important contribution to techniques that use NMR observables as ensemble-averaged restraints to determine the structure and the dynamics of proteins at high resolution.

Methods

Molecular Dynamics Simulations. Structure determinations were performed with a modified version of CHARMM.[52] Simulations were carried out either in vacuo, using a distance-dependent dielectric constant, in a water-box (see below), or in a shell of 4 Å of TIP3 water molecules;[53] for the latter, a boundary potential was used to prevent water molecules from escaping.[54] All calculations used an atom-based truncation scheme with a list cutoff of 14 Å, a nonbond cutoff of 12 Å, and the Lennard-Jones smoothing function initiated at 10 Å. Electrostatic and Lennard-Jones interactions were force switched. Molecular dynamics simulations used a 2 fs integration time step and SHAKE for covalent bonds involving hydrogen atoms.[55] The energy function that we used has the form

$$E_{tot} = E_{CHARMM} + E_{NOE} + E^{h3}{}_{J_{NC'}} + E_{S^2} \quad (1)$$

in which E_{CHARMM} is the CHARMM22 force field[52] and E_{NOE}, $E^{h3}{}_{J_{NC'}}$ and E_{S^2} are the energies of the NOE, $^{h3}J_{NC'}$ and S^2 ensemble-averaged restraints, respectively. The restraint energy is implemented as[56]

$$E_X(\rho, t) = \begin{cases} \frac{\alpha_X}{2}(\rho(t) - \rho_0(t))^2 & \text{if } \rho(t) > \rho_0(t) \\ 0 & \text{if } \rho(t) \le \rho_0(t) \end{cases} \quad (2)$$

where X corresponds either to NOE, $^{h3}J_{NC'}$, or S^2, α_X is the force constant associated with each type of restraint, and $\rho_0(t)$ is defined as[57]

$$\rho_0(t) = \min_{0 \le \tau \le t} \rho(t) \quad (3)$$

The initial value of $\rho_0(t)$ is set to be equal to the ρ value of the equilibrated starting configuration. If the distance between back-calculated and experimental data spontaneously decreases in the simulation step from t to $t + \Delta t$, i.e., if $\rho(t + \Delta t) < \rho_0(t)$, the restraining term vanishes and has no effect on the dynamics. In such a case, $\rho_0(t)$ is updated and $E_X(\rho, t)$ is modified accordingly, i.e., $\rho_0(t)$ is set equal to $\rho(t + \Delta t)$. If $\rho(t)$ is greater than $\rho_0(t)$, the harmonic force acts on ρ to prevent the reaction coordinate from decreasing significantly. Hence, the bias depends on the time through ρ_0. The value of α_X determines the magnitude of the allowed backward fluctuation of the reaction coordinate.[58] In this approach, the term E_X acts as a soft "molecular

(34) Cordier, F.; Grzesiek, S. *J. Am. Chem. Soc.* **1999**, *117*, 5179−5197.
(35) Cornilescu, G.; Hu, J. S.; Bax, A. *J. Am. Chem. Soc.* **1999**, *121*, 2949−2950.
(36) Dingley, A. J.; Grzesiek, S. *J. Am. Chem. Soc.* **1998**, *120*, 8293−8297.
(37) Pervushin, K.; Ono, A.; Fernandez, C.; Szyperski, T.; Kainosho, M.; Wuthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 14147−14151.
(38) Grzesiek, S.; Cordier, F.; Jaravine, V.; Barfield, M. *Prog. Nucl. Magn. Reson. Spec.* **2004**, *45*, 275−300.
(39) Kanelis, V.; Rotin, D.; Forman-Kay, J. D. *Nat. Struct. Biol.* **2001**, *8*, 407−412.
(40) Bonvin, A. M. J. J.; Houben, K.; Guenneugues, M.; Kaptein, R.; Boelens, R. *J. Biomol. NMR* **2001**, *21*, 221−233.
(41) Markwick, P. R. L.; Sprangers, R.; Sattler, M. *J. Am. Chem. Soc.* **2003**, *125*, 644−645.
(42) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1989**, *157*, 289−294.
(43) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. *J. Mol. Biol.* **1990**, *214*, 223−235.
(44) Bonvin, A. M. J. J.; Boelens, R.; Kaptein, R. *J. Biomol. NMR* **1994**, *4*, 143−149.
(45) Kemmink, J.; Scheek, R. M. *J. Biomol. NMR* **1995**, *6*, 33−40.
(46) Bonvin, A. M. J. J.; Brunger, A. T. *J. Mol. Biol.* **1995**, *250*, 80−93.
(47) Fennen, J.; Torda, A. E.; van Gunsteren, W. F. *J. Biomol. NMR* **1995**, *6*, 163−170.
(48) Daura, X.; Antes, I.; van Gunsteren, W. F.; Mark, A. E. *Proteins* **1999**, *36*, 542−555.
(49) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. *J. Am. Chem. Soc.* **1999**, *121*, 2337−2338.
(50) Vendruscolo, M.; Paci, E. *Curr. Opin. Struct. Biol.* **2003**, *13*, 82−87.
(51) Vendruscolo, M.; Dobson, C. M. *Philos. Trans. R. Soc. A* **2005**, *363*, 433−450.
(52) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(53) Jorgensen, W. J.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(54) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050−9063.
(55) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(56) Best, R. B.; Vendruscolo, M. *J. Am. Chem. Soc.* **2004**, *126*, 8090−8091.
(57) Paci, E.; Vendruscolo, M.; Dobson, C. M.; Karplus, M. *J. Mol. Biol.* **2002**, *324*, 151−163.
(58) Paci, E.; Karplus, M. *J. Mol. Biol.* **1999**, *288*, 441−459.

ratchet" that directs the trajectories toward a state that satisfies the experimental restraints.

NOE-derived Distance Restraints. The NOE restraint term, $\rho_{NOE}(t)$ was defined to optimize the agreement between the experimentally derived inter-proton distances d_{NOE}^{exp}[59] and the ensemble-averaged inter-proton distances, d_{NOE}^{ens}

$$\rho_{NOE}(t) = \frac{1}{N_{NOE}} \sum (d_{NOE}^{exp} - d_{NOE}^{ens})^2 \quad (4)$$

where the sum is taken over the number N_{NOE} of experimental NOE distances, and

$$d_{NOE}^{ens} = (N_{rep}^{-1} \sum^{N_{rep}} d_{NOE}^{struct-3})^{-1/3} \quad (5)$$

where N_{rep} is the number of molecules (replicas) used in the ensemble-averaged simulations. The d_{NOE}^{ens} distances are allowed to vary freely between their experimental upper and lower bounds, and they are penalised by a quadratic term when they exceed these bounds. To compute the distances d_{NOE}^{struct} within individual molecules, we consider all atom pairs associated with a given NOE restraint

$$d_{NOE}^{struct} = (\sum_{ij} r_{ij}^{-6})^{-1/6} \quad (6)$$

where the sum is taken over all equivalent atoms. All 2872 NOE-derived distances deposited with the PDB file 1D3Z[60] and 671 NOE-derived distances deposited with the PDB file 1GB1[61] were used in the structure refinements of ubiquitin and protein G, respectively.

