Published in final edited form as:
J Comput Chem. 2013 September 30; 34(25): 2135–2145. doi:10.1002/jcc.23354.

CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data

Jing Huang and **Alexander D. MacKerell Jr.***
Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, 20 Penn St. Baltimore, MD 21201, USA

Abstract

Protein structure and dynamics can be characterized on the atomistic level with both nuclear magnetic resonance (NMR) experiments and molecular dynamics (MD) simulations. Here we quantify the ability of the recently presented CHARMM36 (C36) force field (FF) to reproduce various NMR observables using MD simulations. The studied NMR properties include backbone scalar couplings across hydrogen bonds, residual dipolar couplings (RDCs) and relaxation order parameter, as well as scalar couplings, RDCs and order parameters for side chain amino- and methyl- containing groups. It is shown that the C36 force field leads to better correlation with experimental data compared to the CHARMM22/CMAP force field, and suggest using C36 in protein simulations. While both CHARMM FFs contains the same nonbond parameters, our results show how the changes in the internal parameters associated with the peptide backbone via CMAP and the χ_1 and χ_2 dihedral parameters leads to improved treatment of the analyzed nonbond interactions. This highlights the importance of proper treatment of the internal covalent components in modeling nonbond interactions with molecular mechanics FFs.

Introduction

Molecular simulations are now routinely carried out to study protein structure-dynamics-function relationships at an atomic level of detail. The reliability of simulation methods is largely based on the accuracy of the underlying empirical force field, which can be established based on its ability to reproduce and predict experimental observables. The quality of FFs has been continuously improving. For additive force field such as CHARMM[1,2], AMBER[3] and OPLS[4], recent refinements have largely focused on the torsional degree of freedom. Examples include the reparametrization of backbone torsions of Amber FF99SB[5], which has lead to multiple variants of the AMBER protein FF including FF99SB*[6], optimized for the correct description of the helix-coil equilibrium, FF99SB-ϕ'[7], whose optimization targeted the reproduction of the intrinsic conformational preferences of tripeptides, and FF99sbnmr[8] and FF99SB_$\phi\psi$[9], whose target data during optimization included protein NMR chemical shifts and residual dipolar couplings. Concerning side chain dihedrals, OPLS-AA/L involved reparametrized side chain parameters based on quantum mechanics (QM) energies on small peptides[10] and a recent correction to the Amber series of FFs involved the optimization of four amino acid side chains (ILDN)[11].

With respect to the CHARMM protein additive FF, a recent update yielding the CHARMM36 (C36) FF was presented,[12] where the FF version number (ie. 36) is based on the version of the program CHARMM in which the FF was first released. The major improvements over CHARMM22/CMAP (C22/CMAP, or sometimes referred to as CHARMM27) FF[1,2] include refined backbone CMAP potentials and new side-chain

*alex@outerbanks.umaryland.edu.

dihedral parameters. CMAP is a ϕ,ψ grid-based energy correction map introduced into the CHARMM FF to improve the treatment of the protein backbone.[13] In C36 the CMAPs were refined against experimental NMR data for the peptides $(Ala)_5$ and Ac-$(AAQAA)_3$-NH_2 for the common non-Gly, non-Pro residues or refitted to QM energy surfaces at a higher level of theory for the Gly and Pro CMAPs. The updated CMAP corrects the C22/CMAP FF bias towards alpha-helical conformations. C36 FF was shown to be able to fold Ac-$(AAQAA)_3$-NH_2 with the correct fraction of helix and shows enhanced cooperativity of helix and hairpin formation due to the implicit inclusion of many-body effects in the CMAP potential.[14] Side-chain dihedral parameters associated with χ_1 and χ_2 in each amino acids were fitted to QM potential energy surfaces[15] and then subject to manual adjustment targeting NMR J coupling data of unfolded ubiquitin and G protein B1 domain (GB1). Other minor improvements involve revised parameters for aliphatic hydrogens,[16] the guanidinium moiety of arginine,[17] and tryptophan.[18]

NMR observables contain valuable information about the dynamics and conformational sampling that proteins undergo in solution. The recent advent of powerful computer architectures and novel simulation algorithms brings the time scale accessible to atomistic simulations into the NMR time scale region,[19–21] and thus opens the possibility to systematically benchmark and improve the quality of molecular mechanics FFs.[8,9,22–26] For example, Lange *et al* compared residual dipolar couplings (RDCs) and scalar couplings across hydrogen bonds for ubiquitin and G protein B3 domain (GB3) from 1 μs MD simulations with 10 different FFs.[22] Lindorff-Larsen *et al* computed backbone J couplings, RDCs and order parameters of ubiquitin and GB3 from 10 μs MD simulations to evaluate eight protein force fields including the CHARMM27 and CHARMM22*[27] FFs.[23] Pande and co-workers benchmarked a variety of combinations of FFs and water models against NMR chemical shifts and J couplings of short peptides as well as ubiquitin.[24] It is thus important that any newly developed FF for proteins, such as C36, be validated against NMR data for folded proteins.

While NMR couplings of peptides and unfolded proteins were used in its development, the C36 FF was further validated by comparison with scalar couplings on χ_1 torsion and RDCs for four folded proteins including ubiquitin, GB3, bovine pancreatic trypsin inhibitor (BPTI) and hen egg white lysozyme (HEWL).[12] This data set was similar to that used to validate the FF99SB-ILDN FF.[11] The C36 FF was shown to significantly improve the agreement between computed and experimental side-chain J couplings. An improvement in the reproduction RDCs was also found for GB3 and HEWL, while for ubiquitin a slightly larger Q factor (see below) is reported with the C36 FF versus C22/CMAP; however, it should be noted that a limited RDC data set of only two alignment medium was used.

Here we present a complimentary and more through study on how well the C36 protein FF can reproduce a variety of protein backbone and side-chain NMR observables such as hydrogen bond scalar couplings, RDCs, vicinal J couplings and relaxation order parameters. The proteins investigated here include ubiquitin, GB1, cold-shock protein A (CspA), apo-calmodulin (apoCAM), intestinal fatty acid binding protein (IFABP) and HEWL. After a brief introduction of the simulation protocol and methods to compute NMR properties, computational NMR data from sub-microsecond MD simulations with the C36 FF is presented and compared with experimental values. The agreement between calculation and experiments is further compared with those from C22/CMAP FF as well as compared with results from published computational studies. We emphasize that the present study represents a true validation as none of the experimental NMR data used in this study was used as target data during the previous force field optimization work that yielded C36.

Methods

1. Molecular dynamics simulations

Protein simulation systems were prepared with the CHARMM-GUI.[28] Briefly, protein structures taken from corresponding protein data bank[29] files were solvated in pre-equilibrated cubic TIP3P water boxes of suitable sizes and counter-ions were added to keep systems neutral as detailed in Table 1. Periodic boundary conditions were applied and Lennard-Jones (LJ) interactions were truncated at 12 Å with a force switch smoothing function from 10 Å to 12 Å. The non-bonded interaction lists were generated with a distance cutoff of 16 Å and updated heuristically. Electrostatic interactions were calculated using the particle mesh Ewald method[30] with a real space cutoff of 12 Å on an approximately 1 Å grid with 6th order spline. Covalent bonds to hydrogen atoms were constrained by SHAKE.[31] After a 200 step Steepest Descent (SD) minimization with the protein fixed and another 200 steps without the protein fixed, the systems were first heated to 300 K and then subjected to a 100 ps NVT simulation followed by a 100 ps NPT simulation. The minimization, heating and initial equilibrium was performed with CHARMM,[32] and the resultant structures were used to start simulations in NAMD.[33] After a 1 ns NPT simulation as equilibration, the production simulations were run for 100 ns in the NVT ensemble (see Table 1). For HEWL NPT ensembles were generated to better compare with previous work that found CMAP helps to better reproduced order parameter S^2,[2,34] and simulations were extended to 200 ns to reduce the uncertainty of the computed S^2. Langevin thermostat with a damping factor of 5 ps^{-1} was used for NVT simulation and the Nosé-Hoover Langevin piston method with a barostat oscillation time scale of 200 fs was further applied for the NPT simulation at 300 K and 1 atm. The time step equals 2 fs and coordinates were stored every 10 ps. For each protein the above simulation protocol was applied with the C36 and C22/CMAP FFs, while for ubiquitin an additional 1.2 μs trajectories with C36 was generated. This long simulation is used to check the convergence and also to examine whether computed NMR data deteriorate over a longer simulation time, as it was reported that RDCs significantly deviate from experimental values after approximately 500 ns simulations with the C22 FF.[22]

