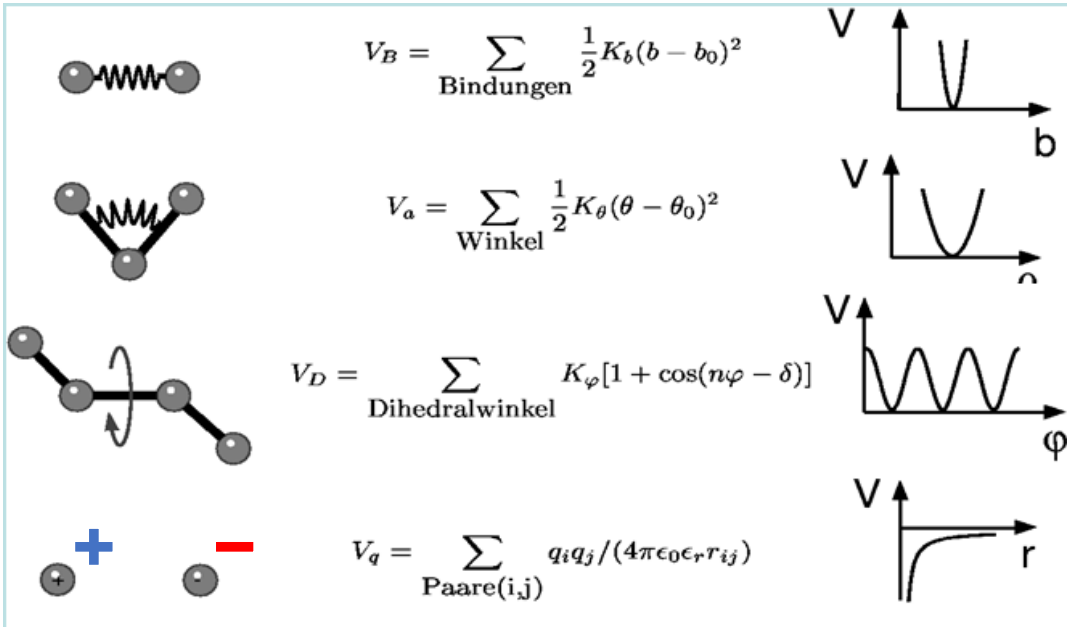
ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB

James A. Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E. Hauser, and Carlos Simmerling

Presented by Andrew Voss

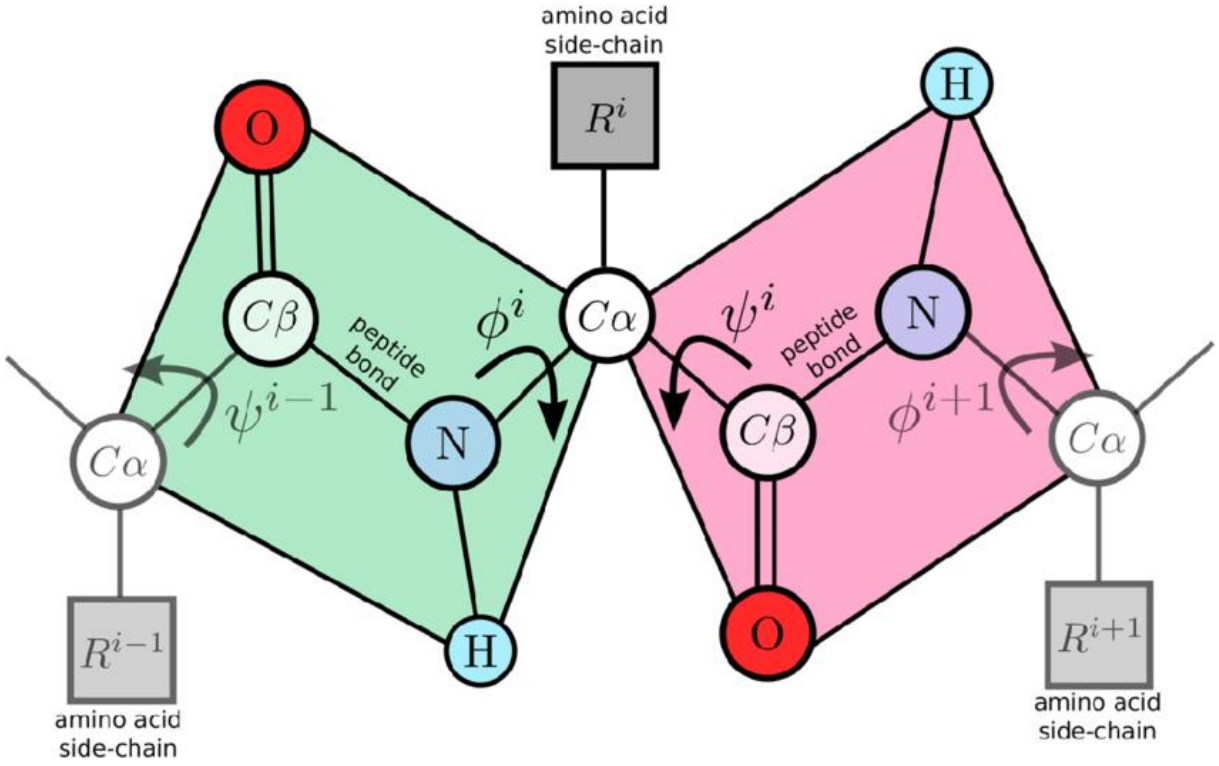
What are force fields?

- Force fields are molecular mechanics (MM) representation of atoms and bonds.
- Atom and bond are better represented by quantum mechanics (QM) but QM calculation are to computationally expensive.