S^2 Order Parameter Restraints. The S^2 restraint term was defined as[56,59]

$$\rho_{S^2}(t) = \frac{1}{N_{S^2}} \sum (S^{2,exp} - S^{2,ens})^2 \quad (7)$$

where the sum is taken over the number N_{S^2} of S^2 restraints and $S^{2,ens}$ values were calculated as described previously.[56,59] For ubiquitin, 112 S^2 order parameters derived from NMR relaxation experiments[62−64] were used in the calculations.

Hydrogen Bond Scalar Coupling Restraints. The hydrogen bond J coupling restraint term, ${}^{\rho h3}J_{NC'}(t)$, was defined as

$${}^{\rho h3}J_{NC'}(t) = \frac{1}{N_{SC}} \sum ({}^{h3}J_{NC'}{}^{exp} - {}^{h3}J_{NC'}^{calc})^2 \quad (8)$$

where N_{SC} is the number of ${}^{h3}J_{NC'}$ scalar coupling restraints. For individual hydrogen bonds, ${}^{h3}J_{NC'}{}^{calc}$ was calculated as[65]

$${}^{h3}J_{NC'}^{calc} = \frac{1}{N_{rep}} \sum^{N_{rep}} [-360.0 \cos^2\theta \exp^{-3.2 r_{HO}} + 0.04] \quad (9)$$

where θ is the H···O═C angle and r_{HO} is the distance between the hydrogen and oxygen atoms (Figure 1). The geometric dependencies of ${}^{h3}J_{NC'}$ couplings calculated with DFT and finite perturbation theory were parametrized to a good approximation with eq 8,[65] where only the H···O═C angle and the r_{HO} distance are used, despite it is well-known that there are additional factors contributing to the ${}^{h3}J_{NC'}$ couplings.[16,66,67] We used ensemble-averaged restraints because there are detectable dynamical contributions to the ${}^{3h}J$ couplings.[38,41] In the structure calculations of ubiquitin and protein G, respectively, 29 and 32 experimentally determined ${}^{h3}J_{NC'}$ scalar couplings[68,69] were used as restraints.

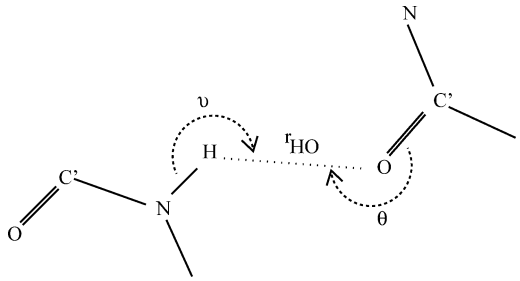

Figure 1. Representation of the parameters characterizing the geometry of a hydrogen bond. The distance between the amide proton and the carboxyl oxygen is defined as r_{HO}, the angle N−H···O as ν, and the angle H···O═C as ϑ.

Structural Ensemble Determination by Molecular Dynamics Simulations with Ensemble-Averaged Restraints. Molecular dynamics simulations were started from a structure equilibrated at 300 K; initial velocities were randomly assigned from a Maxwell−Boltzmann distribution at 300 K with a different random seed for each replica. The ensemble-averaged simulations were implemented by using the Message Passing Interface (MPI) for parallel computation, as described previously for other types of restraints.[56] Simulations at 300 K were carried out for 1.2 ns by increasing progressively the force constant α_X for each restraint to improve the agreement between the experimental and calculated values of each observable. The force constants used for each type of restraint and ensemble are given in the Supporting Information (Tables S1 and S2). Subsequently, cycles of simulated annealing were carried out to sample conformational space efficiently. First, the molecules were heated to 500 K and each α_X constant was reduced to 10% of its initial value (Tables S1 and S2, Supporting Information). Then, structures were cooled back to 300 K and α_X constants gradually increased to the starting value. Each annealing cycle lasted 0.5 ns. This annealing protocol was used in all simulations using experimental data as restraints. Twenty annealing cycles were performed for the restrained simulations with 2 replicas, and 40 annealing cycles for those with only one replica. After each cycle 10 structures were extracted (at 30 ps intervals), resulting in 400 conformations for each type of restraint.

Unrestrained Simulations of Ubiquitin. The minimized X-ray structure (1UBQ)[70] was solvated in a 63 × 50 × 46 Å orthorhombic water box containing 4565 TIP3 water molecules.[53] After heating from 10 K and equilibrating at 300 K for 200 ps, the simulation was carried out for 4 ns and in the constant volume, isothermal (NVT) canonical ensemble via the Nosé-Hoover temperature coupling scheme.

Analysis of the RDCs and Backbone Scalar Couplings. Singular value decomposition to fit the alignment tensor[71] was performed to back-calculate RDCs from the structures, taking ensemble averaging into account so that multiple structures were fitted simultaneously to a single alignment tensor. To test the robustness of this fitting procedure, we analyzed the changes in the directions of the principal axes, the

(59) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433*, 128−132.
(60) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 6836−6837.
(61) Gronenborn, A. M.; Filpula, D. R.; Essig, N. Z.; Achari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M. *Science* **1991**, *253*, 657−661.
(62) Tjandra, N.; Feller, S. E.; Pastor, R. W.; Bax, A. *J. Am. Chem. Soc.* **1995**, *117*, 12562−12566.
(63) Tjandra, N.; Szabo, A.; Bax, A. *J. Am. Chem. Soc.* **1996**, *118*, 6986−6991.
(64) Lee, A. L.; Flynn, P. F.; Wand, A. J. *J. Am. Chem. Soc.* **1999**, *121*, 2891−2902.
(65) Barfield, M. *J. Am. Chem. Soc.* **2002**, *124*, 4158−4168.
(66) Tuttle, T.; Kraka, E.; Wu, A.; Cremer, D. *J. Am. Chem. Soc.* **2004**, *126*, 5093−5107.
(67) Salvador, P.; Kobko, N.; Wieczorek, B.; Dannenberg, J. *J. Am. Chem. Soc.* **2004**, *126*, 14190−14197.
(68) Cordier, F.; Grzesiek, S. *J. Mol. Biol.* **2002**, *317*, 739−752.
(69) Cornilescu, G.; Ramirez, B. E.; Franck, M. K.; Clore, G. M.; Gronenborn, A. M.; Bax, A. *J. Am. Chem. Soc.* **1999**, *121*, 6275−6279.
(70) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531−544.
(71) Losonczi, J. A.; Andrec, M.; Fischer, M. W. F.; Prestegard, J. H. *J. Magn. Reson.* **1999**, *139*, 334−342.

