JCTC Journal of Chemical Theory and Computation

Explicit Hydrogen-Bond Potentials and Their Application to NMR Scalar Couplings in Proteins

Jing Huang and Markus Meuwly*

Department of Chemistry, University of Basel, Klingelbergstrasse 80, 4056 Basel, Switzerland

Received October 27, 2009

Abstract: Hydrogen bonds (H bonds) are fundamental for the stability, structure, and dynamics of chemically and biologically relevant systems. One of the direct means to detect H bonds in proteins is NMR spectroscopy. As H bonds are dynamic in nature, atomistic simulations offer a meaningful way to characterize and analyze properties of hydrogen bonds, provided a sufficiently accurate interaction potential is available. Here, we use explicit H-bond potentials to investigate scalar coupling constants $^{h3}J_{NC'}$ and characterize the conformational ensemble for increasingly accurate intermolecular potentials. By considering a range of proteins with different overall topology a general procedure to improve the hydrogen-bonding potential ("morphing potentials") based on experimental information is derived. The robustness of this approach is established through explicit simulations in full solvation and comparison with experimental results. The H-bond potentials used here lead to more directional H bonds than conventional electrostatic representations employed in molecular mechanics potentials. It is found that the optimized potentials lead to H-bond geometries in remarkable agreement with previous *ab initio* and knowledge-based approaches to H bonds in model systems and in proteins. This suggests that, by combining theory, computation, and experimental data, H-bonding potentials can be improved and are potentially useful to better study coupling, energy transfer, and allosteric communication in proteins.

Introduction

Hydrogen bonds are ubiquitous in chemical and biological systems and are essential for the overall structure, function, and dynamics of proteins and other macromolecules.[1] The role of hydrogen bonds in protein folding,[2] the formation of secondary structural elements,[3,4] molecular recognition,[5,6] and catalysis[7,8] has been established over the past few years. A central feature of H bonds is their directionality, which cannot be easily captured by a superposition of isotropic interactions such as Coulomb interactions, as is done in customary force fields such as CHARMM, AMBER, or OPLS-AA.[9–11] In small molecules, a hydrogen bond can be characterized spectroscopically. For example, in complexes between simple ions (HCO^+, HN_2^+) and rare gas atoms (He, Ne, Ar), it is found that the fundamental infrared transitions in the electronic ground state correspond to $\Sigma-\Sigma$ transitions characteristic for linear molecules.[12–14] This can be inferred from the structure of the ro-vibrational bands (missing Q branch). Also, fitting of a model Hamiltonian[15] allows for a determination of structural constants, which in turn characterize the average geometry of the molecule. For biological macromolecules, it is more difficult to find direct measures for the directionality of H bonds and to locate the positions of the hydrogen atoms. Direct visualization through recording of a structure is impractical, as H atoms can usually not be seen in X-ray crystallography. Structure determination from NMR data, on the other hand, formulates a search problem in structure space which minimizes a cost function that involves the experimental information (usually nuclear Overhauser data) and additional physical information because experimental data are rarely sufficient to determine the three-dimensional structure of a macromolecule.[16]

One experimental signature which recently became more widely available is hydrogen bond scalar couplings, which

* Corresponding author e-mail: m.meuwly@unibas.ch.

10.1021/ct9005695
Published on Web 01/22/2010

can be measured through NMR spectroscopy.[17,18] Scalar couplings across N−H···O=C H bonds in proteins have a typical range of about −0.2 to −1 Hz, and the measurement errors are usually less than 0.05 Hz.[19,20] Scalar $^{h3}J_{NC'}$ couplings have been observed experimentally in peptides,[21] nucleic acids,[22] and a variety of proteins.[18,23–28] Together with other NMR parameters such as relaxation times, residual chemical shift anisotropy, and dipolar couplings, $^{h3}J_{NC'}$ couplings are important in the identification of conformational dynamics taking place on the NMR time scale.[29–31] Further interesting and fundamental aspects of $^{h3}J_{NC'}$ couplings are their sensitivity to H-bonding network dynamics and cooperativity. Such effects are very difficult to probe directly through experiments, and a combined approach including atomistic simulations may prove advantageous. Earlier work established that the explicit dynamics of the solvated protein have to be taken into account to reliably calculate scalar coupling constants from molecular dynamics (MD) simulations.[32,33] This naturally paves the way to improve specific terms in empirical force fields to which the observables are sensitive. In the present case, it is the capability of a force field to correctly describe H bonds.

It has been found that $^{h3}J_{NC'}$ values can be directly correlated with H-bond geometries. Barfield proposed several empirically parametrized formulas which enable the calculation of scalar couplings from the local N−H···O=C structure.[34] As NMR spectroscopy is a time-domain method, measured scalar couplings have to be understood as time averages. From a computational point of view, molecular dynamics simulations are the method of choice for such investigations. In previous work,[32,33] a good correlation between measured $^{h3}J_{NC'}$ couplings and those derived from all-atom simulations was established by carrying out nanosecond MD simulations and averaging $^{h3}J_{NC'}$ values over entire trajectories.

Here, we combine a recently developed explicit hydrogen potential (molecular mechanics with proton transfer - MMPT)[35,36] derived from correlated quantum mechanical calculations with an established force field to characterize $^{h3}J_{NC'}$ couplings in a variety of proteins covering different folds (ubiquitin ($\alpha + \beta$), the GB1 domain of protein G ($\alpha + \beta$), cold-shock protein A (all β), apo-calmodulin (all α), holo-calmodulin (all α), and intestinal fatty acid binding protein (all β), see Figure 1). Compared with conventional MD studies, the deviations between calculated and experimental $^{h3}J_{NC'}$ values are notably lowered. Next, the topology of the potential energy surfaces for the H-bond potentials is modified through morphing transformations[37,38] to best describe experimentally determined couplings for three proteins. This approach is then generalized by applying it to the proteins not belonging to the training set, and very good agreement with measured coupling constants is found. Most notably, the approach pursued here leads to an average separation between the hydrogen atom and the acceptor of 1.93 Å, which agrees with a knowledge-based potential derived from 52 structures and results from electronic structure calculations.[39,40]

