

# Experimental Verification of Force Fields for Molecular Dynamics Simulations Using Gly-Pro-Gly-Gly

Abil E. Aliev\* and Denis Courtier-Murias

Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, U.K.

Received: February 22, 2010; Revised Manuscript Received: July 31, 2010

Experimental NMR verification of MD simulations using 12 different force fields (AMBER, CHARMM, GROMOS, and OPLS-AA) and 5 different water models has been undertaken to identify reliable MD protocols for structure and dynamics elucidations of small open chain peptides containing Gly and Pro. A conformationally flexible tetrapeptide Gly-Pro-Gly-Gly was selected for NMR  $^3J$ -coupling, chemical shift, and internuclear distance measurements, followed by their calculations using 2  $\mu$ s long MD simulations in water. In addition, Ramachandran population maps for Pro-2 and Gly-3 residues of GPGG obtained from MD simulations were used for detailed comparisons with similar maps from the protein data bank (PDB) for large number of Gly and Pro residues in proteins. The MD simulations revealed strong dependence of the populations and geometries of preferred backbone and side chain conformations, as well as the time scales of the peptide torsional transitions on the force field used. On the basis of the analysis of the measured and calculated data, AMBER99SB is identified as the most reliable force field for reproducing NMR measured parameters, which are dependent on the peptide backbone and the Pro side chain geometries and dynamics. Ramachandran maps showing the dependence of conformational populations as a function of backbone  $\phi/\psi$  angles for Pro-2 and Gly-3 residues of GPGG from MD simulations using AMBER99SB, AMBER03, and CHARMM were found to resemble similar maps for Gly and Pro residues from the PDB survey. Three force fields (AMBER99, AMBER99 $\phi$ , and AMBER94) showed the least satisfactory agreement with both the solution NMR and the PDB survey data. The poor performance of these force fields is attributed to their propensity to overstabilize helical peptide backbone conformations at the Pro-2 and Gly-3 residues. On the basis of the similarity of the MD and PDB Ramachandran plots, the following sequence of transitions is suggested for the Gly backbone conformation:  $\alpha_L \rightleftharpoons \beta_{PR} \rightleftharpoons \beta_S \rightleftharpoons \beta_P \rightleftharpoons \alpha$ , where backbone secondary structures  $\alpha_L$  and  $\alpha$  are associated with helices and turns,  $\beta_P$  and  $\beta_{PR}$  correspond to the left- and right-handed polyproline II structures and  $\beta_S$  denotes the fully stretched backbone conformation. Compared to the force field dependence, less significant, but noteworthy, variations in the populations of the peptide backbone conformations were observed. For different solvent models considered, a correlation was noted between the number of torsional transitions in GPGG and the water self-diffusion coefficient on using TIP3P, TIP4P, and TIP5P models. In addition to MD results, we also report DFT derived Karplus relationships for Gly and Pro residues using B972 and B3LYP functionals.

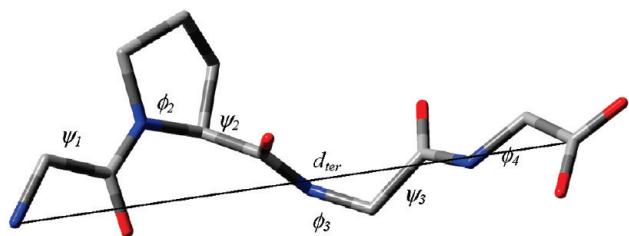
## 1. Introduction

Recently, we have reported a combined approach based on experimental NMR measurements and computational molecular dynamics (MD) and quantum-mechanical (QM) methods for structure and dynamics elucidations of flexible small peptides in solutions.<sup>1</sup> Among the three techniques used, the choice of the suitable protocol is most ambiguous at the MD stage of analysis compared to either QM methods or NMR. In particular, the choice of the force field is of primary concern, as there are many different force fields, optimized using either experimental or quantum mechanical data or a combination of both. Uncertainties of similar nature have been encountered previously and several groups have reported comparative analysis of different biomolecular force fields used in MD simulations.<sup>2–14</sup> However, in some cases the experimentally measured structural and dynamics characteristics in solutions under normal conditions were not available and either solid-state structural data or QM results in the gas phase were used for performance verifications. Besides, these studies have mostly focused on the backbone conformational analysis of Ala containing peptides and little is

known about the performance of different force fields in the case of Gly and Pro containing peptides. For many proteins the accurate prediction of the Pro ring conformations is also desirable, as the overall stability and dynamics of such highly ordered structures as triple helices in collagen, for example, are critically dependent on conformational preference of the pyrrolidine ring.<sup>15–19</sup> Its specific conformational behavior is also known to increase the energy difference between the denatured and native states.<sup>20</sup> Another important aspect of the comparative force field studies is that the overall outcome of the analysis is strongly dependent on the particular geometric or energetic property considered, as well as on the dynamics of the model system chosen. While relatively “static” proteins are expected to show only minor dependence on the choice of the force field, less consistent outcome is expected in the case of disordered proteins or small open chain peptides. Clearly, however, those force fields identified as reliable using highly flexible small peptides are also likely to perform reliably for both dynamically disordered and rigid domains in proteins.

With these considerations in mind, we have compared the performance of 12 different GROMOS,<sup>21,22</sup> AMBER,<sup>23–30</sup> OPLS-AA,<sup>31,32</sup> and CHARMM<sup>33,34</sup> force fields for simulations

\* Corresponding author. E-mail: a.e.aliev@ucl.ac.uk.



**Figure 1.** Unfolded conformation of Gly-Pro-Gly-Gly. Definitions of the backbone torsional angles and the distance between the terminal N and C' atoms ( $d_{ter}$ ) are also shown.

of small peptides in aqueous solutions using 2  $\mu$ s long MD simulations. As a model peptide, we have chosen Gly-Pro-Gly-Gly (GPGG, Figure 1). We have also used QM calculations for deriving Karplus type relationships for  $^3J_{\text{HH}}$ - and  $^3J_{\text{CH}}$ -couplings, which are experimentally measurable NMR parameters with a particular relevance in the case of small molecular weight species,<sup>35–38</sup> for which the NOE measurements are considerably less informative compared to high-molecular-weight proteins. Overall, we have considered three different types of parameters available from natural abundance NMR measurements in solutions (homo- and heteronuclear  $^3J_{\text{HH}}$ - and  $^3J_{\text{CH}}$ -couplings,  $^1\text{H}$  chemical shifts, and NOEs) for detailed comparison of the performance of different force fields. In addition, Ramachandran population maps for Pro-2 and Gly-3 residues of GPGG obtained from MD simulations were used for detailed comparisons with similar data from the protein data bank (PDB) for large number of Gly and Pro residues in proteins.

## 2. Experimental Section

**2.1. NMR Measurements and Calculations.** GPGG was purchased from Sigma-Aldrich and was used without further purification. A 116 mM solution of GPGG in D<sub>2</sub>O and a 125 mM solution of GPGG in H<sub>2</sub>O:D<sub>2</sub>O (20:1) were used for NMR measurements. Except for NOEs determined at 274 K, all other NMR parameters were measured at 298 K. Solution  $^1\text{H}$  NMR spectra were recorded on Bruker NMR spectrometers AVANCE400, AVANCE500, and AVANCE600 equipped with Bruker 5 mm room temperature probes (400 and 500 MHz) or a cryoprobe (600 MHz). A methanol-d<sub>4</sub> sample (99.8 at. %, Sigma-Aldrich) was used for the temperature calibration of the 600 MHz instrument. Data acquisition and processing were performed using standard Bruker XwinNMR (version 2.6) and TopSpin (version 2.1) software.  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts were calibrated using dioxane shifts in D<sub>2</sub>O ( $^1\text{H}$  3.75 ppm,  $^{13}\text{C}$  67.19 ppm).

Two-dimensional NOESY spectra with solvent suppression and varying mixing times were recorded in H<sub>2</sub>O:D<sub>2</sub>O (20:1) on Bruker AVANCE600. For better separation of the solute signals and the suppressed solvent peak, NOESY spectra were measured at 274 K. Two different methods were used for estimating distances from each spectrum: the intensity ratio method using the ratio of cross peak and diagonal peak intensities<sup>39,40</sup> and a simplified version of the growth rates method,<sup>41</sup> in which cross peak intensities measured at a single mixing time are used. The measured distances and their individual uncertainties (estimated as standard deviations) are included in Table S5 (Supporting Information). The largest standard deviation was less than 13% of the measured value of the internuclear distance.

One-dimensional proton-coupled  $^{13}\text{C}$  spectra with selective  $^1\text{H}$  decoupling, as well as two-dimensional selective heteronuclear  $J$ -resolved and  $J$ -HMBC spectra were used for the

**TABLE 1: Selected Properties of Water Models<sup>a</sup>**

	dielectric density, constant	self-diffusion coefficient, $10^{-5} \text{ cm}^2 \text{ s}^{-1}$	melting point, K
SPC	66	973.8	3.85
SPCE	71	998.9	2.49
TIP3P	98	986.0	5.19
TIP4P	52	994.1	3.29
TIP5P	81.5 <sup>76</sup>	999.1 <sup>76</sup>	2.62
experiment	78.3	997.2	2.30
			273

<sup>a</sup> Unless otherwise specified, dielectric constants and densities at 298 K and 1 atm are from ref 6, self-diffusion coefficients at 298 K and 1 atm are from ref 77, and melting points are from ref 78.

measurements of vicinal  $^1\text{H}$ – $^{13}\text{C}$  couplings.<sup>42,43</sup> Uncertainties in measured values of  $J$ -couplings and  $^1\text{H}$  chemical shift differences were less than  $\pm 0.1$  Hz.

The IUPAC labeling of nuclei are used in this work.<sup>44</sup> The least-squares fitting of the vicinal  $^3J$ -couplings of the pyrrolidine ring was carried out using a simulated annealing algorithm,<sup>45,46</sup> and eqs 8C and 8D of ref 36.

**2.2. MD Calculations.** One molecule of zwitterionic GPGG was solvated with 534–544 water molecules in a dodecahedral box with a volume of 17 nm<sup>3</sup> in MD simulations with AMBER and CHARMM force fields. Such a ratio of solute/solvent molecules corresponds to the concentration of ca. 103 mM, which is of the same order of magnitude as that used in NMR measurements (116 mM solution of GPGG in D<sub>2</sub>O and a 125 mM solution of GPGG in H<sub>2</sub>O:D<sub>2</sub>O (20:1)). In the case of the OPLS-AA/L force field, which has been parametrized using relatively long cutoff distances (see below),<sup>31,32</sup> a dodecahedral box with a volume up to 23 nm<sup>3</sup> (738 water molecules) was used. For MD simulations with GROMOS force fields, which also require relatively long cutoff distances,<sup>21,22</sup> a cubic box with a volume of 27 nm<sup>3</sup> was used (877 SPCE water molecules). Additional comparative MD simulations with the length of up to 1.4  $\mu$ s carried out in dodecahedral and cubic boxes showed no differences in the MD averaged parameters of GPGG considered in this work. Periodic boundary conditions were employed in all MD simulations. Various water models and force fields employed for classical MD simulations are listed in Tables 1 and 2. *GROMACS* (version 4)<sup>47</sup> implementations of the force fields were used.<sup>26,34,47</sup>

In simulations using *GROMACS*, neighbor lists were updated every fifth step. An integration step of 2 fs was used. In MD simulations using AMBER, OPLS-AA/L, and CHARMM force fields, the particle mesh Ewald (PME)<sup>48</sup> method was employed for the electrostatics with a grid spacing of 0.12 nm in the  $x$ ,  $y$ , and  $z$  directions and with fourth-order interpolation. Cutoff distances were similar to those used in parametrizations of the original force fields.<sup>21–34</sup> In particular, the neighbor list and the real-space cutoff distances were set to 0.9 nm (AMBER, OPLS-AA/L) and 1.2 nm (CHARMM). In the case of GROMOS force fields, the reaction field method was employed<sup>49</sup> with the neighbor list and Coulomb cutoffs set to 0.8 and 1.4 nm, respectively. The van der Waals interactions in all MD simulations were treated with a twin-range cutoff method using the neighbor list and van der Waals cutoff distances. The values of the neighbor list cutoff distances for different force fields are specified above. The values of the van der Waals cutoff distances were: 0.9 nm (AMBER), 1.2 nm (CHARMM), and 1.4 nm (GROMOS, OPLS-AA/L). Two MD simulations (2  $\mu$ s, AMBER99SB TIP3P) with the three cutoff distances set to 0.9 and 1.2 nm in each case showed negligible or no differences in predicted NMR parameters and other properties.

**TABLE 2: Selected Dynamics and Conformational Properties of GPGG in Water Derived from MD Simulations<sup>a</sup>**

force field	$p_f\%$	$r_g/\text{\AA}^b$	$d_{ter}/\text{\AA}^b$	$N^{\psi^2}$	$N^{\phi^3}$	$N^{\psi^3}$	$N^{\chi^2}$
GROMOS 43A1/SPCE	3	3.71 (0.33)	8.70 (2.37)	20.93	31.33	54.36	99.71
GROMOS 53A6/SPCE	2	3.77 (0.32)	9.03 (2.30)	23.42	31.10	47.82	99.06
AMBER94/TIP3P	58	3.27 (0.26)	5.12 (2.30)	3.63	20.87	37.96	66.96
AMBER96/TIP3P	5	4.01 (0.32)	10.18 (2.16)	7.06	17.36	24.02	61.38
AMBER99/TIP3P	58	3.37 (0.31)	5.45 (2.58)	5.79	26.31	38.54	74.69
AMBER99φ/TIP3P	53	3.27 (0.22)	5.16 (2.16)	3.38	25.24	48.06	79.77
AMBER-GS/TIP3P	47	3.36 (0.36)	6.05 (2.88)	11.02	11.74	27.45	65.03
AMBER-GSs/TIP3P	33	3.45 (0.36)	6.76 (2.89)	12.64	12.60	18.72	53.59
AMBER99SB/TIP3P	20	3.66 (0.40)	8.24 (3.02)	3.90	10.05	18.66	68.83
AMBER03/TIP3P	41	3.50 (0.44)	6.91 (3.38)	3.40	12.42	25.93	68.90
OPLS-AA/L/SPC	44	3.55 (0.44)	7.06 (3.50)	2.28	2.37	9.55	11.92
OPLS-AA/L/SPCE	34	3.65 (0.43)	7.81 (3.45)	1.97	2.64	7.47	14.30
OPLS-AA/L/TIP3P	43	3.56 (0.44)	7.12 (3.50)	2.27	2.60	10.55	13.89
OPLS-AA/L/TIP4P	47	3.55 (0.44)	7.35 (3.43)	2.37	2.42	8.71	12.68
OPLS-AA/TIP5P	38	3.58 (0.42)	7.35 (3.43)	1.16	1.98	8.69	12.26
CHARMM/TIP3P	76	3.21 (0.28)	4.60 (2.42)	1.33	2.57	11.97	50.52

<sup>a</sup> Shown are the population of the folded form ( $p_f$ ); the mean radius of gyration ( $r_g$ ); the mean N(1)–C'(4) distance, the number of  $\psi_2$ ,  $\phi_3$ ,  $\psi_3$ , and  $\chi_2$  torsional transitions per nanosecond ( $N^{\psi^2}$ ,  $N^{\phi^3}$ ,  $N^{\psi^3}$ , and  $N^{\chi^2}$ , respectively). Frames recorded every 1 ps were used in the calculations of  $N^{\psi^2}$ ,  $N^{\phi^3}$ ,  $N^{\psi^3}$ , and  $N^{\chi^2}$ . <sup>b</sup> Standard deviation is included in brackets.

Langevin dynamics, with a reference temperature of 298 K and a frictional constant of 10 ps<sup>-1</sup> was employed in MD simulations using AMBER, OPLS-AA/L, and CHARMM force fields.<sup>50</sup> Canonical sampling through velocity rescaling was used in MD calculations using GROMOS force fields.<sup>51</sup> A Parrinello–Rahman scheme was employed for pressure control at 1 bar using a coupling constant of 0.5 ps.<sup>52</sup> Prior to production runs, the system was minimized using steepest-descent and conjugate gradient algorithms. Minimization steps were followed by 3 steps of equilibration. The system was first equilibrated for 40 ps with the positionally restrained solute molecule to allow water molecules to equilibrate around it, followed by a NVT molecular dynamics for 10 ps without restraints and then by 100 ps of NPT dynamics with an isothermal compressibility of  $4.5 \times 10^{-5}$  bar<sup>-1</sup> and a reference pressure of 1.0 bar. Production simulations were performed for 2  $\mu$ s using NPT ensemble and  $2 \times 10^6$  frames were used for the calculations of averaged NMR parameters and distances.