Table 1. Summary of the Eight Simulations Performed in This Work for Ubiquitin Using Different Types of Solvation (V = in vacuo, E = explicit solvent) and Restraints (N = NOE, $J = {}^{3h}J_{NC'}$, S = S^2)[a]

ensemble label	E_1	VN_1	VNJ_1	VN_2	VNJ_2	EN_2	ENJ_2	$ENJS_2$
number of replicas	1	1	1	2	2	2	2	2
solvation	E	V	V	V	V	E	E	E
restraints applied		NOE	NOE, ${}^{h3}J_{NC'}$	NOE	NOE, ${}^{h3}J_{NC'}$	NOE	NOE, ${}^{h3}J_{NC'}$	NOE, ${}^{h3}J_{NC'}$, S^2
				r.m.s. Deviation				
NOE [Å]	(0.43)	0.02 ± 0.00 (0.01)	0.02 ± 0.00 (0.01)	0.01 ± 0.00 (0.00)	0.02 ± 0.00 (0.01)	0.01 ± 0.00 (0.01)	0.01 ± 0.00 (0.00)	0.02 ± 0.00 (0.01)
${}^{h}3J_{NC'}$ [Hz]	(0.21)	0.32 ± 0.05 (0.19)	0.08 ± 0.01 (0.05)	0.27 ± 0.02 (0.19)	0.07 ± 0.01 (0.05)	0.22 ± 0.03 (0.13)	0.06 ± 0.01 (0.03)	0.06 ± 0.00 (0.03)
S^2	(0.15)	(0.17)	(0.12)	0.19 ± 0.02 (0.18)	0.17 ± 0.01 (0.17)	0.16 ± 0.02 (0.11)	0.16 ± 0.02 (0.12)	0.10 ± 0.00 (0.08)
${}^{3}J_{COCB}$ [Hz]	(0.47)	0.52 ± 0.03 (0.41)	0.50 ± 0.04 (0.40)	0.56 ± 0.09 (0.41)	0.51 ± 0.03 (0.39)	**0.41 ± 0.02 (0.32)**	**0.39 ± 0.05 (0.30)**	**0.40 ± 0.02 (0.31)**
${}^{3}J_{COCO}$ [Hz]	(0.41)	**0.44 ± 0.07 (0.34)**	**0.38 ± 0.05 (0.27)**	0.51 ± 0.07 (0.38)	**0.39 ± 0.04 (0.28)**	**0.43 ± 0.06 (0.33)**	**0.42 ± 0.05 (0.33)**	**0.43 ± 0.07 (0.36)**
${}^{3}J_{COHA}$ [Hz]	(0.79)	0.70 ± 0.18 (0.59)	0.79 ± 0.04 (0.72)	0.81 ± 0.33 (0.68)	0.86 ± 0.10 (0.74)	**0.47 ± 0.10 (0.37)**	**0.40 ± 0.04 (0.33)**	**0.40 ± 0.03 (0.33)**
${}^{3}J_{HNCB}$ [Hz]	(0.53)	0.63 ± 0.04 (0.52)	0.57 ± 0.03 (0.46)	0.66 ± 0.07 (0.51)	0.57 ± 0.04 (0.45)	**0.50 ± 0.03 (0.41)**	**0.47 ± 0.05 (0.38)**	**0.49 ± 0.03 (0.40)**
${}^{3}J_{HNCO}$ [Hz]	(0.81)	0.82 ± 0.07 (0.67)	0.82 ± 0.09 (0.72)	0.83 ± 0.12 (0.68)	0.81 ± 0.07 (0.66)	**0.67 ± 0.07 (0.53)**	**0.67 ± 0.07 (0.56)**	**0.67 ± 0.08 (0.56)**
${}^{3}J_{HNHA}$ [Hz]	(1.62)	1.46 ± 0.11 (1.06)	1.45 ± 0.17 (1.10)	1.64 ± 0.22 (1.29)	1.60 ± 0.09 (1.25)	**1.27 ± 0.08 (0.99)**	**1.23 ± 0.10 (0.96)**	**1.24 ± 0.09 (0.98)**
				Q Factors				
Q_{NH}	0.552 ± 0.038	0.377 ± 0.030	0.347 ± 0.024	0.420 ± 0.003	0.316 ± 0.005	0.236 ± 0.004	**0.225 ± 0.004**	0.235 ± 0.003
Q_{BB}	0.423 ± 0.029	0.339 ± 0.016	0.354 ± 0.013	0.327 ± 0.004	0.289 ± 0.004	**0.207 ± 0.002**	**0.207 ± 0.003**	0.222 ± 0.002
				Backbone Hydrogen-Bonding Potential				
Energy [kcal/mol]	−134.1 ± 12.3	−143.7 ± 13.7	−173.9 ± 12.5	−138.5 ± 13.7	−175.0 ± 12.1	−151.4 ± 6.6	−190.8 ± 9.2	−189.3 ± 8.3
				Backbone Atomic Pairwise RMSD [Å]				
intra-ensemble	1.42 ± 0.39	0.52 ± 0.07	0.48 ± 0.07	1.25 ± 0.5	1.17 ± 0.5	0.97 ± 0.20	0.79 ± 0.10	0.70 ± 0.13
versus X-ray (1UBQ)	1.27 ± 0.16	0.81 ± 0.04	1.02 ± 0.07	1.17 ± 0.30	1.18 ± 0.09	0.83 ± 0.09	0.73 ± 0.08	0.73 ± 0.08
versus NMR refined with RDC (1D3Z)	1.24 ± 0.17	0.80 ± 0.04	0.97 ± 0.07	1.17 ± 0.31	1.19 ± 0.10	0.81 ± 0.10	0.69 ± 0.05	0.68 ± 0.06