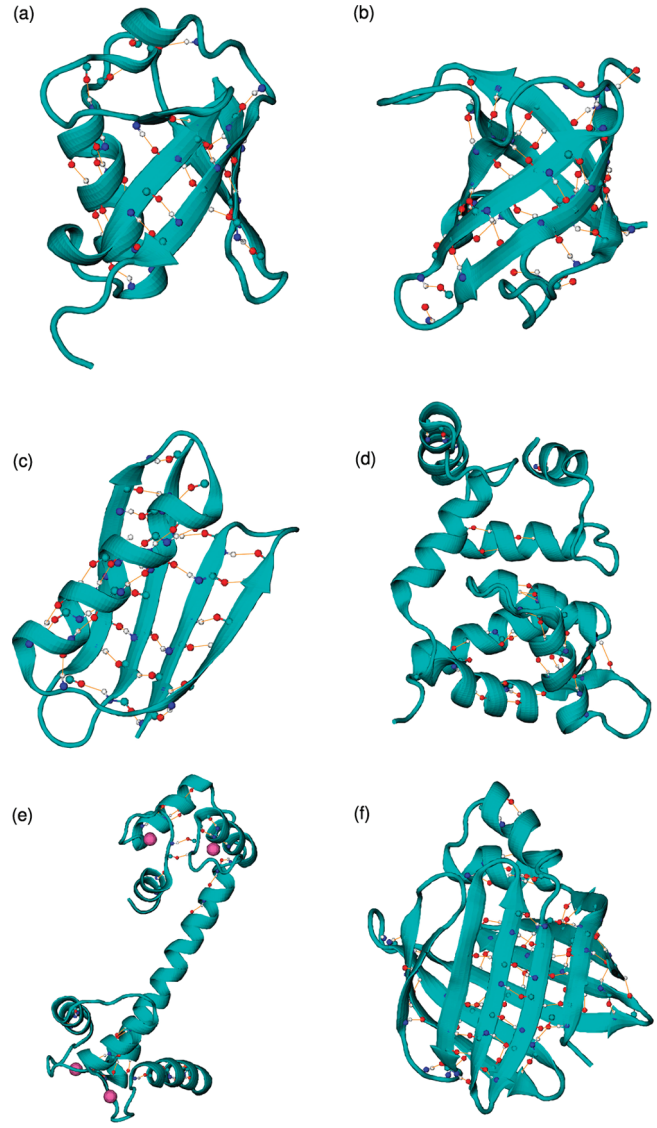

Figure 1. Structure, topology, and H bonds for the six proteins investigated here. (a) Ubiquitin, (b) GB1 domain of protein G, (c) cold-shock protein A, (d) apo-calmodulin, (e) holo-calmodulin, (f) intestinal fatty acid binding protein.

Computational Methods

Molecular Dynamics Simulations. All simulations were carried out with the Charmm program[41] using the CHARMM22 force field[9] and provisions for MMPT.[36] The starting structures were taken from the X-ray structures in the Protein Data Bank[42] (ubiquitin, 1ubq;[43] protein G,[44b] 2qmt;[44a] cold-shock protein A (CspA), 1mjc;[45] apo-calmodulin (apoCAM), 1qx5;[46] holo-calmodulin (holoCAM), 1cll;[47] intestinal fatty acid binding protein (IFABP), 1ifc.[48] Hydrogen atoms were generated with HBUILD,[49] and the structures were relaxed by 3000 steps of steepest descent minimization. Then, the proteins were solvated in pre-equilibrated water boxes of suitable sizes (1ubq, 65.19 Å × 52.77 Å × 49.67 Å; 2qmt, 55.88 Å × 46.56 Å × 40.36 Å; 1mjc, 52.77 Å × 52.77 Å × 46.56 Å; 1qx5, 71.40 Å × 58.98 Å × 58.98 Å; 1cll, 90.03

Å × 62.09 Å × 49.67 Å; 1ifc, 65.19 Å × 55.88 Å × 52.77 Å), and periodic boundary conditions were applied. A cutoff of 14 Å was applied to the shifted electrostatic and switched van der Waals interactions. Before free dynamics simulations, the systems were heated to 300 K and then equilibrated for 10^5 time steps.

For conventional MD simulations, all hydrogen atoms were constrained by SHAKE,[50] whereas for simulations with MMPT, hydrogen atoms involved in $^{h3}J_{NC'}$ couplings were free to move and all other hydrogen atoms were treated with SHAKE. A complete list of H bonds treated by MMPT for all proteins is summarized in Supporting Information S2. In both standard MD and MMPT/MD simulations, the time step was 0.2 fs, and snapshots were taken every 0.02 ps. The hydrogen-bond coordinates were extracted from trajectories and used together with eq 1 to calculate $^{h3}J_{NC'}$ couplings:[32,34]

$$^{h3}J_{NC'} = (-366\ \text{Hz})\exp(-3.2r_{HO'})[\cos^2\theta_1 - (0.47\cos^2\rho + 0.70\cos\rho + 0.11)\sin^2\theta_1] \quad (1)$$

where $r_{HO'}$ is the distance between hydrogen and acceptor atoms, while θ_1 and ρ represent the H···O=C′ angle and the H···O=C′−N′ dihedral angle, respectively.

A simplified formula (eq 2) is also proposed in ref 34 and was used in a previous work:[33]

$$^{h3}J_{NC'} = (-360\ \text{Hz})\exp(-3.2r_{HO'})\cos^2\theta_1 \quad (2)$$

It captures the dominant effects of scalar couplings, while eq 1 provides a better estimate of $^{h3}J_{NC'}$ couplings in protein G because it accounts for the systematic difference between hydrogen bonds along the α helix and β sheet, respectively, by including a term related to the dihedral angle ρ.[34] Equations 1 and 2 can provide the same accuracy as full DFT calculations[32] and have been used to calculate $^{h3}J_{NC'}$ couplings in different proteins.[30,32−34,51] Detailed investigations on small molecules compared the performance of DFT using VWN, BP, or PW91 functionals with results from correlated methods such as the coupled cluster singles and doubles polarization propagator approximation[52] and found that, with the exception of the HF molecule, the performance of DFT is good and provides almost quantitative spin−spin coupling constants.[53] The sensitivity to changes in the parameters of eq 2 has recently been investigated in a systematic fashion.[33] It was found that, overall, a strength factor of $\alpha = -360$ Hz and a decay of $\beta = 3.2$ Å^{-1} provide a good description of most coupling constants. However, for scalar couplings in particular secondary structural elements, the values for α and β could be optimized. As in the present work such aspects are not further pursued, eqs 1 and 2 are used, and the results are virtually identical. Generally, eq 1 leads to slightly smaller deviations between calculated and measured $^{h3}J_{NC'}$ values, so it was used in this work to calculate scalar couplings in all proteins except for CspA, where three backbone−side chain couplings were also included. While distinction between α-helix and β-sheet hydrogen bonds is only relevant for backbone−backbone hydrogen bonds, eq 2 has to be applied for computing these $^{h3}J_{NC'}$ values in CspA.

The quality of the simulations was assessed by comparing root-mean-square deviations (RMSDs) between calculated and experimental $^{h3}J_{NC'}$ couplings:

$$\text{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(J_i^{\text{calcd}} - J_i^{\text{exptl}})^2} \quad (3)$$

MMPT Potential and Morphing Transformations. A detailed account of MMPT has been given in ref 36. Briefly, MMPT uses parametrized three-dimensional potential energy surfaces fitted to high-level *ab initio* calculations (MP2/6-311++G(d,p)) to describe the interactions within a general DH−A motif, where D is the donor, H is the hydrogen, and A is the acceptor atom. Together with a standard force field—here, CHARMM[9] is used—specific rules control how bonded interactions on the donor and acceptor side are switched on and off depending on the position of the transferring H atom (DH−A or D−HA). To adapt the overall shape of the PES to topologically similar, but energetically different, hydrogen bonding patterns—depending on the chemical environment of D and A—the PES can be "morphed".[37,38] Morphing can be a simple coordinate scaling or a more general coordinate transformation depending on whether the purpose of the study and the experimental data justify such a more elaborate approach.

For the present case of hydrogen bonds between an amide (NH) group as the donor and the oxygen atom as the acceptor (NH···O), the MMPT potential depends on R (distance between N and O), ρ (relative position of H for a particular value of R), and θ (angle between unit vectors $\vec{R}$ and $\vec{\rho}$). The relationship between ρ and the N−H distance r is given by

$$\rho = (r - r_{\text{min}})/(R - 2r_{\text{min}}) \quad (4)$$

where $r_{\text{min}} = 0.8$ Å is, in principle, arbitrary but should be sufficiently small to cover the shortest D−A separations. The angular dependence of the potential $V(R, \rho, \theta)$ is harmonic, that is, $V(R, \rho, \theta) = V_0(R, \rho) + k\theta^2$, and a typical PES along R and ρ is shown in Figure 2.