Distances from the MD simulations were calculated in a manner similar to that used in NMR measurements described above: (i) internuclear distances ( $r_i$ ) for pairs of hydrogen atoms were calculated in each MD frame  $i$ ; (ii) a quantity equal to  $r_i^{-6}$  was calculated as a measure of the expected NOE in each frame,  $\eta_i$ ; (iii) the sum of  $r_i^{-6}$  was used as a measure of the expected total NOE over the full length of the MD run; (iv) using  $r = 2.4$  Å as the reference H<sup>α</sup>–H<sup>β3</sup> distance in the Pro residue,<sup>1</sup> internuclear distances for other proton pairs were calculated using the  $\eta \sim r^{-6}$  relationship. As the experimental distances were measured from NOESY spectra recorded at 274 K, for one of the force fields used (AMBER-GSs TIP3P) an additional 2  $\mu$ s long MD simulation with reference temperature of 274 K was carried out. Only negligible differences of 0.001–0.019 Å were found in the predicted distances from the MD simulations with reference temperatures of 274 and 298 K.

The motionally averaged  $^3J_{\text{HH}}$  and  $^3J_{\text{CH}}$ -couplings of the peptide backbone were calculated using either quantum-mechanically derived Karplus relationships described in Results and Discussion (section 4.4) or empirically parametrized Karplus equations described in references 53 and 54.

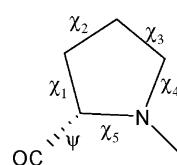
The motionally averaged  $^3J$ -couplings of the pyrrolidine ring were calculated using Karplus eqs 8C and 8D by Haasnoot et al.<sup>36</sup>

<sup>1</sup>H NMR chemical shifts were calculated using empirical correlations as implemented in programs *SHIFTS*<sup>55,56</sup> and *TOTAL*.<sup>57</sup>

For the validation of quantum-mechanically derived Karplus relationships for Gly and Pro residues, a 30 ns MD simulation of *Desulfovibrio vulgaris* flavodoxin (PDB entry 1AKQ)<sup>58</sup> and a 20 ns MD simulation of human erythrocytic ubiquitin (PDB entry 1UBQ)<sup>59</sup> were carried out using the AMBER99SB force field and the TIP3P water model. One molecule of flavodoxin was solvated with 4245 water molecules in a dodecahedral box with a volume of 149 nm<sup>3</sup>, and one molecule of ubiquitin was solvated with 4411 water molecules in a dodecahedral box with a volume of 147 nm<sup>3</sup>. Other conditions and parameters of MD simulations were the same as described above for GPGG.

**2.3. Quantum Mechanical Calculations.** All quantum mechanical calculations were carried out using the *Gaussian 03* program.<sup>60</sup> Water solvent effects were introduced in the quantum mechanical calculations via self-consistent reaction field (SCRF) theory calculations using the IEFPCM method.<sup>61,62</sup> Further details of these calculations are provided in the Results and Discussion.

**2.4. Conformational Notation.** For the pyrrolidine ring conformations, we use the original notation proposed by Haasnoot et al.<sup>15</sup> The downward orientation of the CO group is used as a reference view and the exo- and endo-orientations of the ring carbons are defined relative to the carbonyl group. For endocyclic torsional angles we use the conventional  $\chi$  notation:



The pseudorotation phase angle,  $P$ , which identifies a given conformation on the pseudorotation circle,<sup>63</sup> and the pseudorotation amplitude  $\chi_m$ , which is the maximum value attained by  $\chi_1 - \chi_5$ .<sup>15</sup> The calculations of  $P$  and  $\chi_m$  were done using equations by Westhof–Sundaralingam.<sup>63</sup>

$$P = \tan^{-1}\left(\frac{B}{A}\right) \quad \text{and} \quad \chi_m = (A^2 + B^2)^{1/2}$$

where

$$A = \frac{2}{5} \sum_{i=1}^5 \chi_i \cos\left(\frac{4\pi}{5}(i-2)\right) \quad \text{and}$$

$$B = -\frac{2}{5} \sum_{i=1}^5 \chi_i \sin\left(\frac{4\pi}{5}(i-2)\right)$$

Note that  $180^\circ$  is added to the calculated  $P$  value if  $\chi_2 < 0$ . Hence,  $-90^\circ < P < 90^\circ$  for the C'-exo conformation and  $90^\circ < P < 270^\circ$  for the C'-endo conformation.

**2.5. Ramachandran Plots and Free Energy Surfaces.** Ramachandran-type ( $\phi, \psi$ ) histograms for MD simulations were built by using  $1^\circ \times 1^\circ$  bins ( $360 \times 360$  squares). The fractional population ( $p_i$ ) of each bin ( $\phi_i, \psi_j$ ) was calculated as the number of structures in this bin divided by the total number of structures ( $p_N$ ) and multiplied by 1000. The population fraction of each secondary structure for the midchain Pro-2 and Gly-3 residues was calculated by taking the sum of the populations in defined regions and dividing the result by the total population. The defined regions (as described in section 4.3) were

- (i)  $0^\circ \leq \phi \leq 180^\circ / -90^\circ \leq \psi \leq 80^\circ$  for  $\alpha_L$ ;
- (ii)  $-180^\circ \leq \phi \leq 0^\circ / -90^\circ \leq \psi \leq 80^\circ$  for  $\alpha$ ;
- (iii)  $-135^\circ \leq \phi \leq 0^\circ / 80^\circ \leq \psi \leq 180^\circ$  and  $-135^\circ \leq \phi \leq 0^\circ / -180^\circ \leq \psi \leq -90^\circ$  for  $\beta_P$ ;
- (iv)  $0^\circ \leq \phi \leq 135^\circ / -180^\circ \leq \psi \leq -90^\circ$  and  $0^\circ \leq \phi \leq 135^\circ / 80^\circ \leq \psi \leq 180^\circ$  for  $\beta_{PR}$ ;
- (v)  $-180^\circ \leq \phi \leq -135^\circ / -180^\circ \leq \psi \leq -90^\circ$ ,  $-180^\circ \leq \phi \leq -135^\circ / 80^\circ \leq \psi \leq 180^\circ$ ,  $135^\circ \leq \phi \leq 180^\circ / -180^\circ \leq \psi \leq -90^\circ$  and  $135^\circ \leq \phi \leq 180^\circ / 80^\circ \leq \psi \leq 180^\circ$  for  $\beta_S$ .

In a similar manner, we have calculated the population fraction of each secondary structure using the PDB survey data provided by Hovmöller et al.,<sup>64,65</sup> which employed  $4^\circ \times 4^\circ$  bins ( $90 \times 90$  squares).

Free energy surfaces were also built for MD simulations by calculating relative free energy values for each bin ( $\Delta G_i$ ) using

$$\Delta G_i = -RT \ln(p_i/p_{\max})$$

where  $R$  is the universal gas constant,  $T$  is the absolute temperature (298 K), and  $p_{\max}$  is the population of the most populated bin.

### 3. Background

**3.1. Force Fields Used.** We have compared the performance of biomolecular force fields GROMOS,<sup>21,22</sup> AMBER,<sup>23–30</sup> OPLS-AA,<sup>31,32</sup> and CHARMM.<sup>33,34</sup> These force fields have a similar form of the interaction function. The main difference is in the way they have been parametrized. An extensive overview of the issues associated with the development and application of empirical force fields can be found in refs 5 and 66–69. Here, we consider brief details and differences of the force fields used.

In the original united atom force field GROMOS, nonbonded parameters were obtained from crystallographic data and atomic polarizabilities and adjusted such that experimental distances and interaction energies of individual pairs of functional groups were reproduced.<sup>70</sup> Statistical mechanics was employed to parametrize the nonbonded interactions by reproducing ther-

modynamic properties such as the density and the heat of vaporization of small molecules in the condensed phase at physiological temperatures and pressures. The parameter set 43A1 was developed for simulations of biomolecular systems.<sup>21</sup> The latest parameter set, 53A6, was first optimized to reproduce the thermodynamic properties of pure liquids of a range of small polar molecules and the solvation free enthalpies of amino acid analogs in cyclohexane, followed by the adjustment of the partial charges to reproduce the hydration free enthalpies in water.<sup>22</sup>

In AMBER94, the atomic charges were determined using HF/6-31G\* calculations of appropriately chosen molecules and restrained electrostatic potential (RESP) fitting.<sup>23</sup> The van der Waals parameters were derived from simulations of liquids. The bonded parameters were modified to reproduce experimental vibrational frequencies and structures. For the dihedral parameters of the peptide backbone, a set of  $\phi$  and  $\psi$  parameters was developed to reproduce the energies of the low-energy conformations of glycyl and alanyl dipeptides. AMBER96 represents a minor modification of AMBER94 and includes adjustments of  $\phi$  and  $\psi$  parameters to better reproduce the relative energies of gas-phase alanine tetrapeptide conformations as predicted by QM local MP2 (LMP2) calculations with the cc-pVTZ(-f) basis set.<sup>24</sup> In particular,  $\phi$  and  $\psi$  parameters were adjusted to reproduce the energy difference between extended and constrained  $\alpha$ -helical conformations of the alanine tetrapeptide. However, it was found that AMBER94 overstabilizes helical peptide conformations, whereas AMBER96 favors extended peptide backbones.<sup>27,71</sup> Another modification of the AMBER94 force field is known as AMBER-GS,<sup>27</sup> in which the original  $\phi$  and  $\psi$  torsion potentials are set to zero. Further to this modification, in AMBER-GS<sup>28</sup> force field van der Waals interactions between atoms separated by three covalent bonds (1,4 nonbonded interactions), are not scaled, whereas in other AMBER potentials, including AMBER-GS, they are scaled by a factor of 0.5. In another development, extensive adjustments of the  $\phi$  and  $\psi$  torsions of AMBER94 were undertaken using improved iterative scanning of torsional parameters and QM data for significantly larger number of alanine di- and tetrapeptides, which led to AMBER99.<sup>25</sup> Additional energetic barriers about the  $\phi$  torsion angle were also included in AMBER99. Further modification of AMBER99 was reported, in which these additional  $\phi$  barriers in AMBER99 were removed and  $\phi$  torsion potentials of the original AMBER94 were employed with the goal of better reproducing experimental data for polyalanines.<sup>26</sup> This force field is known as AMBER99 $\phi$ . As shown by Simmerling et al.,<sup>29</sup> however, the existence of the additional sets of backbone  $\phi, \psi$  dihedral terms in AMBER94 for residues other than Gly was neglected in AMBER96, AMBER-GS, AMBER-GSs, AMBER99, and AMBER99 $\phi$ , which led to parameter sets predicting unreasonable conformational preferences for Gly residues. This additional set of dihedral angles include  $\phi' = C-N-C^\alpha-C^\beta$  and  $\psi' = C^\beta-C^\alpha-C-N$ . A new AMBER99SB set of parameters was released, which accounts for the existence of two sets of backbone  $\phi, \psi$  dihedral terms.<sup>29</sup> Dihedral term parameters were optimized on the basis of fitting the energies of multiple conformations of glycine and alanine tetrapeptides from high level QM calculations. NMR *J*-coupling and relaxation data were also used for the verification of the AMBER99SB force field.<sup>9,10,12,14,29</sup> Finally, an extensive modification known as AMBER03 was introduced, in which the partial atomic charges were determined using RESP fits to B3LYP/cc-pVTZ/HF/6-31G\* data.<sup>30</sup> This was followed by readjustment of the main-chain  $\phi, \psi$  dihedral terms to reproduce QM data for alanine and glycine dipeptides obtained at the MP2/

cc-pVTZ//HF/6-31G\* level with a dielectric constant of 4, which was used to mimic the protein interior environment.<sup>30</sup>

In the OPLS-AA force field, both nonbonded and torsional energy parameters were obtained by fittings to HF/6-31G\* gas-phase structures and relative conformational energies of small molecule side chain analogues.<sup>31</sup> In addition, the experimentally observed thermodynamic properties of organic liquids were also used in parameter fittings. The present OPLS-AA/L force field has utilized higher level QM calculations to improve the performance of the OPLS-AA force field for the prediction of the intramolecular conformational energy surface.<sup>32</sup> In particular, refitting the key Fourier torsional coefficients was undertaken with the use of more accurate data generated by geometry optimizations at the HF/6-31G\*\* level followed by single-point LMP2/cc-pVTZ(-f) calculations. The validity of the Coulombic charges and the van der Waals parameters introduced in OPLS-AA/L was proved through reproducing gas-phase energies of complex formation heats of vaporization and densities of pure model liquids.

In the CHARMM22 force field,<sup>33</sup> the partial atomic charges were based on HF/6-31G\* data. As with other force fields, condensed phase simulations of a large set of model compounds were used to determine the Lennard-Jones parameters. Intramolecular parameters for CHARMM22 were derived using fittings to experimental and QM data for small molecules with the emphasis on the reproduction of their vibrational spectra. To remove the bias toward  $\pi$ -helical peptide conformations, the LMP2/cc-pVQZ//MP2/6-31G\* energy surfaces for dipeptides were used to correct the CHARMM22 dipeptide dihedral energy surface using a grid-based difference map (CMAP).<sup>33</sup>

As apparent from the above, the main point of concern of many different force field adjustments has been dihedral angle terms. This is despite the large number of other intra- and intermolecular terms present in the overall interaction function. Partly, the reason for such force field adjustments is driven by the desire to improve the agreement between Ramachandran plots obtained from the X-ray data in the solid state and those from MD simulations in water, even though it is well acknowledged that the relative populations of  $\phi/\psi$  angles in two environments may be different. At the same time, various solution NMR observables are either directly or indirectly dependent on dihedral angles and on the population of conformations, providing an independent means for verification of MD simulations in aqueous solutions. Thus, both the X-ray survey data and NMR parameters are explored in this work in comparing the performance of different force fields.

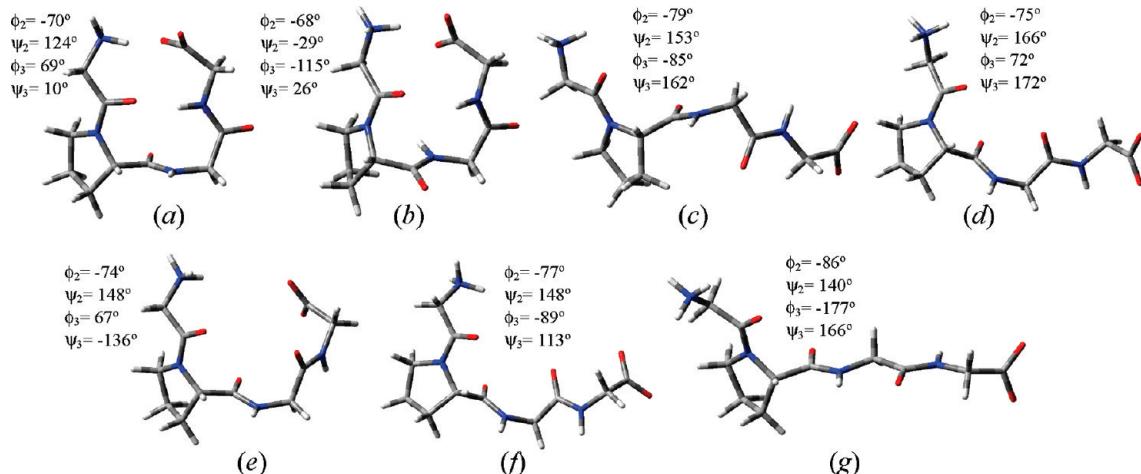
**3.2. Explicit Water Models.** We have used SPC,<sup>72</sup> SPCE,<sup>73</sup> TIP3P,<sup>74</sup> TIP4P,<sup>75</sup> and TIP5P<sup>76</sup> models, which are known to satisfactorily describe water as liquid under ambient conditions. Some selected properties of different explicit water models used in MD simulations are summarized in Table 1. The geometric difference between the SPC and TIP models is that in SPC and SPCE models the OH bond length is 1 Å, whereas in TIP models it is 0.9572 Å. In addition, the SPC and SPCE models assume a tetrahedral geometry with the H–O–H angle of 109.47°, whereas in TIP models this angle is 104.52°, which matches the experimental value. The value of the charge on the H atom and the Lennard-Jones parameters of the SPC and TIP3P models were derived using the density of water and the vaporization enthalpy at ambient conditions. Thus, the two models are expected to be similar, although due to the difference in the H–O–H angle, tetrahedral SPC water molecules are better ordered at the long-range than TIP3P molecules, as revealed by the height of the second peak in the O–O radial distribution

curves. Besides, a diffusion constant of TIP3P is significantly larger than experiment and that of SPC (Table 1). Despite these limitations, TIP3P is used widely and shows a satisfactory performance for biomolecular simulations. Compared to the SPC model, the SPCE model includes an additional polarization energy correction accounting for the neglected many-body polarization forces and is known to better predict density and diffusion constant than the SPC model (Table 1). The difference between the TIP3P and TIP4P models is that the site with the negative charge is not on the oxygen atom and is placed on the H–O–H bisector at 0.15 Å from it. The TIP4P is shown to overcome some of the important deficiencies of the TIP3P model, although at the expense of the increased computational cost due to the additional particle. In the TIP5P model the negative charge is placed at the two positions of the oxygen lone pair electrons at 0.7 Å from the oxygen atom. Unlike TIP3P and TIP4P, this model is optimized to reproduce the density maximum of water at 277 K. TIP5P provides the best prediction of the radial distribution function. In principle, it is better suited for long-range structural predictions of water, though this may not provide a critical improvement for biomolecular simulations. The prediction of the self-diffusion coefficient is also improved in TIP5P compared to TIP3P and TIP4P. Finally, one of the properties illustrating the level of improvement achieved in water models is perhaps the melting point of ice, which is as low as 146 K for TIP3P and 190 K for SPC (Table 1).<sup>78</sup> However, the low melting point of three-center water models appears to be not critical for biomolecular simulations under ambient conditions and many different properties, including hydration thermodynamics of various amino acid residues,<sup>6</sup> are reproduced satisfactorily by these models.