[a] Q_{NH} = Q-factor for the N−HN RDCs; Q_{BB} = Q-factor for all backbone RDCs(N−HN, HN−C, CA-HA, N−C, CA-C). Each row reports the quality in terms of RMSD between experimental and calculated NMR observables in the different ensembles. Bold figures indicate the ensemble for which a particular set of scalar and RDCs is best reproduced within statistical errors. Values averaged over the ensemble of the conformations obtained by pooling all the annealing cycles together are given in parentheses. The energy of hydrogen bonds[76] is also estimated.

rhombicity and D_{ZZ} between the starting structure and the calculated ensembles (for more details see ref 27). The direction of the principal axes changed by less than 6°, and the rhombicities changed by 0.06 or less. D_{ZZ} of the calculated ensembles were scaled relative to the D_{ZZ} of the starting structures by factors between 0.90 and 1.06. These changes in the alignment tensor are within the range of expected changes after a readjustment of the tensor to partially absorb internal motional effects.[27]

RDCs were back-calculated for ensembles containing different numbers of conformations. To do so, structures extracted after each annealing cycle were pooled in bins containing increasing numbers of conformations. The RDCs deposited with the PDB files 1D3Z[60] and 1P7F[72] in the BMRB database were used for fitting the alignment tensors of ubiquitin and protein G, respectively. The quality of the agreement with experimental RDC data was assessed by calculating the Q-factor[73]

$$Q = \frac{\sqrt{\sum (RDC_{calc} - RDC_{exp})^2}}{\sqrt{\sum (RDC_{exp})^2}} \qquad (10)$$

An estimate of the error on the Q-factors was obtained by calculating the standard deviation of the Q-factor of 5 bins containing the same number of randomly selected structures.

Backbone scalar couplings were back-calculated from the structures using a Karplus equation;[74,75] they were linearly averaged over the ensembles of structures.

Hydrogen Bond Energies. We calculated the hydrogen bonding potential by using the method of Kortemme et al.,[76] which is based on geometrical parameters of hydrogen bonds observed in high-resolution protein crystal structures.

WHATCHECK Tests. After an energy minimization of 3000 steps of steepest descent in which the restraints were kept in place (with the force constants listed in Tables S1 and S2), the quality of the ensembles that we generated was assessed by using WHATCHECK.[77]

Results

Eight ensembles of structures were calculated representing the native conformation of ubiquitin using different types of restraints and their structural properties are listed in Table 1.

(72) Ulmer, T. S.; Ramirez, B. E.; Delaglio, F.; Bax, A. *J. Am. Chem. Soc.* **2003**, *125*, 9179−9191.
(73) Bax, A.; Kontaxis, G.; Tjandra, N. *Methods Enzymol.* **2001**, *339*, 127−174.
(74) Karplus, M. *J. Chem. Phys.* **1959**, *30*, 11−15.
(75) Bax, A.; Vuister, G. W.; Grzesiek, S.; Delaglio, F.; Wang, A. C.; Tschudin, R.; Zhu, G. *Methods Enzymol.* **1994**, *239*, 79−105.
(76) Kortemme, T.; Morozov, A. V.; Baker, D. *J. Mol. Biol*. **2003**, *326*, 1239−1259.
(77) Hooft, R. W. W.; Vriend, G.; Sander, C.; Abola, E. E. *Nature* **1996**, *381*, 272−272.

Table 2. Comparison of the Scalar and RDCs for the ENJ_2 and DER Ensembles, the X-ray structure (1UBQ), and the Structures Refined with RDCs (1D3Z) of Ubiquitin

	ensemble ENJ_2	DER (1XQQ)	X-ray (1UBQ)	RDC refined (1D3Z)
$^{h3}J_{NC'}$ [Hz]	0.06 ± 0.00	0.19 ± 0.02	0.14	0.20
		r.m.s. Deviation of Backbone Scalar Couplings		
$^{3}J_{COCB}$	0.39 ± 0.05[a](0.30)	0.35 ± 0.01[b](0.33)	0.25	0.26
$^{3}J_{COCO}$	0.42 ± 0.05(0.33)	0.40 ± 0.02(0.39)	0.30	0.29
$^{3}J_{COHA}$	0.40 ± 0.04(0.33)	0.39 ± 0.01(0.38)	0.30	0.32
$^{3}J_{HNCB}$	0.47 ± 0.05(0.38)	0.43 ± 0.02(0.42)	0.31	0.31
$^{3}J_{HNCO}$	0.67 ± 0.07(0.56)	0.63 ± 0.01(0.61)	0.46	0.47
$^{3}J_{HNHA}$	1.23 ± 0.10(0.96)	1.09 ± 0.04(1.06)	0.76	0.70
		Q Factors		
Q_{NH}	0.225 ± 0.004	0.341 ± 0.004	0.17	0.05
Q_{BB}	0.207 ± 0.003	0.257 ± 0.001	0.23	0.10
		Backbone Hydrogen-Bonding Potential		
energy [kcal/mol]	−190.8 ± 9.2	−144.0 ± 14.6	−214.1	−215.2

[a] Ensembles contain 2 structures, values in parentheses are averages over 20 ensembles. [b] Ensembles contain 16 structures,values in parentheses are averages over 8 ensembles.

None of the ensembles generated with NOE restraints exhibits NOE violations larger than 0.3 Å. Structured refined in explicit water provide better predictions for backbone scalar couplings over those refined in vacuo. The introduction of $^{h3}J_{NC'}$ scalar couplings as additional restraints improves slightly, but systematically, the accuracy of the back-calculated backbone scalar couplings; the best predictions are obtained for ensembles $ENJS_2$ and ENJ_2. The latter ensemble has also better backbone scalar coupling predictions than the recently published 1XQQ ensemble,[59] which was determined using NOE-derived distances and S^2 order parameters as restraints (Table 2). RDC-refined structures (1D3Z), by contrast, show better agreement between back-calculated and experimental backbone scalar couplings data. However, the structures determined using RDC refinement employed more than 200 additional dipolar coupling restraints as well as dihedral restraints derived from ^3J couplings through a Karplus equation.[60] Here instead we use only 29 $^{h3}J_{NC'}$ scalar couplings. Our results shows that the use of 29 $^{h3}J_{NC'}$ scalar couplings as restraints (ensembles VNJ_1, VNJ_2, and ENJ_2) improves the prediction of backbone ^3J scalar couplings with respect to ensembles VN_1, VN_2, and EN_2, and are almost as accurate as in the case in which a large number of RDCs and dihedral restraints is used.[60]

