J|A|C|S ARTICLES

Published on Web 11/04/2009

Toward a Unified Representation of Protein Structural Dynamics in Solution

Phineus R. L. Markwick,*[,†,‡,§] Guillaume Bouvignies,[†] Loic Salmon,[†] J. Andrew McCammon,[§] Michael Nilges,[‡] and Martin Blackledge*[,†]

Protein Dynamics and Flexibility, Institute de Biologie Structurale Jean-Pierre Ebel, CNRS-CEA-UJF UMR 5075, 41 rue Jules Horowitz, 38027-Grenoble Cedex, France, Unite de Bio-informatique Structurale, Institut Pasteur, CNRS, URA 2185, F-75015 Paris, France, and Department of Chemistry and Biochemistry, Howard Hughes Medical Institute, University of California San Diego, 9500 Gilman Drive, Urey Hall, La Jolla, California 92003 0365

Received September 3, 2009; E-mail: pmarkwick@ucsd.edu; martin.blackledge@ibs.fr

Abstract: An atomic resolution description of protein flexibility is essential for understanding the role that structural dynamics play in biological processes. Despite the unique dependence of nuclear magnetic resonance (NMR) to motional averaging on different time scales, NMR-based protein structure determination often ignores the presence of dynamics, representing rapidly exchanging conformational equilibria in terms of a single static structure. In this study, we use the rich dynamic information encoded in experimental NMR parameters to develop a molecular and statistical mechanical characterization of the conformational behavior of proteins in solution. Critically, and in contrast to previously proposed techniques, we do not use empirical energy terms to restrain a conformational search, a procedure that can strongly perturb simulated dynamics in a nonpredictable way. Rather, we use accelerated molecular dynamic simulation to gradually increase the level of conformational sampling and to identify the appropriate level of sampling via direct comparison of unrestrained simulation with experimental data. This constraint-free approach thereby provides an atomic resolution free-energy weighted Boltzmann description of protein dynamics occurring on time scales over many orders of magnitude in the protein ubiquitin.

Introduction

Proteins are inherently flexible, displaying a broad range of dynamics over a hierarchy of time-scales from pico-seconds to seconds.[1] This molecular plasticity enables conformational changes in protein backbone and side chains that play critical roles in biomolecular function.[2,3] Nuclear magnetic resonance (NMR) spectroscopy has emerged as the method of choice for studying biomolecular structure and dynamics in solution. All experimentally measured NMR data are affected by motions occurring with characteristic exchange rates that are faster than the so-called chemical-shift range, giving rise to average peaks that represent a potentially complex dynamic average over relatively long time scales (up to the millisecond range for proteins in solution). In addition spin relaxation experiments reflect motions occurring on time scales faster than the molecular rotation diffusion coefficient τ_c (5−20 ns),[4,5] whereas relaxation dispersion can be used to identify sites of slower (μs−ms) conformational exchange.[6−8]

Although the importance of molecular flexibility is generally recognized, standard NMR-based structure determination protocols ignore the presence of protein dynamics, implying that, in common with X-ray crystallography, rapidly exchanging conformational equilibria are routinely represented in terms of a single static structure.[9] The specific averaging properties of different structurally dependent parameters are rarely incorporated into the structure determination procedure, such that the resulting set of coordinates represent a poorly defined average.

The aim of this study is to actively use the rich dynamic information encoded in motionally averaged NMR parameters to develop a structural, dynamic and statistical mechanical molecular representation of the conformational behavior of proteins in solution. Chemical shifts alone are not yet able to describe the dynamics giving rise to the average spectrum, and interproton cross relaxation rates, although rich in structural information, are dependent upon the time scales of the motional processes, and are therefore difficult to interpret quantitatively unless the time scales of the dynamics are known. However, other interactions such as scalar, and possibly more powerfully,

[†] Institute de Biologie Structurale Jean-Pierre Ebel.
[‡] Pasteur Institute.
[§] University of California San Diego.

(1) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. *Science* **1991**, *254*, 1598–1603.
(2) Karplus, M.; Kuriyan, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6679–6685.
(3) Henzler-Wildman, K.; Kern, D. *Nature* **2007**, *450*, 964–972.
(4) Kay, L. E.; Torchia, D. A.; Bax, A. *Biochemistry* **1989**, *28*, 8972–8979.
(5) Palmer, A. G. *Chem. Rev.* **2004**, *104*, 3623–3640.
(6) Mulder, F. A. A.; Mittermaier, A.; Hon, B.; Dahlquist, F. W.; Kay, L. E. *Nat. Struct. Biol.* **2001**, *8*, 932–935.
(7) Palmer, A. G.; Kroenke, C. D.; Loria, J. P. *Methods Enzymol.* **2001**, *339*, 204–238.
(8) Akke, M. *Curr. Opin. Struct. Biol.* **2002**, *12*, 642–647.
(9) Bouvignies, G.; Markwick, P. R.; Blackledge, M. *Chemphyschem* **2007**, *8*, 1901–1909.

 10.1021/ja907476w CCC: $40.75

Residual Dipolar Couplings (RDCs), are exquisitely sensitive to conformational detail[10,11] and therefore may hold the key to resolving this long-standing problem.

Over the past decade, RDCs have emerged as powerful tools for studying proteins in solution, providing simultaneous information about time- and ensemble averaged structural and dynamic processes occurring up to millisecond time-scales and thereby encoding key information for understanding biomolecular function.[12,13] Numerous approaches have been proposed to characterize protein backbone conformation from RDCs; most notably the direct determination of dynamic amplitudes and anisotropies of bond vectors or structural motifs from multiple RDC measurements.[14−24] Molecular dynamics (MD) simulation can also provide access to slower motions that can be compared to measured RDCs,[25,26] however despite increasing computational power, trajectories are usually restricted to time-scales of hundreds of nanoseconds, and millisecond trajectories are still not viable. Relatively long simulations (up to 1.2 μs) have identified slow dynamic processes occurring on time-scales beyond the range probed by spin relaxation,[27,28] which would affect the RDC data, but such long simulations provide only a single trajectory in phase space and do not avoid the problem of statistical mechanical sampling. A popular alternative to performing long simulations is to implement time- or ensemble-averaged restraints,[29−31] thereby constraining a multiple copy molecular description to reproduce the conformationally averaged RDCs.[32−34] Although efficient for identifying conformational ensembles in agreement with experimental data, adding an arbitrary pseudopotential to a physical force field can perturb the simulated dynamics in a nonpredictable manner, making further analysis of the resulting trajectories uncertain. More importantly, the generation of an ensemble of structures that can reproduce the experimental data do not necessarily include the relative free energy weighting of each member of the ensemble. The potential energy surface of a protein may be rugged and highly structured, resulting in a broad distribution of populations in conformational space. To accurately reproduce RDCs or any other NMR observable, it would be necessary to include an accurate population analysis.