As mentioned above, the MMPT potentials are calculated for model systems (zeroth-order potential) and subsequently morphed to describe the situation in the actual chemical environment. Here, the asymmetric zeroth-order potential for $NH_4^+\cdots OH_2$ is morphed to describe the N−H···O=C motif in proteins. Morphing is achieved by modifying the parameters and thus reshaping the MMPT potential. The original potential has a single minimum $\{R_0 = 2.71$ Å, $\rho_0 = 0.23$, $\theta_0 = 0°\}$ and is mapped to a new one $\{R' = R_0 + \sigma$, $\rho' = \rho_0 - \delta$, $\theta_0 = 0°\}$ where σ and δ are positive because hydrogen bonds in proteins are weaker than in a protonated ammonia−water dimer. Since there is a one-to-one correspondence between morphing parameters $\{\sigma, \delta\}$ and PES minima $\{R', \rho'\}$, only one set—$\{R', \rho'\}$—will be used in the following. The morphed potential has its minimum energy at $\{R', \rho', 0\}$ while maintaining its overall shape, as illustrated in Figure 2.

For most X-ray structures, typically the coordinates of heavy atoms are available since only very rarely can protein crystallography resolve the positions of hydrogen atoms. Therefore, the experimental observable characterizing a

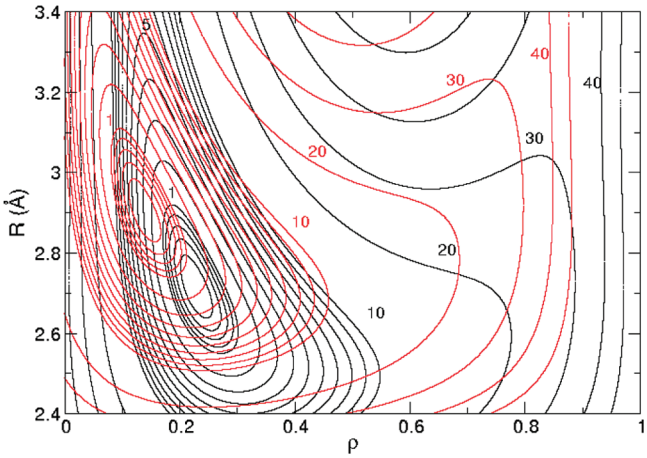

Figure 2. MMPT PES for the NH−O motif and illustration of PES morphing. Black, original PES; red, morphed PES. Contour lines are drawn at intervals of 0.2 kcal/mol for energies below 1 kcal/mol, for energies between 1 and 10 kcal/mol at intervals of 1 kcal/mol, and for higher energies at intervals of 10 kcal/mol. Morphing parameters $\{R', \rho'\} = \{2.92, 0.14\}$.

hydrogen bond is the D−A distance between donor and acceptor. In the Results section, it will be shown that best value for R' corresponds closely to the average D−A distances calculated from the initial X-ray structures. We also establish a relationship between optimized R' and ρ' (eq 5, see below). This leads to the following procedure for optimizing MMPT parameters and calculating $^{h3}J_{NC'}$ couplings by MMPT/MD simulations:

(1) From the X-ray/NMR structure, the average distance R' is calculated.
(2) Compute ρ' by eq 5 (see below).
(3) Morph the MMPT PES to the minima $\{R', \rho'\}$ by coordinate transformations.
(4) Carry out MD simulations with the MMPT potential, and calculate the hydrogen-bond scalar couplings according to eq 1 or 2.

Results

Conventional MD as Benchmarks. Standard MD simulations 1 ns in length were first carried out for all six proteins, and the RMSDs between calculated and experimental couplings were computed as benchmarks for comparison. As shown in Supporting Information S3, $^{h3}J_{NC'}$ couplings converge well within 1 ns. Hence, the RMSD as the average over all $^{h3}J_{NC'}$ couplings is also stable during our simulation time scale; for example, the RMSDs of CspA calculated from 0.5, 1, and 1.5 ns standard MD trajectories are 0.198, 0.195, and 0.197 Hz, respectively.

We also carried out 500 ps MD simulations with CMAP for ubiquitin and CspA. CMAP is an extension of the CHARMM force field and has recently been shown to obtain a more accurate description of the peptide backbone.[54] By including grid-based energy correction maps and empirical corrections, this approach yields improved dynamical and structural properties of proteins in various simulations.[55,56] However, applied to the present simulations of $^{h3}J_{NC'}$ couplings for ubiquitin and CspA, results are very similar to simulations without CMAP, as illustrated in Supporting Information S4.

MD Simulations with MMPT. The zeroth-order MMPT PES is suitable to describe a N−H···O bond in NH_4^+−H_2O and will not be directly applicable to hydrogen bonding in proteins. Therefore, it is expected that MD simulations using the unmorphed MMPT potential are unsuited for quantitative work, and large deviations between observed and calculated $^{h3}J_{NC'}$ couplings should be found, as illustrated in Figure 3. When different morphing parameters are used, the MMPT potentials will have different minimum energy geometries $\{R', \rho'\}$ and lead to different scalar couplings, which is also shown in Figure 3.

The correlation between morphing parameters and RMSDs has been investigated for ubiquitin, CspA, and protein G. First, short (20 ps) test trajectories were run to locate suitable morphing parameters, and then 100 ps MD simulations were carried out on a fine grid ($\Delta = 0.01$ Å) of $\{R', \rho'\}$ and analyzed. For combinations $\{R', \rho'\}$ with low RMSDs, simulations were continued to 500 ps. Longer trajectories (1 ns) were run for ubiquitin (morphing parameters {2.92, 0.14}), protein G ({2.95, 0.16}), and CspA ({2.96, 0.16}), and RMSDs were calculated and are summarized in Table 1, together with results obtained from standard MD simulations. As an illustration, a detailed comparison between measured and calculated $^{h3}J_{NC'}$ in CspA from standard MD and MMPT/MD simulations, and the squared deviations for each individual hydrogen-bond coupling, are shown in Figure 4. By adopting MMPT PES as the explicit hydrogen-bond potential, the correlation between calculated $^{h3}J_{NC'}$ couplings and experimental data has been enhanced for most hydrogen bonds, especially those with large deviations ($|J_{calcd} - J_{exptl}| > 0.3$ Hz). The range of scalar couplings calculated from MMPT/MD simulations, however, is narrower than that from standard MD simulations. For convergence of most scalar couplings, a total of 500 ps is typically sufficient for MD simulations with the MMPT potential (see Supporting Information S3).