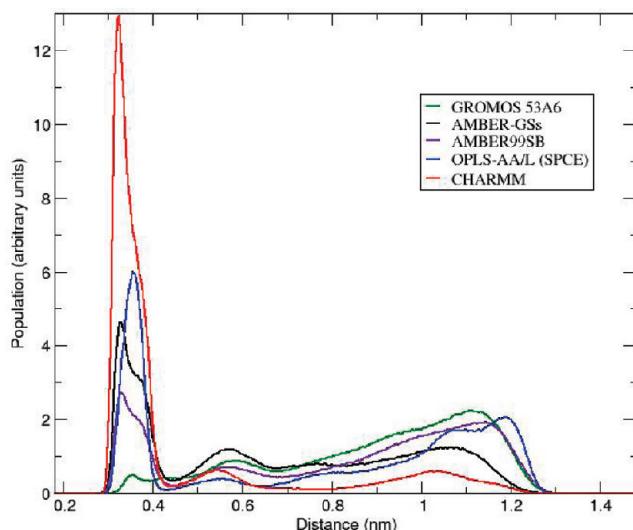
## 4. Results and Discussion

**4.1. Force Field Dependence of MD Results.** Only the major trans-rotamer (86% from NMR measurements in water at 298 K)<sup>79</sup> about the peptide bond preceding the Pro residue is considered further in this report, as the MD force fields are parametrized for this rotamer. The Pro ring dynamics of GPGG can be described in terms of a simple two-site equilibrium model between C'<sup>γ</sup>-endo and C'<sup>γ</sup>-exo conformers,<sup>15,79</sup> whereas a more complex multisite exchange is observed for the tetrapeptide backbone. Examples of folded and unfolded backbone conformations of GPGG extracted from the MD AMBER99SB simulation are shown in Figure 2. Conformations in (a)–(d) are similar to those from the HF/cc-pVDZ geometry optimizations using the IEFPCM solvation model for water, initial structures for which were extracted from the MD OPLS-AA/L TIP4P simulation.<sup>1</sup>

For MD simulations of GPGG using different force fields the distribution of various backbone conformations is best illustrated by histograms of  $d_{ter}$ , as predicted by 2  $\mu$ s long MD simulations using different force fields. From the distribution curves shown in Figure 3, it is apparent that significantly different populations of backbone conformations are predicted by different force fields; i.e., some of them predict the preference of the folded backbone, whereas others give preference to the unfolded backbone conformation. In the two extreme cases, the population of the folded U-shaped backbone conformation ( $p_f$ , in %) is ca. 2% and 76% (uncertainty less than  $\pm 1\%$ ) according to GROMOS 53A6 and CHARMM force fields (Table 2). Similarly, the values of the mean radius of gyration ( $r_g$ ) are significantly different. For example, the  $r_g$  values are 3.27 and 4.01 Å for AMBER94 and AMBER96, suggesting that the AMBER96 force field favors strongly the unfolded backbone



**Figure 2.** Representative examples of backbone conformations of GPGG extracted from the MD AMBER99SB/TIP3P simulation. The distance between the terminal N and C' atoms ( $d_{ter}$ ) is 3.33 Å in (a), 3.30 Å in (b), 11.44 Å in (c), 9.96 Å in (d), 5.71 Å in (e), 8.37 Å in (f), and 12.27 Å in (g). The values of dihedral angles for the midchain Pro-2 and Gly-3 residues are also shown.



**Figure 3.** MD predicted population distributions of conformers by selected force fields as a function of a distance  $d_{ter}$  between the terminal N (Gly-1) and C' (Gly-4) atoms in GPGG. The corresponding curves for all the combinations of force fields and water models are included in Supporting Information (Figures S1 and S2).

conformation. At the same time, from the  $p_f$  values for the same pair of force fields in Table 2, the AMBER94 force field gives strong preference to the folded U-shaped backbone conformation.

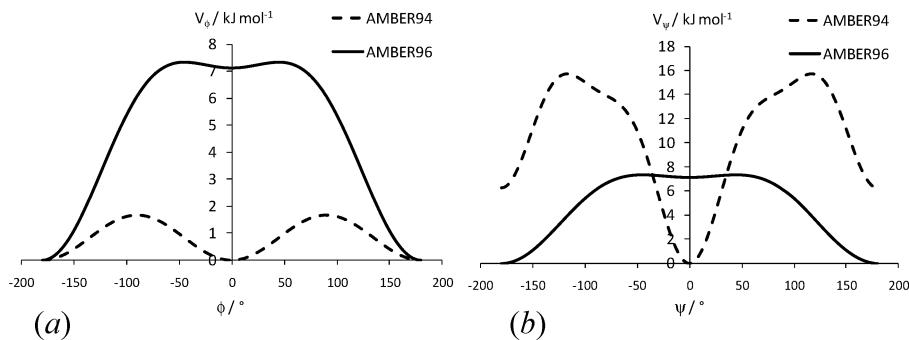
We can verify the sensitivity of the MD predicted populations of the peptide backbone conformations to the change of parameters associated with the backbone torsion using AMBER94 and AMBER96 force fields, since only the backbone dihedral angle parameters  $\phi$  and  $\psi$  were changed in AMBER96 compared to AMBER94.<sup>24</sup> As the populations of conformations are determined by their relative energies, in Figure 4 we compare the dihedral angle potentials  $V_\phi$  and  $V_\psi$  for these two force fields. The energy of conformations with  $\phi \approx \pm 90^\circ$  relative to that with  $\phi \approx \pm 180^\circ$  is significantly higher for AMBER96 compared to AMBER94 (7.3 and 1.7 kJ mol<sup>-1</sup>, respectively), suggesting that unfolded conformations with  $\phi \approx \pm 180^\circ$  are favored by AMBER96 (Figure 4a). Similarly, backbone conformations with  $\psi \approx 0^\circ$  are strongly favored by AMBER94 compared to AMBER96 (Figure 4b). The dihedral angle potential curves shown in Figure 4 also explain the differences in Ramachandran plots for AMBER94 and AMBER96 presented below.

Another important difference is in the number of the  $\psi_2$ ,  $\phi_3$ ,  $\psi_3$  (Figure 1) and  $\chi_2$ (Pro) torsional transitions per nanosecond ( $N^{\psi_2}$ ,  $N^{\phi_3}$ ,  $N^{\psi_3}$ , and  $N^{\chi_2}$  in Table 2). For example,  $N^\phi$  varies between 2 and 31 transitions per nanosecond for OPLS-AA/L and GROMOS force fields. These changes in the number of torsional transitions suggest a significant dependence of the time scale of the intramolecular dynamics on the type of the force field used in MD simulations. As a consequence, significantly longer MD simulations may be needed on using OPLS-AA/L and CHARMM compared to either GROMOS or AMBER force fields to ensure adequate statistical sampling of the conformational space.

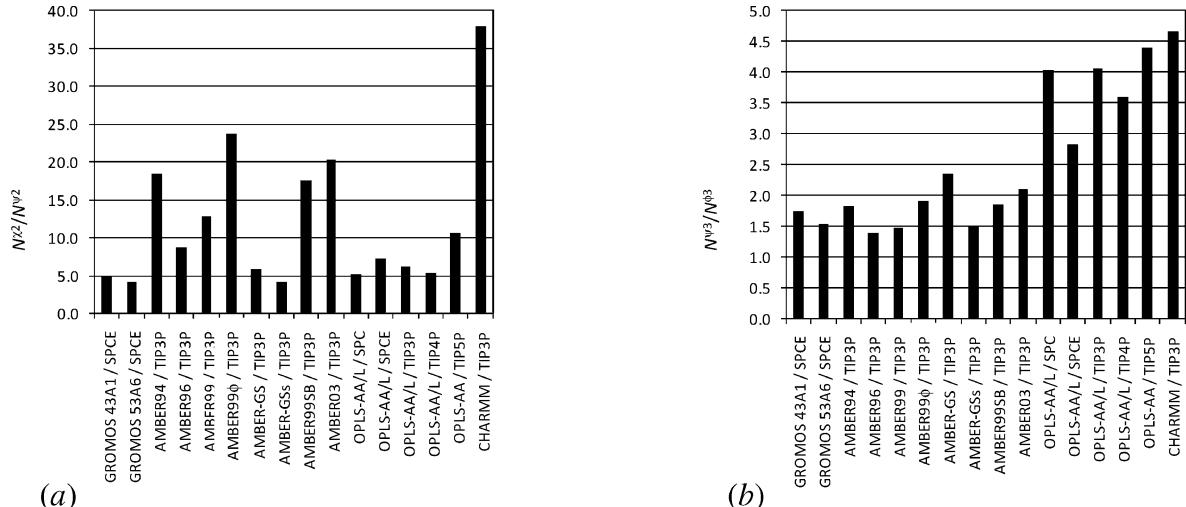
A clear difference in the backbone and the side chain dynamics of Pro-2 is evident from the  $N^{\psi_2}$  and  $N^{\chi_2}$  values shown in Table 2. For all the force fields considered, the ratio  $N^{\chi_2}/N^{\psi_2}$  varied between 4.2 (GROMOS 53A6) and 38.0 (CHARMM), suggesting that the Pro side chain fluctuates at significantly higher frequencies than the peptide backbone (Figure 5a). As for the backbone dynamics, for a given noncyclic amino acid residues, larger number of  $\psi$  transitions are expected compared to  $\phi$  transitions, since the C–C bond is longer than the C–N bond. Indeed, for all the force field considered  $N^{\psi_3}/N^{\phi_3} > 1$ . However, from the comparison of the  $N^{\psi_3}/N^{\phi_3}$  in Figure 5b, OPLS-AA/L and CHARMM are different from GROMOS and AMBER force fields; i.e., OPLS-AA/L and CHARMM give a higher preference to  $\psi$  torsional transitions over  $\phi$  transitions compared to GROMOS and AMBER.

In the case of the OPLS-AA/L force field we have carried out MD simulations using five different water models (Table 2). Compared to the force field dependence, less significant, but noteworthy, variations in the conformational behavior were observed. In particular, the  $p_f$  values for different water models vary between 34% (SPCE) and 47% (TIP4P). The number of backbone and side chain torsional transitions decreases on going from TIP3P to TIP5P ( $N^{\psi_2}$ ,  $N^{\phi_3}$ ,  $N^{\psi_3}$ , and  $N^{\chi_2}$  in Table 2). Also, the preference of  $\psi$  torsional transitions over  $\phi$  transitions is reduced for SPCE compared to other water models (Figure 5b). As expected, the results from the MD simulations using the SPC and TIP3P models were very similar.

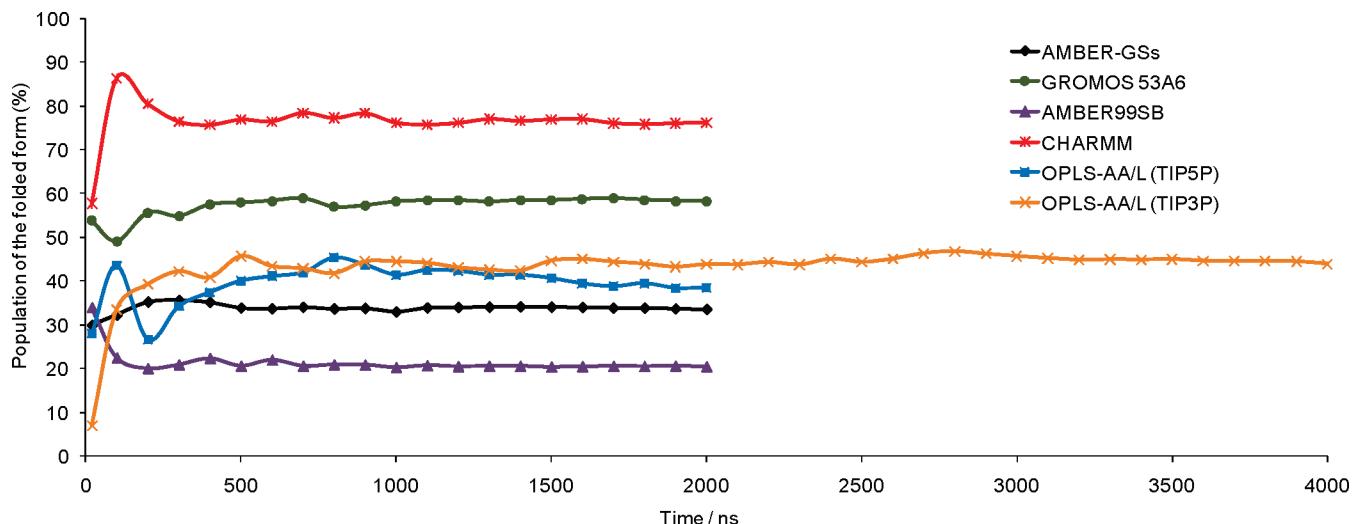
Since the peptide backbone shows slower fluctuation rates than the pyrrolidine ring, we have used the dependence of the population of the folded form (U-shaped backbone conformation,  $p_f$ , in %) on the simulation time to compare different force fields and assess whether 2  $\mu$ s MD simulations are sufficiently



**Figure 4.** Comparison of the dihedral angle potentials  $V_\phi$  (a) and  $V_\psi$  (b) for AMBER94 [ $V_\phi = 1.6736 - 1.6736 \cos^2(\phi - \pi)$  and  $V_\psi = 14.4348 + 3.138 \cos(\psi - \pi) + 2.092 \cos^2(\psi - \pi) - 13.3888 \cos^3(\psi - \pi)$ ] and AMBER96 [ $V_\phi = 6.0688 - 3.5564 \cos(\phi - \pi) - 2.5104 \cos^2(\phi - \pi)$  and  $V_\psi = 6.0688 - 3.5564 \cos(\psi - \pi) - 2.5104 \cos^2(\psi - \pi)$ ] force fields.



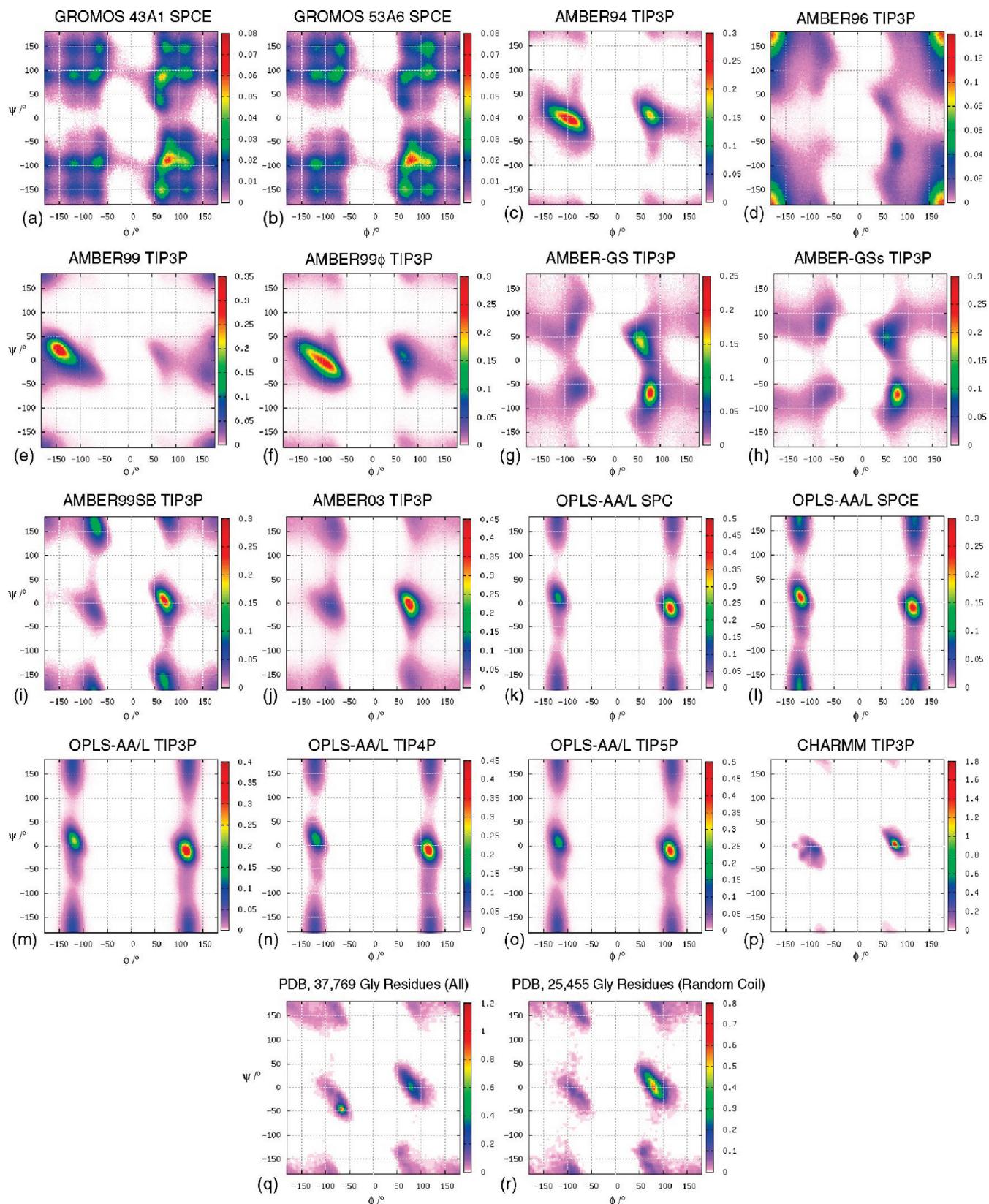
**Figure 5.** Comparison of different force fields using the ratio of the number ( $N$ ) of (a)  $\chi_2$  and  $\psi_2$  transitions for the Pro residue and (b)  $\psi_3$  and  $\phi_3$  transitions for the Gly-3 residue. While  $N^{\chi_2}/N^{\psi_2}$  in (a) reflects the relative frequency of side chain vs backbone transitions,  $N^{\psi_3}/N^{\phi_3}$  in (b) quantifies the preference of  $\psi$  torsional transitions over  $\phi$  transitions in the backbone.