In this paper, we present a novel approach aimed at providing a self-consistent structural dynamic representation of protein conformational sampling using a combination of state-of-the-art MD simulation and a large set of experimental NMR data.[33,35−38] In contrast to previously proposed techniques, we avoid using empirical energy terms to guide the conformational search, a procedure that has the potential to perturb both the nature and the time scale of simulated dynamics in a nonpredictable way. Rather, we sample conformational space in an unrestrained way, such that Boltzmann statistics can be respected. To sample conformational space efficiently, we use a recently proposed accelerated molecular dynamics (AMD) approach,[39−41] which biases the actual potential energy surface of the protein to enhance transition probabilities between low energy conformational substates. The appropriate level of acceleration, and therefore sampling of conformational space, is directly determined by matching the reproduction of millisecond-averaged experimentally measured dipolar and scalar couplings to those predicted from the different ensembles. The method is used to describe conformational dynamics occurring on time scales over many orders of magnitude in the prototypical protein system ubiquitin and validated against experimental data sensitive to the diverse time scales.

Results

Figure 1 shows averaged order parameters for $^{15}N-^{1}H$ vectors obtained from MD trajectories seeded from AMD simulations using increasing levels of acceleration (see Methods). A heterogeneous distribution of long time-scale dynamics is observed, with increasing amplitude motions occurring predominantly in loop regions (residues 8−11 and 46−48, and to a lesser extent, residues 19, 20, 22, 36 and 60−62). Slower motions are generally seen in regions identified from previously

(10) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111–1114.
(11) Hus, J. C.; Salmon, L.; Bouvignies, G.; Lotze, J.; Blackledge, M.; Brüschweiler, R. *J. Am. Chem. Soc.* **2008**, *130*, 15927–15937.
(12) Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Nat. Struct. Biol.* **1997**, *4*, 292–297.
(13) Zhang, Q.; Stelzer, A. C.; Fisher, C. K.; Al-Hashimi, H. M. *Nature* **2007**, *450*, 1263–1267.
(14) Tolman, J. R.; Al-Hashimi, H. M. *Annu. Rep. NMR Spectrosc.* **2003**, *51*, 105–166.
(15) Tolman, J. R. *J. Am. Chem. Soc.* **2002**, *124*, 12020–12030.
(16) Meiler, J.; Prompers, J. J.; Peti, W.; Griesinger, C.; Bruschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*, 6098–6107.
(17) Bernado, P.; Blackledge, M. *J. Am. Chem. Soc.* **2004**, *126*, 4907–4920.
(18) Bernado, P.; Blackledge, M. *J. Am. Chem. Soc.* **2004**, *126*, 7760–7761.
(19) Ulmer, T. S.; Ramirez, B. E.; Delaglio, F.; Bax, A. *J. Am. Chem. Soc.* **2003**, *125*, 9179–9191.
(20) Bouvignies, G.; Bernado, P.; Meier, S.; Cho, K.; Grzesiek, S.; Bruschweiler, R.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13885–13890.
(21) Bouvignies, G.; Markwick, P. R. L.; Brüschweiler, R.; Blackledge, M. *J. Am. Chem. Soc.* **2006**, *128*, 15100–15101.
(22) Salmon, L.; Bouvignies, G.; Markwick, P. R. L.; Lakomek, N.; Showalter, S.; Li, D. W.; Walter, K.; Griesinger, C.; Brüschweiler, R.; Blackledge, M. *Angew. Chem., Int. Ed.* **2009**, *48*, 4154–4157.
(23) Tolman, J. R. *Nature* **2009**, *459*, 1063–1064.
(24) Yao, L.; Vogeli, B.; Torchia, D. A.; Bax, A. *J. Phys. Chem. B* **2008**, *112*, 6045–6056.
(25) Showalter, S. A.; Brüschweiler, R. *J. Am. Chem. Soc.* **2007**, *129*, 4158–4159.
(26) Frank, A. T.; Stelzer, A. C.; Al-Hashimi, H. M.; Andricioaei, I. *Nucleic Acids Res.* **2009**, *37*, 3670–3679.
(27) Nederveen, A. J.; Bonvin, A. M. J. J. *J. Chem. Theory Comput.* **2005**, *1*, 363–374.
(28) Maragakis, P.; Lindorff-Larsen, K.; Eastwood, M. P.; Dror, R. O.; Klepeis, J. L.; Arkin, I. T.; Jensen, M.Ø.; Xu, H.; Trbovic, N.; Friesner, R. A.; Iii, A. G.; Shaw, D. E. *J. Phys. Chem. B,* **2008**, *112*, 6155–6158.
(29) Torda, A. T.; Scheek, R. M.; van Gunsteren, W. F. *J. Mol. Biol.* **1990**, *214*, 223–235.
(30) Kemmink, J.; Scheek, R. M. *J. Biomol. NMR* **1995**, *5*, 33–40.
(31) Bonvin, A. M. J. J.; Rullman, J.; Lamerichs, R.; Boelens, R.; Kaptein, R. *Proteins* **1993**, *15*, 385–400.
(32) Clore, G. M.; Schwieters, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 2923–2938.
(33) Lange, O. F.; Lakomek, N. A.; Fares, C.; Schröder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmüller, H.; Griesinger, C.; De Groot, B. L. *Science* **2008**, *320*, 1471–1475.
(34) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433*, 128–132.
(35) Lakomek, N. A.; Walter, K. F. A.; Fares, C.; Lange, O. F.; de Groot, B. L.; Grubmuller, H.; Brüschweiler, R.; Munk, A.; Becker, S.; Meiler, J.; Griesinger, C. *J. Biomol. NMR* **2008**, *41*, 139–155.
(36) Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 12334–12341.
(37) Ruan, K.; Tolman, J. R. *J. Am. Chem. Soc.* **2005**, *127*, 15032–15033.
(38) Briggman, K. B.; Tolman, J. R. *J. Am. Chem. Soc.* **2003**, *125*, 10164–10165.
(39) Hamelberg, D.; Mongan, J.; McCammon, J. A. *J. Chem. Phys.* **2004**, *120*, 11919–11929.
(40) Hamelberg, D.; McCammon, J. A. *J. Am. Chem. Soc.* **2005**, *127*, 13778–13779.
(41) Markwick, P. R. L.; Bouvignies, G.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 4724–4730.

