The Journal of Chemical Physics

Links between the charge model and bonded parameter force constants in biomolecular force fields

Cite as: J. Chem. Phys. **147**, 161730 (2017); https://doi.org/10.1063/1.4985866
Submitted: 31 May 2017 . Accepted: 02 September 2017 . Published Online: 04 October 2017

David S. Cerutti, Karl T. Debiec, David A. Case, and Lillian T. Chong

Links between the charge model and bonded parameter force constants in biomolecular force fields

David S. Cerutti,[1,a)] Karl T. Debiec,[2] David A. Case,[1] and Lillian T. Chong[3]
[1]*Department of Chemistry and Chemical Biology, Rutgers University, 174 Frelinghueysen Road, Piscataway, New Jersey 08854-8066, USA*
[2]*Epic Systems, 1979 Milky Way, Verona, Wisconsin 53593, USA*
[3]*Department of Chemistry, University of Pittsburgh, 219 Parkman Avenue, Pittsburgh, Pennsylvania 15260, USA*

(Received 31 May 2017; accepted 2 September 2017; published online 4 October 2017)

The ff15ipq protein force field is a fixed charge model built by automated tools based on the two charge sets of the implicitly polarized charge method: one set (appropriate for vacuum) for deriving bonded parameters and the other (appropriate for aqueous solution) for running simulations. The duality is intended to treat water-induced electronic polarization with an understanding that fitting data for bonded parameters will come from quantum mechanical calculations in the gas phase. In this study, we compare ff15ipq to two alternatives produced with the same fitting software and a further expanded data set but following more conventional methods for tailoring bonded parameters (harmonic angle terms and torsion potentials) to the charge model. First, ff15ipq-`Qsolv` derives bonded parameters in the context of the ff15ipq solution phase charge set. Second, ff15ipq-`Vac` takes ff15ipq's bonded parameters and runs simulations with the vacuum phase charge set used to derive those parameters. The IPolQ charge model and associated protocol for deriving bonded parameters are shown to be an incremental improvement over protocols that do not account for the material phases of each source of their fitting data. Both force fields incorporating the polarized charge set depict stable globular proteins and have varying degrees of success modeling the metastability of short (5–19 residues) peptides. In this particular case, ff15ipq-`Qsolv` increases stability in a number of α-helices, correctly obtaining 70% helical character in the K19 system at 275 K and showing appropriately diminishing content up to 325 K, but overestimating the helical fraction of $AAQAA_3$ by 50% or more, forming long-lived α-helices in simulations of a β-hairpin, and increasing the likelihood that the disordered p53 N-terminal peptide will also form a helix. This may indicate a systematic bias imparted by the ff15ipq-`Qsolv` parameter development strategy, which has the hallmarks of strategies used to develop other popular force fields, and may explain some of the need for manual corrections in this force fields' evolution. In contrast, ff15ipq-`Vac` incorrectly depicts globular protein unfolding in numerous systems tested, including Trp cage, villin, lysozyme, and GB3, and does not perform any better than ff15ipq or ff15ipq-`Qsolv` in tests on short peptides. We analyze the free energy surfaces of individual amino acid dipeptides and the electrostatic potential energy surfaces of each charge model to explain the differences. *Published by AIP Publishing.* https://doi.org/10.1063/1.4985866

I. INTRODUCTION

A major focus in the development of biomolecular force fields has been the accurate modeling of electrostatic interactions in the presence of solvent. Models in which the atomic charges are fixed must therefore incorporate the effects of solvent polarization. Such effects have long been incorporated by fitting charges to mimic a molecule's electrostatic potential computed at the Hartree-Fock level of quantum theory with a small 6-31G* basis set: such calculations generally over-polarize molecules, approximating the effect of surrounding water in quantum calculations on the isolated systems.[1] The common HF/6-31G* electrostatics seen in the Cornell line of Amber force fields[2–4] are also a benchmark for charge design in the AM1-BCC approximation[5,6] of the generalized Amber force field.[7,8] Other approaches to capturing the electrostatic character of molecules include the route taken by CHARMM,[9] optimizing partial charges towards interaction energies of amino acids and individual water molecules (again computed by quantum methods), or an empirical route taken by the line of OPLS force fields,[10–12] optimizing partial charges with Lennard-Jones parameters to mimic the bulk properties of neat small molecule liquids. The latter approach harkens to the derivation of the water models; these force fields depend on to perform biomolecular simulations.

Recently, we released ff15ipq,[16] a protein force field built on the implicitly polarized charge model.[25] The goal of the project is not simply to provide new protein force fields, of which there are many[4,12,13] but also to demonstrate a set of physical assumptions and a workflow that converges to robust

[a)]Author to whom correspondence should be addressed: dscerutti@gmail.com. Telephone: (732) 445-0334. Fax: (732) 445-5958.

molecular models by leveraging cluster computation rather than human effort. Validation was more extensive for ff15ipq than its predecessor ff14ipq, and while the physical basis of the two force fields was the same, the ff15ipq parameter set was completely reformulated, save for bond stretching terms and heavy atom van der Waals parameters, using our automated fitting tools. Surpassing its predecessor, ff15ipq gave accurate descriptions of several globular proteins, metastable peptides, and disordered proteins on time scales of several μs. ff15ipq also reproduced Ala-5 J-coupling constants more accurately than a force field fitted directly to reproduce these results, presumably because our data fitting included angle terms rather than dihedrals alone.[4] The force field stands out among contemporary protein models because it comprises two distinct sets of charges for every amino acid: one describing the amino acid's electrostatic field in the gas phase and the other describing the field of the hydrated molecule. Only the latter is used in actual simulations: ff15ipq is a fixed charge model with a mean field description of the polarization, compatible with many simulation packages that do not treat polarization explicitly.[14,15] The vacuum phase charge set is retained for fitting bonded parameters and is central to a workflow that we hope will make force fields less reliant on human intervention.

The choice of pairing bonded parameters derived from vacuum potential energy surfaces with charges derived to approximate solution phase conditions was a compromise aimed at combining two well-defined sets of information in a thermodynamically defensible way. With little approximation, *in vacuo* quantum mechanical (QM) calculations can map the potential energy surface of a molecule in the gas phase by collecting ensembles of conformations and their corresponding single-point energies. Given a set of atomic partial charges and van der Waals parameters, the gas-phase potential energy surface is a rich source of information for developing bond, angle, and torsion parameters. The implicitly polarized charge scheme converges to an optimal fixed charge representation of a molecule in a polarizing medium such as water. However, the interactions of dipoles at either end of the molecule constitute a major part of the potential energy surface, and if these charges are given solution phase character, the bonded parameters may become optimized to counteract the additional polarization and return the overall energy surface towards its gas phase values. In making the complete IPolQ protein force fields, we therefore took the additional approximation that the energetic consequences of immersing the solute in water are primarily electrostatic, expressed the IPolQ charges as a perturbation of the gas phase charges, and fitted bonded parameters in the context of the gas phase charges. The hope was that including the additional polarization after bonded parameter development would correctly model the system's overall behavior in solution.

This study quantifies the effects of our approximations by running additional simulations using versions of the ff15ipq force field as it would have been had we not included the additional polarization in our production simulations (ff15ipq-`Vac`), and as it would have been had, we included the additional polarization while fitting the bonded parameters (ff15ipq-`Qsolv`). The first variant models solvated proteins with gas phase charges, while the second risks having its bonded parameters oppose the solution-phase characteristics of its charge set when sampling the local structure such as side chain conformations or secondary structure preferences. Although there are definitely limits to our sampling abilities, and none of the three force fields is completely successful on this slate of structural and dynamic tests, we find that the IPolQ scheme is more than simply philosophically superior to the other approaches. We make further analyses in terms of the microscopic electrostatics and local free energy surfaces that each scheme produces. Importantly, we find that none of the schemes are well-differentiated on time scales less than 0.5 μs, perhaps indicating the extent to which a high quality data fitting protocol can produce reliable force fields, as well as the subtle effects of the underlying physics. We hope that the results will be pertinent to the ongoing debate over the costs and benefits of more complex molecular models.

II. METHODS

A. Preparation of ff15ipq-`Vac`

As explained in the preceding study,[16] the ff15ipq force field was, like its predecessor ff14ipq,[18] derived with a pair of coupled charge sets appropriate for molecular representations in the gas phase and solution phase, respectively. The gas phase charge set Q^{Vac} was used to fit over 60 new angle and 700 torsion parameters for ff15ipq by optimizing the parameters to reproduce some 265 000 quantum single-point energies of various amino acids and short peptides collected at the MP2/cc-pvTZ level of theory. The angle terms and torsion parameters, fitted to reproduce the gas-phase quantum energy surfaces in the context of the gas-phase charge set, were then combined with the solution phase charge set for production simulations, under the assumption that the energetic consequences of solvating the peptides were predominantly electrostatic. As such, no additional work was necessary to derive ff15ipq-`Vac` for this study: the charge set was simply swapped back to the gas-phase charges that had already been derived, producing a force field that is optimized at multiple levels to reproduce the behavior of amino acids and peptides *in vacuo*.