Application of the Morphed Potentials. After establishing that morphed MMPT potentials lead to improved agreement between calculated and experimental scalar coupling constants compared to those of a conventional force field (Table 1), potential morphing is used to further improve scalar coupling constants starting from X-ray and NMR structures. This is done for the three proteins studied in the previous section: ubiquitin, CspA, and protein G. As might be suspected, somewhat different coordinate transformations are most suitable to best describe the scalar couplings in the three different proteins (a summary of the relationship between RMSD and different MMPT PESs is given in the Supporting Information S5). Due to the nonlinearity between parameters (R', ρ') for the H-bond potentials and the calculated RMSDs between calculated and measured $^{h3}J_{NC'}$ couplings, there is no simple, detectable relationship between the two. However, it is found that deviations are generally small around a certain $\{R', \rho'\}$ combination, and these values are summarized in Table 2. The average N−O distances computed from the initial structure are also reported and are

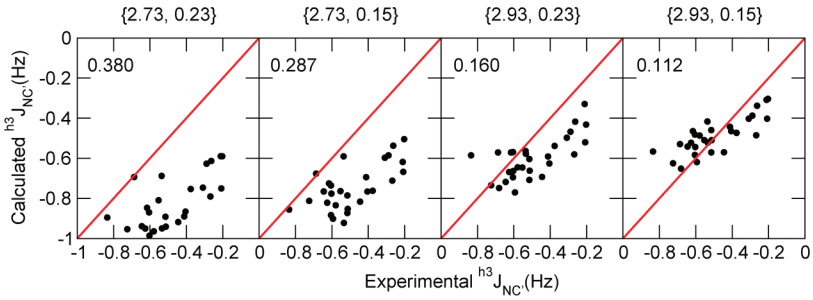

Figure 3. Calculated scalar couplings from 0.2 ns of MMPT/MD simulations compared with experimental data in ubiquitin for different morphing parameters.

Table 1. RMSDs of Ubiquitin, CspA, and Protein G Calculated from 1 ns Trajectory

	ubiquitin	CspA	protein G
standard MD	0.142	0.195	0.134
MD/MMPT	0.118[a]	0.123[b]	0.130[c]

[a] Morphing parameters {2.92,0.14}. [b] Morphing parameters {2.96,0.18}. [c] Morphing parameters {2.95,0.16}.

close to R'. With the use of relationship 4, the actual hydrogen bondlength r'_{OH} is found to be almost identical for all three proteins, namely, 1.93 Å with an average of 1.931 Å ± 0.002 Å. It is worthwhile mentioning that this value is reminiscent of the hydrogen-bond geometry parameter δ_{HA} calculated from a statistical analysis of 52 proteins.[40] In the following, potential morphing for MMPT PESs is further investigated such that the additional constraint $r'_{OH} = 1.93$ Å is fulfilled:

$$\rho' = \frac{R' - 2.73}{R' - 1.6} \tag{5}$$

This equation directly relates the two morphing parameters.

To test the procedure, it was applied to apo-CAM, holo-CAM, and IFABP, which were not part of the training set. The 500 ps MMPT/MD simulations were carried out with MMPT PES minima $\{R', \rho'\}$ found above, and scalar couplings were calculated. RMSDs between calculated and experimentally measured $^{h3}J_{NC'}$ couplings are summarized in Table 3. Compared with results from 1 ns standard MD simulations, considerably better agreement is achieved for all six proteins we investigated.

In all previous MMPT/MD simulations, only hydrogen bonds corresponding to experimentally measured scalar couplings are treated with the explicit hydrogen-bond potential. It would be interesting to test whether $^{h3}J_{NC'}$ can be predicted by MMPT/MD simulations without knowing which couplings can be observed in E.COSY experiments. Visual Molecular Dynamics (VMD)[57] has been used to assign hydrogen bonds in ubiquitin and CspA with a distance cutoff of 3.5 Å and an angle cutoff of 40°. In both proteins, more hydrogen bonds are found with this criterion (see Supporting Information S6), but not all of the previously assigned hydrogen bonds are covered. MD simulations with all of these hydrogen bonds treated by MMPT were carried out. $^{h3}J_{NC'}$ couplings were calculated from 500 ps trajectories and compared to experimental values. The RMSDs (0.122 and 0.160 Hz) are not as good as previous MMPT/MD results

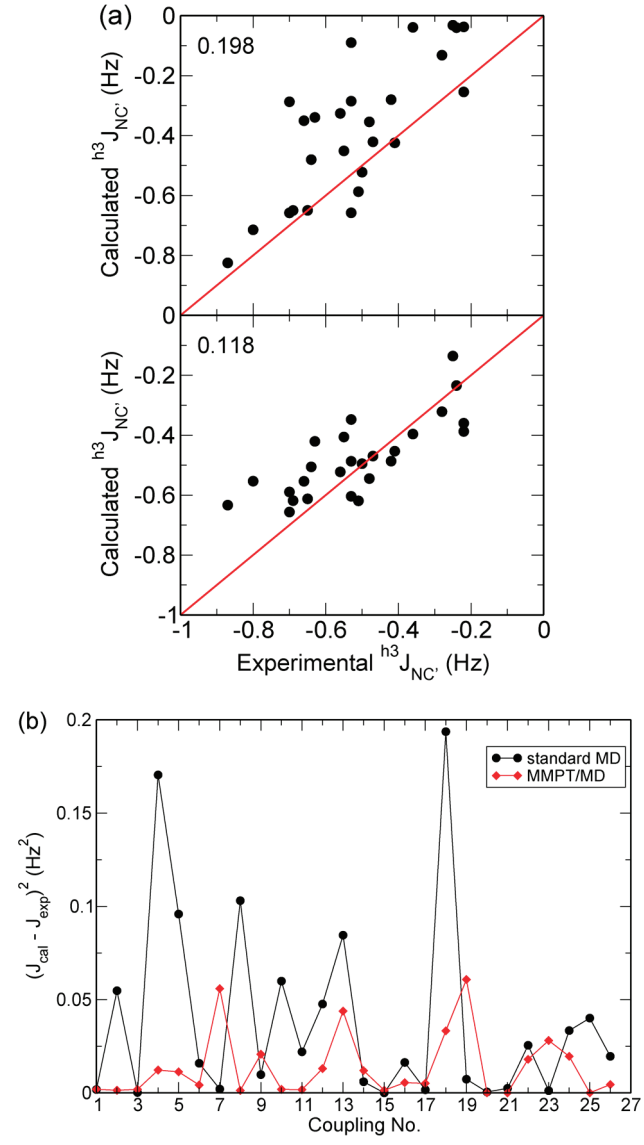

Figure 4. Comparison between scalar couplings calculated by standard MD simulations and MMPT/MD simulations for cold-shock protein A. (a) Comparisons of calculated and experimental $^{h3}J_{NC'}$ couplings. (b) Squared deviations for hydrogen bond scalar couplings.

(0.116 and 0.140 Hz) but are still significant improvements over standard MD simulations (0.140 and 0.195 Hz).

Characterization of the Conformational Ensemble. Once suitable morphing parameters are available, MMPT/MD can also be used to characterize the conformational

Table 2. Overview of PES Morphing Parameters for Ubiquitin, CspA, and Protein G

	R', Å	ρ'	r'_{OH}, Å	$\langle R_{X\text{-}ray}\rangle$, Å
1ubq	2.925	0.145	1.933	2.921
1mjc	2.960	0.170	1.929	2.958
2qmt	2.945	0.160	1.930	2.942

ensemble starting from the X-ray structure. The conformational ensembles generated by MD simulations with and without the MMPT potential are investigated through the distance between hydrogen and acceptor atoms and the angle at the acceptor atom. The respective density distributions (r_{NO}, θ_{HOC}) from 500 ps simulations for 29 hydrogen bonds in ubiquitin are shown in Figures 5 and 6. Using the MMPT potential for the hydrogen bonds in proteins leads to slightly shorter donor−acceptor distances, more pronounced directionality of the H bonds, and significant reductions in the fluctuations of r_{NO} and θ_{HOC}. This is also observed in protein G and CspA (data not shown) and can be explained by the fact that the MMPT potential is stronger and more directional than a conventional superposition of Coulomb terms. Thus, the H bonds are more restricted in the conformational space, which also leads to better stability and convergence of $^{h3}J_{NC'}$ couplings calculated from MMPT/MD compared to standard MD simulations (Supporting Information S3).