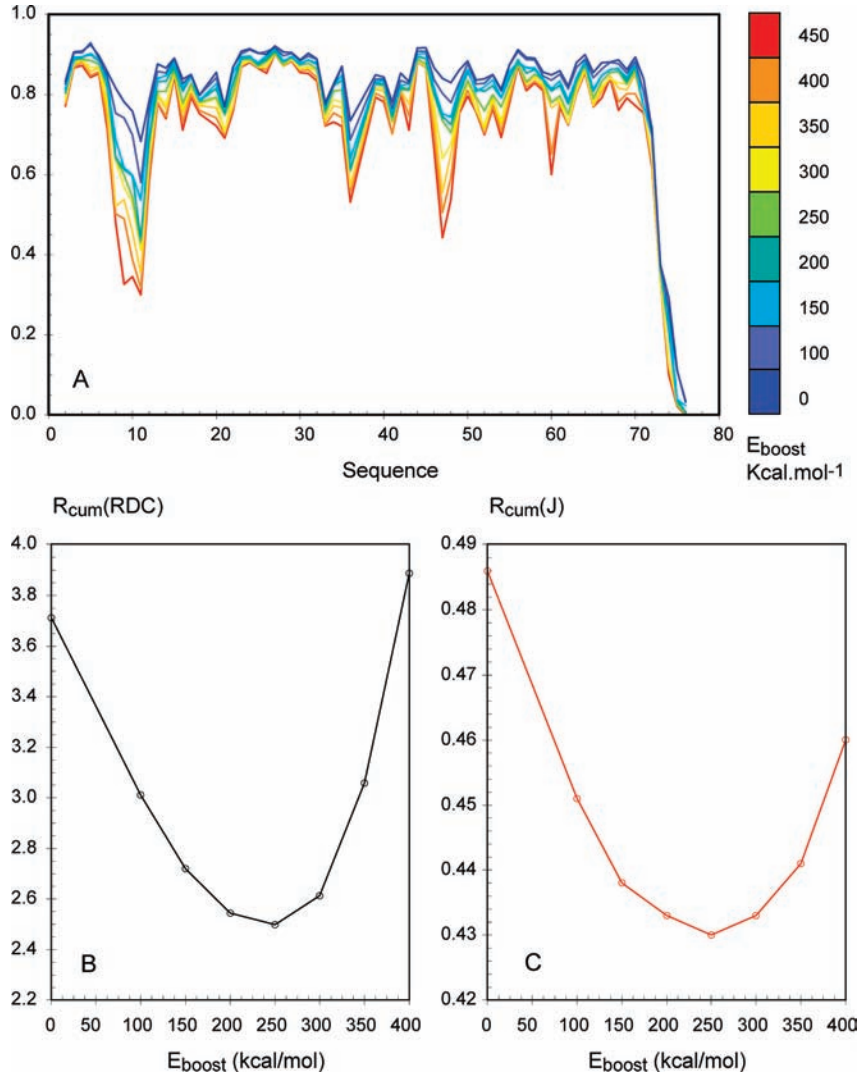

Figure 1. Effect of increasing the acceleration level on $N-H^N$ order parameters in ubiquitin. (A) Order parameters are shown after performing a free energy weighting correction, and are averaged over the trajectories. From top to bottom, the boost energy is 0 (standard 5 ns MD control set), 100, 150, 200, 250, 300, 350, 400, and 450 kcal/mol. The acceleration parameter, α, was fixed at a value of 60 kcal/mol. (B) Change in the trajectory-averaged cumulative R-values for RDCs as a function of the acceleration level. In all cases, the acceleration parameter α = 60 kcal/mol. The boost energy of 0 represents the control set of 5 ns standard MD simulations starting from the X-ray crystal structure[47] using a different random seed generator. (C) Change in the trajectory-averaged cumulative R-values for J-couplings as a function of the acceleration level. In all cases the acceleration parameter α = 60 kcal/mol.

applied long time-scale classical MD simulations of ubiquitin.[25,27,28] A single AMD trajectory, using the equivalent number of steps to 8 ns of classical MD, with relatively low acceleration parameters of ($E_{boost} - V_{dih} = 100$ kcal mol^{-1}, α = 60 kcal mol^{-1}) results in essentially identical conformational sampling to a 150 ns classical MD trajectory ($N-H^N$ order parameters are compared in Figure S1, Supporting Information). This demonstrates that the AMD approach samples meaningful conformational space analogous to standard long MDs.

1. Agreement between Experimental and Theoretical RDCs and Scalar J-Couplings. To assess the most appropriate level of acceleration, we have analyzed the ability of ensembles derived at each value of ($E_{boost} - V_{dih}$) to reproduce experimental RDCs and J-couplings (see Methods). Figure 1 shows the trajectory averaged cumulative R-factor (R_{cum}) for RDCs and scalar J-couplings as a function of the acceleration level, and clearly identifies the "optimum" level at ($E_{boost} - V_{dih}$) = 250 kcal/mol for an acceleration parameter α = 60 kcal/mol. Ensembles generated with less aggressive acceleration sample too little conformational space, while more aggressive acceleration samples too much conformational space to reproduce experimental data.