2. Computation of NMR properties

(i) Scalar coupling across hydrogen bonds $^{h3}J_{NC}$—Hydrogen bond scalar coupling $^{h3}J_{NC}$ represents through-space coupling between N and C nuclei across N--H···O=C hydrogen bonds.[35–37] It has been found that their magnitudes are correlated with hydrogen bond (H-bond) geometries[38] and ensemble averaging is essential to reliably reproduce the experimental measurements.[39] Interesting and fundamental aspects of $^{h3}J_{NC}$ couplings are their sensitivity to hydrogen bonding network dynamics and cooperativity in proteins,[40–42] which makes them valuable for assessing FF quality. Backbone - backbone $^{h3}J_{NC}$ couplings were computed from MD simulations by[38,39]

$$^{h3}J_{NC} = \langle (-357\,\mathrm{Hz}) \exp(-3.2 r_{HO}) \cos^2\theta \rangle \quad (1)$$

, where r_{HO} is the distance between the hydrogen and the acceptor oxygen atom, θ represents the H···O=C angle, and the angular bracket denote ensemble averaging over the MD trajectories. For hydrogen bonds between lysine side chain -NH_3^+ and protein backbone C=O, since the three amino hydrogen atoms are equivalent and the hydrogen bonds are highly transient, we took the coupling value across the most probable hydrogen bond as judged by obtaining the maximum N-- H_i···O=C $^{h3}J_{NC}$ magnitude for each snapshot and then averaging them over all snapshots, *i.e.*

$$^{h3}J_{N_\zeta C}=\langle \max_{i=1,3}\{(-357\,\mathrm{Hz})\exp(-3.2r_{H_iO})\cos^2\theta_i\}\rangle \quad (2)$$

(ii)Residual dipolar couplings—Residual dipolar coupling (RDC) measures the orientation of inter-nuclear vectors. In general this averages out to zero because of the isotropic distribution of orientations of an ensemble of proteins in solution; however, with the help of alignment media such as liquid crystals, anisotropic distributions can be induced which renders the dipolar coupling interaction detectable. Its magnitude is given by

$$D=D_a\{\langle 3\cos^2\theta-1\rangle+1.5R\langle \sin^2\theta\cos 2\phi\rangle\} \quad (3)$$

, where D_a is the axial component of the alignment tensor and R is its rhombicity, θ and ψ are the polar angles of the inter-nuclear vector with respect to the alignment tensor. There are five independent variables in the alignment tensor, which can be determined by least-square fitting between calculated and experimental RDCs with singular value decomposition (SVD). The dipolar coupling constants can then be back calculated and compared with experimental values, as described in Ref. [43].

(iii) Three-bond scalar couplings—Vicinal scalar couplings between C_γ and N_ζ nuclei in lysine side chains were determined from dihedral angle χ_4 using the following Karplus equation:

$$^3J_{C_\gamma N_\zeta}=\langle A\cos^2\chi_4+B\cos\chi_4+C\rangle \quad (4)$$

Karplus parameters of Perez *et al* for χ_1 torsion were used[44] (A=1.29 Hz, B=−0.49 Hz, and C=0.37 Hz) as they represent coupling through N-C-C-C bonds similar to that associated with χ_4 in lysines.

(iv) Relaxation order parameter S^2—Relaxation order parameter S^2 in the model-free formalism can be defined as

$$S^2=\lim_{t\to\infty}C(t)=\lim_{t\to\infty}\langle P_2(\hat{\mu}(0)\cdot\hat{\mu}(t))\rangle \quad (5)$$

, where $\hat{\mu}$ is normalized inter-nuclear vector along the relevant N-H or C-N bond within the protein coordinate frame and P_2 is the second-order Legendre polynomial.[45,46] S^2 can be computed from MD simulation with[34,45]

$$S^2=\frac{3}{2}\mathrm{tr}(\langle\Phi\rangle^2)-\frac{1}{2} \quad (6)$$

, where the 3×3 tensor $\Phi=\hat{\mu}\otimes\hat{\mu}$, and the uncertainty is estimated with[47]

$$\Delta S^2=\pm(1-S^2)\sqrt{\frac{2\tau_e}{T}} \quad (7)$$

, where the correlation time τ_e is determined by integration of the correlation function up to the time when C(t) first crosses S^2:[34]

$$\tau_e = \frac{1}{1-S^2}\int (C(t)-S^2)dt \quad (8)$$

(v)Comparison between calculated and experimental results—The agreement of computed NMR data from MD simulations with experimental data can be quantified by linear correlation coefficients *r* between two data sets, or root mean square deviations (RMSD)

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_i \left(J_{\mathrm{calc}}^{(i)} - J_{\mathrm{exp}}^{(i)}\right)^2} \quad (9)$$

or Q factors.

$$Q = \sqrt{\frac{\sum_i \left(J_{\mathrm{calc}}^{(i)} - J_{\mathrm{exp}}^{(i)}\right)^2}{\sum_i \left(J_{\mathrm{exp}}^{(i)}\right)^2}} \quad (10)$$

The computed uncertainties are estimated by block averages by dividing the 100 ns trajectories into 5 20 ns blocks and calculating the associated average and standard errors.

Results and Discussion

1. Hydrogen bond scalar coupling

Proper treatment of backbone-backbone hydrogen bonding is a central requirement of any protein FF given the central role of such interactions in protein-secondary structure as well as in other interactions. Reproduction of $^{h3}J_{NC}$ couplings across backbone N--H···O=C hydrogen bonds allow for evaluation of these interactions. $^{h3}J_{NC}$ couplings were computed from 100 ns simulations and compared with experimental data for ubiquitin,[35] GB1,[36] CspA,[48] apo-calmodulin[49] and IFABP.[50] Calculated uncertainties of the individual $^{h3}J_{NC}$ coupling was around 0.002 Hz (see Figure 1A and 1B). The deviation between calculated and experimental couplings equals 0.10 Hz for ubiquitin with C36 FF, a modest decrease from a RMSD of 0.11 Hz with C22/CMAP FF. As shown in Figure 1A and C, $^{h3}J_{NC}$ couplings remained the same for most out of the 29 hydrogen bonds in ubiquitin, while the deviation from experimental values for a few hydrogen bond scalar couplings was greatly reduced with the C36 FF, for example the ones between Ile13 - Val5 (absolute error $|^{h3}J_{calc} - {}^{h3}J_{exp}|$ decrease from 0.09 Hz with C22/CMAP to 0.02 Hz with C36 FF), Ile23 - Arg54 (from 0.06 to 0.00 Hz), Val26 - Thr22 (from 0.10 Hz to 0.01 Hz), Ala28 - Glu24 (from 0.14 to 0.00 Hz) and Ile30 - Val26 (from 0.13 to 0.04 Hz). The better description of these hydrogen bonds under the new C36 FF improved the overall correlation between experimental and computational couplings as the correlation coefficient *r* increased from 0.83 to 0.86. Inspection of the protein structure showed that H-bonds Val26 - Thr22, Ala28 - Glu24 and Ile30 - Val26 were all in the α_1 helix, while H-bond Ile13 - Val5 was between β strands but the side-chain of both residues Ile13 and Val5 are adjacent to that of Ile30. The N--H···O=C hydrogen bond between Ile23 and Arg54 connects the α_1 helix and the β_4-α_2 loop and it was proposed to be a key element regulating conformational dynamics in ubiquitin,[51] which makes it particularly meaningful for C36 FF to be able to accurately predict its strength as measured in NMR experiments.