B. Preparation of ff15ipq-`Qsolv`

In developing ff15ipq-`Qsolv`, we did not deem it sufficient to simply reoptimize the same angle terms and torsion parameters in the context of the solution-phase IPolQ charges Q^{IPol} based on the original 265 000 peptide single point energies. We did use the bonded parameters as an initial guess, but another critical aspect of the IPolQ force fields is the notion of parameter convergence by iterative rounds of optimization and force-field guided structure manipulation. This iterative optimization was not necessary in the case of ff15ipq-`Vac` because the parameters were already converged with respect to the vacuum-phase charge set. A complete charge set replacement would dramatically influence the potential energy surface and perhaps result in artificial minima. We therefore took the initial guess for ff15ipq-solv and used it

to reoptimize approximately 5% of the ff15ipq training set, including 1000 Ace-Ala-Ala-Ala-Nme tetrapeptide conformations, 800 Ace-Gly-Gly-Gly-Nme tetrapeptide conformations, various conformations of all individual amino acid dipeptides, and assorted tripeptides. While it may seem like a relatively small portion of the training set to support a complete rederivation of over 700 torsion parameters and 60 backbone angle terms, the original data set already had very good coverage of the conformational space: rather, the purpose was to guard against the minute chance that traps in the new potential energy surface could lead to structures with drastically lower estimated energies than the quantum benchmark would portray. The model obtained by optimizing the ff15ipq bonded parameter set with the additional peptide conformations and single-point energies, in the context of Q^{IPol} as found in the distributed form of ff15ipq itself, is hereafter called ff15ipq-`Qsolv`.

C. Equilibrium molecular dynamics simulations

Equilibrium molecular dynamics (MD) simulations were carried out according to the protocols used in the original ff15ipq study, for similar lengths of time, temperatures, and choices of systems.[16] Briefly, simulations of 4 μs were carried out for all proteins studied with the original ff15ipq; this time employing each of the ff15ipq-`Vac` and ff15ipq-`Qsolv` force fields. For many systems, equilibrium simulations at a range of temperatures were conducted to examine folding stability, a critical aspect of the present work which mirrors the original study. Some simulations of small globular proteins were extended to 10 μs in the original study: here, we chose to extend simulations of two α-helical peptides to obtain better convergence in their structural propensities.

III. RESULTS

A. Secondary structural preferences of short peptides

The purpose of fitting parameters for a force field is to extrapolate details of the potential energy surfaces of small systems to the behavior of much larger ones composed of the same building blocks. The IPolQ charges in each of our force fields were fitted against dipeptides of each natural amino acid, and the fitting data for angle and torsion parameters comprised dipeptides, tripeptides, and some tetrapeptides. A tetrapeptide can form the simplest unit of an α-helix, but none of the systems used to create fitting data could form anything like a β-hairpin. The smallest peptide system that we examined, Ala_5, was scarcely larger than some of the systems in our quantum data set. Larger peptides in the canonical force field's validation set included longer, metastable α-helical oligomers with relatively few unique residues: K19 and $AAQAA_3$. The β-hairpin peptides we examined, in contrast, comprised much richer sequences of different residues.

Reproduction of NMR J-couplings in the Ala_5 pentapeptide system was overall satisfactory by all three force fields, as shown in Table I. Each produced a very low χ^2 score by the Karplus equation using its original coefficients, with the published ff15ipq scoring the best. According to two other sets of Karplus parameters derived by density functional theory (DFT), however, ff15ipq-`Vac` and ff15ipq-`Qsolv` appear to

TABLE I. NMR J-couplings for the Ala_5 peptide computed with three different variants of ff15ipq. Values for the canonical ff15ipq are copied from our previous work. A total of 6 μs aggregate trajectories were collected for each force field in the isothermal, isobaric ensemble. Adapted from K. T. Debiec, D. S. Cerutti, L. Baker, D. A. Case, A. Gronenborn, and L. T. Chong, J. Chem. Theory Comput. **12**, 3926–3947 (2016). Copyright 2016 American Chemical Society.

		Simulation Canonical/ff15ipq-`Vac`/ff15ipq-`Qsolv`			
J-coupling	Residue	Orig.[a]	DFT-1[b]	DFT-2[b]	Experiment
$^1J(\mathrm{N}, \mathrm{C}_\alpha)$	2	11.4 / 11.3 / 11.4	11.4 / 11.3 / 11.4	11.4 / 11.4 / 11.4	11.36 ± 0.59
$^1J(\mathrm{N}, \mathrm{C}_\alpha)$	3	11.1 / 10.9 / 11.1	11.1 / 10.9 / 11.1	11.1 / 10.9 / 11.1	11.26 ± 0.59
$^2J(\mathrm{N}, \mathrm{C}_\alpha)$	2	8.6 / 8.6 / 8.5	8.6 / 8.6 / 8.5	8.6 / 8.6 / 8.5	9.20 ± 0.50
$^2J(\mathrm{N}, \mathrm{C}_\alpha)$	3	8.6 / 8.4 / 8.4	8.5 / 8.4 / 8.4	8.5 / 8.4 / 8.4	8.55 ± 0.50
$^3J(\mathrm{C}, \mathrm{C})$	2	0.7 / 0.8 / 1.0	0.5 / 0.6 / 0.9	0.6 / 0.7 / 1.0	0.19 ± 0.22
$^3J(\mathrm{H}_\alpha, \mathrm{C})$	2	1.6 / 1.7 / 1.7	1.3 / 1.4 / 1.5	1.5 / 1.6 / 1.7	1.85 ± 0.38
$^3J(\mathrm{H}_\alpha, \mathrm{C})$	3	1.8 / 2.1 / 1.8	1.6 / 1.9 / 1.7	1.8 / 2.0 / 1.9	1.86 ± 0.38
$^3J(\mathrm{H}_\mathrm{N}, \mathrm{C})$	2	1.3 / 1.3 / 1.3	1.3 / 1.4 / 1.4	0.9 / 1.0 / 1.1	1.10 ± 0.59
$^3J(\mathrm{H}_\mathrm{N}, \mathrm{C})$	3	1.2 / 1.2 / 1.2	1.2 / 1.3 / 1.3	0.9 / 1.0 / 1.0	1.15 ± 0.59
$^3J(\mathrm{H}_\mathrm{N}, \mathrm{C}_\beta)$	2	2.1 / 2.0 / 1.8	4.1 / 3.8 / 3.3	3.2 / 3.0 / 2.6	2.30 ± 0.39
$^3J(\mathrm{H}_\mathrm{N}, \mathrm{C}_\beta)$	3	2.0 / 1.9 / 1.7	3.8 / 3.4 / 3.1	3.0 / 2.7 / 2.5	2.24 ± 0.39
$^3J(\mathrm{H}_\mathrm{N}, \mathrm{H}_\alpha)$	2	5.3 / 5.5 / 6.0	4.8 / 5.0 / 5.6	5.5 / 5.6 / 6.2	5.59 ± 0.91
$^3J(\mathrm{H}_\mathrm{N}, \mathrm{H}_\alpha)$	3	5.7 / 6.0 / 6.4	5.3 / 5.6 / 6.1	5.9 / 6.1 / 6.6	5.74 ± 0.91
$^3J(\mathrm{H}_\mathrm{N}, \mathrm{C}_\alpha)$	2	0.6 / 0.6 / 0.7	0.6 / 0.6 / 0.7	0.6 / 0.6 / 0.7	0.67 ± 0.10
$^3J(\mathrm{H}_\mathrm{N}, \mathrm{C}_\alpha)$	3	0.6 / 0.6 / 0.7	0.6 / 0.6 / 0.7	0.6 / 0.6 / 0.7	0.68 ± 0.10
Mean χ^2		0.5 / 0.8 / 1.3	2.9 / 2.1 / 1.8	1.1 / 0.9 / 1.2	

[a]Original Karplus coefficients used by Graf.[21]
[b]DFT-based Karplus coefficients from Case.[28]

TABLE II. Secondary structure propensities of the central residues in penta-alanine. Results complement the J-couplings in Table I. All definitions below entail periodicity in the regions, but the range of each angle is shifted to keep the definitions as contiguous as possible. Errors in each population are computed by four-fold block averaging of the 6 μs aggregate trajectories. Adapted from K. T. Debiec, D. S. Cerutti, L. Baker, D. A. Case, A. Gronenborn, and L. T. Chong, J. Chem. Theory Comput. **12**, 3926–3947 (2016). Copyright 2016 American Chemical Society.

	Percent secondary structure type			
Force field	α-helix[a]	β-sheet[b]	Poly-Pro II[c]	$L-\alpha$ helix[d]
Canonical	8.0 ± 0.5	10.7 ± 0.4	71.3 ± 1.3	5.0 ± 0.8
ff15ipq-Vac	14.5 ± 1.0	14.5 ± 0.3	57.0 ± 2.2	8.3 ± 2.3
ff15ipq-Qsolv	14.9 ± 1.2	24.3 ± 0.5	50.8 ± 1.0	1.9 ± 0.6

[a]α-helices are defined with ($-160° \le \phi \le -20°$, $-120° \le \psi \le 50°$).
[b]β-sheets are defined with ($-180° \le \phi \le -110°$, $50° \le \psi \le 240°$) or ($160° \le \phi \le 180°$, $110° \le \psi \le 180°$).
[c]Poly-Pro II conformations are defined with ($-90° \le \phi \le -20°$, $50° \le \psi \le 240°$).
[d]$L-\alpha$ helix conformations consist of ($0° \le \phi \le 90°$, $50° \le \psi \le 240°$).

have improved over the canonical force field. As before, the most significant sources of discrepancy were the 3-J-$H_N C_\beta$ and 3-J-CC terms. Judging by ϕ and ψ distributions of the central residue, Table II shows that the distributions of occupancy in basins corresponding to major secondary structures were also affected by the choice of charge model and torsion parameter development. (Ala_5 is too small to form an α-helix, let alone a β-hairpin. Nonetheless, its backbone is nearly always found in conformations corresponding to one of the major secondary structures of the Ramachandran plot.) All force fields depicted α-helical and β-sheet basin occupancies as minor conformations in the ensemble, but ff15ipq-Vac showed somewhat more of these two structures, apparently at the expense of the Poly-Proline II (PPII) population. The drop in the PPII population was greater than the gain in the populations of other defined regions of the Ramachandran plot, with the increase in $L-\alpha$ helical content accounting for the missing population. The changes were more striking for ff15ipq-Qsolv: β-sheet content more than doubled over the canonical ff15ipq's distribution, and PPII content was down even further. While all three force fields produce strong scores on the J-coupling analysis, PPII has been measured to account for 80% of the Ala_5 population[19,20] and 66% of the centermost residue:[21] the canonical force field is much closer to either of these experimental results than either variant.