To further test the accuracy of the structural ensembles determined we calculated the Q-factors for NH and for several backbone RDCs (Table 1). RDCs contain information about the dynamics of the protein,[27,32] and that structural ensembles provide better predictions of scalar couplings and lower RDC Q-factors than do individual conformations.[59] We therefore calculated Q-factors for RDCs back-calculated from ensembles containing different numbers of structures (see Methods) (Figure 2A). In agreement with previous results,[59] the correlation between predicted and experimental RDCs improves progressively as the number of conformations increases. Moreover, inclusion of $^{h3}J_{NC'}$ scalar couplings as restraints improves the orientation of NH-vectors as indicated by the lowered Q-factor for NH RDCs for ensembles VNJ_2 and ENJ_2 (Table 1). The best correlation between experimental and back-calculated RDCs is achieved by refinement in explicit water using NOEs and $^{h3}J_{NC'}$ scalar couplings as restraints (ensemble ENJ_2). This ensemble has Q-factors close to those calculated for the RAPPER[78] ensemble of ubiquitin (0.24),[59] which is constructed to take explicitly into account the structural variability compatible with diffraction patterns in X-ray crystallography. As 24 of the 29 $^{h3}J_{NC'}$ scalar couplings that were used as restraints are from NH groups located in secondary structural elements, specific Q-factors were back-calculated for RDCs of NH groups in secondary structural elements (Table 3); an improvement of about 0.02 is observed when $^{h3}J_{NC'}$ scalar couplings are used as restraints independent of which solvation model was employed. As Q-factors are worse by at least 0.1 for flexible loop regions than segments located within secondary structure elements, we considered whether the use of $^{h3}J_{NC'}$ scalar couplings for hydrogen bonds connecting such loops with secondary structural elements, or just within the loops, improves the local geometry of flexible regions. In ubiquitin, $^{h3}J_{NC'}$ scalar couplings were measured for four hydrogen bonds located in the loop region of residues 51−64, NH56−O21, NH51−O19, NH61−O56, and NH64−O2.[68] The use of these $^{h3}J_{NC'}$ scalar couplings as restraints clearly improves the Q-factor for the RDCs of NH groups in the region of residues 51−64 when more than one replica is used (Table 3). In the single replica case, the flexibility of the loop region is likely to be much lower because the use of a single replica results in a over-restrained structure,[30] and consequently the Q factors are higher for the one-replica case. A similar result is obtained in the case of S^2 order parameters, which are too large for side chains when only one replica is used. To test whether the observed improvement in local

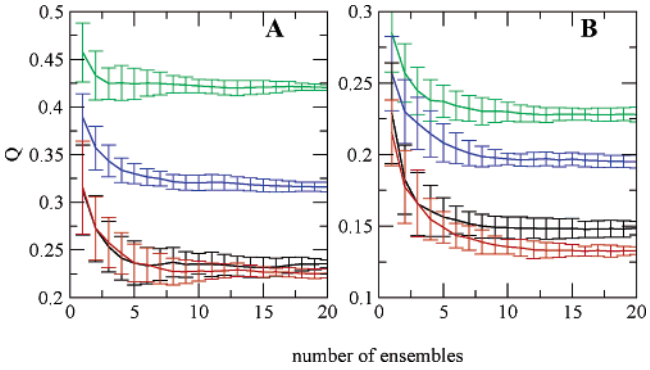

Figure 2. Q-values[73] for N−HN RDCs of (A) ubiquitin and (B) protein G as a function of the number of ensembles in VN_2 (green), VNJ_2 (blue), EN_2 (black), and ENJ_2 (red). Error bars were obtained as described in the Methods section.

(78) DePristo, M. A.; de Bakker, P. I. W.; Blundell, T. L. *Structure* **2004**, *12*, 831−838.

Table 3. Comparison of the Q factors for Different Regions of the Structure of Ubiquitin

ensemble label	E_1	VN_1	VNJ_1	VN_2	VNJ_2	EN_2	ENJ_2	$ENJS_2$
number of replicas	1	1	1	2	2	2	2	2
solvation	E	V	V	V	V	E	E	E
restraints applied		NOE	NOE, $^{h3}J_{NC'}$	NOE	NOE, $^{h3}J_{NC'}$	NOE	NOE, $^{h3}J_{NC'}$	NOE, $^{h3}J_{NC'}$, S^2
Q^{sec}	0.356 ± 0.045	0.216 ± 0.007	0.190 ± 0.020	0.226 ± 0.005	0.211 ± 0.009	0.164 ± 0.010	0.131 ± 0.006	0.137 ± 0.007
Q^{loop}	0.603 ± 0.040	0.408 ± 0.035	0.413 ± 0.010	0.516 ± 0.004	0.369 ± 0.006	0.272 ± 0.006	0.265 ± 0.008	0.266 ± 0.006
$Q^{loop(51-64)}$	0.508 ± 0.057	0.237 ± 0.013	0.241 ± 0.042	0.358 ± 0.010	0.177 ± 0.016	0.135 ± 0.006	0.110 ± 0.005	0.107 ± 0.010

Table 4. Summary of the Results Obtained with the Four Simulations Performed in This Work for Protein G Using Different Types of Restraints

ensemble label	VN_2	VNJ_2	EN_2	ENJ_2
number of replicas	2	2	2	2
solvation	V	V	E	E
restraints applied	NOE	NOE, $^{h3}J_{NC'}$	NOE	NOE, $^{h3}J_{NC'}$
		r.m.s. Deviation		
NOE[Å]	$0.01 \pm 0.00(0.01)$	$0.01 \pm 0.01(0.01)$	$0.01 \pm 0.00(0.00)$	$0.01 \pm 0.00(0.00)$
$^{h}3J_{NC'}$[Hz]	$0.32 \pm 0.02(0.23)$	$0.07 \pm 0.03(0.01)$	$0.23 \pm 0.03(0.16)$	$0.06 \pm 0.00(0.03)$
		Q Factors		
Q_{NH}	0.228 ± 0.005	0.195 ± 0.005	0.149 ± 0.005	0.132 ± 0.003
Q_{BB}	0.216 ± 0.004	0.199 ± 0.005	0.122 ± 0.005	0.119 ± 0.003
		Backbone Hydrogen-Bonding Potential		
energy [kcal/mol]	-114.6 ± 12.1	-155.2 ± 6.7	-136.1 ± 14.6	-167.4 ± 9.6
		Backbone Atomic r.m.s. Differences [Å]		
intra-ensemble	1.03 ± 0.3	0.92 ± 0.3	1.01 ± 0.3	0.87 ± 0.2
versus X-ray structure (1IGD)	1.13 ± 0.2	1.06 ± 0.1	1.06 ± 0.3	0.99 ± 0.2
versus NMR structure refined with RDC (3GB1)	1.08 ± 0.2	1.03 ± 0.1	0.88 ± 0.2	0.81 ± 0.1
versus X-ray refined with RDC (1P7F)	1.06 ± 0.2	1.00 ± 0.1	0.91 ± 0.3	0.84 ± 0.2