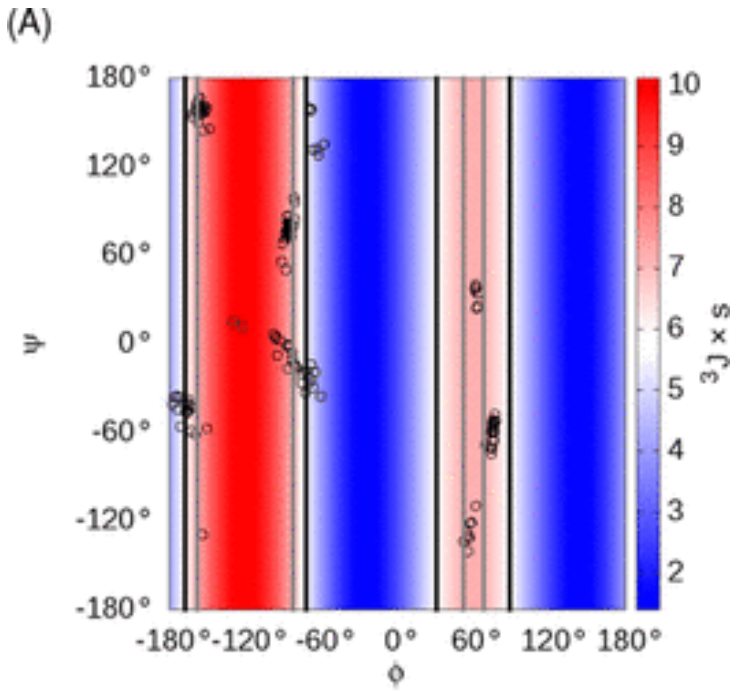
Outline of Presentation

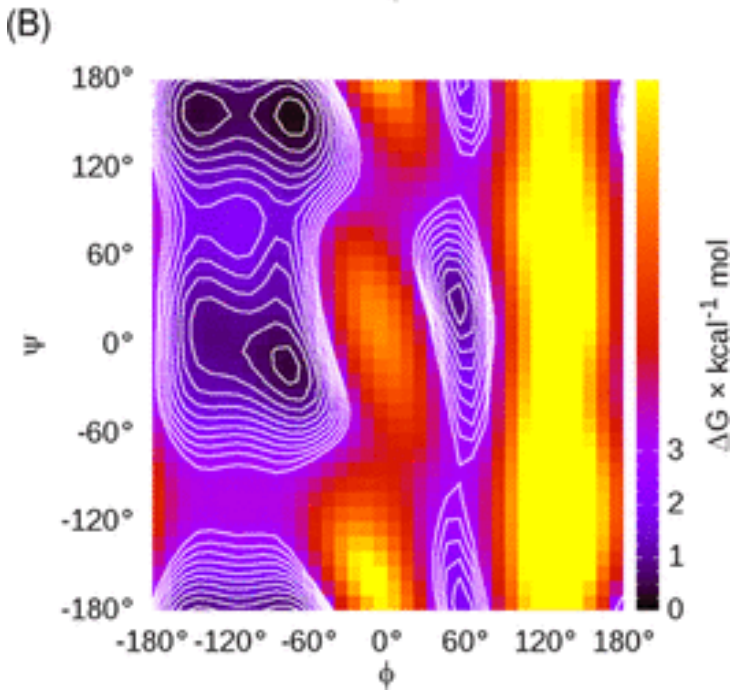
- Problems with the backbone parameters of ff99SB.
- Modification to improve backbone parameters of ff99SB.
- Problems with side chain parameters of ff99SB.
- Improvement strategy for side chain parameters.
- Test for helical stability and hairpin stability using peptides.
- Test ff99SB vs ff14SB with third Igg-binding domain of protein G (GB3), ubiquitin, lysozyme and bovine pancreatic trypsin inhibitor (BPTI) as test systems.

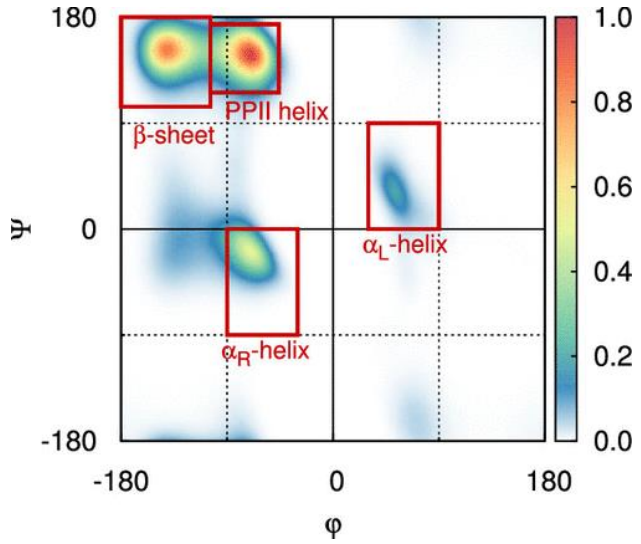


Problems with ff99SB Backbone

- HN−Hα scalar couplings calculated from the MD ensemble were too high, compare to the coupling values observed in the NMR experiment.
- They hypothesized that the main problem is that the β-ppII barrier is too low.
- It has also been noted that helix stability is low with ff99SB







Method for Backbone optimization

- Ala5 peptide as their test model.
- Simulated the Ala5 peptide with in TIP3P explicit water at 300 K.
- HN−Hα scalar coupling (J) of the simulations and NMR were calculated using Karplus relations using the original Karplus parameters (Orig), Density functional theory-based Karplus parameters derived from Ala1 (DFT1), density functional theory-based Karplus parameters derived from Ala2 (DFT2).