At the optimal acceleration level, the trajectory-averaged $N-H^N$ RDC cumulative R-factor across all 23 alignment media is 2.496 (average 0.1085). For individual alignment media cumulative R-factors vary between 0.090 and 0.129. Figure 2 shows correlations between experimental and theoretical $N-H^N$ RDCs for four representative data sets with R-factors of 0.096, 0.098, 0.100, and 0.111. Residue specific trajectory-averaged R_{cum} values are compared to a control set calculated from standard 5 ns MD simulations (Figure 3). Only residue 54 shows any increase in the R_{cum} value. Long time-scale dynamics are predominantly located in residues 8−11 and the residue specific R_{cum} values for these residues show improvement compared to the control set. Significant improvement is also observed in residues that show little or no long time-scale dynamics, (e.g.,

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

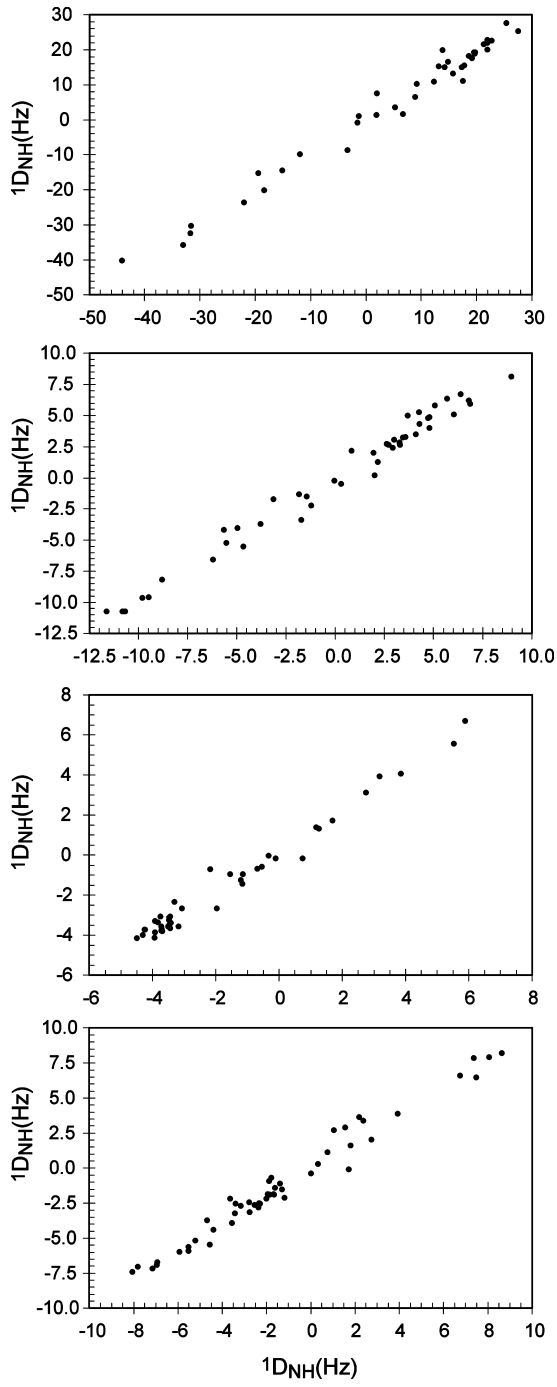

Figure 2. Experimental vs theoretical RDCs for four representative alignment media out of the 23 alignment media. The trajectory averaged cumulative R-factors for the shown alignment media are respectively 0.096, 0.098, 0.100, and 0.111. The trajectory averaged cumulative R-factors across all alignment media varied from 0.090 to 0.129.

15, 17, 34, 42, and 67). The improvement here arises from the more appropriate representation of the time- and ensemble-averaged alignment tensor.

The same level of acceleration also produces the best trajectory-averaged cumulative R-factors for the three scalar J-couplings (H^N-H^α, H^N-C^β and H^N-C'). In comparison to RDCs, scalar J-couplings are less sensitive to the inclusion of long time-scale dynamics, as seen by the relatively small improvement in the R_{cum} values (Figure 2B). This agrees with the previously described phenomenon whereby fitted Karplus

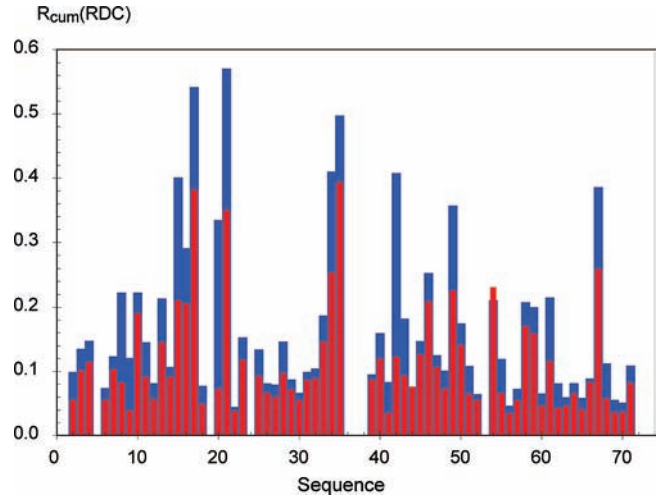

Figure 3. Residue specific trajectory averaged RDC cumulative R-factors. The optimal extended conformational space molecular ensemble [(E_{boost} − V_{dih}) = 250 kcal/mol] is shown in red and compared to a "control set" of standard 5 ns MD simulations shown in blue. Cumulative R-factors for the extended conformational space molecular ensemble are in general lower than those for the control set, confirming the observation of a global improvement in the theoretical RDC data.

parameters can absorb a component of the motion.[42,43] Figure 4 shows the correlation between experimental and theoretical results for the three scalar J-couplings for the molecular ensemble associated with an acceleration level of $E_{boost} - V_{dih} = 250$ kcal/mol. Optimized Karplus curves are compared for $^3J_{NH\text{-}H\alpha}$ obtained from a static structure (1D3Z), standard 5 ns MD simulations and the AMD ensembles and quantum chemistry calculations performed using sum-overstates (SOS) density functional theory (DFT).[44] The curve for the optimal AMD acceleration is almost identical to the DFT-based Karplus curve. A similar effect is observed for the other scalar J-couplings (data not shown).