The changing equilibria of alanine were also pertinent to new simulations of the peptides K19 and $AAQAA_3$. Both systems consist of repeats with four alanine residues punctuated by lysine and glutamine, respectively, and are expected to be meta-stable at room temperature[22,23] with decreasing stability

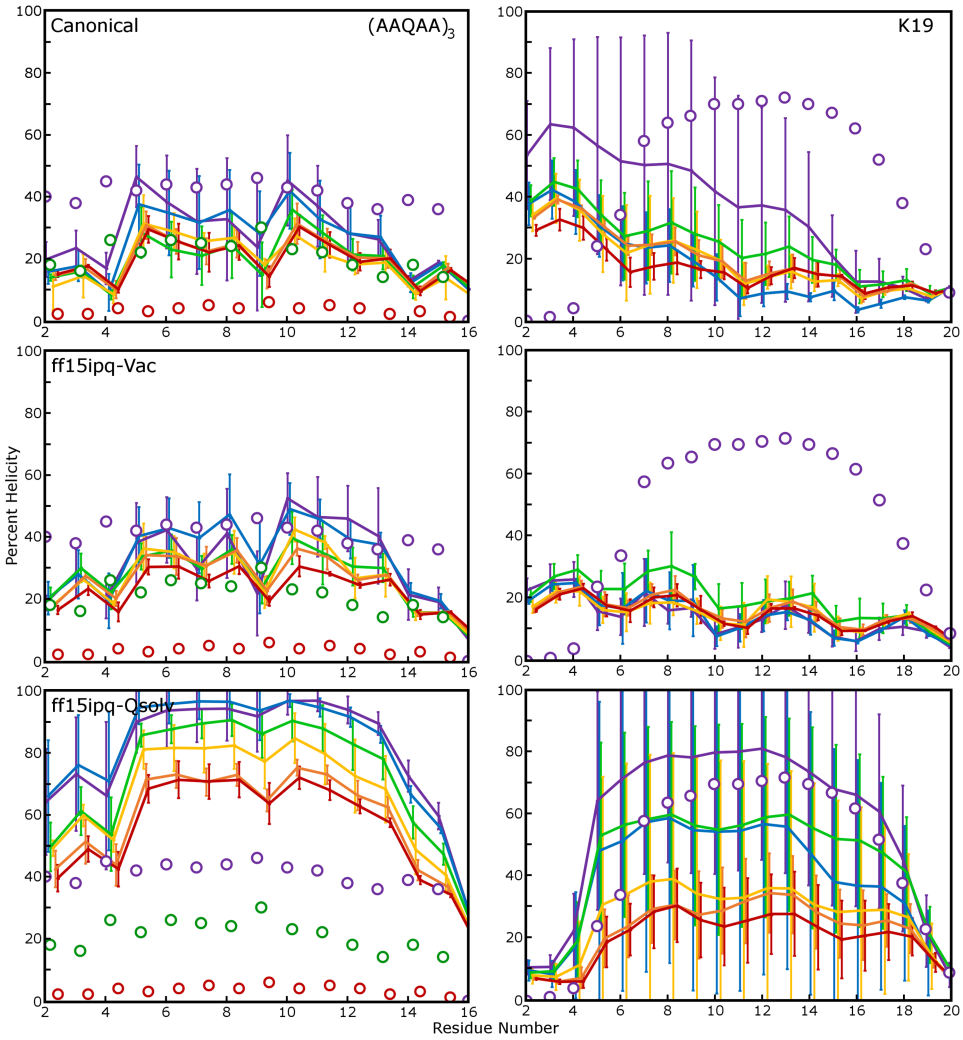

FIG. 1. Helical propensity of $AAQAA_3$ and K19 under the canonical ff15ipq and two alternative force fields. Each system is plotted in its own column while different panels present different force field variants. Error bars represent one standard deviation of the helical propensity of each residue, averaged over four simulations with 3 μs of production each. In $AAQAA_3$, the color scale purple > blue > green > yellow > orange > red indicates simulations at temperatures of 280 K–330 K. In K19, the same color scheme represents temperatures of 275 K–325 K. Circles represent experimental helical propensities at the color-indicated temperatures.

as the temperature rises. In our previous study, we simulated each system at several temperatures spanning a 50 K range but were unable to obtain convergent ensembles for either system in 4 μs of equilibrium molecular dynamics at each temperature.[16] Stronger sampling at the lower end of the temperature range could have been obtained with a method such as temperature replica exchange molecular dynamics (REMD),[24] perhaps by running the six replicas together and swapping their states periodically. However, our interest in folding propensity over the entire range would not have been any better served by this approach, as indicated by the limited sampling in the 325 K run. In order to truly improve the sampling, it would have been necessary to extend the temperature range much higher, perhaps to 425 K, and this would have required roughly three times as much aggregate simulation time to maintain additional replicas above 325 K. We therefore performed quadruplicate 4 μs equilibrium simulations of each peptide using ff15ipq and our two variants, discarding the first 1 μs of each run for equilibration. The results are shown in Fig. 1. The extended simulations lend certainty to the conclusion that ff15ipq-`Qsolv` stabilized α-helical structure in $AAQAA_3$, but the helical content was too great. The results for the lysine-studded K19 parallel those of $AAQAA_3$: the canonical ff15ipq had roughly the same helical propensity as ff15ipq-`Vac`, but ff15ipq-`Qsolv` significantly increased the helical content, in line with results from the penta-alanine simulations. For K19, both the helical content and the melting curve were well depicted by ff15ipq-`Qsolv`.

To further investigate the details of each residue's free energy surface and its contribution to the stability in each

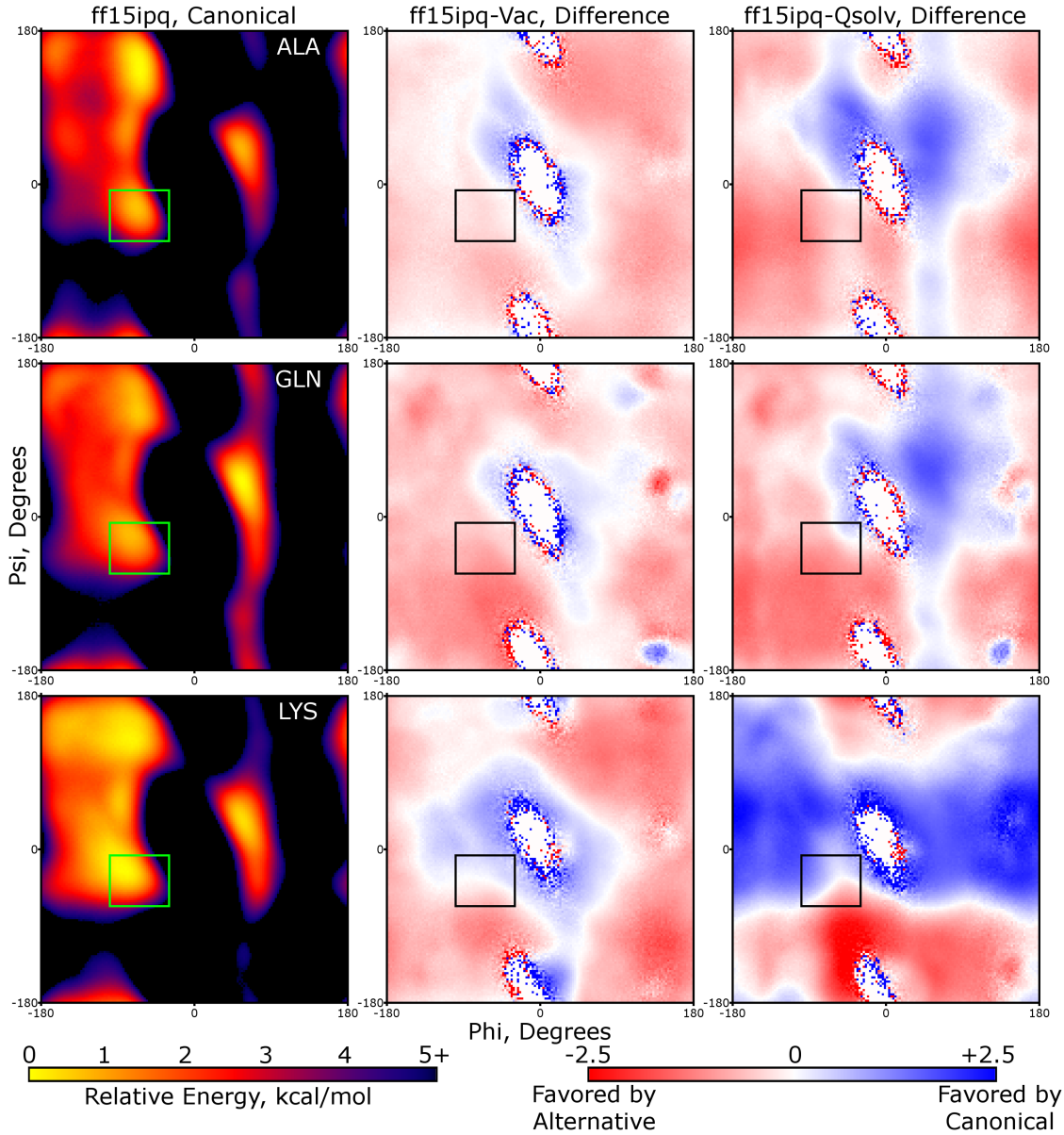

FIG. 2. Potentials of mean force for key amino acids in helical systems. Umbrella sampling on Ace-Ala-RES-Ala-Nme tetrapeptides, where RES = (Ala, Gln, Lys) indicates the free energy surface of the central residue's major backbone dihedrals in plots that can be compared to Ramachandran diagrams. Each residue occupies a row of the above panels, while different force fields are displayed in separate columns. ff15ipq-`Vac` and ff15ipq-`Qsolv` landscapes are displayed as difference plots relative to the canonical model. The inset box in each panel indicates the α-helical region as defined for calculating helical propensity in Fig. 1.