Similar improvement was achieved for protein G, apo-CAM and IFABP (see Table 2, Fig. 1B and Fig. S1) although in general the correlation between experimental and calculated scalar couplings was less pronounced than ubiquitin, while for CspA both FFs gave same correlation. For a total of 182 backbone - backbone hydrogen bond scalar couplings considered here, the RMSD decreases from 0.15 Hz with C22/CMAP to 0.14 Hz with C36 FF, while the Q factor decreased from 0.30 to 0.27 and the correlation coefficient increased from 0.68 to 0.73. Considering H-bonds separately according to their secondary structure showed that H-bonds in α-helices and β-sheets were equally improved (see Table 3) while no improvement was observed for those in the loop region. As the adjustments in the C36 FF versus C22/CMAP did not involve changes in the nonbond parameters, the small, but consistent improvements appear to be due to more accurate sampling of the backbone conformations. As the changes in the backbone gave improved sampling of the PPII versus α-helical equilibrium, the possibility existed that the changes in C36 would also lead to improvements in the treatment of the loop regions as these are the most poorly treated by additive FFs.[52] However, improvements were not obtained although the number of data is too small to allow a statistically significant difference to be measured.

Lange and co-workers reported hydrogen bond scalar couplings for ubiquitin and protein G computed from 1 μs MD simulations with a variety of FFs and simulation protocols.[22] The lowest Q factors they obtained were 0.25 for ubiquitin and 0.23 for protein G, with the Amber99SB FF and PME treatment of electrostatics. Three $^{h3}J_{NC}$ couplings (Ile13 - Val5 and Gly35 - Gln31 in ubiquitin and Asp36 - Gln32 in protein G) were reported as outliners[22] as the absolute error between calculated and experimental data were all above 0.25 Hz for all FFs they considered. Here, C22/CMAP and C36 FF lead to Q factors of 0.21 and 0.18 for ubiquitin, and 0.30 and 0.28 for protein G, respectively. It should be noted that the R factor reported in Reference[22] is related to the Q factor by $Q = \sqrt{2}R$. The absolute error $|^{h3}J_{calc} - {}^{h3}J_{exp}|$ for H-bonds Ile13 - Val5 and Gly35 - Gln31 in ubiquitin is 0.02 Hz and 0.08 Hz with C36 FF, illustrating the possibility of reproducing these weak H-bond couplings using an additive FF. With C36, Asp36 - Gln32 in protein G was the H-bond with the largest deviation from the experimental value, with an absolute error of 0.23 Hz. (see Fig. 1) It's worth noting that both C22/CMAP and C36 lead to better agreement between simulation and experiment results than the original C22 force field without the CMAP revision, which gave a higher Q factor of 0.32 for ubiquitin and 0.37 for protein G.[22] The RMSDs between experimental and computed scalar couplings for the five proteins studied here were all smaller than those from 1 ns MD simulations with C22 FF,[25] illustrating the utility of including the CMAP correction in protein backbone potential energy functions.

Low RMSDs of 0.06 Hz were reported for ubiquitin and protein G from MD simulations.[21,26] However, in these simulations the experimental values of $^{h3}J_{NC}$ couplings were included either explicitly by adding an energy penalty term in the Hamiltonian[21] or implicitly by introducing a hydrogen bonding potential with four fitting parameters.[26] Here $^{h3}J_{NC}$ couplings are computed from MD simulations in an *ab initio* fashion without any prior knowledge of experimental couplings. The improvement of C36 over C22/CMAP is not large but occurs consistently over a collection of proteins in hydrogen bonds in different secondary structures, which suggests that the refined C36 FF leads to conformational ensembles generated with MD simulations that are more representative of the experimentally measured ones.

2. RDCs in ubiquitin

RDCs yield information on the relative orientation of selected bond vectors in the protein by making the overall motion of the protein anisotropic by the use of alignment media that hinders the protein's overall rotation and translation. RDCs, and the correlation between

calculated and experimental RDCs, depend on the alignment environment. For ubiquitin an extensive data set (D44) containing RDCs measured in 44 alignment media[43,53–56] was considered. For this data set the average Q factor was computed to be 0.264 and 0.236 with C22/CMAP and C36, and the standard error was 0.012 and 0.013, respectively. The flexible tail containing residues 72–76 was excluded from the analysis, while residue Ile36, previously reported to be an outlier,[43] was found to give consistent results with the CHARMM FFs and was included in the analysis. To compare with other studies, average Q values of subsets of D44 that contains a portion of alignment measurement data were computed, and the same results were found among these RDC data sets as detailed in Table 4. The D10 data set includes data from 10 alignments as used by Bruschweiler and coworkers [43], which found a <Q> value of 0.222 with 50 ns MD simulations with the Amber99SB FF. The D36M data set contained RDCs from 36 alignment media that was previously used in self-consistent model-free analysis[53], and Lange *et al* reported <Q> values of 0.28 and 0.26 from 1 μs MD simulations with the Amber03 and Amber99SB FFs, respectively.[22]

The improvement of C36 over C22/CMAP in these different RDC data sets was similar to that of GB3 where RDCs from five media were computed.[12] As listed in Table S1, C36 led to lower or the same Q factors compared to C22/CMAP for each individual alignment medium. This illustrates that the optimization of target dihedral parameters in C36 FF provides better characterization of protein dynamics on the RDC time scale.

It was previously reported that during 1 μs MD simulations of ubiquitin with the C22 FF the RDC Q factors sharply increased from around 0.30 to approximately 0.50 after 500 ns of simulation time and remained high afterwards, indicating a sudden shift of the simulated system away from the native state ensemble of conformations. In the present study time-resolved RDC calculations were carried out on a 1.2 μs MD trajectory with C36 FF, and such a phenomenon was not observed. As shown in Fig. S2, <Q> rose to 0.35 and quickly decreased to 0.25, and then fluctuated around 0.25–0.30 during the remainder of the 1.2 μs simulation.

3. Side-chain NMR: Lysines in ubiquitin

Recently, Iwahara and co-workers measured and analyzed NMR properties involving lysine -NH_3^+ groups in ubiquitin with novel NMR pulse sequences.[57,58] Lysine residues in ubiquitin are key to its function and understanding their dynamic properties should lead to insights into the function of the protein. It was found from the ^{15}N relaxation data that the amino group of Lys48, which is the major linkage site in the ubiquitination process, was highly mobile with the lowest order parameter of 0.19 among all lysines in ubiquitin. Long range ^{15}N -- ^{13}C scalar couplings, including six inter-residue $^{3}J_{C\gamma N\zeta}$ couplings and two trans hydrogen bond $^{h3}J_{N\zeta C'}$ couplings, were also measured to high precision. Here we utilized these NMR data to compare C22/CMAP and C36 FF, which contain different parameter sets for lysine χ_1 and χ_2 torsions.