2. Agreement with Fast Motional Amplitudes Sampled by Spin Relaxation. The free energy weighted molecular ensembles at the RDC-optimum acceleration level provide a representation of the conformational space sampled on time-scales up to the milli-second. These molecular ensembles are composed of individual substates, each with a relative free energy weighting, and from which standard MD simulations have been seeded to probe the fast motions occurring in the local conformational vicinity. Figure 5 depicts the $^{15}N-^1H^N$ order parameters for the fast (ps-ns) dynamics and the effective order parameters probing dynamics on the millisecond time-scale calculated from the free energy weighted ensembles. The fast time-scale $^{15}N-^1H^N$ order parameters obtained by averaging the weighted order-parameters from each substate are in very good agreement with experimental spin relaxation data.[45] In agreement with earlier studies on GB3,[41] we observe an improvement in the agreement between experimental and predicted spin relaxation order parameters when averaging over the extended conformational space ensemble, compared to standard 5 ns MD simulations

(42) Case, D. A.; Scheurer, C.; Brüschweiler, R. *J. Am. Chem. Soc.* **2000**, *122*, 10390–10397.
(43) Markwick, P. R. L.; Showalter, S. A.; Bouvignies, G.; Brüschweiler, R.; Blackledge, M. *J. Biomol. NMR* **2009**, *45*, 17–21.
(44) Malkin, V. G.; Malkina, O. L.; Casida, M. E.; Salahub, D. R. *J. Am. Chem. Soc.* **1994**, *116*, 5898–5908.
(45) Lienin, S. F.; Bremi, T.; Brutscher, B.; Brüschweiler, R.; Ernst, R. R. *J. Am. Chem. Soc.* **1998**, *120*, 9870–9879.

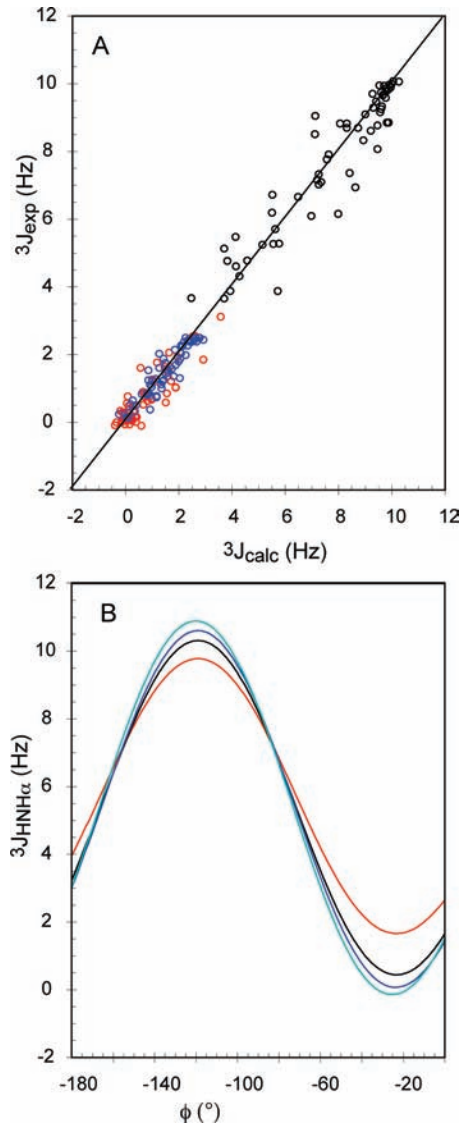

Figure 4. (A) Experimental vs theoretical scalar J-couplings for the optimized extended conformational space molecular ensemble [(E_{boost} − V_{dih}) = 250 kcal/mol]. The three scalar J-couplings are $^3J_{HN\text{-}H\alpha}$ [black circles], $^3J_{HN\text{-}C\beta}$ [red circles], and $^3J_{NH\text{-}C'}$ [blue circles]. (B) NH−H$^\alpha$ Karplus Curves. Red: optimal Karplus curve for ′. Black: optimal Karplus curve for standard 5 ns MD simulation. Blue: optimal Karplus curve for optimal AMD result. Cyan: DFT Karplus curve for NMe-Ala-Ace.

(data not shown). Figure 6 shows a representative bundle of structures for ubiquitin obtained from the molecular ensemble generated from the RDC-optimal acceleration level.

3. Comparison to Single-Copy and Restrained Ensemble Descriptions. Standard NMR structure refinement against experimental observables generates a time- and ensemble-averaged static conformational representation. In the case of ubiquitin, a high resolution static structure has been optimized (1D3Z),[46] against extensive RDCs, nOes, scalar couplings, and hydrogen bonding restraints. The average R-factor per alignment medium for this structure is 0.093 compared to 0.107 for the optimal AMD ensemble. A direct comparison between unrestrained MD and the NMR structure is complicated by the similarity of the different alignment tensors with those used to refine the structure

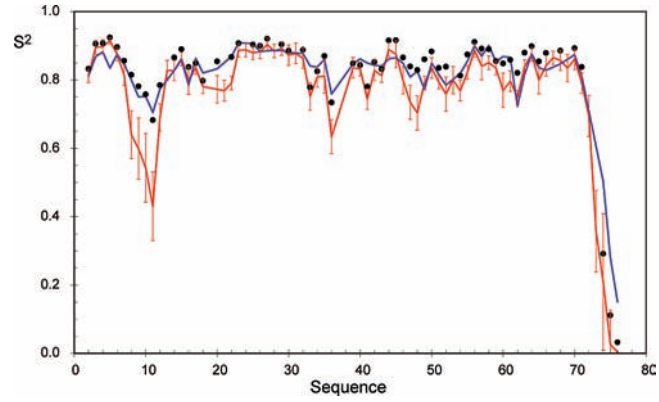

Figure 5. Order parameters for ubiquitin. The ^{15}N spin relaxation experimental data are represented by the blue line.[45] The theoretical fast time-scale (ps−ns) order parameters are shown as black circles and the slow time-scale (RDC-optimized) order parameters are shown in red. The error bars depict the variation in the magnitude of the order parameters for the different molecular ensembles generated from the 20 AMD simulations at the same acceleration level (E_{boost} − V_{dih} = 250 kcal/mol).