The protein dynamics based on using MMPT as an explicit hydrogen-bond potential are also investigated by calculating the root-mean-square fluctuations (or B-factors; Figure 7) and 2D cross-correlation maps (Figure 8). Generally, using the MMPT potential leads to rigidification of the protein, which is consistent with previous efforts to better describe hydrogen bonds in proteins.[58] The cross-correlation maps of ubiquitin show that most correlated motions are caused by hydrogen-bonding structures in the protein. Cross-correlation maps computed from MMPT/MD and standard MD simulations show similar dynamical features, while the comparison indicates that the MMPT potential enhances the correlations between hydrogen-bonding residues.

Discussion

In this work, we present a general method for deriving quantitative potential energy surfaces for H-bonding motifs and demonstrate their ability to accurately calculate scalar couplings across hydrogen bonds in proteins from atomistic simulations. Compared with standard MD simulations, RMSDs between calculated and experimental $^{h3}J_{NC'}$ couplings have been reduced in all six proteins investigated (Table 3). The $^{h3}J_{NC'}$ couplings can be calculated with an average deviation of 0.14 Hz by MMPT/MD simulation. Better agreement between calculated and experimental values are observed for all different secondary structures (Table 4), while the most significant improvements are found in loop regions. As has been noted previously,[59] current molecular mechanics force fields perform most poorly in the loop regions in proteins.

Our calculations are based on a force field treating hydrogen bonds explicitly. The MMPT potential, originally developed to investigate proton transfer reactions, has been shown to be adequate for describing hydrogen bonds in proteins by simple PES morphing techniques. This is consistent with the well-known fact that hydrogen bonds can be regarded as incipient or "frozen stage" proton transfer reactions.[1] It is possible that more sophisticated PES morphing strategies, a more realistic angular dependence (e.g., $V(R, \rho, \theta) = \Sigma_n V_n(R, \rho)P_n(\cos\theta)$, where P_n are Legendre polynomials), or different MMPT parametrizations for H bonds in different secondary structure elements will lead to additional improvements.

The results presented here are based on an average treatment of H bonds in proteins, which means that the same MMPT potential is used for all hydrogen bonds in a certain protein. This is reflected by the fact that the PES morphing parameter R' corresponds to the average D−A distance from the X-ray structures. However, hydrogen bonds in different chemical environments exhibit different strengths so describing them with environment-specific parametrizations is a possibility for improvement. In fact, this has been previously found to be the case when hydrogen bonds in different secondary structures (α helices, β sheets, and loops) were investigated separately.[33]

On the basis of a detailed study of correlation between PES morphing parameters and RMSDs in three proteins (ubiquitin, CspA, and protein G), we propose a generic procedure whereby, starting from X-ray structures, the PES is morphed to a minimum (R'_{NO}, r'_{OH}, θ'_{HNO}). Here, R'_{NO} equals the average N−O distance in the X-ray structure, r'_{OH} = 1.93 Å and $\theta'_{HNO} = 0$. Such an approach enables us to reliably calculate $^{h3}J_{NC'}$ couplings, and it has been applied to a set of six proteins. Due to the nonlinear relationship between the morphing parameters, the dynamics in proteins, and the calculated RMSDs for scalar couplings, morphing parameters $\{R', \rho'\}$ may not always yield the minimal RMSD between calculated and observed couplings. For example, in CspA, the morphing parameters lead to a RMSD of 0.14 Hz, while the minima {2.96,0.16} yield 0.12 Hz. However, differences are small, and both parameter sets are significant improvements over results from standard MD simulations (0.20 Hz), given that experimental errors are usually smaller than 0.05 Hz.[19,20]

In previous work relating NMR observables and MD simulations, biased simulations with an additional restraining penalty function have been used.[51] In this approach, $^{h3}J_{NC'}$ couplings were taken as input information and different dynamical ensembles were generated, which enables the determination of accurate geometries and energetics of hydrogen bonds in the native states of proteins. Here, a different approach is pursued. Instead of biasing simulations, the intermolecular interactions are represented more accurately by explicitly including potentials describing H bonds. The dynamical ensemble for the two methods is comparable in that narrower distributions of hydrogen-bond lengths and more restrictions for hydrogen-bond angles are found. Because scalar coupling constants directly characterize the geometries of H bonds, it is tempting to suggest that better quantitative agreement between calculated and experimentally measured $^{h3}J_{NC'}$ couplings also reflects a better description of the conformational ensemble of the protein.

Table 3. Comparison of RMSDs from Conventional MD Simulations and MMPT/MD Simulations with the Morphed Potentials

		ubiquitin	CspA	protein G	apoCAM	holoCAM	IFABP
RMSD (standard MD)		0.142	0.195	0.134	0.204	0.203	0.175
morphing parameters	R'	2.921	2.958	2.942	2.986	2.937	2.946
	ρ'	0.144	0.168	0.158	0.185	0.155	0.161
RMSD (MD/MMPT)		0.116	0.140	0.134	0.144	0.142	0.164

As the results show, the procedure pursued here is generally applicable and leads to appreciable improvement for all proteins investigated, and predictions for observables can be attempted. This is, in general, not possible with biased simulations for which the bias introduced is only valid for the particular protein under investigation and is not easily transferred to a different protein.

Hydrogen-bonding dynamics between standard MD and MMPT/MD simulations have been also compared in this work. Stronger hydrogen bonding, shorter hydrogen-bond lengths, and more pronounced directionality have been observed in MMPT/MD simulations. This agrees with a statistical analysis of X-ray structures which yields $\delta_{HA} = 1.93$ Å, which is identical to the separation found here and close to results from electronic structure calculations (1.94−1.97 Å). Furthermore, the average NHO angle from all simulations is 166°, which compares with values between 155° and 162° from electronic structure calculations, and 175° from the knowledge-based potential. Analysis of the protein dynamics shows that the MMPT potential rigidifies the entire protein and leads to stronger correlation between residues coupled by hydrogen bonds. This suggests that using explicit hydrogen-bond potentials shifts the conformational ensemble sampled in MD simulation toward the experimentally measured one.[40]

Here, we showed that an explicit, three-dimensional hydrogen-bond potential leads to—sometimes considerably—improved calculation of hydrogen bond scalar couplings from explicit atomistic simulations in full solvation for six proteins with different folds. A general computational strategy is formulated which employs the coordinates from (high-resolution) X-ray structures and leads to suitably