**Figure 6.** MD predicted population of the folded form as a function of the simulation time.

long for this parameter to converge. While the majority of the force fields showed rapid convergence within the first half of the simulation, significantly slower convergence was found for OPLS-AA/L (Figure 6). Nevertheless, for the five MD simulations involving the OPLS-AA/L force field and different water models the variations of the predicted  $p_f$  values from 1.5, 1.6, 1.7, 1.8, 1.9, and 2.0  $\mu\text{s}$  long simulations were within  $\pm 0.8\%$  (SPC),  $\pm 1.2\%$  (SPCE),  $\pm 0.9\%$  (TIP3P),  $\pm 0.8\%$  (TIP4P), and

$\pm 1.1\%$  (TIP5P). In addition, a 4  $\mu\text{s}$  MD OPLS-AA/L TIP3P simulation predicted  $p_f = 43.9\%$  (Figure 6), which compares well with  $p_f = 43.8\%$  from the corresponding 2  $\mu\text{s}$  MD run. Similarly, only negligible or no differences were found between Ramachandran plots or the values of NMR parameters predicted by 2 and 4  $\mu\text{s}$  MD OPLS-AA/L TIP3P simulations. Thus, 2  $\mu\text{s}$  long MD OPLS-AA/L calculations were included into our further analysis.



**Figure 7.** Ramachandran population maps as predicted by MD simulations using different force fields [(a)–(p)] for Gly-3 of GPGG, together with PDB survey data of Hovmöller et al. [(q) and (r)].<sup>64,65</sup>

**4.2. Ramachandran Plots.** The differences in the geometry and the population of various backbone conformations are further illustrated using Ramachandran plots for the Gly-3 residue of GPGG (Figure 7). A density map (rather than a scatter plot) presentation of Ramachandran plots are used to reflect the difference in the populations of different regions.

As expected, there are clear similarities between GROMOS 43A1 and 53A6, AMBER-GS and AMBER-GSs, and OPLS-AA/L calculations using five different water models. Interestingly, the GROMOS plots reproduce the classical Ramachandran map for Gly,<sup>80</sup> which is built on the basis of the consideration of steric clashes and does not account for backbone dipole–dipole

**TABLE 3: Density Maxima Positions (deg, Uncertainties  $\pm 5^\circ$ ) from MD Ramachandran Density Maps for Pro-2 ( $\phi_2$  and  $\psi_2$ ) and Gly-3 ( $\phi_3$  and  $\psi_3$ ) Residues of GPGG**

force field	$\phi_2$	$\psi_2$	$\phi_3$	$\psi_3$
GROMOS 43A1/SPCE	-70	145	80	-90
GROMOS 53A6/SPCE	-70	145	80	-90
AMBER94/TIP3P	-70	-25	-95	-5
AMBER96/TIP3P	-75	145	$\pm 180$	$\pm 180$
AMBER99/TIP3P	-65	-25	-145	25
AMBER99 $\phi$ /TIP3P	-70	-25	-90	-10
AMBER-GS/TIP3P	-70	120	80	-70
AMBER-GSs/TIP3P	-70	120	80	-70
AMBER99SB/TIP3P	-70	150	70	10
AMBER03/TIP3P	-70	120	75	0
OPLS-AA/L/SPC	-75	105	120	-10
OPLS-AA/L/SPCE	-75	160	120	-10
OPLS-AA/L/TIP3P	-75	105	120	-10
OPLS-AA/L/TIP4P	-75	105	120	-10
OPLS-AA/TIP5P	-75	105, 160	120	-10
CHARMM/TIP3P	-60	120	80	5

interactions.<sup>81</sup> The AMBER96 plot reproduces some features of the classical Ramachandran map but differs in giving strong preference to the extended conformation with  $\phi_3 \approx \pm 180^\circ$  and  $\psi_3 \approx \pm 180^\circ$ , which belongs to the excluded region in the classical Ramachandran plot. The AMBER-GS and AMBER-GSs plots are dissimilar to other AMBER plots and show some resemblance with some parts of GROMOS plots. From the color scales shown for each plot, it is apparent that the difference in fractional populations of different conformations is considerably smaller in the case of GROMOS force fields; thus the population is spread over a larger number of conformations, covering >60% of the overall area. In contrast, significantly smaller area is covered in the CHARMM plot, suggesting the predominance of a few well-defined Gly-3 conformations.

The positions of the most populated conformations in the corresponding Ramachandran plots by each force field are summarized in Table 3. For the Gly-3 residue, there are encouraging similarities between AMBER99SB, AMBER03, and CHARMM force fields. These three force fields place the peak with the highest population at  $\phi_3 = 70\text{--}80^\circ$  and  $\psi_3 = 0\text{--}10^\circ$  (Table 3). However, populations of other conformations are clearly different (Figure 7). There is a resemblance in the  $\psi_3$  distribution between AMBER99SB and OPLS-AA/L plots; however, peaks are separated to significantly larger extent (by  $\sim 240^\circ$ ) along the  $\phi_3$  dimension in OPLS-AA/L plots compared to AMBER99SB ( $\sim 150^\circ$ ). The  $\phi_3$  separation of peaks is also  $\sim 150^\circ$  for AMBER94, AMBER99 $\phi$ , AMBER-GS, AMBER-GSs, AMBER03, and CHARMM.

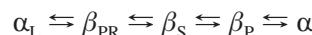
The overall agreement for the Pro-2 residue among different force fields is relatively satisfactory (Figure 8), except for the  $\psi_2$  preference of AMBER94, AMBER99, and AMBER99 $\phi$  (Table 3), which shifts the population density maximum into the  $\alpha$ -helical region of the Ramachandran plot. The characteristic feature of the OPLS-AA/L plots is that there is a clear  $\psi$  splitting into  $105^\circ$  and  $160^\circ$  in the upper left corner and the relative populations of the two peaks show a clear dependence on the water model used.

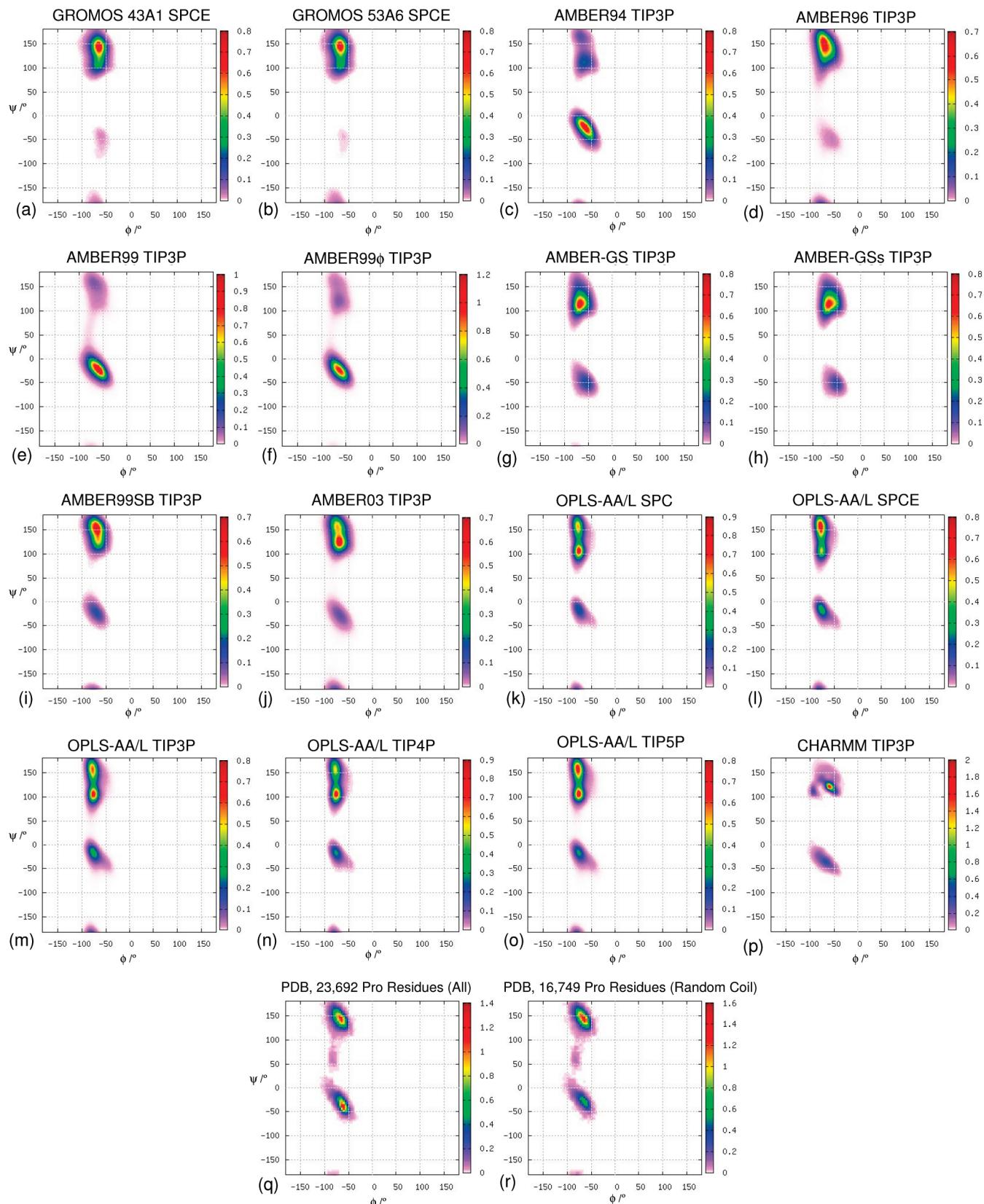
**4.3. Comparison of MD and X-ray Data.** We now compare the MD predicted Ramachandran plots to those from the X-ray diffraction. Hovmöller et al. analyzed 237 384 amino acid residues in 1042 protein subunit structures from the PDB determined by the X-ray diffraction to a resolution of  $\leq 2.0\text{ \AA}$  with  $R \leq 0.2$ . Over 40% of amino acid residues that did not form either helix or sheet conformations were grouped as random coils. Their Ramachandran plots for all Gly residues

and those within the random coil group (Figures 5 and 9 of ref 64, presented here in Figure 7q,r using the same type of presentation as for the MD data) resemble closely the Ramachandran plots for the Gly-3 residue of GPGG by MD AMBER99SB and AMBER03 simulations (Figure 7i,j). The AMBER99SB and CHARMM plots for the Pro-2 residue of GPGG (Figure 8i,p) show better agreement with the PDB survey for all and random coil Pro residues (Figure 8q,r) than other force fields.

Both PDB data sets for all random coil Gly residues, show five distinct  $\phi/\psi$  regions (Figure 7q,r), which have been classified previously as  $\alpha_L$  and  $\alpha$ , associated with helices and turns,  $\beta_P$ , corresponding to the polyproline II structure (left-handed helix),  $\beta_{PR}$ , reflection of the  $\beta_P$  region (right-handed helix) and  $\beta_S$ , associated with the extended conformation of residues in  $\beta$  sheets.<sup>82</sup> These five regions are also present in Ramachandran plots for Gly-3 of GPGG from MD AMBER99SB and AMBER03 simulations (Figure 7i,j). The regions assigned to each secondary structure are illustrated in Figure 9 using the  $\phi/\psi$  free energy surface calculated for the MD AMBER03 simulation. The MD-predicted  $\phi,\psi$  positions of the maximum populations in the five assigned regions are compared to those from the PDB survey<sup>64</sup> in Table 4. For the three regions ( $\alpha_L$ ,  $\alpha$ , and  $\beta_P$ ) the agreement between the MD and the random coil PDB data is satisfactory, whereas for the remaining  $\beta_{PR}$  and  $\beta_S$  regions a larger number of experimental data may be required for more accurate determination of the positions of density maxima in these regions of the PDB data. Nevertheless, the center of mass positions for the  $\beta_{PR}$  and  $\beta_S$  regions in the PDB plots compare well with those from MD AMBER99SB and AMBER03 simulations (Figure 7). By calculating the sum of the population in each of the regions and dividing the results by the total population, we determined the relative population of each of the regions. From the calculated populations shown in Table 4, there is a satisfactory agreement between MD and PDB data, with the MD AMBER03 simulation showing almost a quantitative agreement with the PDB random coil data. For the Pro-2 residue, the relative populations of the  $\alpha$  and  $\beta_P$  regions from the PDB data are reproduced best by the MD CHARMM simulation, while the density maxima positions of these two regions are reproduced best by the MD AMBER99SB simulation (Table 4).

By converting the population maps into free energy surfaces, we can estimate the energetic characteristics of the conformations and their transitions. Shown in Table 5 are the MD-predicted relative free energies (in  $\text{kJ mol}^{-1}$ ) of the minima in each of the five assigned regions for Gly-3 in GPGG. For comparison, the PDB survey data was also converted in the same way on the assumption that the statistical distribution of “frozen” Gly conformations in different proteins in the solid state represents a reasonable approximation to the distribution of possible conformations of a midchain Gly residue. As the AMBER03 force field reproduces best the PDB data for a large number of Gly residues in different environments (Figure 7j,r), we use the free energy map for this force field (Figure 9) to deduce the likely backbone transition path for a Gly residue in a peptide backbone. The PDB survey data cannot be used for this purpose as the transition areas are not well-defined in the corresponding maps (Supporting Information, Figure S3). In particular, Figure 9 suggests that the following sequence of transitions is most likely for Gly residues:

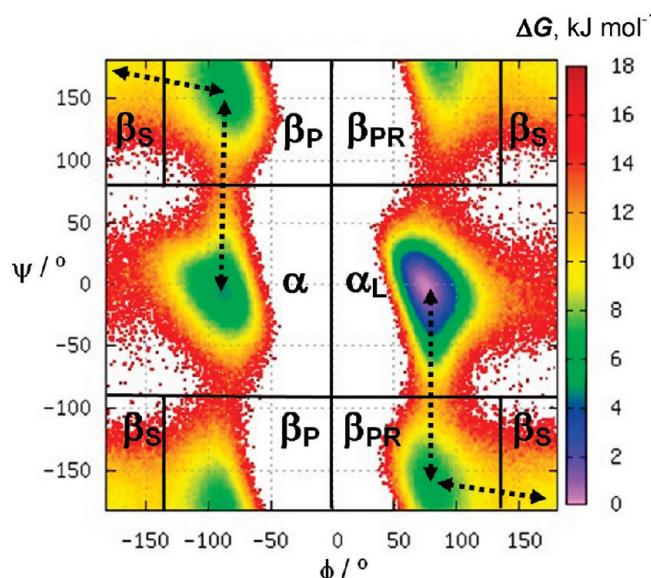




**Figure 8.** Ramachandran population maps as predicted by MD simulations using different force fields [(a)–(p)] for Pro-2 of GPGG, together with PDB survey data of Hovmöller et al. [(q) and (r)].<sup>64,65</sup>

with free energies of activation ( $\Delta G^\ddagger$ ) on the order of 4–12 kJ mol<sup>-1</sup>. A simplified schematic presentation of Gly backbone transitions is shown in Figure 10. To illustrate the flexibility of Gly residues, the  $\Delta G^\ddagger$  values are ca. 7 and 20 kJ mol<sup>-1</sup> for the  $\beta_P \rightarrow \alpha$  transition in Gly-3 (Figure 9) and Pro-2 residues (Figure

S4, Supporting Information), respectively, according to the MD AMBER03 simulation. The geometries of transition states (TS) can also be deduced from the free energy surface shown in Figure 9 (e.g.,  $\phi_{TS} \approx 77^\circ$  and  $\psi_{TS} \approx -90^\circ$  for the  $\alpha_L \rightleftharpoons \beta_{PR}$  equilibrium). The “staircase” pattern of transitions is also



**Figure 9.** Free energy  $\phi/\psi$  map for Gly-3 of GPGG as predicted by the MD AMBER03 simulation. Ten rectangular boxes within the map identify regions corresponding to the five secondary structures  $\alpha_L$ ,  $\alpha$ ,  $\beta_P$ ,  $\beta_{PR}$ , and  $\beta_S$ . Broken arrows indicate the likely transition path for the backbone conformations.

noteworthy:  $\alpha_L \rightleftharpoons \beta_{PR}$  and  $\beta_P \rightleftharpoons \alpha$  transitions involve only the change of the  $\psi$  angle (with the highest  $\Delta G^\ddagger$  of ca.  $12 \text{ kJ mol}^{-1}$  for the  $\alpha_L \rightarrow \beta_{PR}$  transition), whereas mainly the  $\phi$  angle changes on  $\beta_{PR} \rightleftharpoons \beta_S$  and  $\beta_S \rightleftharpoons \beta_P$  transitions (with the highest  $\Delta G^\ddagger$  of ca.  $5 \text{ kJ mol}^{-1}$  for the  $\beta_P \rightarrow \beta_S$  transition).