geometries is independent of the system investigated, protein G was refined by using the same protocol as for ubiquitin with 32 $^{h3}J_{NC'}$ scalar couplings[69] as restraints. The properties of the resulting protein G ensembles are given in Table 4. Consistent with the results on ubiquitin, the most accurate results were obtained by using NOEs and $^{h3}J_{NC'}$ scalar couplings as restraints (ensemble ENJ_2). While NOE-restrained simulations in explicit water (ensemble EN_2) provide already a good description of the local geometry, as shown by the close agreement of the calculated and the experimental RDCs, a further improvement can be obtained by using $^{h3}J_{NC'}$ scalar couplings (Figure 2B). The simulations in this case were started from a NMR structure (1GB1) obtained without the use of RDCs as restraints.[61] The Q-factors for this structure are 0.35 and 0.32 for the NH and all backbone RDCs, respectively. After refinement with $^{h3}J_{NC'}$ scalar coupling restraints, Q-factors for both NH and all backbone RDCs are very close to those of the X-ray (1IGD)[79] and one of the structures refined by RDCs (3GB1)[49] (Table 5), although the Q-factor of the structure refined with five sets of RDCs measured in different alignment media (1P7F) are even lower.[72]

To understand the differences in geometry and dynamics that give rise to the observed differences in RDCs among the different ensembles that we determined, we investigated the orientation and distributions of the NH vectors in the different ensembles. For ubiquitin, the refinement in explicit water reduces significantly the deviations of the average NH vector positions from the orientations they have in the RDC-refined structure (1D3Z). Within elements of secondary structure, this deviation decreases from 8.0° in VN_2 to 6.2° in EN_2, and with the inclusion of the $^{h3}J_{NC'}$ scalar coupling restraints it is further reduced to 5.5° (ENJ_2). Similar results are observed for protein G, for which the average deviations of NH vector positions with respect to those of the RDC-refined structure (1P7F) in ENG_2 is 5.6°, while it decreases to 4.8 in $ENJG_2$. The accuracy of the representation of the NH vectors is presented for two residues of ubiquitin and of protein G in Figure 3. Figure 3B,D also illustrates that the inclusion of $^{h3}J_{NC'}$ scalar couplings as restraints reduces the amplitude of the motions of restrained NH vectors. Significant reductions in these fluctuations are achieved also by the refinement using an explicit solvent model. The fluctuations around the average orientation are reduced from 11.3° in VN_2 to 9.4° in EN_2, while they are about 8.8° in ENJ_2. In the case of protein G, the use of $^{h3}J_{NC'}$ scalar couplings as restraints reduces the fluctuations from 9.4° in ENG_2 to 8.7° in $ENJG_2$. Hence, if RDCs are back-calculated from an ensemble of structures rather than a single structure, better results are obtained (see also Figure 2) because fluctuations in the orientation of the NH vectors are taken into account.[32,59] Simulations in explicit water appear also capable of defining

Table 5. Comparison of the Scalar and RDCs for the ENJ_2 Ensemble, the X-ray Structure (1IGD), and the Structures Refined with RDCs (3GB1 and 1P7F) of Protein G

	Ensemble ENJ_2	X-ray (1IGD)	RDC refined (3GB1)	RDC refined (1P7F)
$^{h}3J_{NC'}$[Hz]	0.06 ± 0.00	0.17	0.30	0.13
		Dipolar Coupling Agreement		
Q_{NH}	0.132 ± 0.003	0.10	0.11	0.02
Q_{BB}	0.119 ± 0.003	0.12	0.14	0.05
		Backbone Hydrogen-Bonding Potential		
energy [kcal/mol]	-167.4 ± 9.6	−201.9	−139.95	−201.5

(79) Derrick, J. P.; Wigley, D. B. *J. Mol. Biol.* **1994**, *243*, 906−918.