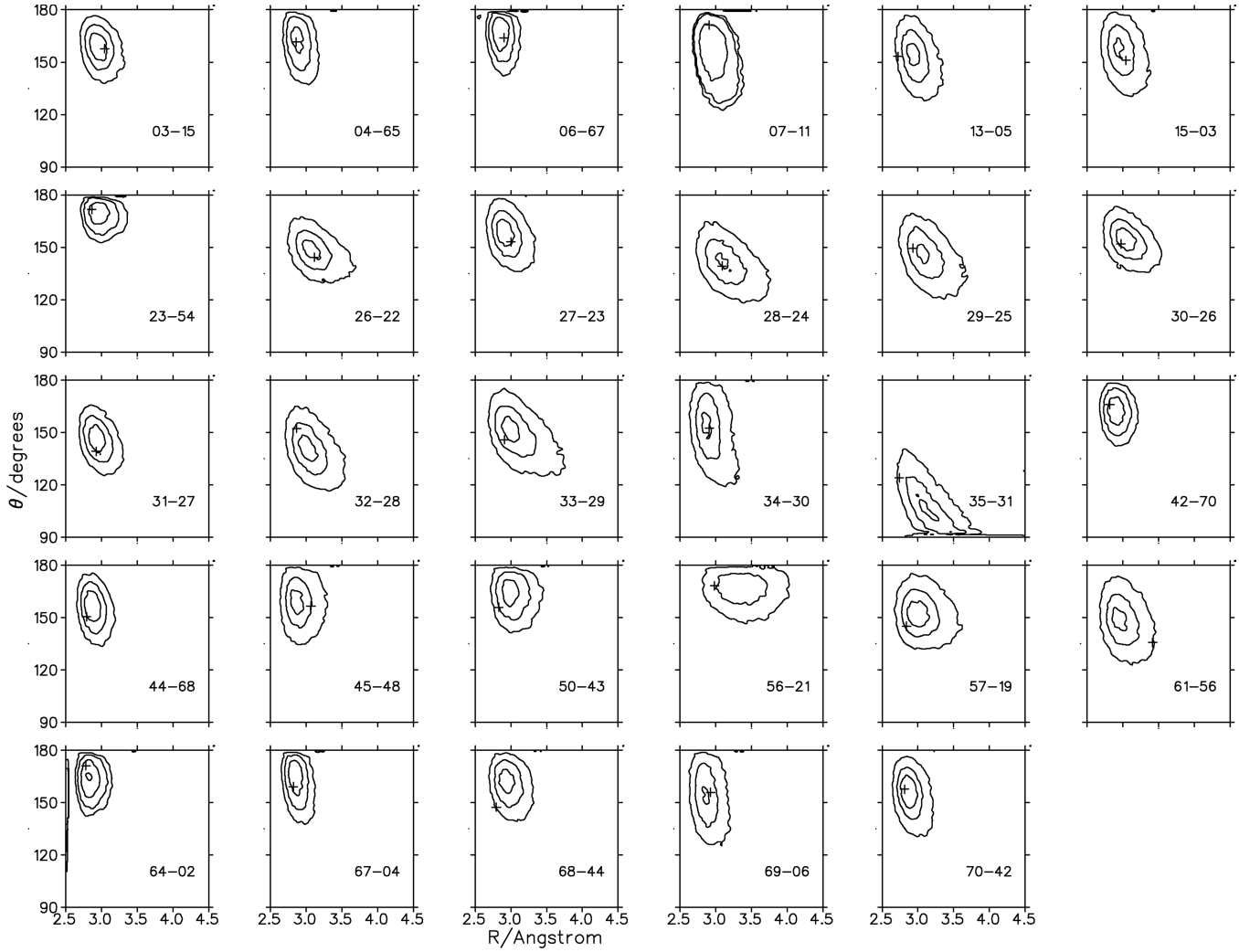

Figure 5. Distributions of hydrogen-bond geometries (r_{NO}, θ_{HOC}) populated during 500 ps standard MD simulation for 29 H bonds in ubiquitin.

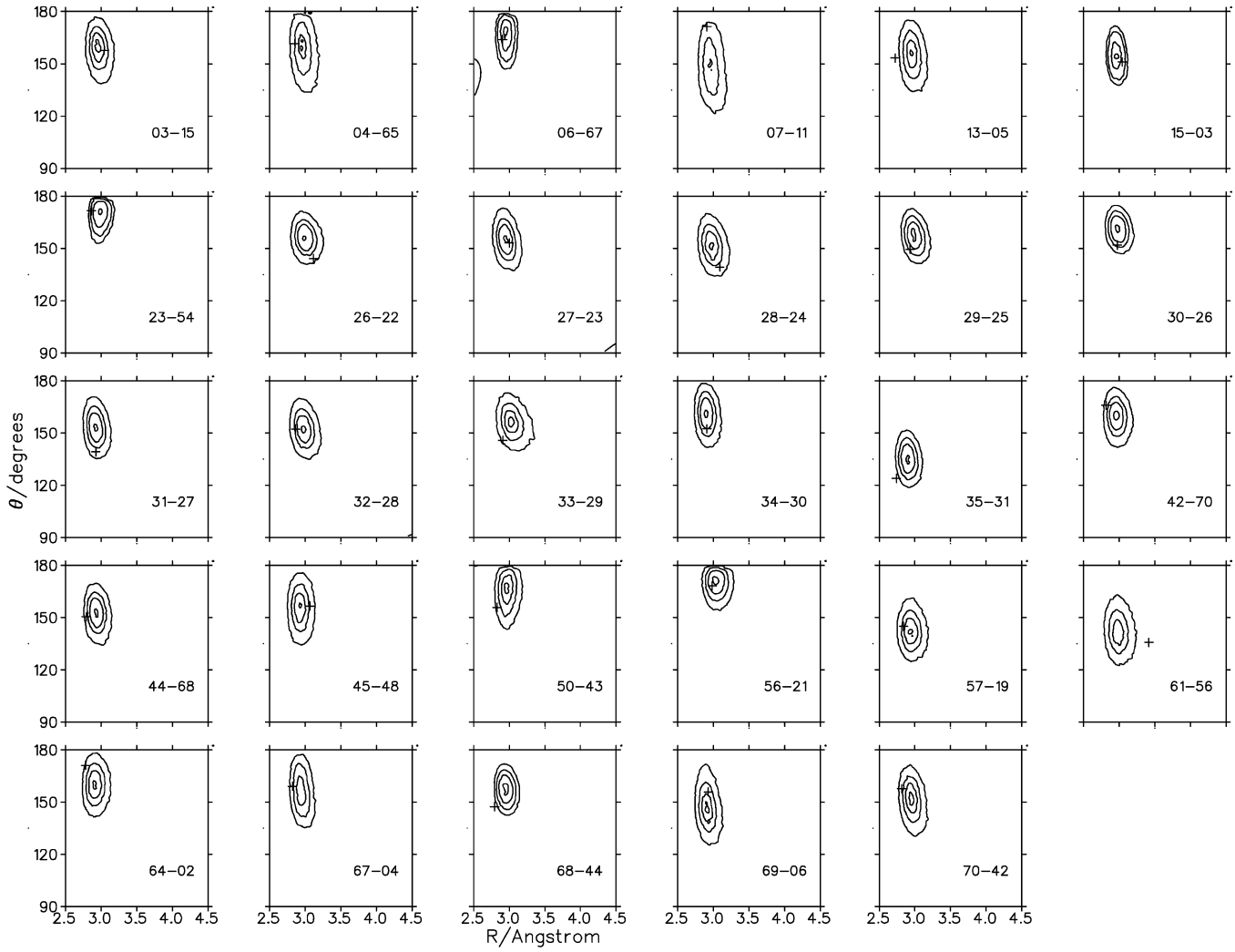

Figure 6. Distributions of hydrogen-bond geometries (r_{NO}, θ_{HOC}) populated during 500 ps MMPT/MD simulation for 29 H bonds in ubiquitin.