From the above comparison of the X-ray diffraction survey and the MD data for GPGG, satisfactory agreement with the experimental solid-state data is observed for AMBER03 and AMBER99SB force fields in the case of the Gly residue. For the Pro residue, satisfactory agreement with the experimental solid-state data is observed for CHARMM and AMBER99SB. As the MD simulations are for aqueous solutions, it is important to verify whether the same or other force fields will agree with solution state NMR data. Thus, we use experimentally measured NMR parameters of GPGG in water for the verification of the force fields considered. In this type of comparative study, it is essential to minimize the influence of uncertainties associated with the measurement of experimental parameters, as well as those arising from the approximate nature of the dependence of NMR parameters on molecular geometry. Thus, we have

**TABLE 4: MD-Predicted Populations (%) and  $\phi/\psi$  Positions (deg, in Parentheses) of the Density Maxima in the Five Assigned Regions for Gly-3 and Pro-2 in GPGG<sup>a,b</sup>**

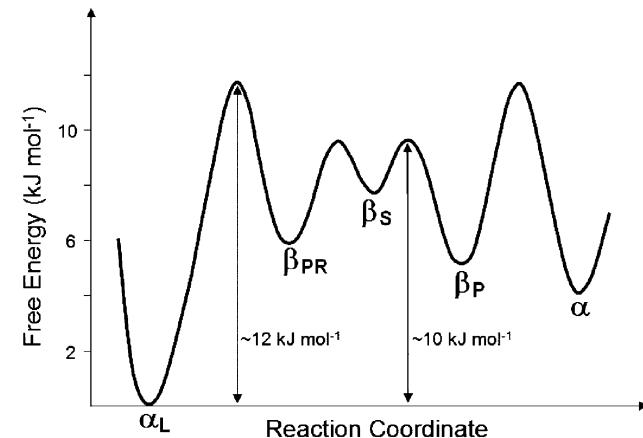
	residue	$\alpha_L$	$\alpha$	$\beta_P$	$\beta_{PR}$	$\beta_S$
MD, AMBER99SB	Gly-3	30 (70/10)	9 (-74/-12)	27 (-70/161)	23 (71/-159)	11 (176/179)
MD, AMBER03	Gly-3	51 (75/0)	15 (-89/-1)	15 (-81/153)	13 (78/-166)	6 (-179/-173)
MD, CHARMM	Gly-3	58 (79/4)	33 (-86/-5)	5 (-68/175)	4 (73/-179)	0
X-ray, all <sup>c</sup>	Gly-3	33 (80/0)	24 (-64/-44)	15 (-80/172)	14 (56/-136)	14 (-176/176)
X-ray, random coil <sup>d</sup>	Gly-3	47 (80/0)	11 (-88/0)	17 (-80/172)	17 (56/-136)	8 (156/176)
MD, AMBER99SB	Pro-2	0	13 (-66/-23)	87 (-68/151)	0	0
MD, AMBER03	Pro-2	0	9 (-67/-24)	91 (-68/123)	0	0
MD, CHARMM	Pro-2	0	32 (-65/-34)	68 (-57/122)	0	0
MD, OPLS-AA/L (TIP4P)	Pro-2	0	18 (-72/-18)	82 (-74/106)	0	0
X-ray, alle <sup>e</sup>	Pro-2	0	43 (-60/-36)	57 (-68/148)	0	0
X-ray, random coil <sup>f</sup>	Pro-2	0	32 (-64/-24)	68 (-60/144)	0	0

<sup>a</sup> For comparison, corresponding data for Gly and Pro residues from the PDB survey of Hovmöller et al.<sup>64,65</sup> are also shown. <sup>b</sup> Details of the assigned regions are provided in the Experimental Section, Ramachandran plots (section 2.5). <sup>c</sup> A total of 37 769 Gly residues were included in group “all”. <sup>d</sup> A total of 25 455 Gly residues were included in group “random coil”. <sup>e</sup> A total of 23 692 Pro residues were included in group “all”. <sup>f</sup> A total of 16 749 Pro residues were included in group “random coil”.

**TABLE 5: MD-Predicted Relative Free Energies (kJ mol<sup>-1</sup>) of the Minima in Each of the Five Assigned Regions for Gly-3 in GPGG<sup>a,b</sup>**

	$\alpha_L$	$\alpha$	$\beta_P$	$\beta_{PR}$	$\beta_S$
MD, AMBER99SB	0	4.0	1.5	1.9	4.8
MD, AMBER03	0	4.2	5.1	5.7	8.3
MD, CHARMM	0	3.9	9.2	9.5	15.8
X-ray, all <sup>c</sup>	2.0	0	5.8	5.7	6.8
X-ray, random coil <sup>d</sup>	0	4.4	4.0	4.4	6.4

<sup>a</sup> For comparison, corresponding data calculated for Gly residues in the PDB survey of Hovmöller et al.<sup>64,65</sup> are also shown. <sup>b</sup> Details of the assigned regions are provided in the Experimental Section, Ramachandran plots (section 2.5). <sup>c</sup> A total of 37 769 Gly residues were included in group “all”. <sup>d</sup> A total of 25 455 Gly residues were included in group “random coil”.



**Figure 10.** Sequence of backbone transitions for the midchain Gly-3 residue of GPGG as deduced from the free energy surface for the MD AMBER03 simulation. The relative free energies are 0 kJ mol<sup>-1</sup> ( $\alpha_L$ ), 5.7 kJ mol<sup>-1</sup> ( $\beta_{PR}$ ), 8.3 kJ mol<sup>-1</sup> ( $\beta_S$ ), 5.1 kJ mol<sup>-1</sup> ( $\beta_P$ ), and 4.2 kJ mol<sup>-1</sup> ( $\alpha$ ).

considered up to eight different NMR parameters and their dependences as merit criteria to identify force fields which satisfy all or the majority of them.

In the following, we first consider theoretical Karplus relationships for Gly and Pro residues, which are needed for the calculation of  $^3J_{HH}$  and  $^3J_{CH}$ -couplings in GPGG based on MD simulations, followed by comparative NMR vs MD analysis of vicinal couplings, NOEs and chemical shifts.

**TABLE 6: Vicinal  $^1\text{H}$ – $^1\text{H}$  and  $^1\text{H}$ – $^{13}\text{C}$  Couplings (Hz) Calculated Using Different DFT Functionals at the  $\phi = 0^\circ$  Geometry of AceAlaNHMe<sup>a</sup>**

DFT functional	$\text{C}'\text{--H}^N$	$\text{H}^N\text{--H}^\alpha$
B3LYP	7.88	1.86
B971	7.64	1.71
PBE1PBE	7.63	1.81
B98	7.63	1.72
PW91PW91	7.57	1.78
PBEPBE	7.54	1.76
B972	7.29	1.65
B972 <sup>b</sup>	7.32	1.53

<sup>a</sup> The triple- $\zeta$  basis set of EPR-III was used in all cases. Water solvent effects were accounted for using the IEFPCM method. Unless otherwise specified, the B3LYP/6-31G\*\* optimized geometry of AceAlaNHMe was used. <sup>b</sup> The geometry of AceAlaNHMe was optimized at the B972/6-31G\*\* level of theory.

#### 4.4. Parameterizations of Karplus-Type Equations for Gly and Pro Residues.

For the backbone conformational analysis, a Karplus relationship of the following form was used for calculations of  $^3J$ -couplings:

$$J = A \cos^2(\phi + \Delta\phi) + B \cos(\phi + \Delta\phi) + C \quad (1)$$

where  $\phi$  is a backbone dihedral angle formed by  $\text{C}^{i-1}\text{--N}^i\text{--C}_\alpha^i\text{--C}^i$ , and  $\Delta\phi$  is used to account for an offset angle for a given pair of nuclei.

Initially, we compared  $J$ -coupling calculations for AceAlaNHMe using seven different DFT functionals (Table 6). The  $\phi = 0^\circ$  geometry was chosen, as the  $^3J(\text{C}',\text{H}^N)$ -coupling reaches its maximum value at this angle, which in turn leads to the largest positive difference of  $^3J_{\text{calc}} - ^3J_{\text{expt}}$  (see, for example, Figure 2a in ref 83). Only a weak DFT functional dependence was observed (Table 6), with B3LYP and B972 predicting the highest and the lowest  $^3J$  values, respectively, for  $\text{C}'\text{--H}^N$  and  $\text{H}^N\text{--H}^\alpha$  pairs. To minimize possible sources of systematic errors, we have selected these two functionals for quantum-mechanical derivations of Karplus parameters. We note that both B972 and B3LYP have been recognized as sufficiently accurate functionals for the calculation of  $J$ -couplings.<sup>84–86</sup> As with AceAlaNHMe (Table 6), the triple- $\zeta$  basis set of EPR-III was used, which is optimized for the computation of magnetic properties and includes an enhanced s-part for better description of the electron density at nuclear region.<sup>87</sup>

One set of equations was derived for Gly residues in the middle of the peptide chain (e.g.,  $-\text{NH}^N\text{--CH}_2\text{--C}'\text{ONH}-$ ), and the other was optimized for either C-terminal (e.g.,  $-\text{NH}^N\text{--CH}_2\text{--C}'\text{OO}^-$ ) or the N-terminal Gly residues (e.g.,  $\text{H}_3\text{N}^+\text{--CHR--OC}'\text{--NH--CH}_2^\alpha-$ ). Two model molecules (diglycine and triglycine) were sufficient to include all possible coupling pathways in glycines (Figure 11).

To derive parameters of Karplus relationships, we have first optimized the linear conformations of two molecules using either B972 6-31G(d,p) or B3LYP 6-31G(d,p) levels of theory. Two types of calculations were considered subsequently for 37 values of the dihedral angle  $\phi$  fixed at  $-180^\circ$ ,  $-170^\circ$ , ...,  $+180^\circ$ : (i) no other geometric parameter was fixed, and (ii) in addition to  $\phi$ , the dihedral angle formed by  $\text{H}^N$ ,  $\text{N}$ ,  $\text{C}'$ , and  $\text{C}^\alpha$  atoms was fixed at the value obtained from the full geometry optimization of the linear conformations without any constraints:  $-178.2^\circ$  and  $+179.3^\circ$  for di- and triglycine, respectively, from B972 6-31G(d,p) geometry optimizations, and  $178.6^\circ$  and  $179.5^\circ$  for di- and triglycine, respectively, from B3LYP 6-31G(d,p)

geometry optimizations. The remaining degrees of freedom were then optimized using B972 6-31G(d,p) and B3LYP 6-31G(d,p) levels of theory. These optimized geometries were used for the B972 EPR-III and B3LYP EPR-III calculations of the  $J$ -couplings. We then used least-squares fittings to derive unknown parameters  $A$ ,  $B$ ,  $C$ , and  $\Delta\phi$  of eq 1. A simulated annealing algorithm was used for nonlinear fittings of  $^3J(\phi)$  dependences.<sup>45,46</sup> The root-mean-square (rms) deviations of the fittings for the set of calculations (ii) was significantly (ca. 7.5-fold) smaller than that for (i). Thus, in our further analysis we used only the results of type (ii) calculations requiring planarity of the peptide fragment on geometry optimizations with fixed  $\phi$  values.

The best-fit values of the Karplus parameters together with rms deviations of fittings are shown in Tables 7 and 8. For two of the vicinal couplings considered, the B972 EPR-III and B3LYP EPR-III derived Karplus curves are compared to those derived empirically for ubiquitin<sup>53</sup> and flavodoxin<sup>54</sup> in Figure 12. Due to the equal value of the  $^3J$ -couplings involving  $\text{H}^{\alpha 2}$  and  $\text{H}^{\alpha 3}$  protons of Gly with both  $\text{C}'$  and  $\text{H}^N$ , modified equations of the following general form were used for  $^3J$ -couplings involving both  $\text{H}^\alpha$ -protons of Gly in the case of empirical relationships:

$$J = \frac{1}{2}[A \cos^2(\phi + \Delta\phi) + A \cos^2(\phi - \Delta\phi) + B \cos(\phi + \Delta\phi) + B \cos(\phi - \Delta\phi)] + C \quad (2)$$

where the values of  $A$ ,  $B$ , and  $C$  were derived for chiral amino acid residues in ubiquitin<sup>53</sup> and flavodoxin.<sup>54</sup> The observed difference between the DFT and empirical curves in Figure 12 is largely due to the motional averaging effects reflected in the optimized values of the Karplus coefficients for ubiquitin and flavodoxin, while the DFT derived curves correspond to the “zero-motion” limit.<sup>88</sup> Indeed, as shown in Figure 13, the DFT B972 and B3LYP Karplus curves for the  $^3J(\text{H}^N,\text{H}^{\alpha 2})$ -coupling [similar to  $^3J(\text{H}^N,\text{H}^\alpha)$  in Ala] and  $^3J(\text{H}^N,\text{H}^\alpha)$ -coupling in midchain Gly residues show better agreement with the extrapolated “zero-motion” empirical curve<sup>88</sup> compared to two other empirical curves, in which the motional averaging effects are not accounted for. The motional averaging effect is further illustrated in Figure S5 in Supporting Information for  $^3J_{\text{HH}}$ -couplings in ethane.

For further validation of the DFT B972 and B3LYP Karplus parameters, we have compared 15 experimental  $^3J(\text{C}',\text{H}^N)$ -couplings for Gly residues in flavodoxin (varying between  $-0.6 \pm 0.35$  and  $+3.5 \pm 0.35$  Hz)<sup>54</sup> and 36 experimental  $^3J(\text{H}^N,\text{H}^{\alpha 2})$  and  $^3J(\text{H}^N,\text{H}^{\alpha 3})$ -couplings for 18 Gly residues in flavodoxin (varying between  $1.26 \pm 0.35$  and  $7.94 \pm 0.35$  Hz)<sup>54</sup> with those calculated from 30 ns MD AMBER99SB/TIP3P simulations of flavodoxin. The experimental  $^3J(\text{C}',\text{H}^N)$ -couplings for Gly residues in flavodoxin were reproduced with satisfactory rms deviations of 0.05 and 0.06 Hz by B972 and B3LYP Karplus parameters (Tables 7 and 8), respectively. Considering that  $^3J_{\text{HH}}$ -couplings are usually larger than  $^3J_{\text{CH}}$ -couplings, the experimental  $^3J(\text{H}^N,\text{H}^{\alpha 2})$ - and  $^3J(\text{H}^N,\text{H}^{\alpha 3})$ -couplings for Gly residues in flavodoxin were also reproduced satisfactorily with rms deviations of 0.48 and 0.36 Hz by B972 and B3LYP Karplus parameters, respectively. For comparison, the corresponding rms deviation for the extrapolated “zero-motion” empirical curve<sup>88</sup> modified for Gly using eq 2 was 0.44 Hz.