- Quantified the deviations between simulations and NMR using χ2 metric.

$$\chi^2 = \frac{1}{N}\sum_{i}^{N}\frac{(\langle J_i\rangle_{\text{sim}} - J_{i,\text{NMR}})^2}{\sigma_i^2}$$

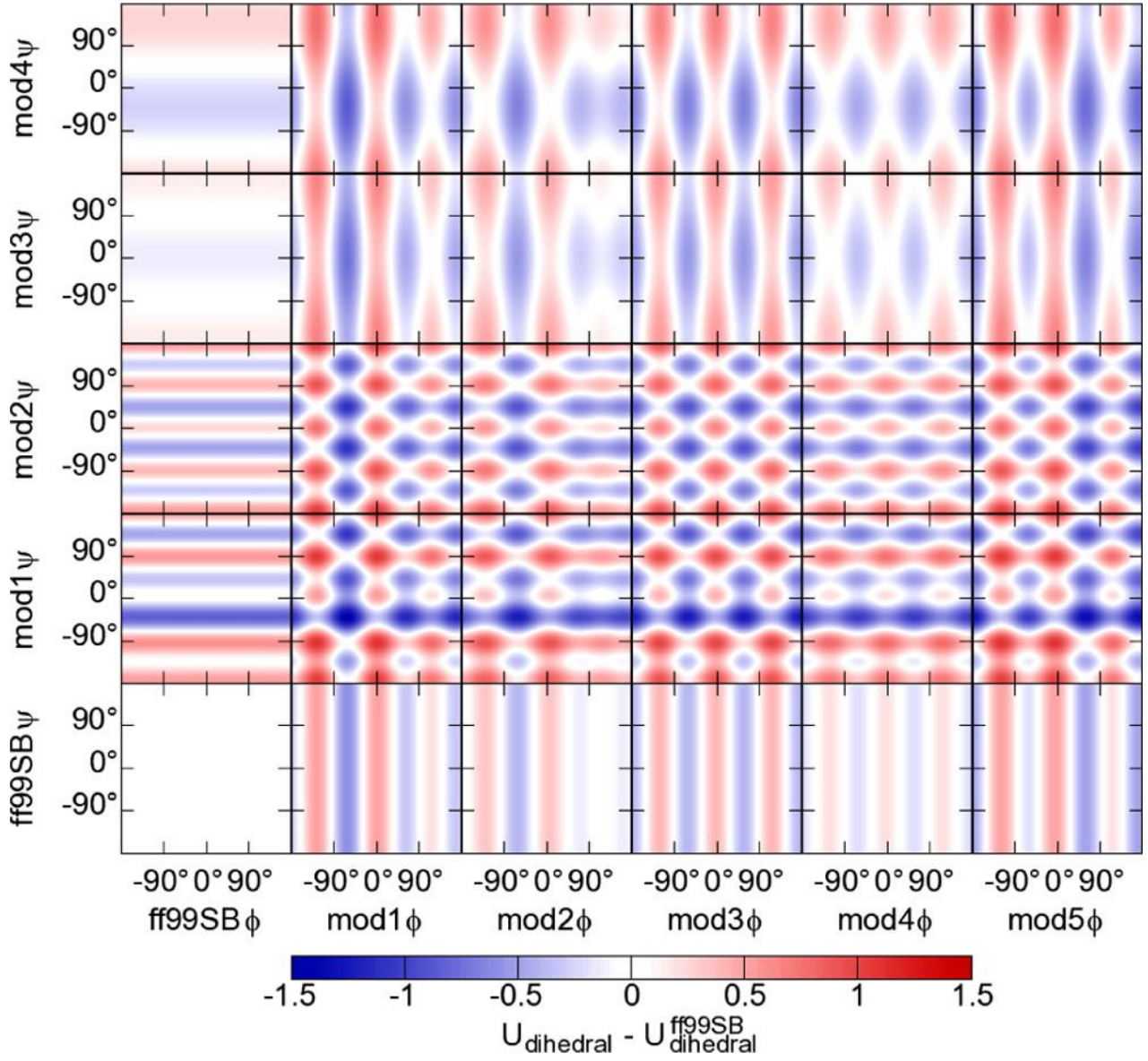
- σi is the estimated systematic error in experimental constant i

Backbone Modificatios

- Some modifications raise the barriers between β and ppII basins or between ppII and α basins.
- Other modifications stabilize ppII and α relative to β or stabilize α relative to β and ppII.

Table S5: Revisions to ϕ'

Force field	Correction term $V1'/2$	$V2'/2$	$V3'/2$
ff99SB	2.0	2.0	0.4
mod1φ	2.0	1.8	0.8
mod2φ	2.0	1.8	0.6
mod3φ	2.0	2.0	0.8
mod4φ	2.0	2.0	0.6



Modification that showed improvement

- Mod1ϕ and mod1ϕ2ψ showed improvement in scalar coupling when using the Orig and DFT2 parameters.
- Mod3ϕ had the best scalar coupling improvement with the DFT2 parameters.
- No mods showed improvement in scalar coupling with the DFT1 parameters.