helical peptide, we computed two-dimensional potentials of mean force (PMF) for alanine, glutamine, and lysine in Ace-Ala-X-Ala-Nme tetrapeptides at 298 K using 36 windows in each of the ϕ and ψ backbone angles with all three force fields. Each PMF and the differences in either variant relative to the canonical ff15ipq are shown in Fig. 2. The energy landscapes were more sensitive to determining bonded parameters in the context of the polarized charge set than omitting the polarized charges from simulations and in line with the observation that ff15ipq-Qsolv produced more dramatic changes in $AAQAA_3$, K19, and the populations of major secondary structures in penta-alanine. The α-helical basin, indicated by boxes in the figure, was deepened in alanine as depicted by ff15ipq-Qsolv but not as much by ff15ipq-Vac. A similar effect was evident in glutamine, but the changes in the lysine energy landscape, though more severe, did not clearly indicate an overall change in the α-helical propensity. The concerted effects of alanine and glutamine, in contrast to having only alanine driving helical folding in K19, could explain why ff15ipq-Qsolv overestimated helicity in $AAQAA_3$ but predicted the correct amount in K19.

For completeness, we attempted to quantify the β-sheet character of each force field using the GB1 hairpin system. In contrast to the helical peptides, GB1 hairpin contains ten different residues (sequence GEWTYDDATKTFTVTE) and its behavior is therefore harder to reduce to individual amino acids. As shown in Fig. 3, all of the force fields depicted hairpin unfolding and refolding at various temperatures, but the 4 μs simulations are far from converged. Even in this system of many different residues, there appeared to be a greater population of α-helix, and of the β-hairpin's native fold, in the simulations run with ff15ipq-Qsolv than with each of the other models. The canonical ff15ipq did exhibit some α helical contents at 285 K and 325 K, but simulations with ff15ipq-Qsolv showed more residues forming the helices, and the helices persisted for more than 1.5 μs as opposed to 300 or 500 ns under the canonical force field. The β-hairpin could also be very long lived and stable under ff15ipq-Qsolv, but once a significant amount of α-helix formed, there were no examples in our simulations of the hairpin refolding. It is possible that both secondary structures were stabilized in ff15ipq-Qsolv and that α-helix was overall dominant, but we did not have the resources to pursue additional tests on the hairpin or potentials of mean force for all the residues composing it.

B. Results for globular proteins

Small globular protein systems simulated on the time scale of several μs further distinguish the canonical and variant ff15ipq models. Of the proteins GB3, ubiquitin, lysozyme, and villin, all except ubiquitin show notably higher backbone positional root mean squared deviation (rmsd) when simulated with ff15ipq-Vac as opposed to the canonical ff15ipq, but the differences tend to manifest themselves only after hundreds of ns to 1 μs. The differences between ff15ipq-Qsolv and the canonical force field were not as pronounced, as shown in Fig. 4, with the exception that ff15ipq-Qsolv appeared to cause lysozyme to deviate slightly more from its X-ray structure than the canonical force field.

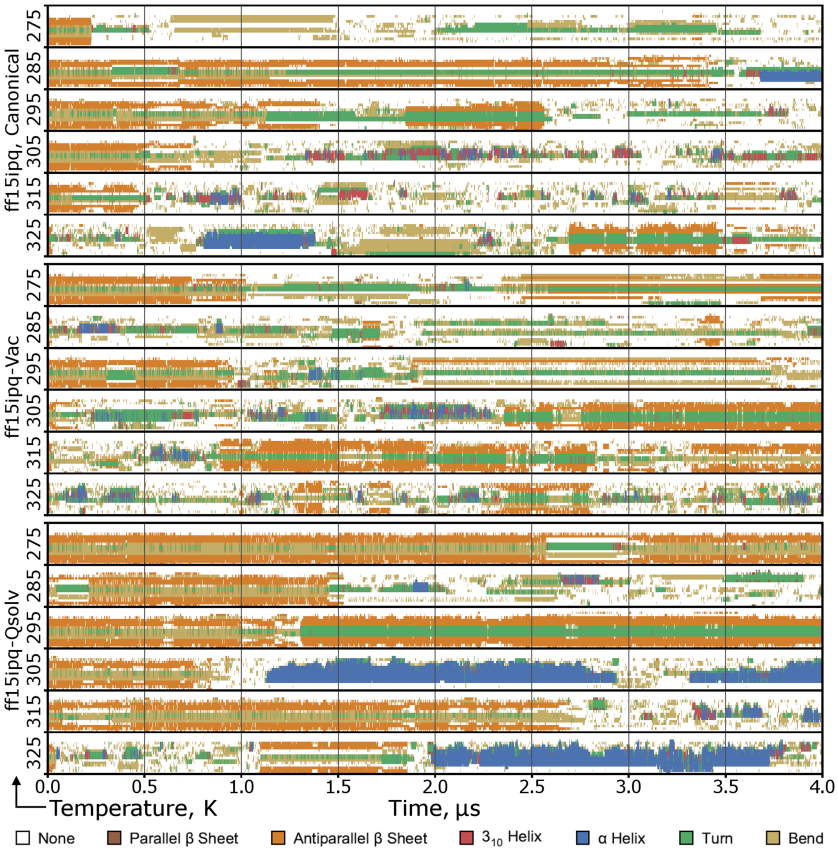

FIG. 3. Secondary structures of the GB1 hairpin in simulations at six different temperatures. The color scheme is drawn from our previous publication.[16] Each of the 18 panels has an implicit scale, starting with residue 1 on the bottom and proceeding to residue 16 on the top to show the secondary structure of each residue at any time in the simulation. Adapted from K. T. Debiec, D. S. Cerutti, L. Baker, D. A. Case, A. Gronenborn, and L. T. Chong, J. Chem. Theory Comput. **12**, 3926–3947 (2016). Copyright 2016 American Chemical Society.

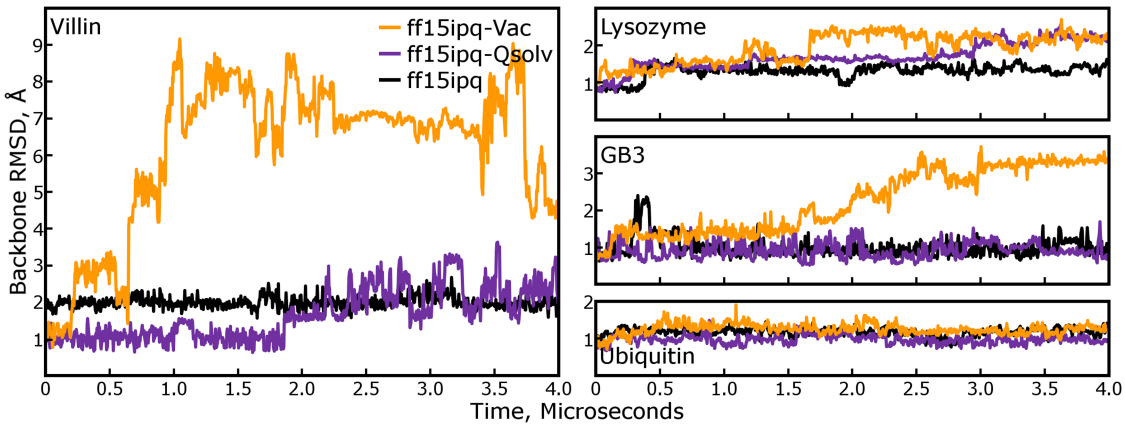

FIG. 4. Backbone positional rmsd of four small proteins at or around 300 K. Results for the canonical ff15ipq and two variants, ff15ipq-`Vac` and ff15ipq-`Qsolv`, are shown in black, gold, and purple, respectively. Adapted from K. T. Debiec, D. S. Cerutti, L. Baker, D. A. Case, A. Gronenborn, and L. T. Chong, J. Chem. Theory Comput. **12**, 3926–3947 (2016). Copyright 2016 American Chemical Society.