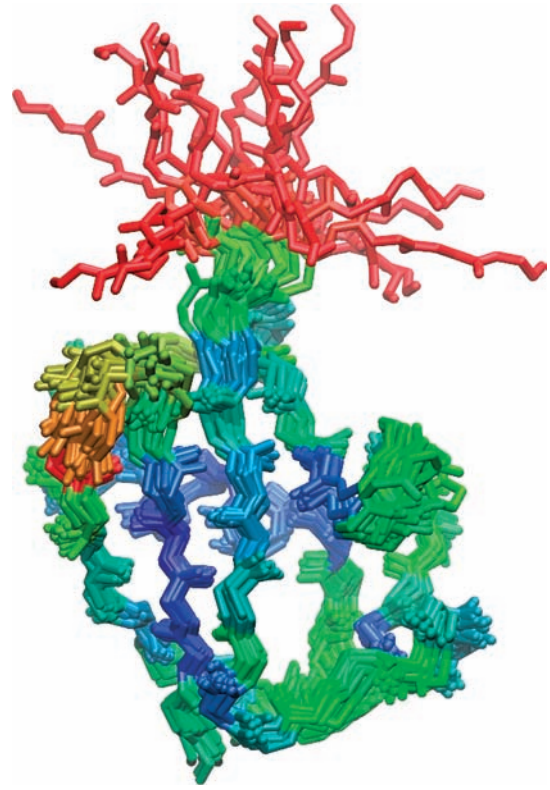

Figure 6. Twenty-four representative structures taken from an RDC-optimized molecular ensemble. The residues are color-coded according to the value of the RDC order parameters (blue: 1.0, red: 0.0).

of 1D3Z. The average RDC R-factor for the optimal AMD ensemble is better than the X-ray crystal structure for ubiquitin (1UBQ)[47] (0.116). Similarly, the AMD approach provides a mean trajectory-averaged J-coupling R-factor (0.143) that reproduces the couplings better than 1UBQ (0.153), and identically to 1D3Z (0.143) despite the fact that these scalar J-couplings were used in the refinement of 1D3Z.

Average backbone coordinates obtained over all free energy-weighted molecular ensembles generated at the RDC-optimal acceleration level are much closer (0.35 Å) to the 1D3Z structure than those obtained from a control set of 5 ns standard MD

(46) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 6836–6837.

simulations (0.55 Å). Thus although the RDC-optimal AMD trajectories sample broader conformational space they are distributed about a mean conformation that resembles the time- and ensemble-averaged static structure. Average structures obtained from the optimal molecular ensembles also exhibit negligible violations to the NOE upper- and lower-bounds. Although a quantitative analysis of cross-relaxation rates would require a more rigorous analysis,[48] this demonstrates that the RDC-optimal AMD ensembles are in qualitative agreement with available structural data.

We have also compared the results of the AMD approach to a recent ensemble restrained molecular dynamics (EROS)[33] description of ubiquitin using extensive NOE data and, in this case, all of the 23 RDC data sets treated in our AMD study. Not surprisingly, as the RDCs were used to directly restrain the EROS ensemble, an SVD analysis of the $N-H^N$ RDCs using this ensemble gives a lower average R-factor (0.066).

Principal component projections of the conformational sampling (Supporting Information Figure S2) reveal that the AMD/SVD approach described here samples essentially the same conformational space as the EROS ensemble, without the need to employ ensemble-averaged restraints. This is remarkable considering that no structural restraints are applied, and the conformational space is only defined via global agreement with the entire data set. Closer inspection reveals that the AMD approach further refines the available conformational space sampling following free energy weighting, therefore providing a more realistic structural dynamic representation of the system.

Finally in the recent EROS study, the order parameters derived from the 116 member ensemble were further scaled by a factor of 0.93, on the basis that, in the opinion of the authors, the ensemble may not include a sufficient representation of librational motions. In our case no an additional scaling is applied, so that the AMD ensemble can be considered as a true molecular representation of the dynamic ensemble giving rise to the experimental data. We note that the extent and nature of the dynamics determined using the AMD approach are quantitatively very similar to that determined using the recently developed three-dimensional Gaussian Axial Fluctuation analysis of the experimental RDCs, where, in comparison to spin relaxation derived order parameters, ubiquitin was shown to be essentially rigid on ns-ms time scales, except for the 8−11 hairpin region and some additional surface loops.[22] A more detailed comparison of our results to those obtained from alternative representations, such as 1D3Z and EROS is provided in the Supporting Information.

Discussion

All NMR parameters are affected by motions occurring on time-scales that are faster than the so-called chemical-shift time-scale, resulting in resonance peaks that represent potentially complex dynamic averages over relatively long times (up to the millisecond range for proteins in solution). In this paper, we have combined a novel AMD/SVD approach with extensive experimental NMR data, to provide an accurate description of the structural dynamic behavior of the protein ubiquitin on time-scales ranging from the picosecond to the millisecond. The results of the AMD simulations performed at different acceleration levels confirm that this method can efficiently and accurately sample extended conformational space explored by globular proteins. The SVD analysis allows the model-free determination of the optimal RDC alignment tensor and the optimal J-coupling Karplus parameters, for a given molecular ensemble, obviating the need for calibration against external references or rescaling of order parameters.

The problems of statistical mechanical sampling associated with the incorporation of additional terms into a hybrid potential energy force field, and thereby perturbing the simulated dynamics, are avoided by using restraint-free trajectories seeded at different points of conformational space sampled by the accelerated MD. The accuracy of the resulting RDCs and scalar couplings is however hardly compromised by this procedure, with a similar level of reproduction compared to state-of-the-art single-structure or restrained-ensemble approaches. Importantly fast motional (ps−ns) order parameters derived from experimental spin relaxation data are well reproduced by the population weighted average over MD simulations performed within the different conformational substates. This important result nicely illustrates the potential, inherent to this approach, of resolving the time scales of different motions for comparison with appropriately sensitive experimental data. The optimal AMD molecular ensemble is therefore in agreement with all available experimental data, giving excellent reproduction of RDCs and scalar J-couplings, experimentally determined nOes, as well as ^{15}N spin relaxation data.