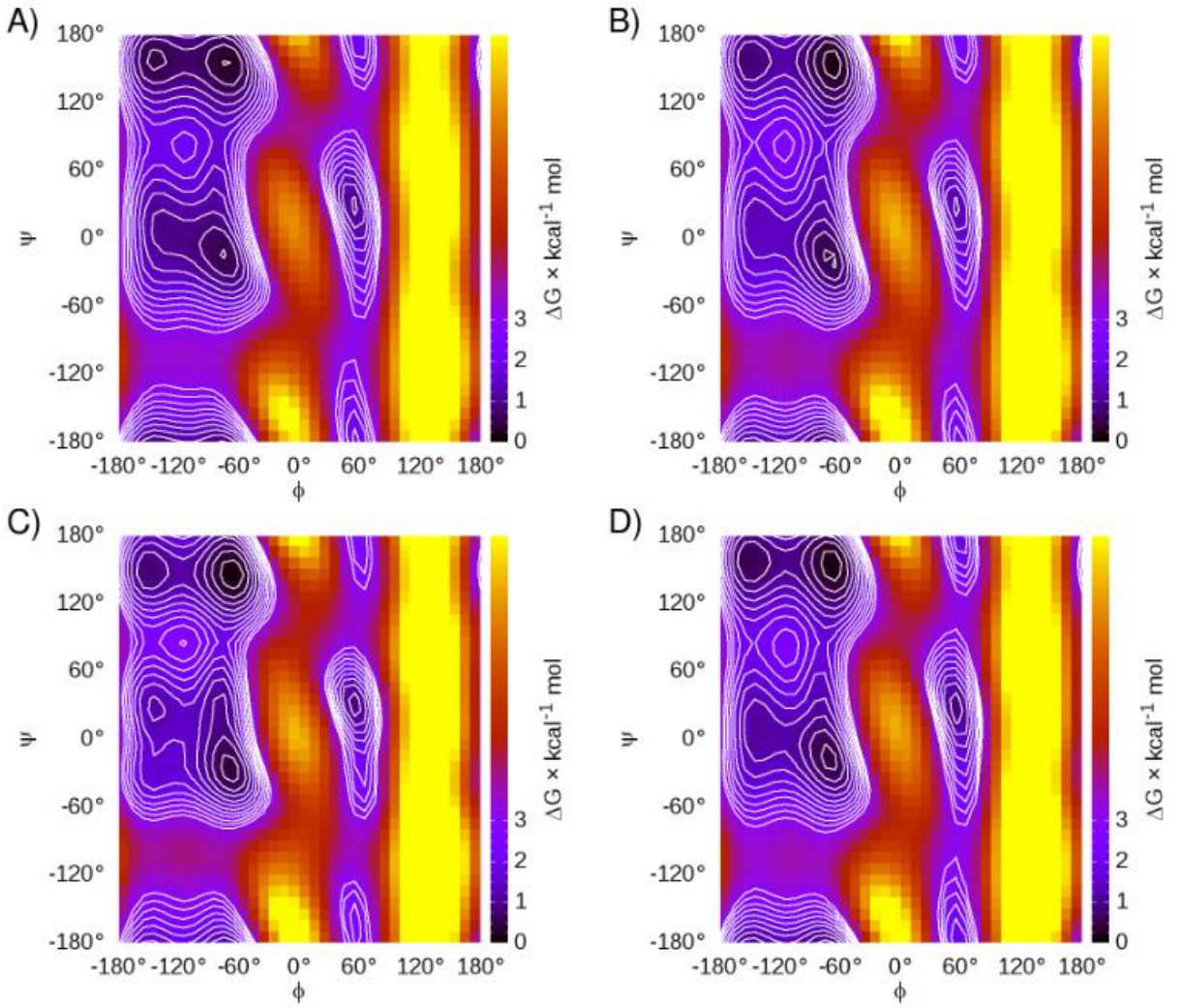
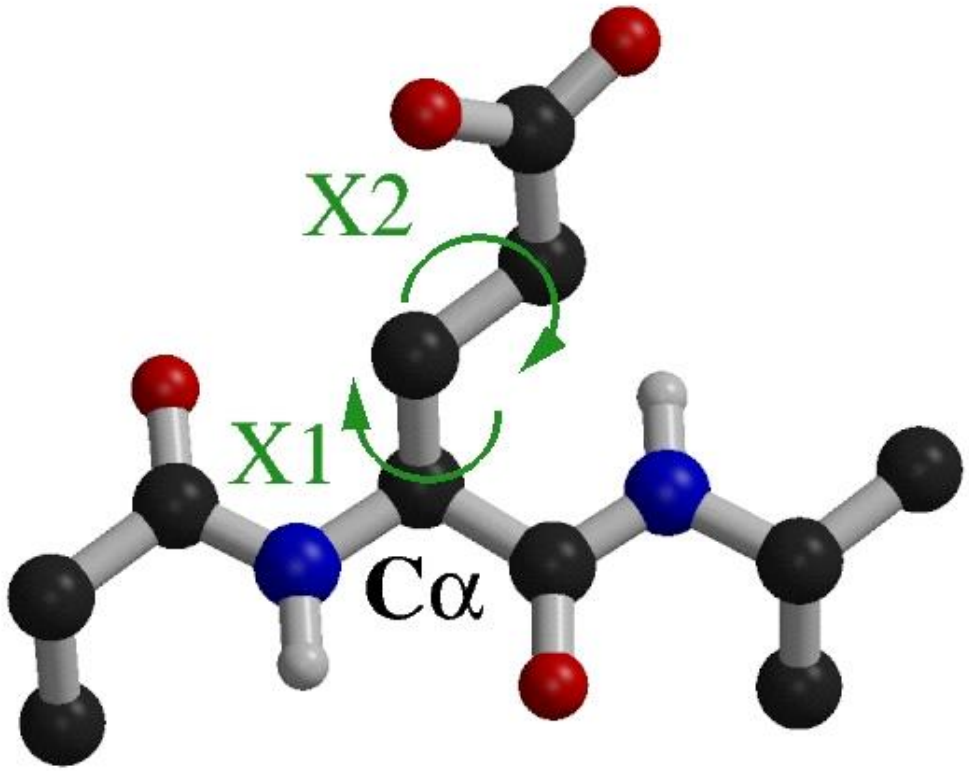


Figure S1: PMFs of alanine dipeptide in TIP3P water simulated with A) ff99SB, B) mod1ϕ, C) mod1ϕ2ψ, and D) mod3ϕ.