Simulations of the Trp cage miniprotein at a variety of temperatures indicate that ff15ipq-`Vac` allows the protein to unfold at most temperatures via a similar pathway to that observed in our earlier study on the canonical form at higher temperatures (Fig. 5). The proline cage surrounding the tryptophan residue opens up, and further unfolding occurs when the N-terminal α-helical segment unwinds, a process which appears to be much slower to reverse or perhaps irreversible. This process occurred in simulations with all three force fields but with different temperature dependence. Just as ff15ipq-`Qsolv` depicted the correct room-temperature stability for the other small proteins, it also tracked the canonical ff15ipq in terms of the folding behavior of the Trp cage over the 275-325 K temperature range. Two exceptions to the trends in either variant force field were seen: the ff15ipq-`Qsolv` simulation unfolded for 600 ns at 275 K and the ff15ipq-`Vac` simulation appeared mostly stable at 295 K. The unfolding at 275 K by ff15ipq-`Qsolv` mostly resolved itself by the end of the simulation, and the ff15ipq-`Vac` simulation at 295 K finished with signs of unfolding in the proline cage like those that eventually led to unfolding at the other temperatures. While there remains some ambiguity in the results at 4 μs of elapsed simulation time, none of the simulations anywhere on the temperature range would have yielded conclusive results after only 1 μs.

One other system simulated during validation of the original ff15ipq had both a small peptide and a large protein: the N-terminal peptide from p53 and its binding protein mDm2. Results for the protein are not plotted but consistent with the result from the canonical force field: the backbone did not change shape significantly over the course of 4 μs simulations using either force field variant when simulated with the peptide bound, and backbone positional rmsd from the

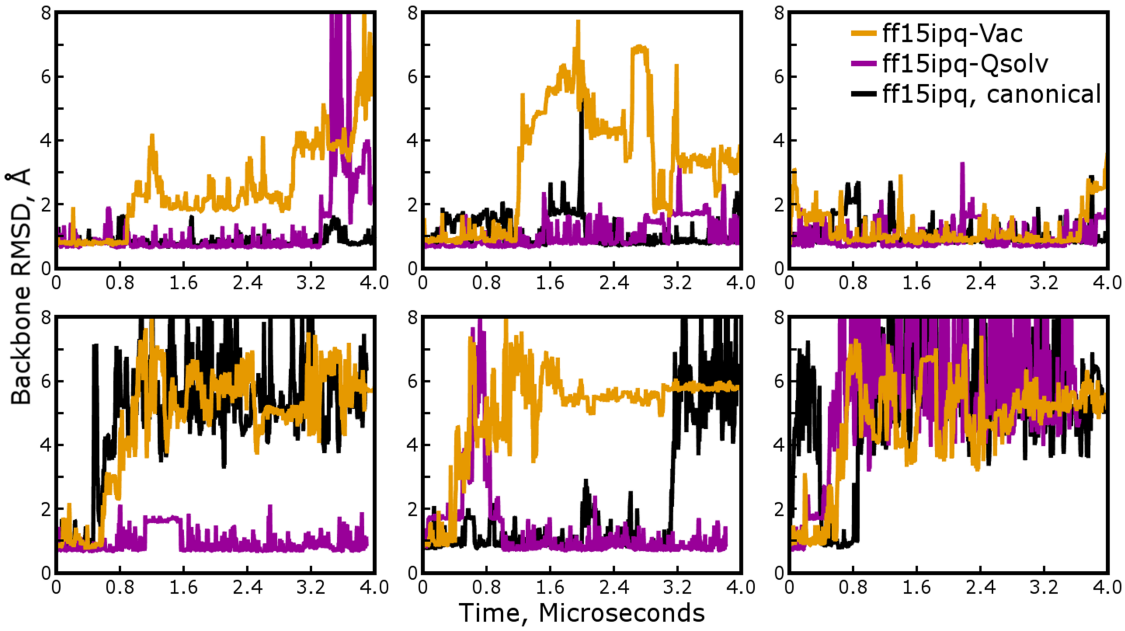

FIG. 5. Backbone positional rmsd for the Trp cage miniprotein over a range of temperatures and force field variants. The color scheme follows Fig. 4. Adapted from K. T. Debiec, D. S. Cerutti, L. Baker, D. A. Case, A. Gronenborn, and L. T. Chong, J. Chem. Theory Comput. **12**, 3926–3947 (2016). Copyright 2016 American Chemical Society.

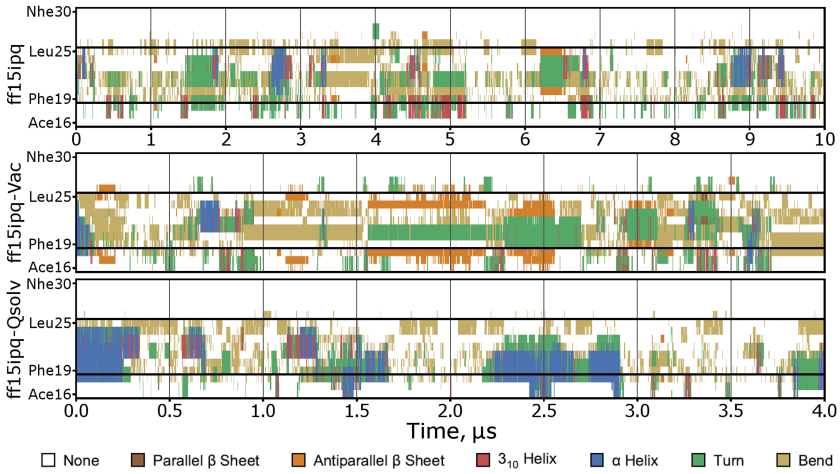

FIG. 6. Secondary structures of p53 *N*-terminal peptide under each force field. The scale on each panel is explicitly labeled in this figure as there is only one simulation temperature (298 K). Secondary structure of any residue on the *y*-axis at any time on the *x*-axis may be read from the color in the plot. Simulations concern the free peptide. When bound to mDm2, residues 19-26 form a stable α-helix. Adapted from K. T. Debiec, D. S. Cerutti, L. Baker, D. A. Case, A. Gronenborn, and L. T. Chong, J. Chem. Theory Comput. **12**, 3926–3947 (2016). Copyright 2016 American Chemical Society.

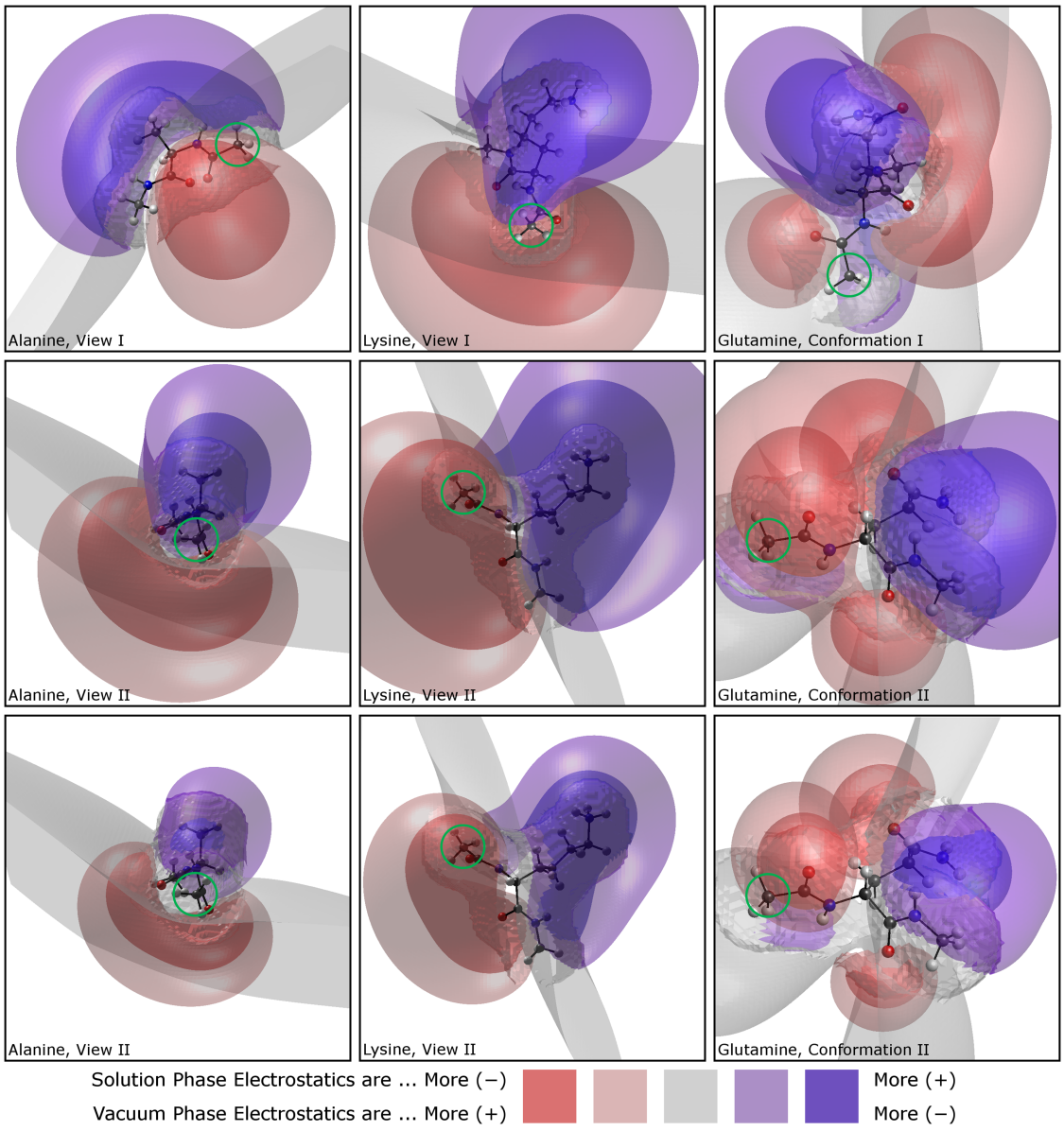

FIG. 7. Solvent-induced polarization effects for three amino acids as predicted by MP2/cc-pvTZ calculations. The color scale in the figure spans a range of ±5 kcal/mol in the top two rows of panels and ±10 kcal/mol in the lowest row of panels. Green circles in each panel highlight the methyl group of the leading Ace residue of each amino acid dipeptide: the side chain and tailing Nme residues are then easy to identify. Two *views* are shown for alanine and lysine while two *conformations* of glutamine are presented. Electrostatic potential isosurfaces are plotted throughout the solvent accessible volume, including that volume which might be accessible to water hydrogen atoms. The roughened surfaces surrounding each molecule trace the solvent accessible boundary.