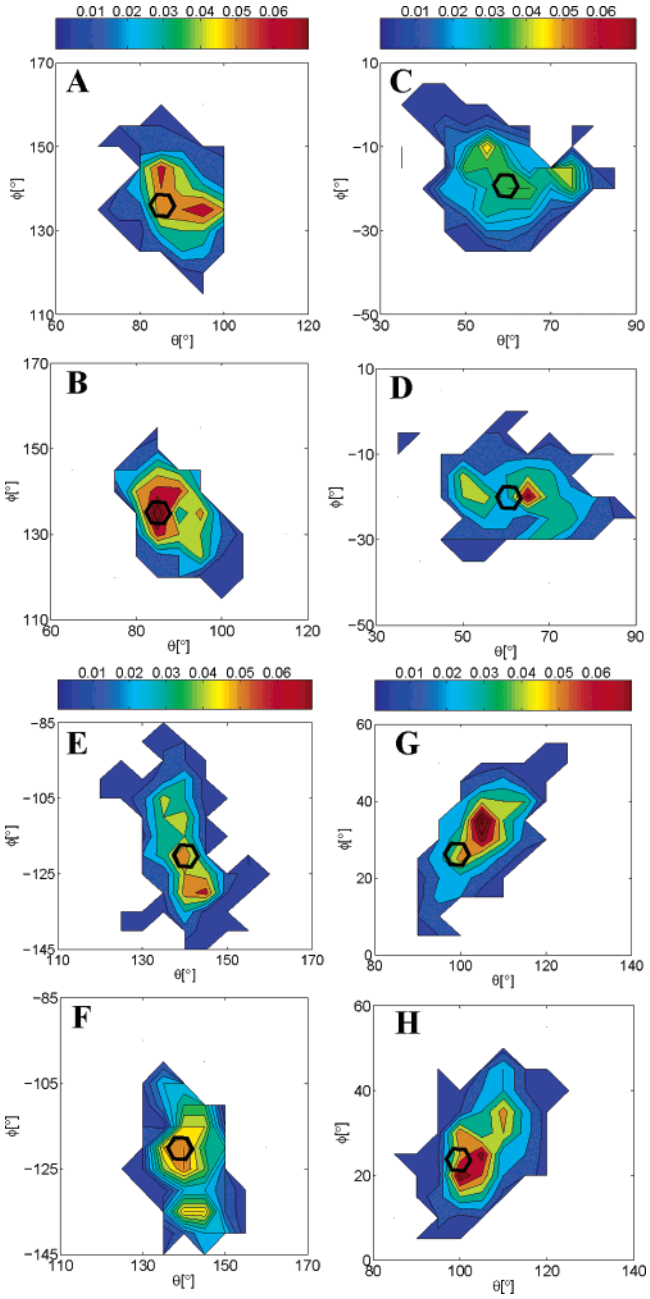

Figure 3. Distributions in polar coordinates of the NH vectors of residues (A,B) 29 and (C,D) 69 in ubiquitin, and of (E,F) 3 and (G,H) 30 in protein G. The results of the simulations with NOE restraints in explicit water with two replicas (EN_2 and EN_2G) are shown in panels A, C, E, and G. The results of the NOE and $^{h3}J_{NC'}$ restraint simulations in explicit water with two replicas (ENJ_2 and ENJ_2G) are shown in panels B, D, F, and H. Black polygons indicate the positions of the vectors in the structures of ubiquitin and protein G refined with RDCs (1D3Z and 1P7F).

accurately the accessible conformational space, so that better predictions of RDCs are possible (ensemble EN_2, Table 1), and further improvement can be obtained by using $^{h3}J_{NC'}$ scalar couplings (ensemble ENJ_2) and S^2 values (ensemble $ENJS_2$) as restraints.

We have assessed the quality of macromolecular structures by determining their Z-scores and RMS Z-score with WHATCHECK[77] (see Methods). We found that RMS Z-scores were distributed around 1 for all the structures determined in this study. Regarding the structure Z-scores, only the second generation packing quality and the Ramachandran plot appearance showed differences between the refinements with and without $^{h3}J_{NC'}$ scalar coupling restraints. As shown in Figure 4, refinement in water significantly improves the structure Z-scores. However, only if $^{h3}J_{NC'}$ scalar coupling are used as restraints,

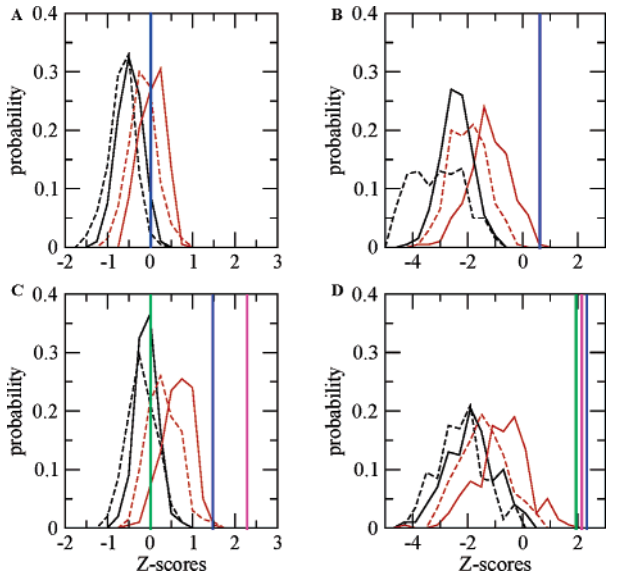

Figure 4. Distribution of the Z-scores of the packing quality and the Ramachandran plot appearance in different ensembles. (A,C) Z-scores of the packing quality in ubiquitin and protein G. (B,D) Z-scores of the Ramachandran plot appearance in ubiquitin and protein G. ENJ_2 and ENJ_2G (red solid line), EN_2 and EN_2G(red dashed line), VNJ_2 and VNJ_2G (black solid line), VN_2 and VN_2G (black dashed line). The blue line in A and B indicates both the Z-score of the X-ray structure of ubiquitin (1UBQ) and the structure refined with RDCs (1D3Z). In C and D, the Z-scores of the X-ray structure of protein G (1IGD), the RDC-refined structure 1P7F, and the RDC-refined structure 3GB1 are indicated by magenta, blue, and green lines, respectively.

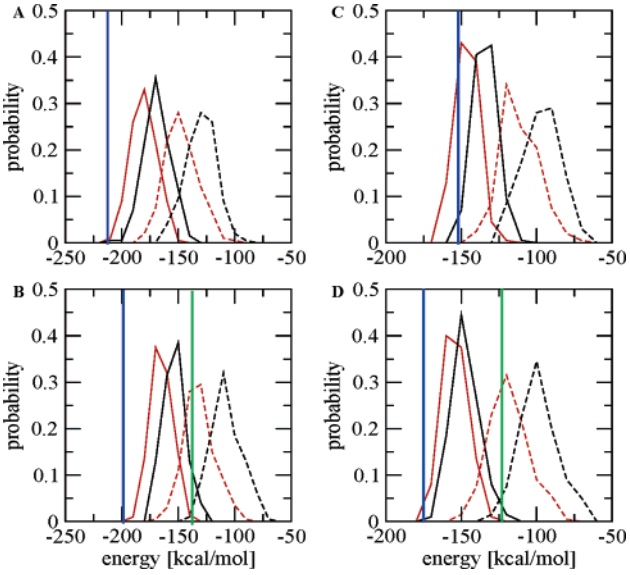

Figure 5. Distributions of the energies[76] of the hydrogen bonds: (A) ubiquitin, (B) protein G; ENJ_2 (red solid line), EN_2 (red dashed line), VNJ_2 (black solid line), VN_2 (black dashed line). Total energies are shown in Table 2. The blue line in (A) indicates both the energy of the X-ray structure of ubiquitin (1UBQ) and the structure refined with RDCs (1D3Z). The blue line in (B) indicates the energy of the X-ray structure of protein G (1IGD) and the RDC-refined structure 1P7F (their difference in hydrogen bond energy is only about 1 kcal/mol); the green line indicates the energy of the RDC-refined structure 3GB1. For comparison, we show also the distributions of the energies of the hydrogen bonds that were restrained in the simulations: (C) ubiquitin; (D) protein G.