Problems with ff99SB Side Chains

- The side chain training data for ff99SB used small organic compounds and not complete amino acids in their QM calculations.
- The ff99SB MM force field's rotatable bond parameters are independent of the rest of the molecule.



Method for Side Chains Optimization

- Acetyl and N-methyl capped dipeptide natural amino acids, except proline, alanine, and glycine, were built at α (−60°, −45°) and β (−135°, 135°) backbone conformations and conformations were generated by rotating $\vec{\chi}$ in 10° increments.
- Quantum mechanics energies for conformers were calculated with MP2/6-31+G**.
- Molecular mechanics reoptimizations were performed in the gas phase with ff99SB and Molecular mechanics energies were calculated from the last step of ff99SB minimization.

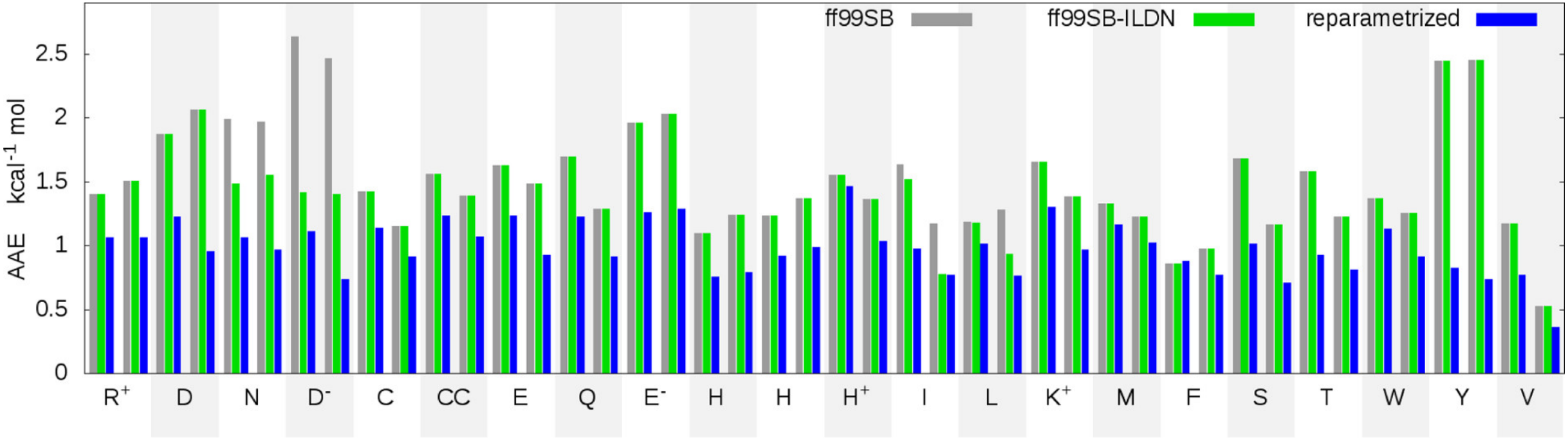
- Calculate the relative energy error (REE) between a single pair of conformations i and j:

$$\mathrm{REE}(i, j) = (E_{\mathrm{QM},i} - E_{\mathrm{QM},j}) - (E_{\mathrm{MM},i} - E_{\mathrm{MM},j})$$

- where EQM,i and EMM,i are the quantum and molecular mechanics energies of conformation i.
- Then able to sum the magnitudes of the REE for a given protonation state in a specific backbone conformation of N side chain conformations obtain the average absolute error

$$\mathrm{AAE} = \frac{2}{N(N-1)} \sum_{i} \sum_{j<i} |\mathrm{REE}(i, j)|$$

Average Absolute Error between MD and QM



AAE of each force field for each amino acid (single letter codes), with data for both α and β backbone conformation. For ionizable residues, the ionic form is indicated by a charge superscript. CC indicates the disulfide bridge. Data are shown for ff99SB, ff99SB-ILDN, and ff99SB with the reparametrized side chain corrections obtained using the procedure described in the text. ff99SB-ILDN is the ff99SB force field with rotamer corrections for amino acids Isoleucine, Leucine, Aspartic Acid and Asparagine from Lindorff−Larsen et al.

Combining Backbone and Side Chain Testing

- Test for helical stability using hydrogen bond surrogate peptide (HBSP) and K19.
- HSBP sequence
  - Ac-GQVARQLAEIY-NH
- K19 sequence
  - Ac–GGG(KAAAA)$_3$K–NH
- Both HBSP and K19 were simulated with TIP3P water molecules and simulated for 1.6 μs with two independent runs.
- Circular dichroism was used to experimentally determine the helical content of HSBP and chemical shift deviation was used for K19.