The $^{3}J_{C\gamma N\zeta}$ couplings were computed from the 100 ns MD simulations with the C22/CMAP and C36 FFs. As shown in Fig. 2A, the correlation with experimental measurements is greatly enhanced with C36 FF (correlation coefficient r=0.97) compared to C22/CMAP (r=0.81). The sole dependence of $^{3}J_{C\gamma N\zeta}$ couplings on the χ_4 dihedral angle indicates that the difference in computed J couplings reflects the difference in the χ_4 distribution. For example, the χ_4 rotamer distribution in Lys27 and Lys48 are the same between the two FFs, as shown in Fig. 2C, as are the corresponding $^{3}J_{C\gamma N\zeta}$ couplings. Alternatively, the $^{3}J_{C\gamma N\zeta}$ values differ significantly for Lys11 and Lys33, even though all three rotamers (trans, gauche+ and gauche−) are being sampled by both force fields. The difference in their $^{3}J_{C\gamma N\zeta}$ values is due to relatively small changes in the relative populations of those rotamers,

indicating the sensitivity of $^{3}J_{C\gamma N\zeta}$ couplings to force field parameters. Interestingly, as the χ_4 dihedral parameters were not changed between C22/CMAP and C36, the differences are actually associated with differences in the rotamer distribution of the side chain dihedral angles χ_3, χ_2, χ_1, as well as backbone ϕ and ψ. This is illustrated in Fig. S3, where 2 dimensional χ_4-χ_3, χ_3-χ_2, χ_2-χ_1, χ_1-ψ and χ_1- ϕ distributions for Lys33 are plotted. The better correlation between experimental and computed $^{3}J_{C\gamma N\zeta}$ couplings with the C36 FF indicates improved sampling of these dihedrals and significant difference in the sampling of χ_1 between C22/CMAP and C36 occurs. This is consistent with the optimization performed in C36 on χ_1 and χ_2, where it was noted that lysine was among residues most improved in C36 with respect to χ_1 J couplings.[12] This result also emphasizes that care must be taken in interpreting experimental observables with respect to which aspect of a FF should be adjusted to improve the particular agreement with experiment.

From the lysine J coupling analysis it was observed that a systematic underestimation of the computational coupling was obtained, as shown in Fig. 2A. This was mainly due to the Karplus coefficient that was used. For example, $^{3}J_{C\gamma N\zeta}$ in Lys27 was measured to be 2.45 Hz, however, based on the Karplus coefficients (A=1.29 Hz, B=−0.49 Hz and C=0.37 Hz) the largest possible coupling magnitude is 2.15 Hz when χ_4=180°. As shown in Fig. 2C, the χ_4 angle of Lys27 in the simulations were fluctuating in a narrow range around χ_4=180° and correctly predicted its $^{3}J_{C\gamma N\zeta}$ coupling to be the highest (2.09 Hz) among lysine residues. The Karplus coefficient we adopted was originally developed for χ_1 torsion in lysines, and to account for difference in chemical environment leading to the observed systematic shifts in couplings it was decided to alter the parameters, especially the angle-independent parameter C as also performed by Perez et al.[44] Incremental Karplus parameter ΔC was estimated to be 0.26 Hz by comparing the average coupling magnitudes of the experimental measurement and calculations. Computed scalar couplings using this new set of Karplus parameter for lysine χ_4 scalar couplings (A=1.29 Hz, B=−0.49 Hz and C=0.63 Hz) were plotted in Fig. 2B. The RMSD between experimental and computed $^{3}J_{C\gamma N\zeta}$ couplings was 0.27 Hz and 0.09 Hz for C22/CMAP and C36 FF, respectively.

Two scalar couplings across the sidechain - backbone hydrogen bonds, Lys29 -NH_3^+ to Glu16 C=O and Lys33 -NH_3^+ to Thr14 C=O have been reported.[58] MD simulations with both CHARMM FFs can satisfactorily reproduce the experimental measurements as listed in Table 5. C36 leads to slightly a larger coupling magnitude compared with C22/CMAP, while both captured the weak nature of the interaction. To better understand the related interactions H-bond analysis was carried out with a cutoff criteria of 2.4 Å, with the results listed in Table S2. As expected the three protons in the lysine NH_3^+ groups were approximately equivalent during the nanosecond simulations. Both H-bonds were weak as reflected in their low occupancy (34% for Lys29-Glu16 and 6% for Lys33-Thr14), and of a highly transient nature with mean residence times of 20 ps. It was found that H-bonding to backbone C=O groups were competing with other acceptors such as side chain C=O groups. The Lys33 -NH_3^+ group formed H-bonds with the Thr14 backbone C=O group for 6% of the simulation time and with the Glu34 sidechain Cδ=Oε_2 group for 7% of the time. Increasing the occupancy of the Lys33-Thr14 H-bonds from 6% to 18% should bring the computed $^{h3}J_{N\zeta C'}$ coupling from 0.06 Hz closer to the experimental value of 0.17 Hz. Such shifts in H-bond population will involve a complicated interplay between lysine side chain dynamics and its electrostatics environment. The changes in the C36 dihedral parameters would only affect the side chain dynamics. In contrast, Piana *et al* revised the C22 charges of the Arg, Asp and Glu side chains to improve the description of salt-bridge interactions, in particular they proposed altering the partial charge for the Glu Cδ and Oε atoms from 0.62 *e* and −0.76 *e* to 0.56 *e* and −0.69 *e*,[27] which would in our case reduce the H-bond occupancy between Lys33 NH_3^+ and Glu34 Cδ=Oε and thus increase the occupancy between Lys33

NH_3^+ and Thr14 C=O groups. This may lead to larger scalar couplings across Lys33-Thr14 H-bonds and potentially better agreement with the experimental values.

^{15}N relaxation order parameters S^2 around the $C_\gamma N_\zeta$ axis of lysine -NH_3^+ groups were also computed from the MD trajectories and compared with experimental values.[57] The flexible nature of the amino groups was captured in the simulations as reflected in the low order parameters, which also yielded quite large computational uncertainties (Table 6). The uncertainties of S^2 from 100 ns MD simulations were about 0.1 while those from the 1.2 μs simulation with C36 were around 0.02 as listed in Table 6. RMSD between experimental S^2 and those from the 1.2 μs MD trajectory equaled 0.09, and the correlation coefficient is 0.92. Qualitatively, all simulations correctly identified the lysine residues with the largest (Lys27), second largest (Lys11) and lowest (Lys48) order parameters. The largest deviation between simulation and experiment was Lys29, where simulations significantly underestimated the NH_3^+ relaxation order parameter.

4. Side-chain NMR analysis of methyl groups in ubiquitin

NMR spectroscopy has also been used to study the side chain methyl (CH_3) groups in ubiquitin.[59–61] Methyl-containing residues in ubiquitin include 16 leucines, 14 isoleucines, 7 valines, 6 threonines, 2 alanines and 1 methionine. Order parameters of the side-chain methyl 3-fold axis motions in these amino acids were computed and compared with experimental measurement by Lee *et al*.[59] The magnitude of experimental S^2 varies dramatically among different residues, and simulations reproduce the trend as shown in Fig 3. C36 FF leads to better correlation (r=0.84) than C22/CMAP (r=0.69). Leu and Val were found to improve the most in C36 with RMSDs half of those of C22/CMAP as illustrated in Table 7, while no improvement was observed for Thr residues. C36 predicted too flexible $C_\gamma H_3$ groups in Thr (too low S^2), especially for Thr7 and Thr9, as shown in Fig. 3A. The vast difference in the methyl S^2 of Thr7 between experiment (0.75), C22/CMAP (0.86) and C36 (0.30) was the main reason C36 FF yielded larger RMSD and Q factors. It was noted that considering protein χ_1 J couplings threonine was among residues that C22/CMAP provided better prediction than C36.[12]

Additional side-chain NMR data are methyl RDCs. Axial RDCs of methyl groups in ubiquitin were computed with MD simulations using the scaled alignment tensor[60,62] fitted with NH RDCs in the same alignment medium as detailed in the methods section, and compared with experimental RDCs for nine alignment media (listed in Table S1). It was found that 100 ns was too short for convergence, as also observed in Ref. [63]. Results from the 1.2 μs MD trajectory with C36 FF were plotted in Fig. 4. Correlation between experimental and computational RDCs was good for Ile (correlation coefficient r=0.89), Val (r=0.84), Thr (r=0.96) and Ala (r=0.94), while the quality of agreement was lower with Leu (r=0.61). The average Q factor for the all the methyl RDCs equals 0.55, similar to the recent calculation by Bruschweiler and coworkers based in a 1 μs MD simulation with the ff99SBnmr1-ILDN force field and with the ubiquitin EROS ensemble.[63]