Interestingly, the average backbone structure of the optimal molecular ensemble compares very closely with that of the experimentally refined 1D3Z structure, indicating that although these ensembles sample more conformational space, they appear to be distributed about a mean that resembles the experimentally determined time- and ensemble-averaged structure. In all cases, the free energy weighted extended conformational space ensembles reproduce the experimental observables to a substantially greater degree of accuracy than a control set of 5 ns standard MD simulations and provide better reproduction compared to the static X-ray crystal structure (1UBQ).

Conclusions

The ability to provide an explicit description of protein dynamics in terms of conformational substates and associated populations will undoubtedly improve our understanding of the molecular basis of their biological function, and simultaneously provide an essential basis for interpreting dynamically averaged NMR spectra of proteins. A full characterization of protein dynamics requires an integrated experimental and computational approach. In this study we have therefore used enhanced sampling from biased potential molecular dynamics simulation, combined with extensive experimental dipolar and scalar coupling data, to define a self-consistent representative molecular ensemble for solution state protein conformational dynamics. This approach presents a unified structural dynamic representation of the motional properties of proteins in solution that will provide the basis for furthering our understanding of molecular stability, folding, and function, while simultaneously proposing a new methodology for the interpretation of NMR data in terms of molecular ensembles that will be applicable to a wide range of experimental systems.

Methods

Accelerated Molecular Dynamics. The AMD approach involves adding a continuous non-negative bias potential to the potential energy surface of the protein to raise and flatten the potential energy landscape, thereby enhancing the escape rate between low energy

(48) Brüschweiler, R.; Roux, B.; Blackledge, M.; Griesinger, C.; Karplus, M.; Ernst, R. R. *J. Am. Chem. Soc.* **1992**, *114*, 2289–2302.

conformational substates.[39−41] On increasing the level of acceleration, the simulation probes more conformational space. The essential idea behind accelerated molecular dynamics is to define a reference, or "boost energy", E_b, which is fixed above the minimum of the potential energy surface. At each step in the AMD simulation, if the potential energy of the system lies below this boost energy, a continuous, non-negative bias is added to the actual potential. If the potential energy is greater than the boost energy, it remains unaltered. This results in a raising and flattening of the potential energy landscape, decreasing the magnitude of the energy barriers between low energy states, and therefore enhancing the escape rate from one low energy conformational state to another, while maintaining the essential details of the underlying potential energy surface. The extent to which the potential energy surface is modified depends on the difference between the boost energy and the actual potential. Explicitly, the modified potential, $V^*(r)$, is defined as:

$$V^*(\vec{r}) = V(\vec{r}) \tag{1}$$

if the potential energy, V(r), is equal to or greater than the boost energy, and

$$V^*(\vec{r}) = V(\vec{r}) + \Delta V(\vec{r}) \tag{2}$$

if the potential energy is less than the boost energy. The energy modification, or "bias" is given by:

$$\Delta V(\vec{r}) = \frac{(E_b - V(\vec{r}))^2}{\alpha + (E_b - V(\vec{r}))} \tag{3}$$

The extent of acceleration (i.e., how aggressively we enhance the conformational space sampling) is determined by the choice of the boost energy and the acceleration parameter, α. Conformational space sampling can be enhanced by either increasing the boost energy, or decreasing the acceleration parameter. In the present work, the extent of conformational space sampling was controlled by systematically increasing the boost energy using a fixed acceleration parameter. During the course of the simulation, if the potential energy is modified, the forces on the atoms are recalculated for the modified potential. The use of the bias potential defined above ensures that the derivative of the modified potential will not be discontinuous at points where $V(r) = E_b$.

A series of 20, 8 ns, accelerated molecular dynamics (AMD) simulations of ubiquitin were performed at increasing levels of acceleration using the program AMBER8.[49] In each case α was fixed at 60-kcal/mol and the boost energy for the eight acceleration levels was set at 100, 150, 200, 250, 300, 350, 400, and 450-kcal/mol above the dihedral angle energy (estimated from the average dihedral angle energy from the unbiased 5-ns MD simulations).

In all simulations, a time-step of 1 fs and periodic boundary conditions were used with a Langevin thermostat and a Berendsen weak-coupling pressure-stat. Electrostatic interactions were treated using the Particle Mesh Ewald[50] method with a direct space sum limit of 10 Å. The recently developed ff99SB force field was used.[51]

After reweighting the conformational space to obtain the correct canonical Boltzmann distribution, a clustering protocol was implemented to identify low energy conformational substates. A series of short 3 ns standard MD simulations were then seeded from the AMD simulations, to sample the low energy substates. The initial 0.5 ns were discarded, and a MMPB/SA[52] analysis on the resulting MD simulations was used to confirm the AMD free energy weighting protocol. Using these approximate free energies, a set of large (free energy weighted) structural ensembles was generated from the seeded MD simulations for each acceleration level. Resulting ensembles represent free energy weighted trajectories, sampling the conformational space explored by the AMD trajectories at the relevant acceleration level. This method represents an efficient equivalent to performing numerous long time-scale MD simulations. As a control, a series of molecular ensembles were generated from standard 5 ns MD simulations.

The next step is to identify which ensembles can best reproduce the experimental RDC data. At each increasing acceleration level, we have sampled an increasingly large amount of conformational space. Ideally there should exist an optimum sampling on the time-scale relevant to the RDC data. However, for each molecular ensemble, the optimum alignment tensor for a given alignment medium must be calculated. This is achieved in a model free way using a singular value decomposition (SVD) approach,[53,24] and the analysis is performed for available $N-H^N$ RDCs in 23 different alignment media (see details below). Using the optimized alignment tensors, theoretical RDCs were calculated for each molecular ensemble associated with a given acceleration level. As each molecular ensemble represents a single long time-scale trajectory, the theoretical RDCs for each ensemble associated with the same acceleration level were averaged and the agreement between experiment and theory was monitored using the trajectory averaged cumulative R-factor. A similar protocol was performed to calculate scalar J-couplings as outlined below.