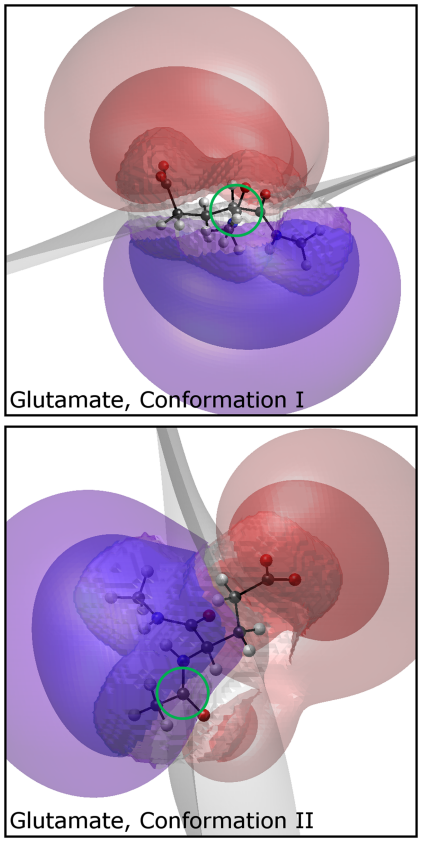

FIG. 8. Solvent-induced polarization effects for glutamate as predicted by MP2/cc-pvTZ calculations. The color scheme follows from Fig. 7 with a range of ±10 kcal/mol-*e*. Two conformations of the dipeptide are presented.

crystallographic structure was consistent between 1.1 Å and 1.2 Å with standard deviation 0.1 Å. When simulated alone, however, the peptide did show divergent behavior when simulated with the variant force fields. When simulated with the canonical ff15ipq in our earlier study, the peptide showed no propensity to form β-sheets, and only transient occupancy of the α-helical structure. In contrast, ff15ipq-`Vac` depicted some propensity for residues at either end of the peptide to adopt anti-parallel β-sheet character for approximately 20% of the simulation (see Fig. 6): the p53 peptide took on a fold like a β-hairpin for part of the simulation. Echoing the results for K19, ff15ipq-`Qsolv` depicted additional α-helical content: up to 25% in four of the central residues. In contrast to K19 and $AAQAA_3$, the p53 N-terminal peptide comprises 10 distinct amino acids in its 15-residue sequence. While they continue a trend in α-helical content among the various force fields, these differences are difficult to quantify, and the most important result for all three models is that the peptide remained predominantly disordered in the absence of mDm2.

C. Analysis of the vacuum and solution-phase charge sets

An alternative method of assessing the viability of each force field development protocol is to analyze the quality of data fitting obtained in each case. Force field studies typically report metrics such as the standard deviation of molecular mechanics energies in their final model, relative to the quantum benchmark. As parameter sets have grown and matured, these errors have diminished, but subtle differences in the

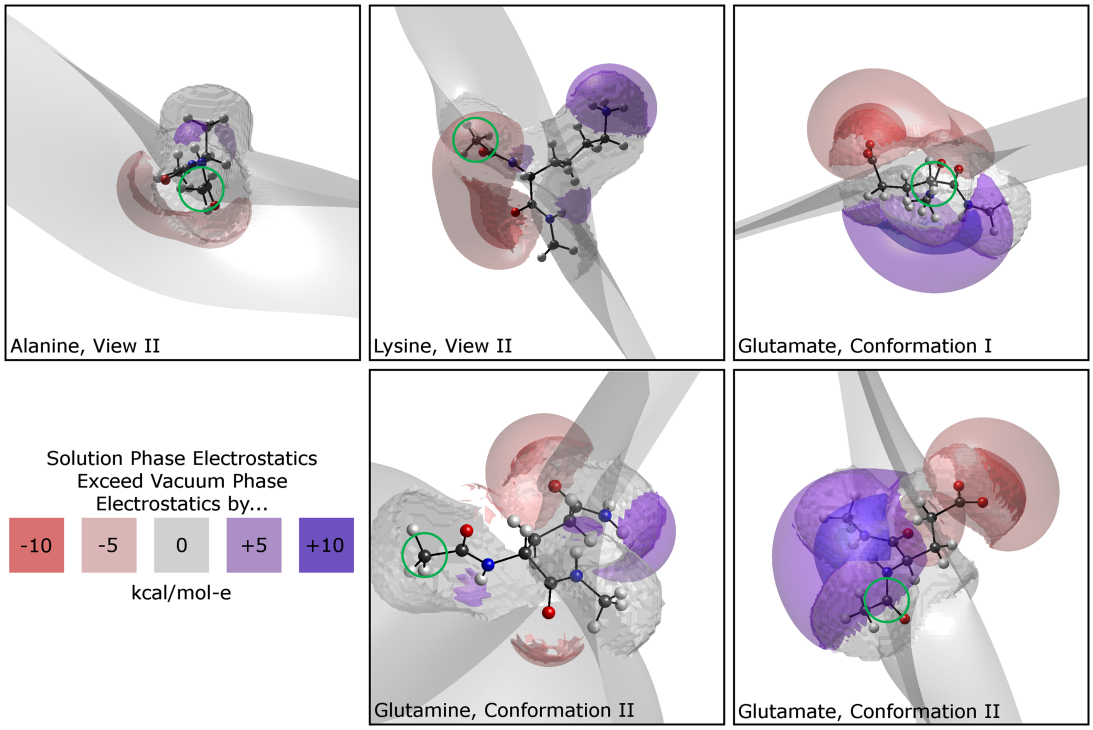

FIG. 9. Solvent-induced polarization by molecular mechanics partial charges in ff15ipq. Plots indicate the difference in electrostatic potentials projected by solution phase charges used for simulations with ff15ipq and ff15ipq-`Qsolv` and that projected by gas phase charges in ff15ipq-`Vac`. The color scheme matches Figs. 7 and 8 with a range of ±10 kcal/mol-*e*. A rough correspondence of the ±5 kcal/mol-*e* lobes of this figure with the ±10 kcal/mol-*e* lobes of Figs. 7 and 8 can be seen. All poses and view zooms are consistent between the figures.

protocols make such numbers hard to compare across different force field families. Here, we have three different force fields with the same parameter space, similar data sets, and a self-consistent development protocol that is deterministic in the limit of many refinement cycles. Correlating specific results in simulations with the details of the underlying physical approximations is one of the most difficult problems in force field development, but the three force fields we have prepared for this study present a rare opportunity.

The difference in vacuum and solution phase electrostatic fields around any of the amino acids can be mapped by subtracting the grids obtained from our MP2/cc-pvTZ calculations for any of the amino acids. We will first focus on alanine, glutamine, and lysine, as shown in Fig. 7—these amino acids compose the K19 and $AAQAA_3$ systems discussed earlier and include non-polar, polar, and charged amino acids. The method is detailed in our previous publications:[18,25] briefly, the polarized charge set Q^{Vac} is obtained from MP2/cc-pvTZ calculations in the presence of an extensive set of point charges taken directly from, or if very close to the molecular surface fitted to mimic the electrostatic field of, a time-averaged distribution of explicit water molecules obtained from molecular dynamics simulations as they move about the rigid solute. Due to the electronic polarization induced by these charges, a proton test charge will have a ±2.5 kcal/mol difference in potential energy as far as 6 Å from heavy atoms of the each dipeptide we tested: well into the second shell of water molecules. Differences of ±5 kcal/mol fill the volume of the first shell of water molecules, and protons in the first hydration layer could have their potential energies altered by up to ±10 kcal/mol due to electronic polarization. Another striking feature of Fig. 7 is that the differences in electrostatic potential follow the orientations of carbonyl groups, as is evident in the lower right panel when the backbone of glutamine takes on a β-sheet arrangement and the side chain head group is decorrelated.

The influence of polarization in the carbonyl groups is reinforced by findings for glutamate. Figure 8 shows an overlap of the carboxylate head group orientation and an α-helical backbone arrangement: electronic polarization changes the potential energy surface by more than 10 kcal/mol for a proton test charge throughout the first hydration shell. Solvent-induced polarization in the carboxylate head group creates perhaps the strongest differences in the electrostatic field of any natural amino acid: comparing the lower panel of Fig. 8 to the lower row of panels in Fig. 7 (note that the color scale of these panels matches, plotting isosurfaces of ±5.0 and ±10.0 kcal/mol-*e*, but the top two rows of Fig. 8 use the color scale to plot isosurfaces of only ±2.5 and ±5.0 kcal/mol-*e*). In contrast, the solvent-induced polarization of the lysine amino head group projects strong +10 kcal/mol-*e* potential differences only slightly beyond the molecular surface. It must be understood that differences in the electrostatic potential affect water molecules as correlated *dipoles*, not an ideal gas of ions, but the plots establish a clear order of the strength of effects projected by several chemical groups.