5. Backbone N-H Order parameters

Hen egg white lysozyme has been used as a model system to validate the CHARMM22 force field with and without CMAP.[34] It was found that the CMAP extension to C22 provided more accurate dynamic properties of HEWL, especially better reproduction of backbone N-H relaxation order parameters. Here we studied the effect of the new CMAP and other revisions made in C36 on backbone N-H S^2 from 200 ns MD simulations. Order parameters computed from MD simulation with the C22/CMAP FF were first compared with those presented in Ref. [34], and the same S^2 values were obtained within the calculated uncertainty as illustrated in Fig. S5. As shown in Fig. 5, C36 leads to lower S^2 values than

C22/CMAP while the RMSDs between experimental and computed S^2 (excluding flexible tails) equaled 0.09 for both FFs. Significantly smaller order parameters were observed in loop regions with C36. The dip of S^2 at residue 85 and residues 101–104 in the C36 simulations reproduced very well the experimental data, while the lower S^2 for residues 18–20 deviated further from the experimental values. Smaller S^2 values were also observed in residues 109–114, which constitutes a short helix in the x-ray structure. This helical element was well maintained during the MD simulation with C22/CMAP while it was disrupted in the C36 trajectory (see Fig. S4). The average fraction helix of residues 109–114 during MD simulations was found to be 90% with the C22/CMAP FF and 41% with the C36 FF by DSSP analysis.[64] C36 was developed in part to decrease the C22/CMAP FF's general bias towards helical conformations.[65] However, the new FF in this particular region yielded backbone motions with increased flexibility over those obtained from the model-free analysis of the N-H relaxation experiments. C36 was also observed to be slightly more flexible as compared to C22/CMAP in other proteins such as ubiquitin (see Fig. S5 and Table S3). The mean S^2 value of all six proteins studied in this work decreased from 0.81 with C22/CMAP to 0.79 with C36 force field. For ubiquitin the C36 simulation leads to lower backbone S^2 values as compared to the experimental values for residues 8–11 (see Fig. S5), which form the loop region of the N-terminal β-hairpin. Thus, while this loop is the most mobile region in ubiquitin except for the C-terminal tail, the C36 FF overestimates its flexibility. We note that the experimental S^2 order parameters were derived from NMR relaxation raw data (T_1, T_2, and NOE) assuming an overall tumbling correlation time, therefore the comparison between computed and experimental S^2 can only be approximate. [66]

Conclusion

In this work we validated the most recent CHARMM protein FF for its ability to reproduce a range of protein NMR properties calculated from MD simulations. The studied NMR properties were not used in the parametrization of the C36 FF so the results presented here serve as an independent validation of the model. The C36 FF was found to provide better correlation between experimental and computed hydrogen bond scalar couplings in five proteins, implying a better description of H-bonds in proteins. It also led to better reproduction of backbone RDCs in ubiquitin when considering an extensive set of alignment medium. The improvement observed was not large, but it reflected a general improvement in the total quality of the underlying FF given the fact that both H-bond J couplings and RDCs are not directly related to one or a few FF terms. It was also found that C36 yielded slightly more flexible internal dynamics with respect to backbone N-H relaxation order parameters.

NMR observables characterizing side chain -CH_3 and -NH_3^+ groups in protein were also studied. In particular we showed that MD simulations with the C36 FF were able to reproduce a series of NMR data related to lysine -NH_3^+ moieties in ubiquitin. The lysine sidechain-backbone NH_3^+--CO hydrogen bond scalar couplings have been, to our knowledge, for the first time computed from molecular simulation. Better correlation between experimental and computational NMR data for long side chains, such as $^3J_{C\gamma N\zeta}$ couplings and methyl S^2 was observed with C36 compared to C22/CMAP, and this should mainly be attributed to the optimization of the χ_1 and χ_2 torsional parameters in the C36 FF. The general ability of the revised C36 FF to better reproduce NMR data that are largely influenced by nonbond interactions emphasizes the importance of proper treatment of the internal, covalent aspect of a FF in modeling nonbond interactions.

The results presented in this work highlighted the advantage of using the C36 force field in protein simulations. The present work emphasizes the ability of the FF to more accurately treat folded proteins as well as equilibrium of conformations observed in partially disorded

peptides.[12,14] In addition, the experimental NMR data considered here are expected to be also useful in validating and benchmarking other molecular mechanics force field, especially the Drude polarizable force field currently under extensive development.[16,67–70]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support from the NIH to ADM (GM051501 and GM072558) and computational support from the University of Maryland Computer-Aided Drug Design Center, and the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575, are acknowledged. JH is supported by a Swiss National Science Foundation Fellowship (PBBSP2_144301).

References

1. MacKerell AD Jr, Bashford D, Bellott, Dunbrack, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. J Phys Chem B. 1998; 102:3586–3616.
2. MacKerell AD Jr, Feig M, Brooks CL. J Comput Chem. 2004; 25:1400–1415. [PubMed: 15185334]
3. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ. J Comput Chem. 2005; 26:1668–1688. [PubMed: 16200636]
4. Jorgensen WL, Maxwell DS, Tirado-Rives J. J Am Chem Soc. 1996; 118:11225–11236.
5. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Proteins. 2006; 65:712–725. [PubMed: 16981200]
6. Best RB, Hummer G. J Phys Chem B. 2009; 113:9004–9015. [PubMed: 19514729]
7. Nerenberg PS, Head-Gordon T. J Chem Theory Comput. 2011; 7:1220–1230.
8. Li DW, Bruschweiler R. Angew Chem. 2010; 49:6778–6780. [PubMed: 20715028]
9. Li DW, Brueschweiler R. J Chem Theory Comput. 2011; 7:1773–1782.
10. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. J Phys Chem B. 2001; 105:6474–6487.
11. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Proteins. 2010; 78:1950–1958. [PubMed: 20408171]
12. Best RB, Zhu X, Shim J, Lopes P, Mittal J, Feig M, ADM. J Chem Theory Comput. 2012
13. MacKerell AD, Feig M, Brooks CL. J Am Chem Soc. 2003; 126:698–699. [PubMed: 14733527]
14. Best RB, Mittal J, Feig M, MacKerell AD. Biophys J. 2012; 103:1045–1051. [PubMed: 23009854]
15. Zhu X, Lopes PEM, Shim J, Mackerell AD Jr. J Chem Inf Model. 2012; 52:1559–1572. [PubMed: 22582825]
16. Vorobyov IV, Anisimov VM, MacKerell AD. J Phys Chem B. 2005; 109:18988–18999. [PubMed: 16853445]
17. Mason PE, Neilson GW, Enderby JE, Saboungi ML, Dempsey CE, MacKerell AD, Brady JW. J Am Chem Soc. 2004; 126:11462–11470. [PubMed: 15366892]
18. Macias AT, MacKerell AD. J Comput Chem. 2005; 26:1452–1463. [PubMed: 16088926]
19. Lindorff-Larsen K, Trbovic N, Maragakis P, Piana S, Shaw DE. J Am Chem Soc. 2012; 134:3787–3791. [PubMed: 22339051]
20. Lindorff-Larsen K, Best RB, Depristo MA, Dobson CM, Vendruscolo M. Nature. 2005; 433:128–132. [PubMed: 15650731]
21. Gsponer J, Hopearuoho H, Cavalli A, Dobson CM, Vendruscolo M. J Am Chem Soc. 2006; 128:15127–15135. [PubMed: 17117864]
22. Lange OF, van der Spoel D, de Groot BL. Biophys J. 2010; 99:647–655. [PubMed: 20643085]