Singular Value Decomposition (SVD) and Calculation of RDCs and Scalar J-Couplings. The principal difficulty concerned with the direct calculation of RDCs and scalar J-couplings arises in the determination of parameters defining the strength of the interaction. In the case of RDCs, five unknown parameters are required to explicitly define the alignment tensor. As our simulations are performed in explicit solvent, in the absence of any alignment medium, it is not possible to define explicitly from the simulation alone the preferential alignment of the molecule in a given alignment medium. The issue is further complicated by the fact that the structure, dynamics and preferential alignment of the molecule are mutually dependent: The alignment tensor depends on the shape and anisotropy of the molecule, which is specifically related to the structure. Dynamic motions on different time-scales result in small changes in the shape and anisotropy of the molecule, which, in turn result in small changes in the preferential alignment tensor for a given alignment medium. The approach taken in this work involves the use of an SVD analysis to determine the optimal alignment tensor for each molecular ensemble directly from the experimental data.[53,24] SVD is an exquisite method for solving a set of simultaneous equations. Explicitly, the optimal alignment tensor and RDCs for each molecular ensemble were calculated in a reduced form as:

$$\begin{bmatrix} \langle y^2 - x^2 \rangle & \langle z^2 - x^2 \rangle & \langle 2xy \rangle & \langle 2xz \rangle & \langle 2yz \rangle \\ \langle y^2 - x^2 \rangle & \langle z^2 - x^2 \rangle & \langle 2xy \rangle & \langle 2xz \rangle & \langle 2yz \rangle \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \langle y^2 - x^2 \rangle & \langle z^2 - x^2 \rangle & \langle 2xy \rangle & \langle 2xz \rangle & \langle 2yz \rangle \end{bmatrix} \begin{bmatrix} A_{yy} \\ A_{zz} \\ A_{xy} \\ A_{xz} \\ A_{yz} \end{bmatrix} = \begin{bmatrix} D_{1\text{red}} \\ D_{2\text{red}} \\ \cdots \\ \cdots \\ D_{N\text{red}} \end{bmatrix} \tag{4}$$

where x,y,z are the Cartesian components of the normalized bond vector of interest (in the case of N−H RDCs, the N−H bond vector), A_{ij} is a vector containing the five components necessary to completely define the 3×3 alignment tensor (bearing in mind that this tensor is symmetric and traceless) and $D_{i\text{red}}$ is a vector containing the experimental RDCs for the particular alignment medium. The matrix on the left-hand side of the equation, which

(49) Case, D. A.; et al. *AMBER 8*; University of California: San Francisco, CA, 2004.
(50) Cheatham, T. E.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4193–4194.
(51) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712–725.
(52) Massova, I.; Kollman, P. A. *J. Am. Chem. Soc.* **1999**, *121*, 8133–8143.
(53) Losonczi, J. A.; Andrec, M.; Fischer, M. W.; Prestegard, J. H. *J. Magn. Reson.* **1999**, *138*, 334–342.

describes the bond vector fluctuations, has dimensions (N,5), where N is the number of RDCs in the given alignment medium. This matrix is formulated from the molecular ensemble where the brackets $\langle .. . \rangle$ represent ensemble averages. SVD of the matrix of bond vector fluctuations liberates the optimal alignment tensor components, from which the theoretical RDCs can be calculated. The principle behind such an analysis is that there should exist some optimal ensemble which represents the conformational space sampled by the system over the time-scales to which RDCs are sensitive (i.e., up to 10-ms for N−H RDCs). For this optimal molecular ensemble, and its optimal SVD-calculated alignment tensor, the resulting theoretical RDCs will be in best agreement with the experimental observables. For molecular ensembles that sample too little or too much conformational space, the SVD analysis will attempt to find the best possible alignment tensor for that particular ensemble, but the resulting RDCs will not be optimal. As mentioned above, we perform a series of AMD simulations at increasing acceleration levels to obtain a set of free energy weighted molecular ensembles that systematically sample an increasing amount of conformational space. By using the SVD analysis to obtain the optimal alignment tensor and hence the theoretical RDCs for each molecular ensemble, we can identify the most appropriate acceleration level (ie. the optimal conformational space sampling) to reproduce the experimental RDCs.

A similar approach was also applied to calculate the backbone scalar J-couplings: We calculated three backbone scalar J-couplings, $^3J(H^N,H\alpha)$, $^3J(H^N,C\beta)$, and $^3J(H^N,C')$. The magnitude of all these J-couplings is strongly related to the backbone ϕ angle and can in general be described using the well-known Karplus equation:[54]

$$^3J(i,j) = A\cos^2(\varphi + \theta) + B\cos(\varphi + \theta) + C \qquad (5)$$

where A, B, and C are the Karplus parameters, and θ is an offset angle, which typically has a value of 180° for $^3J(H^N,C')$, −60° for $^3J(H^N,H\alpha)$ and 60° for $^3J(H^N,C\beta)$. To calculate these scalar J-couplings, we used the SVD analysis to obtain the optimal Karplus parameters for each molecular ensemble:

$$\begin{bmatrix} \langle\cos^2(\varphi_1 + \theta)\rangle & \langle\cos(\varphi_1 + \theta)\rangle & 1 \\ \langle\cos^2(\varphi_2 + \theta)\rangle & \langle\cos(\varphi_2 + \theta\rangle & 1 \\ \ldots & \ldots & \ldots \\ \langle\cos^2(\varphi_N + \theta)\rangle & \langle\cos(\varphi_N + \theta)\rangle & 1 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} J_1 \\ J_2 \\ \ldots \\ J_N \end{bmatrix} \qquad (6)$$

In each case, the analysis was initially performed using the typical θ offset angles defined above. The θ-offset values were then optimized by changing the θ-offset value in 1° steps and repeating the SVD analysis until the best reproduction of the experimental scalar J-couplings was achieved.

Acknowledgment. This work was supported by the EU through EU-NMR JRA3 and French Research Ministry through ANR Protein-Motion PCV (2007) ANR-07-PCVI-0013 and ANR-06-CIS6-012-01. The work at UCSD was supported in part by NSF, NIH, HHMI, CTBP, and NBCR.

Supporting Information Available: Detailed description of the data analysis and of the comparison of different ensemble descriptions of ubiquitin. Complete ref 49. This material is available free of charge via the Internet at http://pubs.acs.org.

JA907476W

(54) Karplus, M. *J. Chem. Phys.* **1959**, *30*, 11. Karplus, M. *J. Am. Chem. Soc.* **1963**, *85*, 2870–2871.