While it is trivial to look at the differences in point charges between Q^{Vac} and Q^{IPol}, the familiar problem of buried charges and uncertainty in the solutions to REsP calculations[17] confounds meaningful conclusions based on the partial charges themselves. Rather, the collective action of the monopole distribution—the electrostatic field it creates—is essential. Figure 9 shows how the *molecular mechanics* representation $Q^{\mathrm{IPol}} - Q^{\mathrm{Vac}}$ depicts the effects of electronic polarization in alanine, lysine, glutamine, and glutamate. Each panel contains a view of the molecule in a pose matching a panel from Fig. 7 or 8. The IPolQ charge model does not depict as extensive a difference between the electrostatic fields in a vacuum and in solution for an obvious reason: Q^{IPol} is fitted to the *average* of the electrostatic potentials in each phase, not simply the solution phase electrostatic potential, implicitly estimating the energy penalty of electronic polarization with a mean field approximation. The process of adding restraints to weakly couple Q^{IPol} and Q^{Vac} also has the effect of depressing differences between the electrostatics in each phase (see our previous publications for more details).[18] Virtually anywhere, the

TABLE III. Error in fitting electrostatic potential energy surfaces around the canonical amino acid dipeptides and some protonated variants. Details of the charge fitting can be found in the original ff15ipq publication:[16] 20 conformations of each dipeptide were used in a simultaneous least squares optimization of 250 unique charges for the main-chain amino acids.

	Error,[a] kcal/mol-*e*	
Residue	Vacuum	IPolQ
Non-polar amino acids		
ALA	1.80	1.73
GLY	1.93	1.74
ILE	2.02	1.75
LEU	1.98	1.82
MET	2.33	2.18
PHE	1.96	1.65
PRO	1.80	1.89
TRP	2.03	1.74
VAL	1.89	1.68
Polar amino acids		
ASH	2.34	1.75
ASN	2.01	1.74
CYS	2.64	2.27
GLH	2.13	1.79
GLN	1.97	1.76
HID	2.10	1.92
HIE	2.19	1.94
LYN	2.30	2.14
SER	2.14	1.84
THR	2.11	1.81
TYR	1.94	1.73
Ionic amino acids		
ARG	2.40	1.93
ASP	2.63	2.02
GLU	2.47	1.89
LYS	2.41	1.88
HIP	2.43	1.80

[a]Root mean squared error fitting the electrostatic potential (ESP) target throughout the volume surrounding each molecule but not closer than water hydrogens might reasonably penetrate during a normal simulation. For vacuum phase charges, the target is just the ESP of the molecule in vacuum. For IPolQ charges, the target is the average of the vacuum ESP and the ESP obtained in the presence of a polarizing charge density based on the time-averaged distribution of SPC/E-b water.

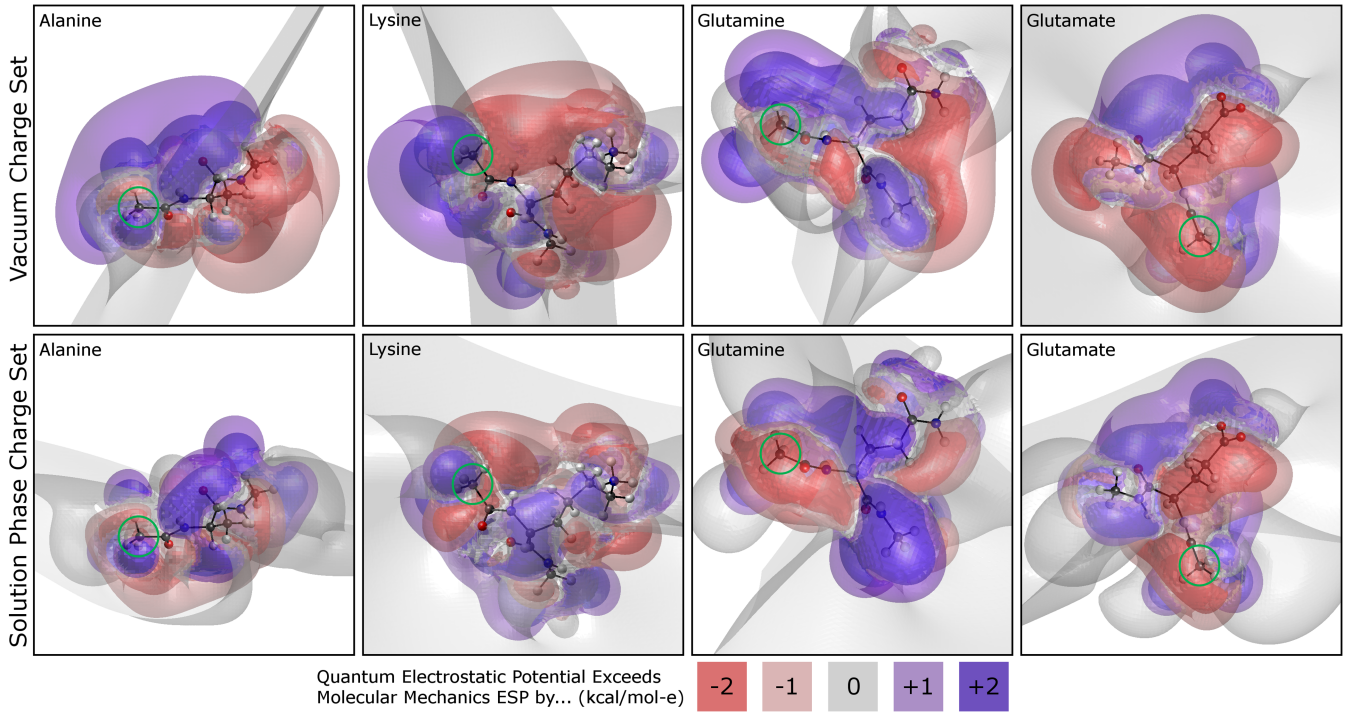

FIG. 10. Error in reproduction of the quantum electrostatic potential surface by Q^{Vac} and Q^{IPol}. The color scale spans only ±2 kcal/mol-*e* in this figure. Poses of the dipeptide systems do not correspond to previous figures, but the green circles again indicate the locations of the Ace methyl group for orientation.

polarized quantum mechanical (QM) field was ±10 kcal/mol-*e* stronger than the QM field in a vacuum, the molecular mechanics representation Q^{IPol} depicted a field ±5 kcal/mol-*e* stronger than Q^{Vac}, but close inspection shows that the differences between fields created by Q^{IPol} and Q^{Vac} were never quite as extensive as their QM counterpart.

The overall error in charge fitting for each of the main chain amino acids is shown in Table III. Clearly, the monopole distribution in ff15ipq and its variants (partial charges on all atomic nuclei) were better able to reproduce the electrostatic potential we posited for the amino acids in solution (the average of potentials due to the polarized and unpolarized wavefunctions) than the electrostatic potential due to the amino acids' unpolarized wavefunctions. Underneath the mean values, the error in each molecular mechanics approximation possessed its own structure, as illustrated in Fig. 10. For all of the amino acids presented, there were numerous correspondences: both Q^{Vac} and Q^{IPol} make many of the same compromises, over- and under-estimating their respective electrostatic potential targets in the same way throughout the space around each molecule. Some of this must have occurred because of the weak coupling between Q^{Vac} and Q^{IPol} but partial charges situated on nuclei can systematically fail to reproduce many details of molecular electrostatics.[25]

The primary motivation for building the canonical ff15ipq bonded parameters with a set of gas phase charges was to properly handle the gas phase quantum potential energy surfaces that formed the basis of the bonded parameter fitting. The error with which the molecular mechanics representation could reproduce these potential energy surfaces (PESs), scanning torsion profiles and angle flexing, was also dependent on the charge set. As expected, the gas phase charges Q^{Vac} afforded better reproduction of the quantum PES for systems containing polar and non-polar residues (including the terminal residues Ace and Nme). However, this trend was not absolute: bonded parameter fitting in systems containing any ionic residues tended to improve with Q^{IPol} as shown in Fig. 11. At present, we have no explanation as to why, but the

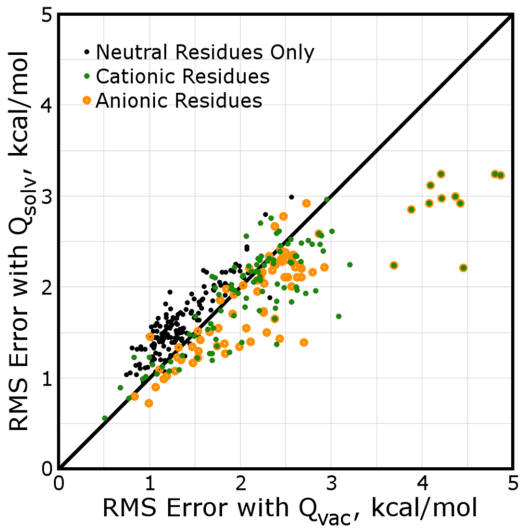

FIG. 11. Error in bonded parameter fitting in the context of either charge model. The root mean squared error by which the molecular mechanics model mimics the quantum potential energy surface was calculated for each of 330 unique systems in the ff15ipq bonded parameter training. Points lying above the 1:1 trend line indicate that the error in data fitting was lower when using Q^{Vac} in the molecular mechanics energy function. Points below the trend line indicate that bonded parameters were easier to fit in the context of Q^{IPol}. About a dozen systems contained both cationic and anionic residues: these are the donut points with a green dot inside a yellow circle.

fact that the error in the electrostatic PES also tends to be the worst for ionic residues in a vacuum may offer a clue.