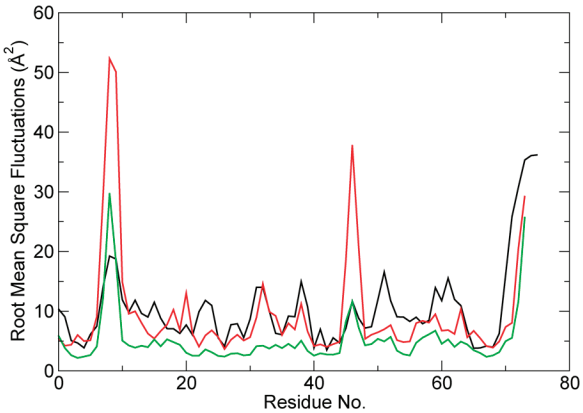

Figure 7. Root mean square fluctuations of backbone atoms calculated from 1 ns standard MD (red) and MMPT/MD (green) simulations of ubiquitin. Experimental B factors are plotted on the black line.

morphed H-bonding potential that can be used to investigate the nuclear dynamics in proteins. It is further illustrated that hydrogen-bonding potentials which lead to better agreement between calculated and measured $^{h3}J_{NC'}$ couplings are those with physically meaningful (PES morphing) parameters. This opens the possibility to further improve force fields by combining NMR data and atomistic simulations, which is of particular relevance in characterizing conformational ensembles and in studies of signal transduction in proteins. Recently, a detailed analysis of the signaling pathway of rhodopsin led to the proposition that signals in proteins can be conducted through salt bridges and hydrogen bonds because they are more directional and the residues involved can act as molecular switches.[60] For such studies, which will most likely be intensified in the near future due to the fundamental interest in unraveling the means by which signaling occurs at a molecular level, accurate H-bonding potentials will be particularly important. The additional computational effort involved in using MMPT is minimal because, instead of a few harmonic potentials (conventional force field), the same number of anharmonic (Morse) terms have to be evaluated. What currently limits the standard use of MMPT is the fact that a time step of $\Delta t \approx 0.2$ fs is used to propagate the equations of motion. However, multi-time-step procedures are being considered which will largely circumvent this problem. As has been shown in a recent study on CO relaxation in myoglobin, conventional force fields based on harmonic bonded potentials which accurately describe vibrational spectra can be inappropriate when

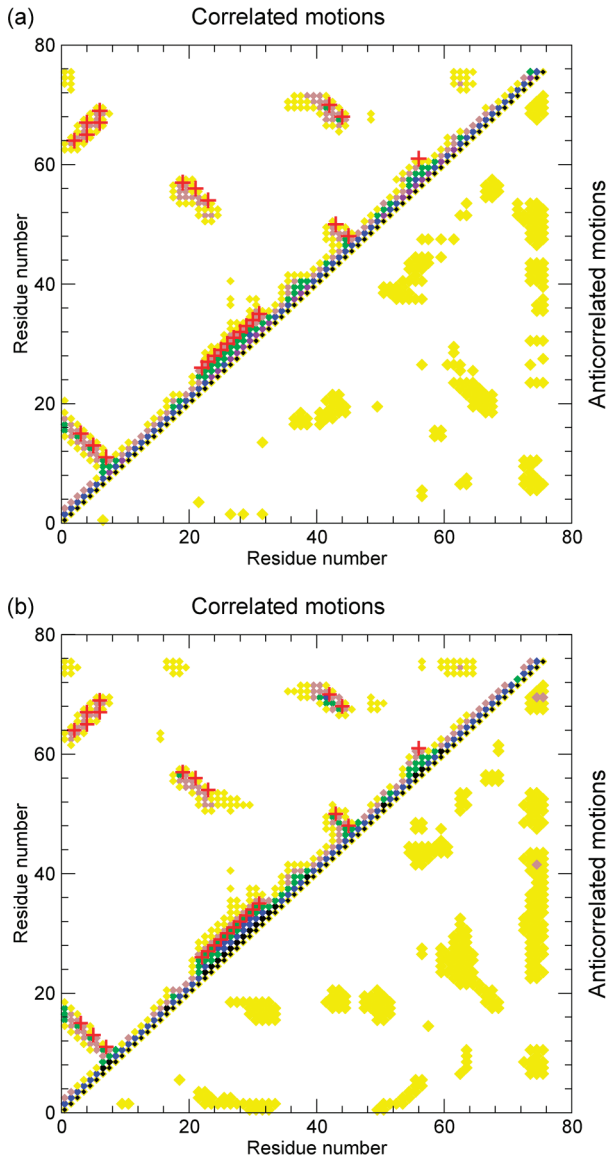

Figure 8. Dynamical cross-correlation maps for ubiquitin calculated from the 1 ns (a) standard MD and (b) MMPT/MD simulations. Positive cross-correlated coefficients are collected in the upper-left triangle and negative ones in the lower-right triangle. Only cross-correlation coefficients C_{ij} larger than 0.20 are shown. The intensity is represented as follows: yellow squares, $0.2 < C_{ij} < 0.35$; brown squares, $0.35 < C_{ij} < 0.5$; green squares, $0.5 < C_{ij} < 0.65$; blue squares, $0.65 < C_{ij} < 0.8$; and black squares, $0.8 < C_{ij} < 1.0$. H bonds with nonvanishing scalar couplings are marked by the red plus.

Table 4. Summary of RMSDs Calculated by Standard MD Simulations and MMPT/MD Simulations for All Couplings and for Couplings in Particular Secondary Structural Elements

	all	α helix	β sheet	loop
number of H bonds	214	86	105	23
standard	0.174	0.168	0.169	0.213
MMPT/MD	0.137	0.128	0.145	0.128

considering energy transfer between vibrational modes with widely separated frequencies.[61] Thus, when energy transfer between modes is studied, details of the interaction potentials may become important. The fundamental role of H bonds, the sensitivity of $^{h3}J_{NC'}$ couplings to their dynamics, and the possibility to compute couplings from meaningful atomistic simulations provide an ideal stage to further develop and extend the range and applicability of simulations.

Acknowledgment. The authors gratefully acknowledge financial support from the Swiss National Science Foundation under grant 200021-117810.

Supporting Information Available: Tables of all hydrogen bonds, scalar couplings and RMSDs, a comparison of different starting structures for protein G, a comparison of MD simulation with and without CMAP correction, and figures of $^{h3}J_{NC'}$ convergence. This information is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Steiner, T. *Angew. Chem., Int. Ed.* **2002**, *41*, 48–76.

(2) Dill, K. A. *Biochemistry* **1990**, *29*, 7133–7155.

(3) Deechongkit, S.; Nguyen, H.; Powers, E. T.; Dawson, P. E.; Gruebele, M.; Kelly, J. W. *Nature* **2004**, *430*, 101–105.

(4) Wang, M.; Wales, T. E.; Fitzgerald, M. C. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 2600–2604.

(5) Gray, M.; Cuello, A. O.; Cooke, G.; Rotello, V. M. *J. Am. Chem. Soc.* **2003**, *125*, 7882–7888.

(6) Aruksankunwong, O.; Hannongbua, S.; Wolschann, P. *J. Mol. Struct.* **2006**, *790*, 174–182.