Helical Stability Is Improved

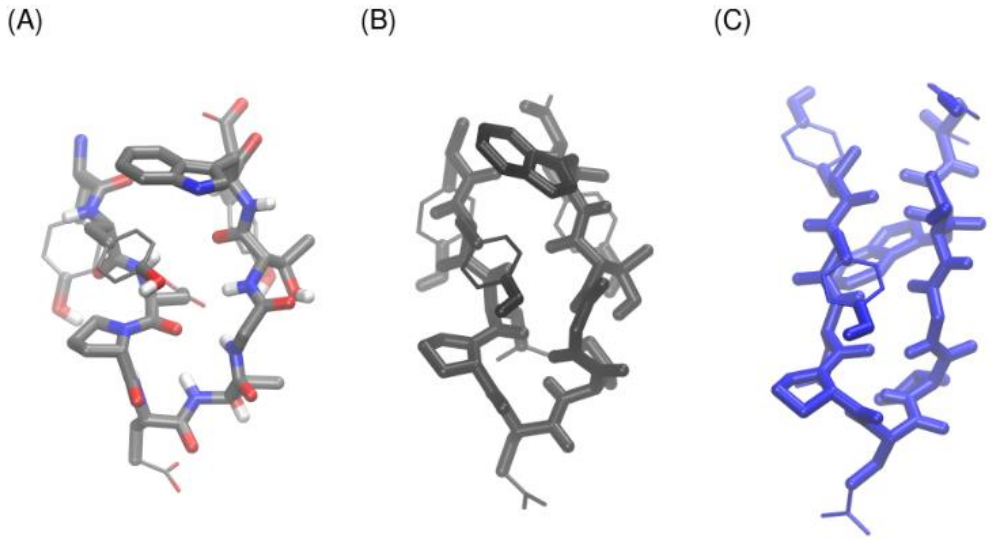
- mod1ϕ was selected as the backbone parameter update for ff14SB
  - It provides the best reproduction of Ala5 scalar coupling data among the combinations tested.
  - When combined with the QM-based side chain parameters, the helical content also reasonably matches the experiment for two different systems.

Table 4. Helical Content of HBSP and K19, from Experiments and Force Fields, namely ff99SB and Our Modifications Chosen Based on Ala_5 Results[a]

	HBSP (only updated backbone parameters)	HBSP (adding updated side chain parameters)	K19
experimental	0.65	0.65	0.31
ff99SB	0.17 ± 0.01	0.26 ± 0.01	0.08 ± 0.01
mod1ϕ	0.51 ± 0.01	0.60 ± 0.01	0.26 ± 0.05
mod1ϕ2ψ	0.72 ± 0.01	0.79 ± 0.01	0.87 ± 0.03
mod3ϕ	0.26 ± 0.01	0.46 ± 0.01	0.10 ± 0.01

Testing Hairpin Stability and Structure

- CLN025 an engineered fast-folding hairpin
- CLN025 sequence
  - YYDPETGTWY
- For ff99SB and ff14SB, they performed four MD runs starting from the average NMR-based structure and four additional runs starting from fully linear structures to quantify convergence.



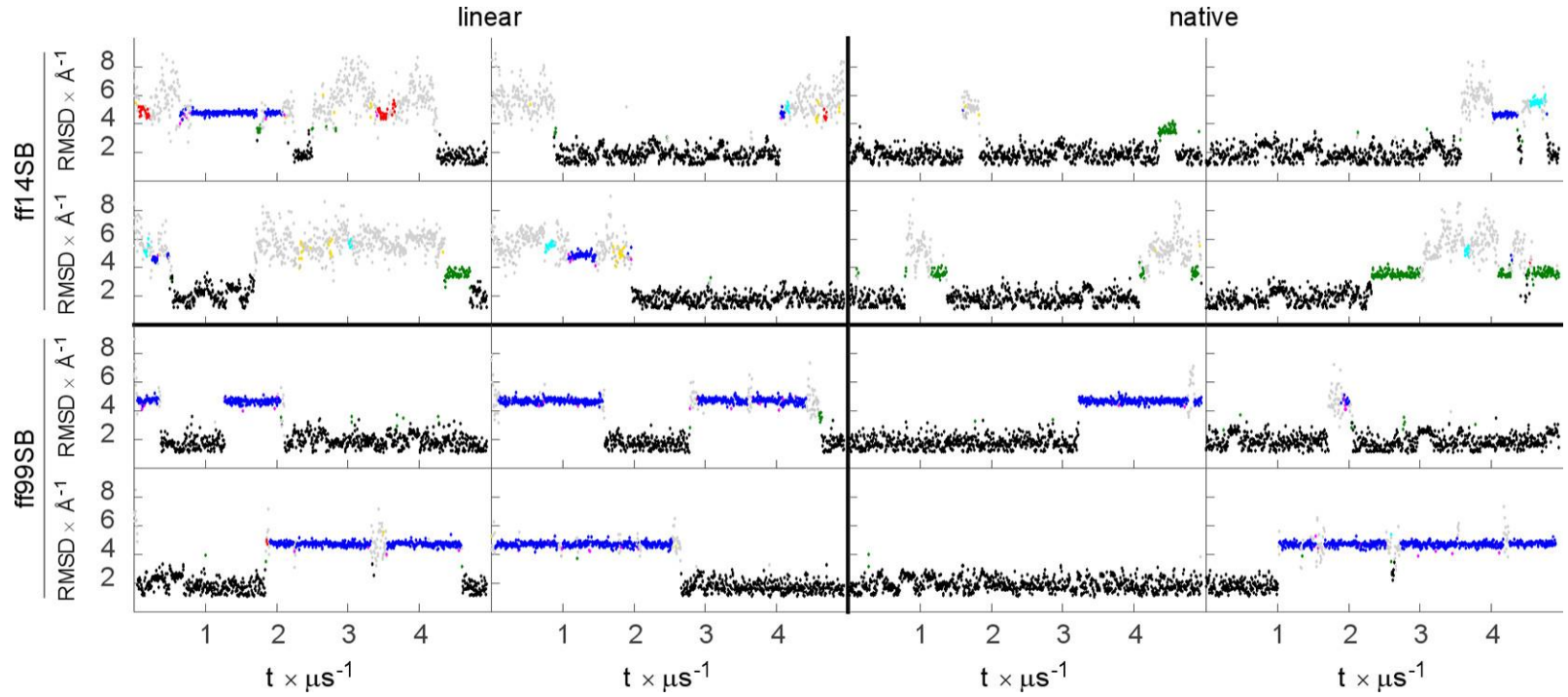
- Structures of CLN025. The NMR structure closest to the ensemble average colored by atom (A); the centroid of cluster 0 colored black (B); the centroid of cluster 1 colored blue (C).