23. Lindorff-Larsen K, Maragakis P, Piana S, Eastwood MP, Dror RO, Shaw DE. PLoS One. 2012; 7:e32131. [PubMed: 22384157]
24. Beauchamp KA, Lin YS, Das R, Pande VS. J Chem Theory Comput. 2012; 8:1409–1414. [PubMed: 22754404]
25. Huang J, Meuwly M. J Chem Theory Comput. 2010; 6:467–476.
26. Huang J, Meuwly M. Chem Phys. 2012; 396:116–123.
27. Piana S, Lindorff-Larsen K, Shaw DE. Biophys J. 2011; 100:L47-49–L47-49. [PubMed: 21539772]
28. Jo S, Kim T, Iyer VG, Im W. J Comput Chem. 2008; 29:1859–1865. [PubMed: 18351591]
29. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. Nucleic Acids Res. 2000; 28:235–242. [PubMed: 10592235]
30. Darden T, York D, Pedersen L. J Chem Phys. 1993; 98:10089–10092.
31. Ryckaert JP, Ciccotti G, Berendsen HJC. J Comput Phys. 1977; 23:327–341.
32. Brooks BR, Brooks CL, MacKerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. J Comput Chem. 2009; 30:1545–1614. [PubMed: 19444816]
33. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. J Comput Chem. 2005; 26:1781–1802. [PubMed: 16222654]
34. Buck M, Bouguet-Bonnet S, Pastor RW, MacKerell AD Jr. Biophys J. 2006; 90:L36–38. [PubMed: 16361340]
35. Cordier F, Grzesiek S. J Am Chem Soc. 1999; 121:1601–1602.
36. Cornilescu G, Ramirez BE, Frank MK, Clore GM, Gronenborn AM, Bax A. J Am Chem Soc. 1999; 121:6275–6279.
37. Grzesiek S, Cordier F, Jaravine V, Barfield M. Prog Nucl Magn Reson Spectrosc. 2004; 45:275–300.
38. Barfield M. J Am Chem Soc. 2002; 124:4158–4168. [PubMed: 11942855]
39. Sass H Jr, Schmid FF-F, Grzesiek S. J Am Chem Soc. 2007; 129:5898–5903. [PubMed: 17429967]
40. Nisius L, Grzesiek S. Nat Chem. 2012; 4:711–717. [PubMed: 22914191]
41. Sundd M. Biochemistry. 2012; 51:8111–8124. [PubMed: 23035694]
42. Kong Y, Karplus M. Structure. 2007; 15:611–623. [PubMed: 17502106]
43. Showalter SA, Bruschweiler R. J Am Chem Soc. 2007; 129:4158–4159. [PubMed: 17367145]
44. Pérez C, Löhr F, Rüterjans H, Schmidt JM. J Am Chem Soc. 2001; 123:7081–7093. [PubMed: 11459487]
45. Best RB, Clarke J, Karplus M. J Mol Biol. 2005; 349:185–203. [PubMed: 15876377]
46. Lipari G, Szabo A. J Am Chem Soc. 1982; 104:4559–4570.
47. Zwanzig R, Ailawadi NK. Phys Rev. 1969; 182:280–283.
48. Alexandrescu AT, Snyder DR, Abildgaard F. Protein Science. 2001; 10:1856–1868. [PubMed: 11514676]
49. Juranic N, Atanasova E, Streiff JH, Macura S, Prendergast FG. Protein Science. 2007; 16:1329–1337. [PubMed: 17567747]
50. Juranic N, Moncrieffe MC, Likic VA, Prendergast FG, Macura S. J Am Chem Soc. 2002; 124:14221–14226. [PubMed: 12440921]
51. Sidhu A, Surolia A, Robertson AD, Sundd M. J Mol Biol. 2011; 411:1037–1048. [PubMed: 21741979]
52. Case DA. Acc Chem Res. 2001; 35:325–331. [PubMed: 12069616]
53. Lakomek NA, Walter KF, Fares C, Lange OF, de Groot BL, Grubmuller H, Bruschweiler R, Munk A, Becker S, Meiler J, Griesinger C. J Biomol NMR. 2008; 41:139–155. [PubMed: 18523727]
54. Ottiger M, Bax A. J Am Chem Soc. 1998; 120:12334–12341.
55. Ruan K, Tolman JR. J Am Chem Soc. 2005; 127:15032–15033. [PubMed: 16248635]

56. Hus JC, Peti W, Griesinger C, Bruschweiler R. J Am Chem Soc. 2003; 125:5596–5597. [PubMed: 12733874]
57. Esadze A, Li DW, Wang T, Bruschweiler R, Iwahara J. J Am Chem Soc. 2011; 133:909–919. [PubMed: 21186799]
58. Zandarashvili L, Li DW, Wang T, Bruschweiler R, Iwahara J. J Am Chem Soc. 2011; 133:9192–9195. [PubMed: 21591797]
59. Lee AL, Flynn PF, Wand AJ. J Am Chem Soc. 1999; 121:2891–2902.
60. Fares C, Lakomek NA, Walter KF, Frank BT, Meiler J, Becker S, Griesinger C. J Biomol NMR. 2009; 45:23–44. [PubMed: 19652920]
61. Liao X, Long D, Li DW, Bruschweiler R, Tugarinov V. J Phys Chem B. 2012; 116:606–620. [PubMed: 22098066]
62. Ottiger M, Bax A. J Am Chem Soc. 1999; 121:4690–4695.
63. Long D, Li DW, Walter KF, Griesinger C, Bruschweiler R. Biophys J. 2011; 101:910–915. [PubMed: 21843482]
64. Kabsch W, Sander C. Biopolymers. 1983; 22:2577–2637. [PubMed: 6667333]
65. Best RB, Buchete NV, Hummer G. Biophys J. 2008; 95:L07–L09. [PubMed: 18456823]
66. Showalter SA, Bruschweiler R. J Chem Theory Comput. 2007; 3:961–975.
67. Lamoureux G, MacKerell JAD, Roux B. J Chem Phys. 2003; 119:5185–5197.
68. Lopes PEM, Lamoureux G, Roux B, MacKerell AD. J Phys Chem B. 2007; 111:2873–2885. [PubMed: 17388420]
69. Lopes PEM, Lamoureux G, Mackerell AD. J Comput Chem. 2009; 30:1821–1838. [PubMed: 19090564]
70. Zhu X, Lopes PEM, MacKerell AD. WIREs Comput Mol Sci. 2012; 2:167–185.

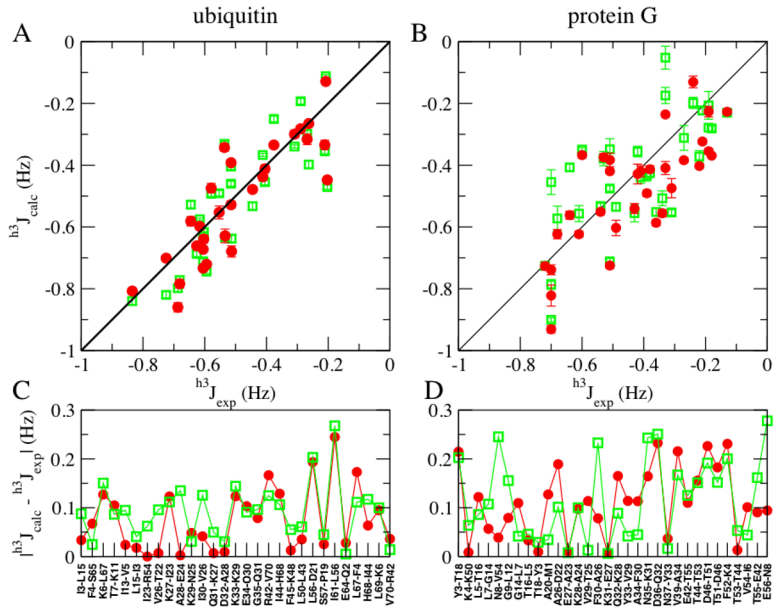

Figure 1.

Comparison between experimental and calculated $^{h3}J_{NC}$ couplings in ubiquitin (A) and GB1 (B), and the absolute deviations of $^{h3}J_{NC}$ coupling for each H-bond in ubiquitin (C) and GB1 (D). Green open squares and red filled circles denote results from the C22/CMAPandC36 FFs, respectively. The uncertainties of calculated $^{h3}J_{NC}$ couplings plotted as the error bars in A and B were estimated by block average analysis.