(7) Taylor, M. S.; Jacobsen, E. N. *Angew. Chem., Int. Ed.* **2006**, *45*, 1520–1543.

(8) Huang, Y.; Unni, A. K.; Thadani, A. N.; Rawal, V. H. *Nature* **2003**, *424*, 146.

(9) MacKerell, A. D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(10) Ponder, J.; Case, D. *Adv. Protein Chem.* **2003**, *66*, 27–85.

(11) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(12) Nizkorodov, S. A.; Dopfer, O.; Ruchti, T.; Meuwly, M.; Maier, J. P.; Bieske, E. J. *J. Phys. Chem.* **1995**, *99*, 17118–17129.

(13) Meuwly, M.; Nizkorodov, S. A.; Maier, J. P.; Bieske, E. J. *J. Chem. Phys.* **1995**, *104*, 3876–3885.

(14) Nizkorodov, S.; Dopfer, O.; Meuwly, M.; Bieske, E.; Maier, J. *J. Chem. Phys.* **1996**, *105*, 1770–1777.

(15) Watson, J. K. G. *Mol. Phys.* **1968**, *15*, 479–490.

(16) Nilges, M.; Bernard, A.; Bardiaux, B.; Malliavin, T.; Habeck, M.; Rieping, W. *Structure* **2008**, *16*, 1305–1312.

(17) Dingley, A. J.; Grzesiek, S. *J. Am. Chem. Soc.* **1998**, *120*, 8293–8297.

(18) Cornilescu, G.; Ramirez, B. E.; Frank, M. K.; Clore, G. M.; Gronenborn, A. M.; Bax, A. *J. Am. Chem. Soc.* **1999**, *121*, 6275–6279.

(19) Grzesieka, S.; Cordiera, F.; Jaravinea, V.; Barfield, M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2004**, *45*, 275–300.

(20) Alkorta, I.; Elguero, J.; Denisov, G. S. *Magn. Reson. Chem.* **2008**, *46*, 599–624.

(21) Eberstadt, M.; Mierke, D. F.; Kock, M.; Kessler, H. *Helv. Chim. Acta* **1992**, *75*, 2583–2592.

(22) Dingley, A. J.; Masse, J. E.; Feigon, J.; Grzesiek, S. *J. Biomol. NMR* **2000**, *16*, 279–289.

(23) Cordier, F.; Grzesiek, S. *J. Am. Chem. Soc.* **1999**, *121*, 1601–1602.

(24) Ahn, H.-C.; Juranic, N.; Macura, S.; Markley, J. L. *J. Am. Chem. Soc.* **2006**, *128*, 4398–4404.

(25) Markwick, P. R. L.; Sprangers, R.; Sattler, M. *J. Am. Chem. Soc.* **2003**, *125*, 644–645.

(26) Alexandrescu, A. T.; Snyder, D. R.; Abildgaard, F. *Protein Sci.* **2001**, *10*, 1856–1868.

(27) Juranic, N.; Atanasova, E.; Streiff, J. H.; Macura, S.; Prendergast, F. G. *Protein Sci.* **2007**, *16*, 1329–1337.

(28) Juranic, N.; Moncrieffe, M. C.; Liki, V. A.; Prendergast, F. G.; Macura, S. *J. Am. Chem. Soc.* **2002**, *124*, 14221–14226.

(29) Eberstadt, M.; Gemmecker, G.; Mierke, D. F.; Kessler, H. *Angew. Chem., Int. Ed.* **1995**, *34*, 1671–1695.

(30) 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.

(31) Grzesiek, S.; Sass, H.-J. *Curr. Opin. Struct. Biol.* **2009**, *19*, 585–595.

(32) Sass, H.-J.; Schmid, F. F.-F.; Grzesiek, S. *J. Am. Chem. Soc.* **2007**, *129*, 5898–5903.

(33) Schmid, F. F.-F.; Meuwly, M. *J. Chem. Theory Comput.* **2008**, *4*, 1949–1958.

(34) Barfield, M. *J. Am. Chem. Soc.* **2002**, *124*, 4158–4168.

(35) Lammers, S.; Meuwly, M. *J. Phys. Chem. A* **2007**, *111*, 1638–1647.

(36) Lammers, S.; Lutz, S.; Meuwly, M. *J. Comput. Chem.* **2008**, *29*, 1048–1063.

(37) Bowman, J. M.; Gazdy, B. *J. Chem. Phys.* **1991**, *94*, 816–817.

(38) Meuwly, M.; Hutson, J. M. *J. Chem. Phys.* **1999**, *110*, 8338–8347.

(39) Kortemme, T.; Morozov, A. V.; Baker, D. *J. Mol. Biol.* **2003**, *326*, 1239–1259.

(40) Morozov, A. V.; Kortemme, T.; Tsemekhman, K.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6946–6951.

(41) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.

(42) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.

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

(44) (a) Schmidt, H. L. F.; Sperling, L. J.; Gao, Y. G.; Wylie, B. J.; Boettcher, J. M.; Wilson, S. R.; Rienstra, C. M. *J. Phys. Chem. B* **2007**, *111*, 14362–14369. (b) 1pga and 2igd are also used as initial structures, and the results are compared and discussed in the Supporting Information S1.

(45) Schindelin, H.; Jiang, W.; Inouye, M.; Heinemann, U. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 5119–5123.

(46) Schumacher, M.; Crum, M.; Miller, M. *Structure* **2004**, *12*, 849–860.

(47) Chattopadhyaya, R.; Meador, W. E.; Means, A. R.; Quiocho, F. A. *J. Mol. Biol.* **1992**, *228*, 1177–1192.

(48) Scapin, G.; Gordon, J.; Sacchettini, J. *J. Biol. Chem.* **1992**, *267*, 4253–4269.

(49) Brunger, A. T.; Karplus, M. *Proteins* **1988**, *4*, 148–156.

(50) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comp. Phys.* **1977**, *23*, 327–341.

(51) Gsponer, J.; Hopearuoho, H.; Cavalli, C. M.; Dobson, A.; Vendruscolo, M. *J. Am. Chem. Soc.* **2006**, *128*, 15127–15135.

(52) Raynes, W. T.; Geertsen, J.; Oddershede, J. *Int. J. Quantum Chem.* **1994**, *52*, 153–163.

(53) Malkin, V. G.; Malkina, O. L.; Salahub, D. R. *Chem. Phys. Lett.* **1994**, *221*, 91–99.

(54) Alexander, D.; Mackerell, J.; Feig, M.; Charles, L.; Brooks, I. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(55) Buck, M.; Bouguet-Bonnet, S.; Pastor, R. W.; Alexander, D.; MacKerell, J. *Biophys. J.* **2006**, *90*, L36–L38.

(56) Bastug, T.; Kuyucak, S. *Biophys. J.* **2009**, *96*, 4006–4012.

(57) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.

(58) Ji, C.; Mei, Y.; Zhang, J. Z. H. *Biophys. J.* **2008**, *95*, 1080–1088.

(59) Case, D. A. *Acc. Chem. Res.* **2002**, *35*, 325–331.

(60) Kong, Y.; Karplus, M. *Structure* **2007**, *15*, 611–623.

(61) Devereux, M.; Meuwly, M. *J. Phys. Chem. B* **2009**, *113*, 13061–13070.

CT9005695