Testing Hairpin Stability and Structure

- The ff14SB simulations sampled cluster 0 with similar frequency (57 ±14%) to ff99SB, but sampled cluster 1 much less than ff99SB (5 ± 3%)
- ff14SB simulations are more diverse when unfolded

$$\Delta\Delta E = (\langle U_{\text{ff14SB}}\rangle_{\text{cluster0}} - \langle U_{\text{ff14SB}}\rangle_{\text{cluster1}}) - (\langle U_{\text{ff99SB}}\rangle_{\text{cluster0}} - \langle U_{\text{ff99SB}}\rangle_{\text{cluster1}})$$

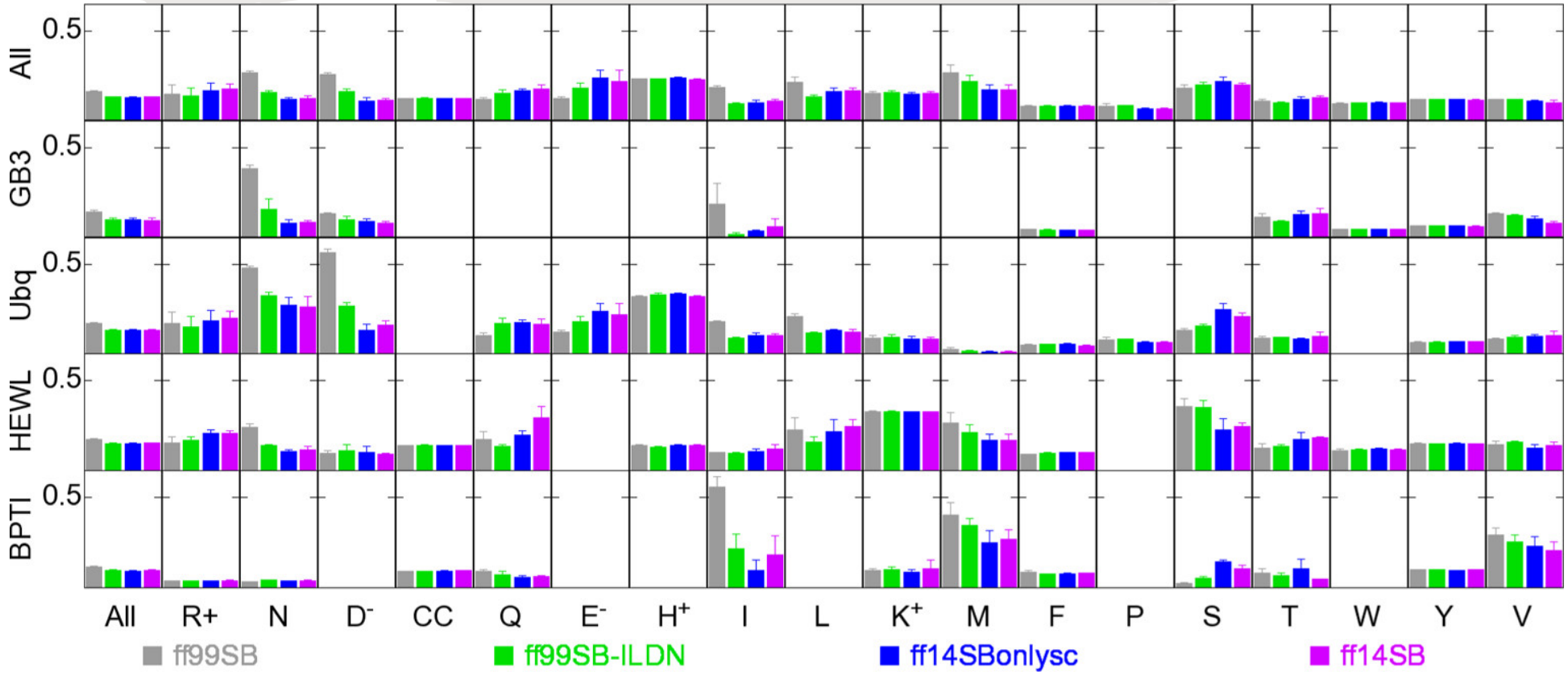
- f14SB favor the cluster 0 over the alternate by 2.9 kcal mol−1 relative to ff99SB.
- Simulations together with energy analysis suggest that ff14SB is comparable to ff99SB at hairpin modeling.



Agreement with Side Chain NMR Scalar Couplings Is Improved with ff14SB.