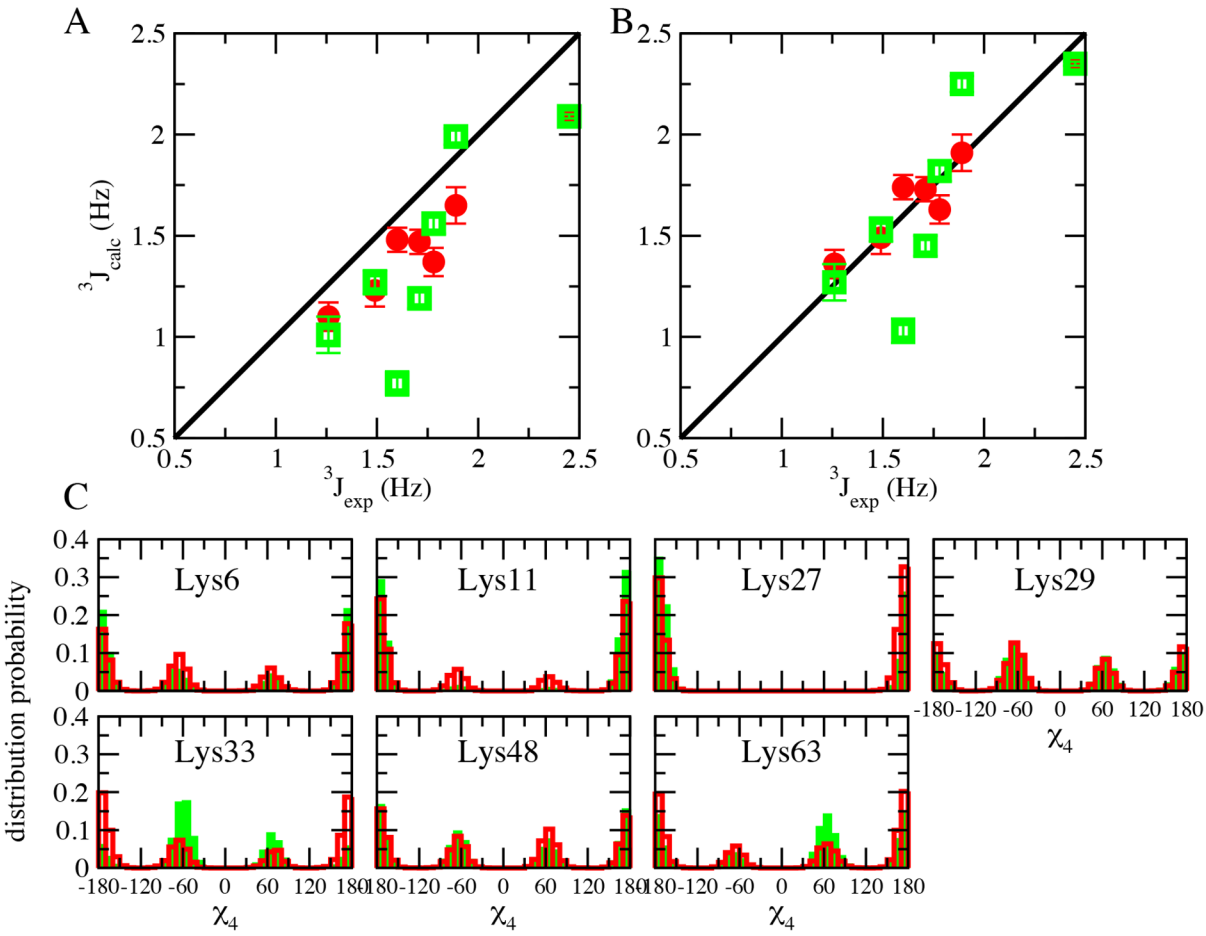

Figure 2.

Lysine $^3J_{C\gamma N\zeta}$ couplings in ubiquitin. A: Correlation between experimental and computed $^3J_{C\gamma N\zeta}$ couplings with Perez's Karplus coefficient (A=1.29 Hz, B=−0.49 Hz and C=0.37 Hz); B: Correlation between experimental and computed $^3J_{C\gamma N\zeta}$ couplings with revised Karplus coefficient (A=1.29 Hz, B=−0.49 Hz and C=0.63 Hz); C: Histograms of theχ_4 distribution of individual lysine side chains from the MD ensembles. Green open squares and red filled circles denote for results from the C22/CMAPand C36 FFs, respectively.

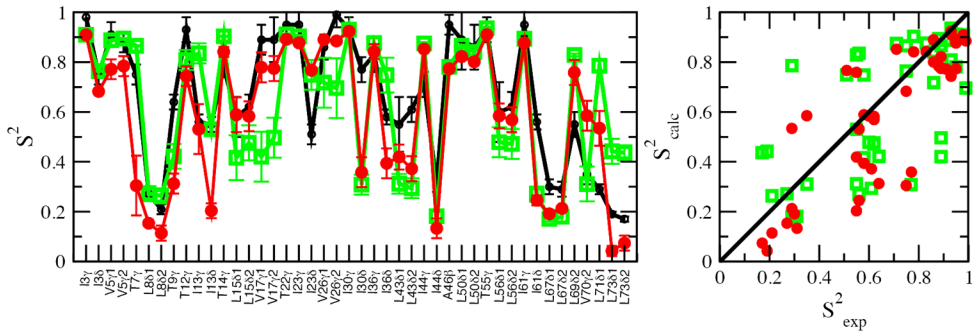

Figure 3.

Side-chain methyl order parameterS^2 in ubiquitin. Black line denotes experimental values, while green open squares and red filled circles denote for results from the C22/CMAP and C36 FFs, respectively.

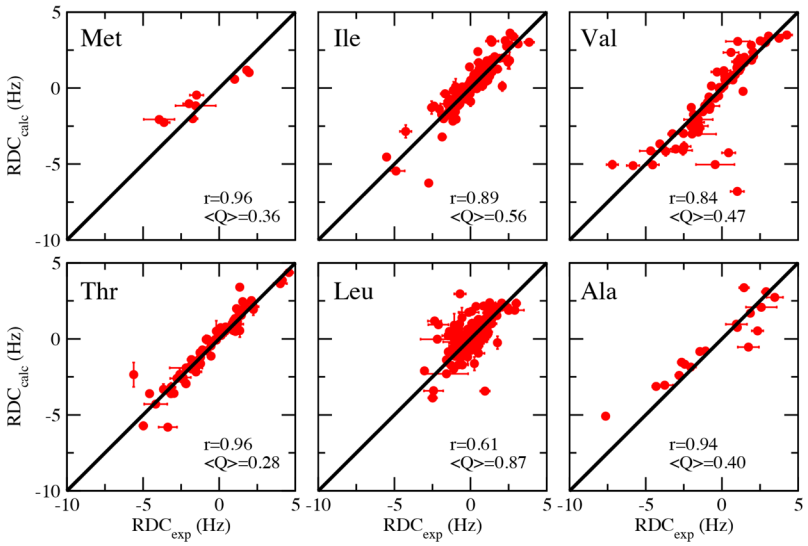

Figure 4.

Comparison between experimental and computed side-chain methyl RDCs in ubiquitin from 1.2 μs MD simulation with the C36 FF. The correlation coefficient and Q factors are listed for each individual methyl-containing residue types.

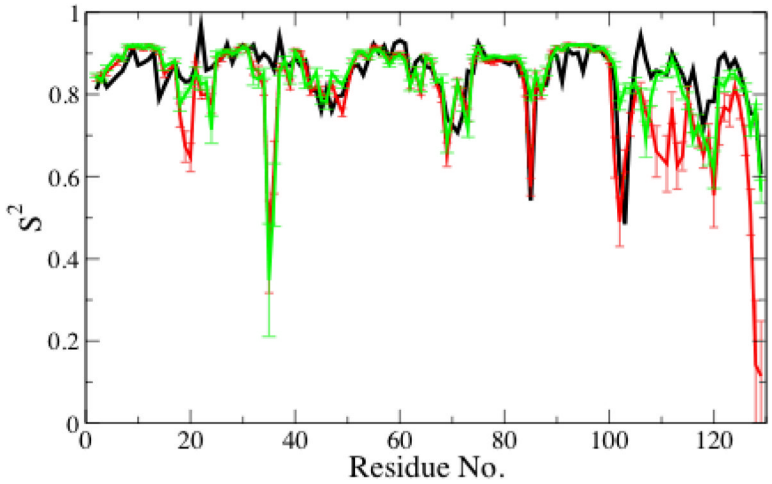

Figure 5.
Comparison between N-H order parameter S^2 in HEWL derived from relaxation experiment (solid black line)and 200 ns MD simulations with C22/CMAP (green) and C36 (red) FFs, respectively.