We have also derived the best-fit values of the Karplus parameters for the  $^3J(\text{C}',\text{H}^\alpha)$ -coupling path in Pro (Tables 7 and 8) using 2-[acetyl(methyl)amino]propanoic acid as an open chain model (Figure 11), in which the change of the  $^3J_{\text{CH}}$ -coupling

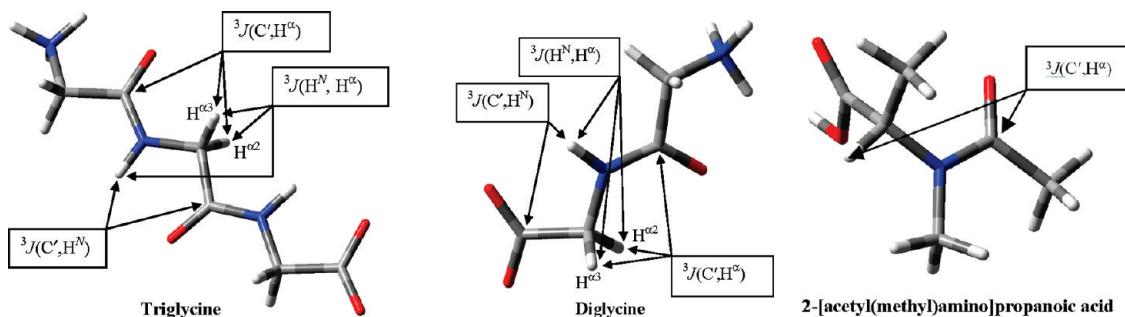


Figure 11. Model molecules used for the derivation of Karplus-type relationships for Gly and Pro residues.

**TABLE 7: B972 EPR-III Predicted Coefficients of Karplus Equations  $J = A \cos^2(\phi + \Delta\phi) + B \cos(\phi + \Delta\phi) + C$** 

residue	position	<i>A</i>	<i>B</i>	<i>C</i>	$\Delta\phi$	rms
C'-H <sup>N</sup>	Gly midchain	5.90	-1.56	-0.22	-179.4	0.079
	Gly C-terminal	5.45	-1.43	-0.24	178.4	0.085
C'-H <sup>a2</sup>	Gly midchain	4.80	-1.85	1.31	118.9	0.207
C'-H <sup>a3</sup>	Gly midchain	4.73	-1.88	1.31	-120.7	0.187
C'-H <sup>a</sup>	Gly midchain	-2.46	-1.00	4.92	179.5	0.083
C'-H <sup>a2</sup>	Gly N-terminal	4.46	-2.17	1.52	120.5	0.158
C'-H <sup>a3</sup>	Gly N-terminal	4.51	-2.17	1.52	-119.0	0.184
C'-H <sup>a</sup>	Gly N-terminal	-2.34	-1.16	4.94	-179.4	0.080
H <sup>N</sup> -H <sup>a2</sup>	Gly midchain	8.95	-1.22	0.33	-58.4	0.217
H <sup>N</sup> -H <sup>a3</sup>	Gly midchain	8.98	-1.29	0.28	58.6	0.232
H <sup>N</sup> -H <sup>a</sup>	Gly midchain	-4.02	-0.54	6.80	0.1	0.141
H <sup>N</sup> -H <sup>a2</sup>	Gly C-terminal	8.64	-1.33	0.58	-60.2	0.325
H <sup>N</sup> -H <sup>a3</sup>	Gly C-terminal	8.72	-1.29	0.57	57.7	0.331
H <sup>N</sup> -H <sup>a</sup>	Gly C-terminal	-3.97	-0.46	6.90	-1.5	0.145
C'-H <sup>a</sup>	Pro midchain	3.88	-1.21	0.85	116.0	0.216

**TABLE 8: B3LYP EPR-III Predicted Coefficients of Karplus Equations  $J = A \cos^2(\phi + \Delta\phi) + B \cos(\phi + \Delta\phi) + C$** 

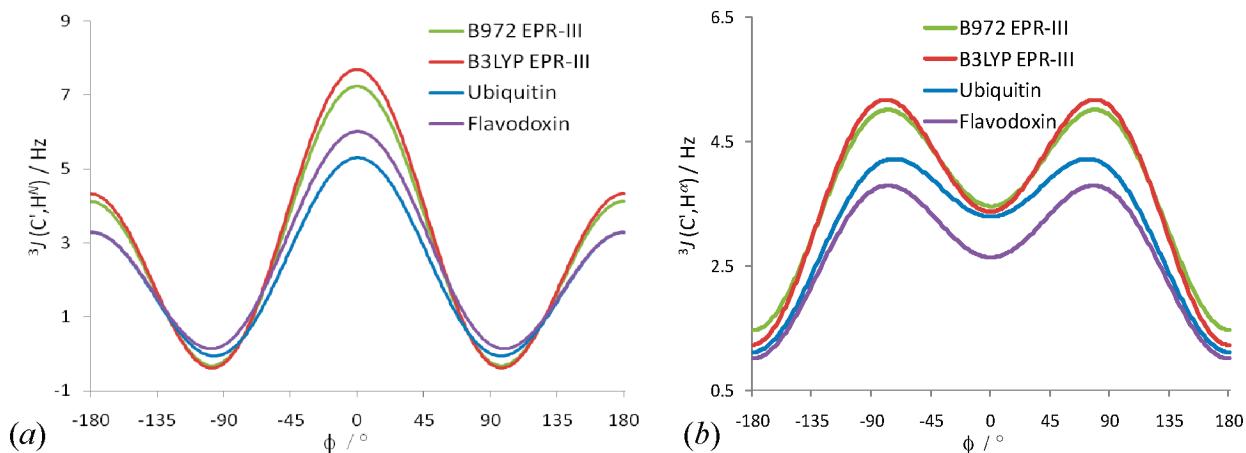
residue	position	<i>A</i>	<i>B</i>	<i>C</i>	$\Delta\phi$	rms
C'-H <sup>N</sup>	Gly midchain	6.28	-1.68	-0.27	-179.4	0.081
	Gly C-terminal	5.77	-1.54	-0.22	178.6	0.097
C'-H <sup>a2</sup>	Gly midchain	5.23	-2.01	1.08	119.2	0.199
C'-H <sup>a3</sup>	Gly midchain	5.15	-2.04	1.11	-119.3	0.179
C'-H <sup>a</sup>	Gly midchain	-2.77	-1.07	5.08	-179.7	0.076
C'-H <sup>a2</sup>	Gly N-terminal	5.03	-2.29	1.25	119.7	0.194
C'-H <sup>a3</sup>	Gly N-terminal	5.06	-2.29	1.26	-119.4	0.193
C'-H <sup>a</sup>	Gly N-terminal	-2.67	-1.23	5.12	-180.0	0.089
H <sup>N</sup> -H <sup>a2</sup>	Gly midchain	9.86	-1.50	0.42	-58.6	0.244
H <sup>N</sup> -H <sup>a3</sup>	Gly midchain	9.77	-1.53	0.43	59.6	0.254
H <sup>N</sup> -H <sup>a</sup>	Gly midchain	-4.60	-0.69	7.63	0.8	0.173
H <sup>N</sup> -H <sup>a2</sup>	Gly C-terminal	9.58	-1.58	0.65	-57.7	0.389
H <sup>N</sup> -H <sup>a3</sup>	Gly C-terminal	9.57	-1.59	0.73	60.3	0.370
H <sup>N</sup> -H <sup>a</sup>	Gly C-terminal	-4.44	-0.56	7.68	1.6	0.161
C'-H <sup>a</sup>	Pro midchain	4.32	-1.53	0.59	115.9	0.222

between the methine proton and the carbon atom of the amide carbonyl group was monitored. For the validation of these parameters, we have compared three experimental  $^3J(C',H^a)$ -couplings of 1.22 Hz (Pro-19), 1.71 Hz (Pro-37), and 1.06 Hz (Pro-38) in ubiquitin with those calculated from 20 ns MD AMBER99SB/TIP3P simulations of ubiquitin<sup>59</sup> using the B972 and B3LYP Karplus parameters. Both sets of parameters showed satisfactory agreement with experiment, although the B3LYP calculated couplings of 1.24 Hz (Pro-19), 1.69 Hz (Pro-37), and 0.97 Hz (Pro-38) were in significantly better agreement with

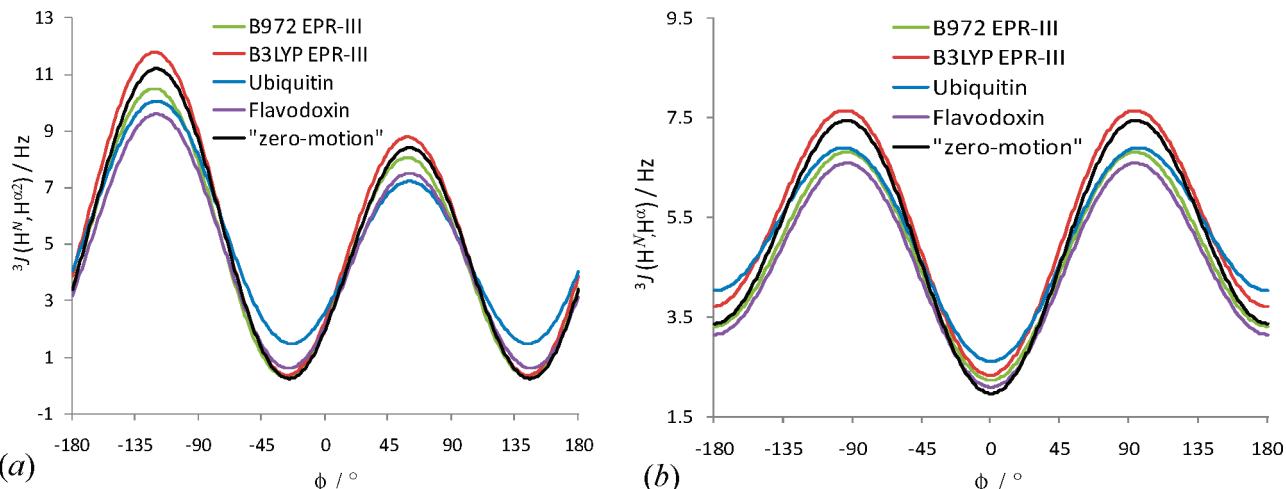
experiment compared to the B972 calculated couplings of 1.53 Hz (Pro-19), 1.95 Hz (Pro-37), and 1.27 Hz (Pro-38).

**4.5. Comparison of Experimental and Calculated Parameters: Vicinal *J*-Couplings.** Four  $^3J_{CH}$  and two  $^3J_{HH}$  were available from NMR measurements for the GPGG backbone (Supporting Information, Table S1), which were used for “experiment vs MD” comparisons (Table 9). Two empirical (corresponding to  $\text{rms}_{J_{el}}$  and  $\text{rms}_{J_{el}}$  in Table 9)<sup>53,54</sup> and two quantum-mechanically derived equations (corresponding to  $\text{rms}_{J_{q1}}$  and  $\text{rms}_{J_{q2}}$  in Table 9) were used to exclude possible model dependent deficiencies. For  $\text{rms}_{J_{q1}}$  and  $\text{rms}_{J_{q2}}$ , we used B972- and B3LYP-predicted Karplus type relationships described above (Tables 7 and 8). Using four sets of Karplus relationships (Supporting Information, Tables S1–S4), the best performing force fields are identified as those with  $\text{rms}_{J_{q1}} < 0.6$  Hz,  $\text{rms}_{J_{q2}} < 0.6$  Hz,  $\text{rms}_{J_{el}} < 0.6$  Hz, and  $\text{rms}_{J_{el}} < 0.9$  Hz (Table 9). The same set of force fields (OPLS-AA/L, AMBER-GSs, and AMBER99SB) was found to reproduce experimentally measured values better than others by four different types of calculations. The minimum values of  $\text{rms}_{J_{q1}}$  (0.36 Hz for OPLS-AA/L) and  $\text{rms}_{J_{q2}}$  (0.38 Hz for AMBER-GSs) are at least 4-fold smaller than the corresponding maximum values of  $\text{rms}_{J_{q1}}$  (1.62 Hz for AMBER99) and of  $\text{rms}_{J_{q2}}$  (1.55 Hz for AMBER99). All four types of rms calculations using *J*-couplings identified AMBER99, AMBER96, and AMBER03 as the worst performing force fields.

Ten vicinal  $^3J_{HH}$ -couplings of the pyrrolidine ring (reported in ref 79) were also analyzed using an empirically derived Karplus relationship by Haasnoot et al.,<sup>36</sup> which has been shown to satisfactorily reproduce experimental  $^3J_{HH}$ -couplings in prolines.<sup>89</sup> The NMR-predicted conformational characteristics of the pyrrolidine ring in GPGG (Table 10) showed only minor changes compared to simple L-prolines.<sup>15,79,89–93</sup> The level of agreement between NMR and MD simulations is reflected in  $\text{rms}_{J_p}$  values included in Table 10. AMBER-GSs and AMBER99SB were among the best performing force fields with  $\text{rms}_{J_p}$  less than 0.8 Hz (Table 10). There is a significant difference between the minimum (0.69 Hz for AMBER94 and AMBER99SB) and the maximum values of  $\text{rms}_{J_p}$  (2.67–2.72 Hz for OPLS-AA/L). Other conformational characteristics of the Pro side chain, such as the population of the C<sup>γ</sup>-endo conformation and the ring pseudorotation parameters (Table 10), were also in satisfactory agreement with NMR for these two force fields. The MD calculations based on the GROMOS 53A6 force field were in particularly good agreement with NMR in predicting the population ratio of C<sup>γ</sup>-endo/C<sup>γ</sup>-exo conformations and their geometries. The worst among the force fields used was OPLS-AA/L, which predicts almost complete predominance of the C<sup>γ</sup>-endo conformation. The endocyclic torsional angles predicted by OPLS-AA/L ( $\chi_m = 30^\circ$ , Table 10) were significantly smaller than those from NMR or other force fields (Table 10).



**Figure 12.** Comparison of Karplus curves for (a)  ${}^3J(C', H^N)$ -couplings in midchain Gly residues and (b)  ${}^3J(C', H^\alpha)$ -couplings in midchain Gly residues.



**Figure 13.** (a) Comparison of B972 and B3LYP Karplus curves for the  ${}^3J(H^N, H^{\alpha 2})$ -coupling in midchain Gly residues with empirical Karplus curves based on data for ubiquitin<sup>53</sup> and flavodoxin<sup>54</sup> and a "zero-motion" empirical curve.<sup>88</sup> (b) Comparison of B972 and B3LYP Karplus curves for the  ${}^3J(H^N, H^\alpha)$ -coupling (calculated as  $\frac{1}{2}[{}^3J(H^N, H^{\alpha 2}) + {}^3J(H^N, H^{\alpha 3})]$ ) in midchain Gly residues with empirical Karplus relationships<sup>53,54,88</sup> modified for Gly using eq 2.

**TABLE 9: MD-Predicted Conformational Characteristics and the RMS Deviations for "NMR vs MD" for GPGG in Water<sup>a</sup>**

	rms <sub>d</sub> ( $\text{\AA}$ )	rms <sub>Jq1</sub> (Hz)	rms <sub>Jq2</sub> (Hz)	rms <sub>Jel1</sub> (Hz)	rms <sub>Jel2</sub> (Hz)	rms <sub><math>\Delta\delta_1</math></sub> (ppm)	rms <sub><math>\Delta\delta_2</math></sub> (ppm)
GROMOS 43A1/SPCE	0.43	0.71	0.59	0.77	1.05	<b>0.036</b>	<b>0.038</b>
GROMOS 53A6/SPCE	0.43	0.76	0.67	0.82	1.09	<b>0.038</b>	<b>0.046</b>
AMBER94/TIP3P	<b>0.32</b>	0.89	0.87	0.90	1.17	0.135	0.064
AMBER96/TIP3P	0.37	1.49	1.38	1.46	1.78	<b>0.036</b>	<b>0.030</b>
AMBER99/TIP3P	0.43	1.62	1.55	1.56	1.87	0.169	0.115
AMBER99 $\phi$ /TIP3P	0.34	0.85	0.82	0.88	1.15	0.170	0.127
AMBER-GS/TIP3P	<b>0.23</b>	<b>0.58</b>	<b>0.49</b>	0.63	0.91	<b>0.044</b>	<b>0.035</b>
AMBER-GSs/TIP3P	<b>0.22</b>	<b>0.50</b>	<b>0.38</b>	<b>0.58</b>	<b>0.87</b>	<b>0.044</b>	<b>0.041</b>
AMBER99SB/TIP3P	<b>0.30</b>	<b>0.57</b>	<b>0.48</b>	<b>0.57</b>	<b>0.86</b>	<b>0.020</b>	<b>0.026</b>
AMBER03/TIP3P	<b>0.31</b>	1.44	1.46	1.31	1.55	<b>0.028</b>	<b>0.030</b>
OPLS-AA/L/SPC	<b>0.32</b>	<b>0.36</b>	<b>0.41</b>	<b>0.55</b>	<b>0.72</b>	0.072	0.025
OPLS-AA/L/SPCE	<b>0.32</b>	<b>0.37</b>	<b>0.42</b>	<b>0.55</b>	<b>0.71</b>	<b>0.043</b>	0.025
OPLS-AA/L/TIP3P	<b>0.32</b>	<b>0.36</b>	<b>0.42</b>	<b>0.55</b>	<b>0.71</b>	0.061	0.023
OPLS-AA/L/TIP4P	<b>0.31</b>	<b>0.37</b>	<b>0.41</b>	<b>0.55</b>	<b>0.72</b>	0.073	<b>0.026</b>
OPLS-AA/L/TIP5P	<b>0.32</b>	<b>0.36</b>	<b>0.43</b>	<b>0.54</b>	<b>0.70</b>	0.056	<b>0.021</b>
CHARMM/TIP3P	<b>0.32</b>	1.28	1.34	1.18	1.41	<b>0.046</b>	0.079

<sup>a</sup> Shown are the rms deviations between experiment and MD predictions for distances (rms<sub>d</sub>),  ${}^3J$ -couplings using Karplus coefficients derived from the B972 EPR-III (rms<sub>Jq1</sub>) and B3LYP EPR-III calculations (rms<sub>Jq2</sub>),  ${}^3J$ -couplings using Karplus coefficients derived empirically for ubiquitin (rms<sub>Jel1</sub>)<sup>53</sup> and flavodoxin (rms<sub>Jel2</sub>),<sup>54</sup> the  ${}^1\text{H}$  chemical shift differences for  $\alpha$ -Gly and  $\delta$ -Pro methylene protons by SHIFTs (rms <sub>$\Delta\delta_1$</sub> )<sup>55,56</sup> and TOTAL (rms <sub>$\Delta\delta_2$</sub> ).<sup>57</sup> The near-minimal RMS values are highlighted using bold letters (RMS<sub>d</sub> < 0.33  $\text{\AA}$ , RMS<sub>Jq1</sub> < 0.6 Hz, RMS<sub>Jq2</sub> < 0.6 Hz, RMS<sub>Jel1</sub> < 0.6 Hz, RMS<sub>Jel2</sub> < 0.9 Hz, RMS <sub>$\Delta\delta_1$</sub>  < 0.05 ppm, and RMS <sub>$\Delta\delta_2$</sub>  < 0.05 ppm).