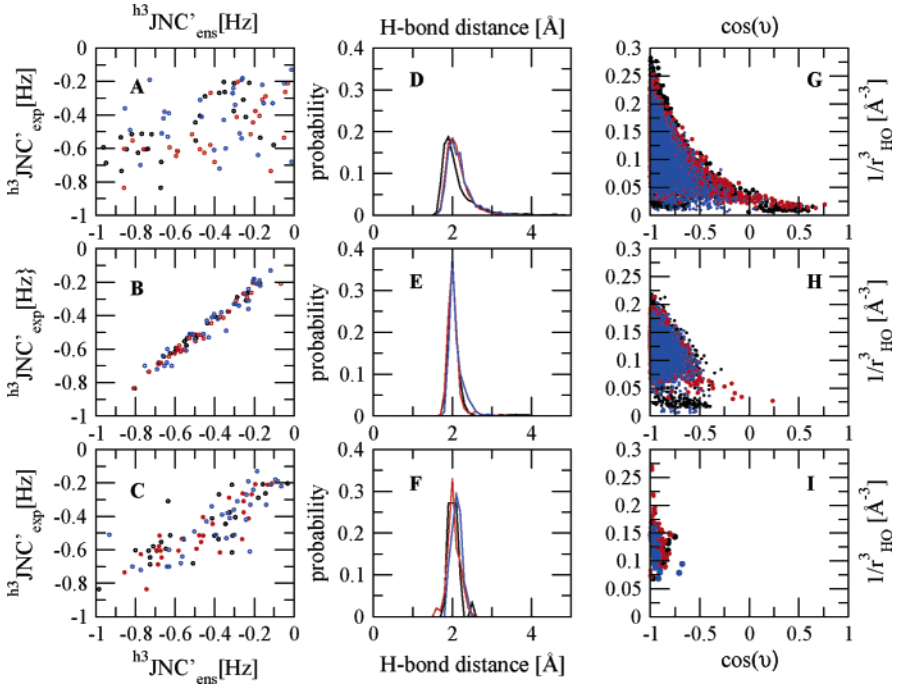

Figure 6. (A, B, C) Correlation between experimental and back-calculated values of $^{h3}J_{NC'}$ of ubiquitin and protein G; (D, E, F) distributions of backbone hydrogen bond distances, r_{HO}; (G, H, I) correlation between r_{HO} and ν. (A, D, G) ensembles VN_2, EN_2, and EN_2G in black, red, and blue, respectively; (B, E, H) ensembles VNJ_2, ENJ_2, and ENJ_2G in black, red, and blue, respectively; (C, F, I)) X-ray structure of ubiquitin (1UBQ), RDC-refined structure of ubiquitin (1D3Z), and RDC-refined structure of protein G (1P7F) in black, red, and blue, respectively.

structures are generated which have Z-scores similar to the ones calculated for high-resolution X-ray and RDC-refined solution structures.

A variety of studies involving surveys of the PDB[76,80] or DFT calculations[66,67] showed that the energies of hydrogen bonds depend on their geometry and dynamics in a complex way. Moreover, highly accurate force fields were recently proposed, based on the analysis of the PDB, to improve the quality of hydrogen bond geometries in NMR-derived structures.[14,76] The ensembles that we report here were also validated by calculating their energies according to a recently developed hydrogen bonding potential based on geometrical parameters of hydrogen bonds observed in high-resolution protein crystal structures.[76] Independent of the solvation model, the use of ensemble-averaged $^{h3}J_{NC'}$ scalar couplings as restraints improves the calculated hydrogen-bond energies significantly (Table 1). The ENJ_2 ensemble is characterized by hydrogen-bond energies approaching those calculated for the structures refined by using RDC restraints (Table 2). Indeed, these energies are equal or even slightly better in the ENJ_2 ensemble if only restrained hydrogen bonds are used in the calculations (Figure 5). These results, together with the corresponding ones for protein G, show that the use of $^{h3}J_{NC'}$ scalar couplings as restraints may provide a description of hydrogen-bond energies of quality comparable to those obtained through the use of RDCs. These results are particularly interesting in the view that the parameters of Baker and collaborators[76] were obtained by a fitting to X-ray structures and they may therefore contain an implicit bias as a result of the construction of hydrogen atoms according to idealized geometries.

The $^{h3}J_{NC'}$ couplings have been reported to depend strongly on three descriptors of the geometry of the hydrogen bonds, r_{HO}, θ and the N−C═O···H dihedral angle.[41,65] In Figure 6, the distributions of backbone hydrogen-bond distances and hydrogen-bond geometries are compared between different ensembles determined here, and the X-ray structure of ubiquitin (1UBQ), the RDC-refined structure of ubiquitin (1D3Z) and the RDC-refined structure of protein G (1P7F). In refinements without $^{h3}J_{NC'}$ coupling restraints, the distances between the hydrogen and the oxygen atoms in secondary structural elements exhibit a rather broad distribution with some values being as large as 4.5 Å. In the X-ray and RDC-refined structures, by contrast, all backbone hydrogen and oxygen atoms in secondary structural elements are separated by less than 2.5 Å. The same conclusion emerges from the analysis of structural ensembles refined with $^{h3}J_{NC'}$ scalar couplings (ensembles VNJ_2, ENJ_2 and ENJ_2G). Moreover, in the latter ensembles the ν angle is rarely smaller than 120°. Hence, the use of $^{h3}J_{NC'}$ scalar coupling restraints, whose magnitude depends in the formalism adopted here (Figure 1 and eq 8) on the distances and angles between the proton and the donor and acceptor atoms, has a significant impact on accuracy of the description of the geometry of hydrogen bonds. In agreement with previous DFT calculations,[14,17] structures refined with $^{h3}J_{NC'}$ scalar couplings indicate that for small hydrogen-bond angles there is a sharp distance limit for a hydrogen bond to maintain the optimal geometry, while linear hydrogen bond angles have much less restriction on the hydrogen-bond length. The existence of a relationship between r_{HO} and ν is the hallmark of highly accurate descriptions of hydrogen bonds.[17]

Conclusions

We have shown that the use of $^{h3}J_{NC'}$ scalar couplings as ensemble-averaged restraints in molecular dynamics simulations enables the determination of highly accurate geometries and

(80) Koch, O.; Bocola, M.; Klebe, G. *Proteins* **2005**, *61*, 310−317.

energetics of hydrogen bonds in the native states of proteins. These couplings therefore represent a valuable complement to NOE-derived inter-proton distances, as do restraints provided by RDC measurements. The results of both these approaches provide a description of hydrogen bonds that is of comparable accuracy to that obtained by high-resolution X-ray crystallography.

Acknowledgment. We thank Dr. Peter Varnai and Dr. Xavier Salvatella for many helpful discussions. J.G. and A.C. are supported by the Swiss National Science Foundation. H.H. is supported by the Finnish Cultural Foundation, the Helsingin Sanomat Centennial Foundation, and the Instrumentarium Science Foundation. M.V. is supported by the Royal Society and by the Leverhulme Trust.

Supporting Information Available: Tables S1 and S2. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0614722