Table 1

MD simulations carried out in this study.

Protein	PDB ID	box size (Å)	ions	Simulation time and ensemble
ubiquitin	1ubq	58.4	none	100 ns NVT simulation with C36 FF 100 ns NVT simulation with C22/CMAP FF 1.2 μs NPT simulation with C36 FF
protein G B1 domain	2qmt	56.7	K^+	100 ns NVT simulation with C36 FF 100 ns NVT simulation with C22/CMAP FF
cold-shock protein A	1mjc	52.4	K^+	100 ns NVT simulation with C36 FF 100 ns NVT simulation with C22/CMAP FF
apo-calmodulin	1qx5	70.0	Ca^{2+}	100 ns NVT simulation with C36 FF 100 ns NVT simulation with C22/CMAP FF
Intestinal fatty acid binding protein	1ifc	63.4	none	100 ns NVT simulation with C36 FF 100 ns NVT simulation with C22/CMAP FF
hen lysozyme	6lyt	60.0	Cl^-	200 ns NPT simulation with C36 FF 200 ns NPT simulation with C22/CMAP FF

Table 2

Correlation between experimental and calculated $^{h3}J_{NC}$ couplings in five proteins. The 100 ns MD trajectories were partitioned into five 20 ns blocks, and the correlation coefficient, RMSD and Q factors for each block were computed. Average correlation coefficients, RMSD and Q factors and their standard errors are listed for the C22/CMAP and C36 FFs, respectively.

proteins	correlation coefficient		RMSD (Hz)		Q factor	
	C22/CMAP	C36	C22/CMAP	C36	C22/CMAP	C36
1ubq	0.83 ± 0.00	0.86 ± 0.01	0.110 ± 0.002	0.104 ± 0.004	0.21 ± 0.00	0.20 ± 0.01
2qmt	0.66 ± 0.03	0.76 ± 0.01	0.156 ± 0.008	0.134 ± 0.002	0.34 ± 0.02	0.29 ± 0.01
1mjc	0.76 ± 0.01	0.76 ± 0.01	0.146 ± 0.003	0.143 ± 0.004	0.25 ± 0.01	0.25 ± 0.00
1qx5	0.19 ± 0.01	0.29 ± 0.06	0.181 ± 0.002	0.171 ± 0.004	0.42 ± 0.01	0.40 ± 0.01
1ifc	0.62 ± 0.02	0.70 ± 0.01	0.187 ± 0.004	0.170 ± 0.001	0.33 ± 0.01	0.30 ± 0.00

Table 3

Correlation between experimental and calculated $^{h3}J_{NC}$ couplings from the C22/CMAP and C36 FFs for different secondary structure elements.

	No. of H-bonds	correlation coefficient		RMSD (Hz)		Q factor	
		C22/CMAP	C36	C22/CMAP	C36	C22/CMAP	C36
total	182	0.68	0.73	0.153	0.143	0.30	0.27
α-helix	59	0.44	0.52	0.151	0.142	0.37	0.34
β-sheet	113	0.66	0.72	0.154	0.143	0.27	0.25
loop	10	0.60	0.65	0.156	0.155	0.39	0.39

Table 4

Average Q factors describing the agreement between calculated and experimental RDCs over different sets of alignment medium. D10 represented the RDC data set that was previously used by Bruschweiler and coworkers in Ref. [42], D36M represented RDC data set from Lakomek *et al* for self-consistent model-free analysis[52], D23M contained RDCs from 23 alignment media that increased homogeneity in the self-consistent model-free analysis[52], while D44 is the union of D10 and D36M data sets. The standard errors were given in parentheses.

alignment data set	No. of medium	<Q> from C22/CMAP	<Q> from C36
D10[42]	10	0.271 (0.022)	0.240 (0.024)
D23M[52]	23	0.263 (0.022)	0.236 (0.021)
D36M[52]	36	0.259 (0.013)	0.231 (0.014)
D44	44	0.264 (0.012)	0.236 (0.013)

Table 5

Through-bond and through-space J couplings in lysines in ubiquitin. Experimental values were taken from Ref. [57] and calculated results were from 100 ns MD simulations with the C22/CMAP and C36 FFs, respectively. Computational uncertainties were estimated from block analysis. For hydrogen bond scalar couplings, the residues containing the acceptor C=O group are given in parentheses.

		J_{exp}(Hz)	J_{calc} (Hz)	
			C22/CMAP	C36
$^{3}J_{C\gamma N\zeta}$	Lys6	1.78 ± 0.25	1.82 ± 0.05	1.63 ± 0.07
	Lys11	1.89 ± 0.03	2.25 ± 0.04	1.91 ± 0.09
	Lys27	2.45 ± 0.03	2.35 ± 0.01	2.35 ± 0.00
	Lys29	1.26 ± 0.03	1.27 ± 0.09	1.36 ± 0.07
	Lys33	1.60 ± 0.01	1.03 ± 0.06	1.74 ± 0.06
	Lys48	1.49 ± 0.01	1.53 ± 0.06	1.49 ± 0.08
	Lys63	1.71 ± 0.01	1.45 ± 0.05	1.73 ± 0.06
$^{h3}J_{N\zeta C'}$	Lys29 (Glu16)	−0.23 ± 0.03	−0.24 ± 0.03	−0.27 ± 0.04
	Lys33 (Thr14)	−0.17 ± 0.02	−0.02 ± 0.01	−0.06 ± 0.02

Table 6

Relaxation order parameter S^2 for the lysine NH_3^+ groups in ubiquitin. The experimental values were taken from Ref. [56], and the calculated results were obtained from 100 ns NVT simulation with C22/CMAP, 100 ns NVT simulation with C36 FF and 1.2 μs NPT simulation with C36 FF, respectively.

S^2	exp	calc		
		C22/CMAP	C36	C36 (1.2 μs)
Lys6	-	0.18 ± 0.10	0.15 ± 0.08	0.22 ± 0.03
Lys11	0.415 ± 0.039	0.50 ± 0.10	0.40 ± 0.10	0.44 ± 0.02
Lys27	0.709 ± 0.021	0.68 ± 0.08	0.69 ± 0.08	0.73 ± 0.01
Lys29	0.378 ± 0.017	0.11 ± 0.07	0.14 ± 0.08	0.19 ± 0.03
Lys33	0.248 ± 0.005	0.23 ± 0.10	0.22 ± 0.11	0.30 ± 0.03
Lys48	0.192 ± 0.005	0.11 ± 0.08	0.10 ± 0.07	0.11 ± 0.02
Lys63	0.267 ± 0.006	0.13 ± 0.09	0.14 ± 0.09	0.19 ± 0.01

Table 7

Correlation between experimental and calculated methyl order parameter S^2 in ubiquitin from the C22/CMAP and C36 FFs for different methyl-containing residue types.

	number of data	correlation coefficient *r*		RMSD (Hz)		Q factor	
		C22/CMAP	C36	C22/CMAP	C36	C22/CMAP	C36
total	44	0.69	0.84	0.21	0.17	0.29	0.24
Leu	16	0.56	0.89	0.21	0.13	0.41	0.24
Ile	14	0.67	0.79	0.19	0.20	0.26	0.27
Val	7	0.62	0.87	0.26	0.13	0.31	0.16
Thr	6	0.73	0.82	0.12	0.24	0.15	0.29