#### 4.6. Internuclear Distances from Nuclear Overhauser Effects.

Solution structure determinations of proteins by NMR rely mainly on large negative nuclear Overhauser effects (NOEs)

which are common for high molecular weight species.<sup>40</sup> In the case of small molecular weight species, NOEs are usually small and positive, which limits their use for structure determinations.

**TABLE 10: Conformational Populations and Geometries of the Pro Ring in GPGG in Water As Predicted by NMR and by MD Simulations Using Various Force Fields<sup>a</sup>**

method	$P^{\text{endo}}/\chi_m$ (deg)	$P^{\text{exo}}/\chi_m$ (deg)	$x^{\text{endo}}$	rms <sub>J<sub>HH</sub></sub> (Hz)
GROMOS 43A1/SPCE	183/38	5/38	0.47	0.88
GROMOS 53A6/SPCE	181/38	7/38	0.52	<b>0.71</b>
AMBER94/TIP3P	184/35	5/35	0.52	<b>0.69</b>
AMBER96/TIP3P	180/35	8/35	0.68	1.07
AMBER99/TIP3P	184/35	4/35	0.47	0.89
AMBER99 $\phi$ /TIP3P	184/35	5/35	0.52	<b>0.71</b>
AMBER-GS/TIP3P	182/34	6/34	0.61	<b>0.77</b>
AMBER-GSs/TIP3P	181/34	6/34	0.57	<b>0.74</b>
AMBER99SB/TIP3P	180/35	14/35	0.59	<b>0.69</b>
AMBER03/TIP3P	179/34	13/34	0.72	1.34
OPLS-AA/L/SPC	170/30	-8/30	0.96	2.72
OPLS-AA/L/SPCE	170/30	-7/30	0.95	2.67
OPLS-AA/L/TIP3P	170/30	-8/30	0.96	2.70
OPLS-AA/L/TIP4P	170/30	-8/30	0.96	2.70
OPLS-AA/L/TIP5P	170/30	-7/30	0.96	2.70
CHARMM/TIP3P	183/36	4/36	0.44	0.98
experiment (NMR)	189/41	11/41	0.54	

<sup>a</sup> Shown are the pseudorotation phase angle ( $P$ ) and the pseudorotation amplitude ( $\chi_m$ ) for the C'-endo and C'-exo conformations of the pyrrolidine ring, population of the C'-endo conformer ( $x^{\text{endo}}$ ,  $x^{\text{endo}} + x^{\text{exo}} = 1$ ) and the rms deviation (rms<sub>J<sub>HH</sub></sub>) between experiment and MD predictions for  $^3J_{\text{HH}}$ -couplings of the pyrrolidine ring using the Karplus equation of Haasnoot et al.<sup>36</sup> The near-minimal RMS values (<0.8 Hz) are highlighted using bold letters.

However, the availability of a small number of key NOEs combined with computational methods may be sufficient for complete structure predictions in certain cases.<sup>94</sup> NOESY cross-peaks for 7 proton pairs were measured for GPGG (Supporting Information, Table S5), volume integrations of which were used for internuclear distance estimates based on the  $r^{-6}$  dependence of NOEs.<sup>40</sup> In a similar manner we have estimated averaged values of internuclear distances from MD runs using  $2 \times 10^6$  frames over  $2 \mu\text{s}$  time length for each force field. The root-mean-square deviation between experiment and the MD predictions of distances (rms<sub>d</sub>) are shown in Table 9. The minimum value of rms<sub>d</sub> (0.22 Å for AMBER-GSs) is almost 2-fold smaller than its maximum value (0.43 Å for AMBER99, GROMOS 53A6, and 43A1). In terms of comparative performance of force fields, the best performing force fields with rms<sub>d</sub>  $\leq 0.33$  Å were AMBER-GSs, AMBER-GS, AMBER99SB, AMBER03, OPLS-AA/L, AMBER94, and CHARMM.

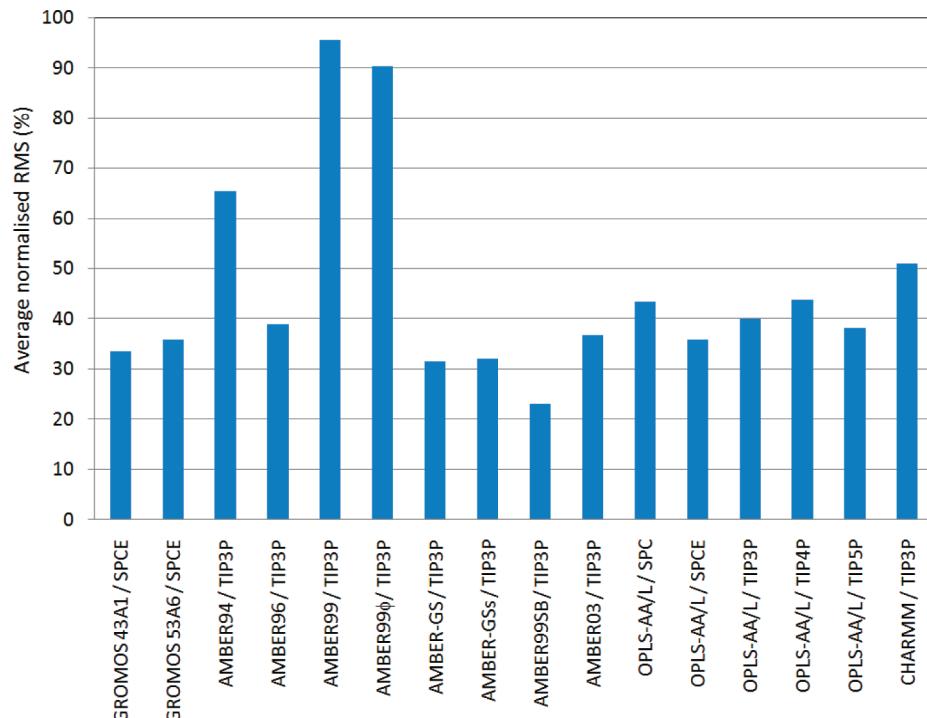
**4.7. Chemical Shifts.** Compared to  $^3J$ -couplings and NOEs, NMR chemical shifts provide less well-defined dependence on molecular geometries and dynamics, as they exhibit a more complex structural and electronic dependence.<sup>95</sup> Nevertheless, there has been a renewed interest in their applications for structural elucidations, as chemical shifts are relatively easy to measure.<sup>95,96</sup> We calculated averaged  $^1\text{H}$  NMR chemical shifts using empirical correlations (as implemented in SHIFT<sup>55,56</sup> and TOTAL<sup>57</sup>). Calculations of rms over 13 different  $^1\text{H}$  shifts for protons bonded to carbon atoms revealed no difference in the performance of different force fields, with rms varying between 0.21–0.23 ppm (SHIFT) and 0.22–0.28 ppm (TOTAL). The chemical shift difference for a pair of methylene protons ( $\Delta\delta$ , ppm) is expected to be more sensitive to the motional averaging effects. Besides, the use of the difference of the chemical shifts (as opposed to the actual values of the chemical shifts) allows us to eliminate possible sources of errors due to chemical shift calibration. Thus, we used the values of  $\Delta\delta$  for  $\alpha$ -Gly and  $\delta$ -Pro (the nearest to the peptide backbone CH<sub>2</sub> of Pro) methylene

protons. Several force fields, including AMBER-GSs and AMBER99SB, showed satisfactory performance with rms <sub>$\Delta\delta$</sub>   $< 0.05$  ppm by both SHIFT and TOTAL calculations (Table 9). It is noteworthy that AMBER99SB shows significantly smaller rms <sub>$\Delta\delta$</sub>  values than AMBER-GSs.

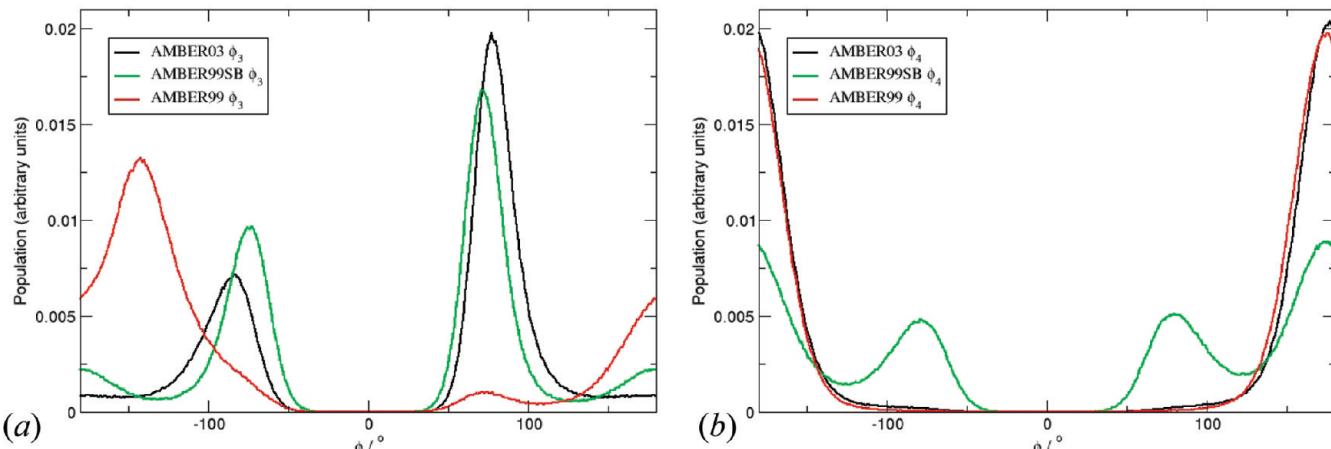
**4.8. Concluding Analysis of Results.** Comparisons of different force fields using rms thresholds (Tables 9 and 10) identify AMBER-GSs and AMBER99SB as the most reliable force fields for reproducing NMR measured parameters, which are sensitive to both the peptide backbone and the Pro side chain conformations and dynamics. In addition to rms threshold values, we have also considered normalized rms values (in %), which are calculated as the rms values divided by the corresponding range of observed values (between 2.6–4.1 Å for distance measurements, 1.1–6.3 Hz for backbone  $J$ -couplings, 5.4–8.5 Hz for the Pro side chain  $J$ -couplings, and 0–0.05 ppm for  $^1\text{H}$  chemical shift differences). We calculated the average values of normalized rms over  $N = 8$  different NMR parameters and dependences. From the result shown in Figure 14, AMBER99SB shows the best agreement with NMR, followed by AMBERGS and AMBERGSs. The average normalized rms for AMBER99SB was less than that for AMBER-GS or AMBER-GSs by 9%. The worst agreement was found for AMBER99, followed by AMBER99 $\phi$  and AMBER94.

Surprisingly, force fields showing very different Ramachandran plots show relatively good overall agreement with NMR measurements. The agreement for AMBER99SB, AMBER-GS, and AMBER-GSs with NMR is particularly noteworthy. A closer inspection of plots shown in Figure 7 reveals that the main features of the Ramachandran plots for AMBER-GS and AMBER-GSs can be reproduced from the AMBER99SB plot by  $\psi$  shifting of the peaks. In other words, the peaks of approximately equal density appear in Ramachandran plots of AMBER99SB and AMBER-GSs at approximately the same  $\phi_3$  values, but at significantly different  $\psi_3$  values. Thus, it is likely that NMR parameters considered above are less sensitive to variations of  $\psi_3$ , which is expected, as the  $^3J_{\text{HH}}$ - and  $^3J_{\text{CH}}$ -couplings used depend on  $\phi$  only. In principle, the  $\psi$  dependence of  $^{15}\text{N}$   $^3J$ -couplings can be used; however, measurements of these couplings require  $^{15}\text{N}$ -isotope labeling and, more importantly, the values of these couplings vary over a very narrow range and their Karplus dependence is not straightforward.<sup>83</sup> Another important factor complicating the NMR analysis for GPGG is the small chemical shift differences for the methylene protons of Gly residues in GPGG, which did not allow us to extract more selective structural information from experiment. For non-Gly amino acid residues, NOEs and chemical shifts are expected to be more sensitive to variations in  $\psi$ . Despite these shortcomings, comparison of NMR and MD results identify AMBER99SB as the force field showing the best overall agreement with experiment. This conclusion is also supported by the above-discussed similarity of MD AMBER99SB Ramachandran plots of Gly-3 and Pro-2 in GPGG with those for Gly and Pro residues from the PDB survey.

For the Gly-3 residue, AMBER03 matched almost exactly the random coil PDB survey for Gly residues (Table 4). However, from the large differences between the experimental and MD-predicted  $^3J$ -couplings for H<sup>a</sup>(4)–C'(3), H<sup>N</sup>(4)–C'(4), and H<sup>N</sup>(4)–H<sup>a</sup>(4) pairs of nuclei (Tables S1–S4, Supporting Information), it is apparent that the AMBER03 force field fails to correctly predict the population distribution of the backbone angle  $\phi$  of the Gly-4 residue. In Figure 15 we compare the population distribution curves for  $\phi$  angles in the midchain Gly-3 ( $\phi_3$ ) and the C-terminal Gly-4 ( $\phi_4$ ) residues obtained from MD



**Figure 14.** Comparison of the average normalized rms (in %) values calculated over eight different NMR parameters and dependences for various force fields used in MD simulations. Lower values indicate the better overall agreement between experimental and MD-predicted NMR parameters.



**Figure 15.** Backbone  $\phi$  angle distributions for (a) the midchain Gly-3 and (b) the C-terminal Gly-4 residues in GPGG in water from MD simulations using AMBER99 (red), AMBER99SB (green), and AMBER03 (black).

simulations using AMBER99, AMBER99SB, and AMBER03. While the  $\phi_3$  population distribution for AMBER03 resembles that of AMBER99SB, the  $\phi_4$  population distribution for AMBER03 is almost identical to that of AMBER99. Similar  $\phi_3$  and  $\phi_4$  distribution curves for AMBER03 were obtained from a 10 ns long MD simulation on using the AMBER (version 10)<sup>97</sup> implementation of this force field. The failure of AMBER03 to correctly predict the population distribution of the backbone angle  $\phi$  of the Gly-4 residue arises from the fact that the newly derived C–N–C <sub>$\alpha$</sub> –H <sub>$\alpha$</sub>  and N–C–C <sub>$\alpha$</sub> –H <sub>$\alpha$</sub>  torsion parameters of ref 30 are used for the midchain Gly residues only, whereas the terminal Gly residues are treated in the same way as in AMBER99. Thus, a revision of the parameters used for the terminal residues is needed to improve the performance of AMBER03.