- Simulated third Igg-binding domain of protein G (GB3), ubiquitin, lysozyme, and bovine pancreatic trypsin inhibitor (BPTI) to compare against experimental scalar couplings
- TIP3P was used to solvate the protein inside a truncated octahedra with a minimum 8.0 Å buffer between protein and box boundary.
- Minimized for up to 10000 cycles with 100 kcal mol^{-1} $Å^{-2}$ positional restraints on protein heavy atoms.
- System was then heated at constant volume from 100 to 300 K over 100 ps, followed by another 100 ps at 300 K.
- Pressure was equilibrated for 100 ps with 100 kcal mol^{-1} $Å^{-2}$ and then 250 ps with 10 kcal mol^{-1} $Å^{-2}$ to get the system to 1 bar and 300 K.
- The system was again minimized, with 10 kcal mol^{-1} $Å^{-2}$ force constant Cartesian restraints on only the protein backbone for 10000 cycles.
- Three 100 ps simulations, with backbone restraints of 10 kcal mol^{-1} $Å^{-2}$, 1 kcal mol^{-1} $Å^{-2}$, and then 0.1 kcal mol^{-1} $Å^{-2}$.
- System was simulated for 500 ps with a 2 fs time step.

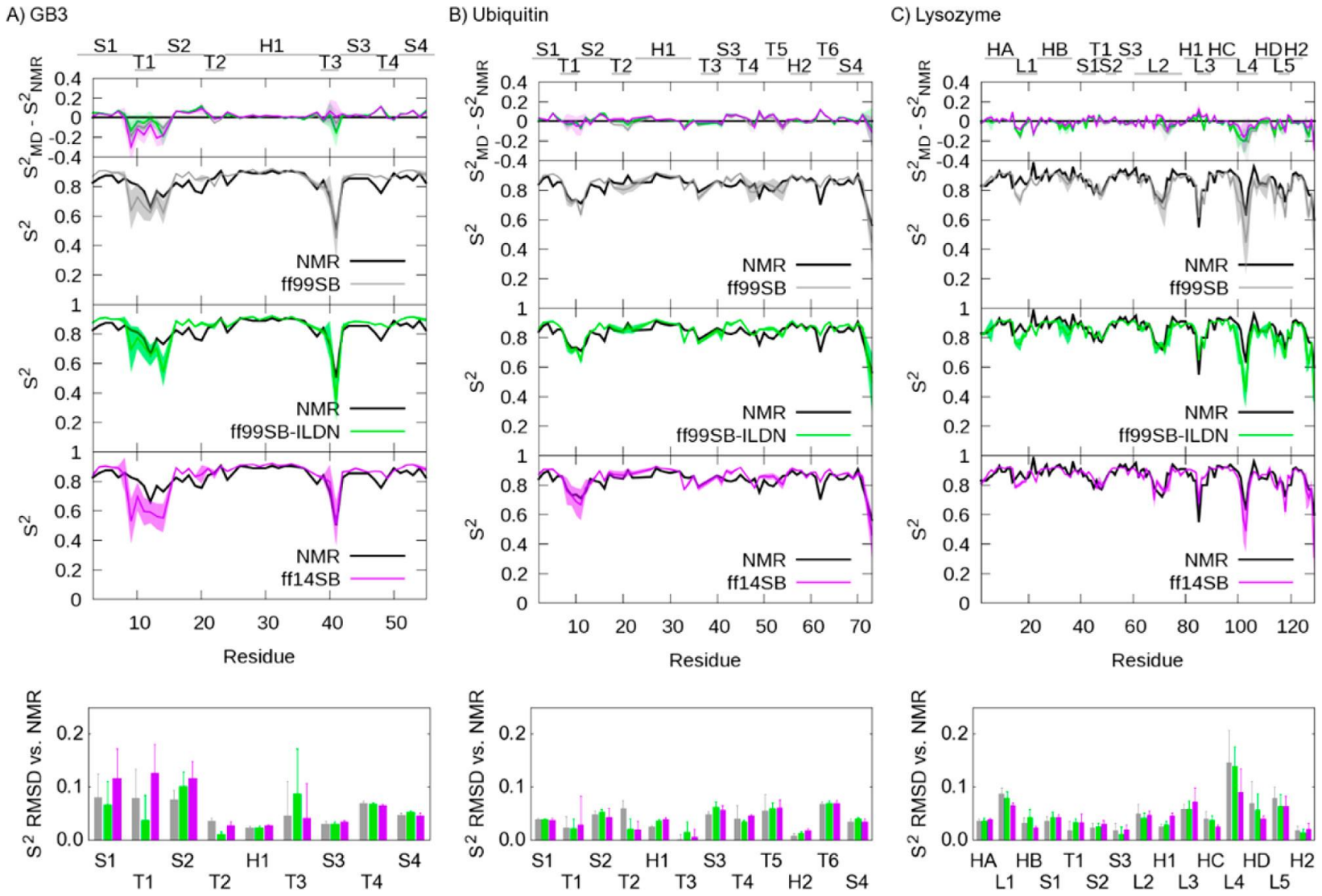
Average normalized errors (ANE) in side chain scalar couplings for all amino acids in GB3, ubiquitin (Ubq), lysozyme (HEWL), and bovine pancreatic trypsin inhibitor (BPTI)



tested ff99SB and ff99SB-ILDN as references, ff14SBonlysc includes the side chain only parameter updates and ff14SB which includes the backbone and side chain parameters updates.

Backbone NH S2 Lipari–Szabo order parameters

- NMR-derived order parameters provide site specific measures of the degree of motion on the ps-ns timescale.
- Loop 4 in lysozyme is better reproduced with ff14SB on average potentially because L4 connects two helices.
- However, the first hairpin, residues 1–20, in GB3 is reproduced less with ff14SB.



CONCLUSION

- The corrections to ff99SB for ff14SB are less than ideal
- Helical stability was improved with the mod1ϕ, and the side chain parameters update without sacrificing to much hairpin structure characteristic of ff99SB.
- Was able improve side chain preferences when compared against QM or against NMR scalar couplings.