Preliminarily, our modified MD AMBER03/TIP3P simulation, in which the N–C–C <sub>$\alpha$</sub> –H <sub>$\alpha$</sub>  and C–N–C <sub>$\alpha$</sub> –H <sub>$\alpha$</sub>  torsion

parameters (Table 1 of ref 30) are used for terminal Gly-1 and Gly-4 residues, respectively, showed a significant improvement in the performance of the AMBER03 force field (Supporting Information, Tables S1–S8 and Figure S6) with the average normalized rms of 24%, which is only slightly higher than that of AMBER99SB (23%, Figure 14). No changes in the Pro ring characteristics were found compared to the original AMBER03 simulation (Table 10), whereas an improved agreement with the PDB data for the Pro-2 residue was observed (Supporting Information, Table S8 and Figure S6). The MD-predicted populations for Gly-3 and Pro-2 residues (Supporting Information, Table S8) and the population of the folded conformation ( $p_f = 23\%$ ) showed improved agreement with AMBER99SB compared to the unmodified AMBER03 simulation. While a more systematic parametrization for the terminal residues is desirable, these preliminary results nevertheless verify the validity of the approach undertaken in ref 30, provided both

the midchain and terminal residues are parametrized in a consistent manner. In principle, further improvement can be achieved if parameters used for the Pro ring are also revised, as both modified and unmodified AMBER03 simulations predict a significantly higher population of the C<sup>γ</sup>-endo conformer ( $x^{\text{endo}} = 0.72$ ) compared to NMR ( $x^{\text{endo}} = 0.54$ ) and AMBER99SB ( $x^{\text{endo}} = 0.59$ ).

Remarkably, there is also a satisfactory agreement between the results of NMR and Ramachandran analyses for the worst performing force fields. In particular, the force fields showing the worst agreement with NMR data (AMBER99, AMBER99 $\phi$ , and AMBER94; Figure 14) populate predominantly the Pro geometry with  $\psi_2 = -25^\circ$  ( $\alpha$ -helix), whereas all other force fields give preference to  $\psi_2$  between  $105^\circ$  and  $160^\circ$  ( $\beta_P$  polyproline II structure, Table 3).

Seemingly, the performance of the OPLS-AA/L force field is satisfactory for the backbone conformations based on the rms values obtained (Table 9). However, the OPLS-AA/L Ramachandran plots reveal significant differences in Gly backbone geometries compared to AMBER99SB and the PDB survey data. In particular,  $\phi$  values for both  $\alpha$  and  $\beta$  secondary structures (except  $\beta_S$ ) of Gly are near  $+75^\circ$  and  $-75^\circ$  for AMBER99SB and PDB data, whereas OPLS-AA/L places  $\phi$  near  $+120^\circ$  and  $-120^\circ$  (Figure 7). From the corresponding Karplus curves (see Figures 12 and 13b),  $^3J$ -couplings of similar magnitudes are expected at  $\phi \approx \pm 75^\circ$  and  $\phi \approx \pm 120^\circ$ , leading to similar rms<sub>j</sub> values for OPLS-AA/L and AMBER99SB force fields (Table 9). This observation necessitates a detailed analysis of Ramachandran plots together with the Karplus dependence of  $^3J$ -couplings. The less hidden drawback of the OPLS-AA/L force field is for the Pro side chain (Table 10), where by far the worst performance was found for OPLS-AA/L compared to other force fields. Furthermore, our results identify OPLS-AA/L as computationally more expensive force field than others, as fewer torsional transitions were observed for OPLS-AA/L compared to other force fields, leading to slower convergence of the conformational populations. However, we note that a detailed experimental verification may be needed to identify force fields that best approximate experimentally measured frequencies of backbone and side chain dynamics.

MD simulations using five different water models (SPC, SPC/E, TIP3P, TIP4P, and TIP5P) were also compared for the OPLS-AA/L force field. In terms of the differences in predicted NMR parameters and the preferred solute conformational geometries, the dependence on the water model was found to be less critical compared to the force field dependence, justifying the use of computationally less expensive SPC/E and TIP3P models. However, as shown previously, the choice of water model may be important when the calculated thermodynamic properties are of interest.<sup>6</sup> It may also be important when the actual frequencies of motions (rather than time-averaged parameters) are of interest, since the number of peptide torsional transitions depends on the water model considered (Table 2). In particular, on going from TIP3P to TIP5P, the number of the  $\psi_2(\text{Pro})$ ,  $\phi_3(\text{Gly}-3)$ ,  $\psi_3(\text{Gly}-3)$ , and  $\chi_2(\text{Pro})$  transitions decreases (Table 2), which correlates with the decrease of the self-diffusion coefficient of the water model (Table 1); i.e., the higher rate of the translational diffusion of the solvent molecules leads to higher frequency of exchange of bound and bulk water in the vicinity of peptide, which in turn induces a higher number of backbone and side chain transitions in the peptide molecule. Such environment-dependent dynamics is in agreement with experimental findings in hydrated solid proteins.<sup>98</sup> Thus, in terms of statistical sampling of peptide conformations and calculations

of time-averaged parameters, the TIP3P model is computationally more efficient than TIP4P or TIP5P, while the TIP5P model is likely to offer a more accurate choice for the assessment of the motional time scales and related parameters (e.g., diffusion coefficients and relaxation times) since it reproduces best the diffusion constant of water measured experimentally.

**Acknowledgment.** We thank University College London for the provision of the NMR and computational facilities and EPSRC for financial support. We are grateful to Brian Alston, Jeremy Yates, and Clare Gryce from the UCL Legion Research Computing team for their support. Prof. S. Hovmöller and Dr. T. Zhou are thanked for the provision of their PDB survey data.<sup>64,65</sup> Dr. J. Schmidt is thanked for the provision of the details of their *J*-coupling measurements.<sup>54</sup> Helpful and stimulating suggestions by the reviewers are gratefully acknowledged.

**Supporting Information Available:** Further results of NMR measurements (tables of chemical shifts and  $^3J$ -couplings), MD simulations (predicted population distribution graphs, free energy maps, and Ramachandran population maps; tables of chemical shifts,  $^3J$ -couplings, mean atomic distances, and predicted populations and  $\phi/\psi$  positions), and QM calculations of vicinal *J*-couplings (Karplus curves). This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References and Notes

- (1) Aliev, A. E.; Courtier-Murias, D.; Bhandal, S.; Zhou, S. *Chem. Commun.* **2010**, 46, 695.
- (2) Price, D. J.; Brooks, C. L., III. *J. Comput. Chem.* **2002**, 23, 1045.
- (3) Hu, H.; Elstner, M.; Hermans, J. *Proteins* **2003**, 50, 451.
- (4) Mu, Y.; Kosov, D. S.; Stock, G. *J. Phys. Chem. B* **2003**, 107, 5064.
- (5) Mackerell, A. D., Jr. *J. Comput. Chem.* **2004**, 25, 1584.
- (6) Hess, B.; van der Vegt, N. F. A. *J. Phys. Chem. B* **2006**, 110, 17616.
- (7) Kaminský, J.; Jensen, F. *J. Chem. Theory Comput.* **2007**, 3, 1774.
- (8) Rueda, M.; Ferrer-Costa, C.; Meyer, T.; Pérez, A.; Camps, J.; Hospital, A.; Gelpí, J. L.; Orozco, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, 104, 796.
- (9) Showalter, S. A.; Brüschweiler, R. *J. Chem. Theory Comput.* **2007**, 3, 961.
- (10) Best, R. B.; Buchete, N.-V.; Hummer, G. *Biophys. J.* **2008**, 95, L07.
- (11) Li, Z.; Yu, H.; Zhuang, W.; Mukamel, S. *Chem. Phys. Lett.* **2008**, 452, 78.
- (12) Wickstrom, L.; Okur, A.; Simmerling, C. *Biophys. J.* **2009**, 97, 853.
- (13) Seabra, G. M.; Walker, R. C.; Roitberg, A. E. *J. Phys. Chem. A* **2009**, 113, 11938.
- (14) Project, E.; Nachliel, E.; Gutman, M. *J. Comput. Chem.* **2010**, 31, 1864.
- (15) Haasnoot, C. A. G.; DeLeeuw, F. A. A. M.; DeLeeuw, H. P. M.; Altona, C. *Biopolymers* **1981**, 20, 1211.
- (16) Wedemeyer, W. J.; Welker, E.; Scheraga, H. A. *Biochemistry* **2002**, 41, 14637.
- (17) Mooney, S. D.; Kollman, P. A.; Klein, T. E. *Biopolymers* **2002**, 64, 63.
- (18) Park, S.; Radmer, R. J.; Klein, T. E.; Pande, V. S. *J. Comput. Chem.* **2005**, 26, 1612.
- (19) Aliev, A. E. *Biopolymers* **2005**, 77, 230.
- (20) Watanabe, K.; Masuda, T.; Ohashi, H.; Miura, H.; Suzuki, Y. *Eur. J. Biochem.* **1994**, 226, 277.
- (21) Daura, X.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **1998**, 19, 535.
- (22) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, 25, 1656.
- (23) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, 117, 5179.
- (24) Kollman, P.; Dixon, R.; Cornell, W.; Fox, T.; Chipot, C.; Pohorille, A. The development/application of a “minimalist” force field using a combination of ab initio and experimental data. In *Computer simulation of biomolecular systems: Theoretical and experimental applications*; Van Gunsteren, W. F., Weiner, P. K., Eds.; Escom: Leiden, The Netherlands, 1997; pp 83–96.

- (25) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.
- (26) Sorin, E. J.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2472.
- (27) Garcia, A. E.; Sanbonmatsu, K. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2782.
- (28) Nymeyer, H.; Garcia, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13934.
- (29) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712.
- (30) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P. J. *Comput. Chem.* **2003**, *24*, 1999.
- (31) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
- (32) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474.
- (33) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. *J. Comput. Chem.* **2004**, *25*, 1400.
- (34) Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. *J. Chem. Theory Comput.* **2010**, *6*, 459.
- (35) Karplus, M. *J. Am. Chem. Soc.* **1963**, *85*, 2870.
- (36) Haasnoot, C. A. G.; DeLeeuw, F. A. A. M.; Altona, C. *Tetrahedron* **1980**, *36*, 2783.
- (37) Contreras, R. H.; Peralta, J. E. *Prog. Nucl. Magn. Reson. Spectrosc.* **2000**, *37*, 321.
- (38) Yongye, A. B.; Lachele Foley, B.; Woods, R. J. *J. Phys. Chem. A* **2008**, *112*, 2634.
- (39) Reggelin, M.; Hoffman, H.; Köck, M.; Mierke, D. F. *J. Am. Chem. Soc.* **1992**, *114*, 3272.
- (40) Neuhaus, D.; Williamson, M. P. *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, 2nd ed.; Wiley-VCH: New York, 2000.
- (41) Huster, D.; Arnold, K.; Gawrisch, K. *J. Phys. Chem. B* **1999**, *103*, 243.
- (42) Bax, A.; Freeman, R. *J. Am. Chem. Soc.* **1982**, *104*, 1099.
- (43) Meissner, A.; Sørensen, O. W. *Magn. Reson. Chem.* **2001**, *39*, 49.
- (44) Markley, J. L.; Bax, A.; Arata, Y.; Hilbers, C. W.; Kaptein, R.; Sykes, B. D.; Wright, P. E.; Wüthrich, K. *Pure Appl. Chem.* **1998**, *70*, 117.
- (45) Press, W. H.; Flannery, B. P.; Teukolsky, S. A. *Numerical Recipes in FORTRAN: the Art of Scientific Computing*; Cambridge University Press: Oxford, U.K., 1992.
- (46) Aliev, A. E.; Harris, K. D. M. *Magn. Reson. Chem.* **1998**, *36*, 855.
- (47) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435.
- (48) Essman, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.
- (49) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *102*, 5451.
- (50) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Simul.* **1988**, *1*, 173.
- (51) Bussi, G.; Donald, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 14101.
- (52) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182.
- (53) Hu, J.-S.; Bax, A. *J. Am. Chem. Soc.* **1997**, *119*, 6360.
- (54) Schmidt, J. M.; Blümel, M.; Löhr, F.; Rüterjans, H. *J. Biomol. NMR* **1999**, *14*, 1 (BMRB entry 15724, [http://www.bmrb.wisc.edu/ftp/pub/bmrb/entry\\_lists/nmr-star2.1/bmrb15724.str](http://www.bmrb.wisc.edu/ftp/pub/bmrb/entry_lists/nmr-star2.1/bmrb15724.str)).
- (55) Osapay, K.; Case, D. A. *J. Am. Chem. Soc.* **1991**, *113*, 9436.
- (56) Xu, X. P.; Case, D. A. *Biopolymers* **2002**, *65*, 408.
- (57) Williamson, M. P.; Asakura, T. *J. Magn. Reson. B* **1993**, *101*, 63.
- (58) McCarthy, A. A.; Walsh, M. A.; Verma, C. S.; O'Connell, D. P.; Reinhold, M.; Yalloway, G. N.; D'Arcy, D.; Higgins, T. M.; Voordouw, G.; Mayhew, S. G. *Biochemistry* **2002**, *41*, 1095.
- (59) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531.
- (60) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazayev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision E.01; Gaussian, Inc.: Wallingford, CT, 2004.
- (61) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669.
- (62) Cancès, E.; Mennucci, B. *J. Math. Chem.* **1998**, *23*, 309.
- (63) Westhof, G.; Sundaralingam, M. *J. Am. Chem. Soc.* **1983**, *105*, 970.
- (64) Hovmöller, S.; Zhou, T.; Ohlson, T. *Acta Crystallogr. D* **2002**, *58*, 768.
- (65) Hovmöller, S.; Zhou, T. Personal communication.
- (66) Hünenberger, P. H.; van Gunsteren, W. F. In *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications*; van Gunsteren, W. F., Weiner, P. K., Wilkinson, A. J., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1997; p 3.
- (67) Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211.
- (68) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27.
- (69) Guvench, O.; MacKerell, A. D., Jr. *Methods Mol. Biol.* **2008**, *443*, 63.
- (70) Hermans, J.; Berendsen, H. J. C.; van Gunsteren, W. F.; Postma, J. P. M. *Biopolymers* **1984**, *23*, 1513.
- (71) Kamiya, N.; Higo, J.; Nakamura, H. *Protein Sci.* **2002**, *11*, 2297.
- (72) Berendsen, H. J.; C. Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction models for water in relation to protein hydration. In *Intermolecular Forces*; Pullman, B., Eds.; Reidel: Dordrecht, The Netherlands, 1981; p 331.
- (73) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269.
- (74) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- (75) Jorgensen, W. L.; Madura, J. D. *Mol. Phys.* **1985**, *56*, 1381.
- (76) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, 8910.
- (77) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2001**, *114*, 363.
- (78) Vega, C.; Abascal, J. L. F.; Conde, M. M.; Aragones, J. L. *Faraday Discuss.* **2009**, *141*, 251.
- (79) Aliev, A. E.; Bhandal, S.; Courtier-Murias, D. *J. Phys. Chem. A* **2009**, *113*, 10858.
- (80) Ramakrishnan, C.; Ramachandran, G. N. *Biophys. J.* **1965**, *5*, 909.
- (81) Maccallum, P. H.; Poet, R.; Milner-White, E. J. *J. Mol. Biol.* **1995**, *248*, 361.
- (82) Ho, B. K.; Brasseur, R. *BMC Struct. Biol.* **2005**, *5*, 14.
- (83) Case, D. A.; Scheurer, C.; Brüschweiler, R. *J. Am. Chem. Soc.* **2000**, *122*, 10390.
- (84) Helgaker, T.; Jaszuński, M.; Pecul, M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2008**, *53*, 249.
- (85) Suardíaz, R.; Pérez, C.; García de la Vega, J. M.; San Fabián, J.; Contreras, R. H. *Chem. Phys. Lett.* **2007**, *442*, 119.
- (86) Maximoff, S. N.; Peralta, J. E.; Barone, V.; Scuseria, G. E. *J. Chem. Theory Comp.* **2005**, *1*, 541.
- (87) Rega, N.; Cossi, M.; Barone, V. *J. Chem. Phys.* **1996**, *105*, 11060.
- (88) Brüschweiler, R.; Case, D. A. *J. Am. Chem. Soc.* **1994**, *116*, 11199.
- (89) Aliev, A. E.; Courtier-Murias, D. *J. Phys. Chem. B* **2007**, *111*, 14034.
- (90) Kapitán, J.; Baumruk, V.; Kopecký, V.; Pohl, R.; Boř, P. *J. Am. Chem. Soc.* **2006**, *128*, 13451.
- (91) Buděšínský, M.; Daněček, P.; Bednárová, L.; Kapitán, J.; Baumruk, V.; Bouř, P. *J. Phys. Chem. A* **2008**, *112*, 8633.
- (92) Kang, Y. K. *J. Phys. Chem. B* **2006**, *110*, 21338.
- (93) Kang, Y. K. *J. Phys. Chem. B* **2007**, *111*, 10550.
- (94) Lafitte, V. G. H.; Aliev, A. E.; Horton, P. N.; Hursthouse, M. B.; Hailes, H. C. *Chem. Commun.* **2006**, *20*, 2173.
- (95) Shen, Y.; Bax, A. *J. Biomol. NMR* **2007**, *38*, 289.
- (96) Li, D.-W.; Brüschweiler, R. *J. Phys. Chem. Lett.* **2010**, *1*, 246.
- (97) Case, D. A.; Darden, T. A.; Cheatham, T. E. I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Ross, W. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlik, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *AMBER 10*; University of California: San Francisco, 2008.
- (98) Aliev, A. E. *Chem. Phys. Lett.* **2004**, *398*, 522, and references therein.