IV. DISCUSSION

Through extensive collection of long molecular dynamics trajectories, we have tested the contribution that each charge set makes to the quality of simulations performed with ff15ipq. While there is still uncertainty in the results, it seems fair to claim that a polarized charge model is essential for simulating proteins accurately, while accounting for the vacuum environment of the target quantum single-point energies yields an incremental improvement when developing non-polarizable molecular models. Almost nowhere does running simulations with an unpolarized, gas phase charge set give better results than simulations with polarized charges (regardless of how the bonded parameters were developed), and while there are select cases where either canonical ff15ipq or ff15ipq-`Qsolv` performs better than the other, it is more evident that both models fail to reproduce challenging quantities such as the melting profiles of helical peptides.

The most striking success of ff15ipq-`Qsolv` over the canonical force field is in the helical content of K19, but this must be taken along with the fact that it vastly overestimated the helical content of the similar $AAQAA_3$ system and shows signs of increased helicity in p53 N-terminal peptide and the GB1 hairpin system. The potentials of mean force in Fig. 2 help explain both results: the region of α-helicity is favored in the case of ff15ipq-`Qsolv` for both alanine and glutamine amino acids but not lysine, conceivably stabilizing $AAQAA_3$. However, if the tetrapeptide PMFs are indicative of the behavior in oligomers, the results hinge on changes of about 0.5 kcal/mol. (The color axis on the difference plots spans ±2.5 kcal/mol, but the limits are only reached in high-energy regions of each landscape.) Such differences are within the error of the data fitting underlying each of the force fields tested.

The over-stabilization of α-helices in force fields developed in the manner of ff15ipq-`Qsolv` has been observed before, but the results are not consistent.[26] Both of the other major Amber force fields, including ff99-SB and its progeny as well as ff03,[27] were found to emphasize helicity in a disordered fragment of the amyloid β protein. CHARMM-27 with CMAP corrections was found to do the same. With the exception of ff03, all of these force fields take what is more or less and polarized charge set and use it as the basis for fitting torsion parameters against gas-phase quantum data at the MP2 level. However, the Amber ff03 model harkens more to the ff15ipq-`Vac` variant we tested: a charge set designed for the interiors of proteins that would be expected to have relatively little solvent-induced polarization was used as the basis to develop torsion parameters against quantum data created in a similar continuum dielectric environment. Furthermore, the α-helical content of ff99-SB (a model with polarized charges and backbone torsion parameters fitted in the context of those charges against quantum data at the MP2/cc-pVTZ level of theory) was found to be too low specifically on the K19 system.[4] It was raised to nearly the experimental value in ff14-SB with *ad-hoc* adjustments to backbone dihedrals. While it remains difficult to attribute any of this behavior to any one aspect of each biomolecular model, a fully automated protocol offers the chance to systematically improve the accuracy of the data fitting with more sophisticated force fields and hopefully one day to see the effects of each physical assumption through the noise that we must contend with today.

While it is in principle better to account for the phases of the charge model and quantum data in bonded parameter development, error of the data fitting was also one of the initial motivations to pursue the dual charge models in IPolQ force fields. Early in production of ff14ipq (data not published), it was found that the Cornell charges, which were somewhat less polar along the backbone than the ff14ipq IPolQ charge set, gave better fits to our backbone ϕ/ψ scans. A more thorough investigation of the ff15ipq charges reveals that this picture is incomplete: while the vacuum phase charge set does improve the bonded parameter data fitting for the neutral amino acids, the data fitting worsens for charged amino acids. This behavior parallels data fitting in the charges themselves: the electrostatic fields from vacuum phase quantum calculations are consistently harder to fit, at least with nuclear-centered charges, than fields from calculations with polarizing charge densities surrounding the quantum system. It will be interesting to test whether this trend holds when including additional monopoles. Either set of molecular mechanics charges Q^{Vac} or Q^{IPol} deviates from its quantum benchmark in a convoluted fashion, but Fig. 10 suggests that an improved monopole distribution which improves the accuracy of either charge set will also benefit the other.

Analysis of the quantum mechanical electrostatic potentials shows that the most significant polarization occurs in carbonyl and carboxylate groups, but that peptide and amide groups show polarization throughout the C═O and N—H moities. Increasingly negative potentials around the carbonyl and positive potentials around the secondary amine lend critical stability to interactions between protein backbones in all of the major secondary structures, but the polarization cannot be over-emphasized lest the proteins become over-stabilized together in their initial configurations or coaxed to prefer one secondary structure over another. By construction, the IPolQ molecular mechanics model only shows up to half of the polarization evident in the underlying quantum calculations, to account for the polarization energy penalty with attenuated non-bonded interactions.

The stability of globular proteins and distributions of structures seen in smaller polypeptides indicate that the polarized Q^{IPol} charge set can strike a good balance depending on how the angle and torsion parameters are developed to complete the force field. If the bonded parameters are developed from gas-phase quantum potential energy surfaces in the context of the appropriate charge set Q^{Vac}, the resulting force field models approximately the same stability in small peptides regardless of which charge model is finally included in simulations, as shown in Figs. 1 and 3. This finding is consistent with Fig. 2, showing that the energy landscapes of individual amino acids do not hinge on the charge model as much as the manner in which the bonded parameters were derived. However, tertiary structures cannot remain stable without the correct charge set for simulations in water, and if Q^{IPol} is necessary, the

effects of polarized charges on the bonded parameter fitting are influential in the energy landscapes of individual amino acids and up to the level of small peptides, making significant changes in the balance of α-helical and β-sheet populations. Whether these effects would also extend to the tertiary structures is an open question until more strenuous sampling methods and faster computer hardware extend simulations to this level.

SUPPLEMENTARY MATERIAL

See supplementary material for parameters of ff15ipq-`Vac` and ff15ipq-`Qsolv` provided in Amber file formats. This information is available free of charge via the internet at aip.jcp.org.

ACKNOWLEDGMENTS

The authors thank the staff in the Rutgers Chemistry Department for help with computations and curating the data set. Replicated simulations on the short peptides were conducted with the use of the Michigan State University Institute for Cyber-Enabled Research. This research was supported in part by NIH Grant No. GM122086.

[1] L. Kuyper, D. Ashton, K. M. Merz, and P. A. Kollman, J. Phys. Chem. **95**, 6661–6666 (1991).
[2] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179–5197 (1995).
[3] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, Proteins: Struct., Funct., Bioinf. **65**, 712–725 (2006).
[4] J. A. Maier, C. Martinex, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, J. Chem. Theory Comput. **11**, 3696–3713 (2015).
[5] A. Jakalian, B. L. Bush, D. B. Jack, and C. I. Bayly, J. Comput. Chem. **21**, 132–146 (2000).
[6] A. Jakalian, D. B. Jack, and C. I. Bayly, J. Comput. Chem. **23**(16), 1623–1641 (2002).
[7] J. Wang, P. Cieplak, and P. A. Kollman, J. Comput. Chem. **21**, 1049–1074 (2000).
[8] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, J. Comput. Chem. **25**, 1157–1174 (2004).
[9] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, I. W. E. Reiher, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, J. Phys. Chem. B **102**, 3586–3616 (1998).
[10] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. **118**, 11225–11236 (1996).
[11] G. A. Kaminski, R. A. Friesner, and J. Tirado-Rives, J. Phys. Chem. B **105**, 6474–6487 (2001).
[12] E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. X. Xiang, L. Wang, D. Lupyan, M. K. Dahlgren, J. Knight, J. W. Kaus, D. S. Cerutti, G. Krilov, W. L. Jorgensen, R. Abel, and R. A. Friesner, J. Chem. Theory Comput. **12**, 281–296 (2016).
[13] J. Huang and A. D. MacKerell, Jr., J. Comput. Chem. **34**, 2135–2145 (2013).
[14] D. A. Case, T. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang, and R. Woods, J. Comput. Chem. **26**, 1668–1688 (2005).
[15] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten, J. Comput. Chem. **26**, 1781–1802 (2005).
[16] K. T. Debiec, D. S. Cerutti, L. R. Baker, D. A. Case, A. Gronenborn, and L. T. Chong, J. Chem. Theory Comput. **12**, 3926–3947 (2016).
[17] C. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman, J. Phys. Chem. **97**, 10269–10280 (1993).
[18] D. S. Cerutti, W. C. Swope, J. E. Rice, and D. A. Case, J. Chem. Theory Comput. **10**, 4515–4534 (2014).
[19] R. Schweitzer-Stenner, J. Phys. Chem. B **113**, 2922–2932 (2009).
[20] S. Woutersen, R. Pfister, P. Hamm, Y. Mu, D. S. Kosov, and G. Stock, J. Chem. Phys. **117**, 6833–6840 (2002).
[21] J. Graf, P. H. Nguyen, G. Stock, and H. Schwalbe, J. Am. Chem. Soc. **129**, 1179–1189 (2007).
[22] K. Song, J. M. Stewart, R. M. Fesinmeyer, N. H. Andersen, and C. Simmerling, Biopolymers **89**, 747–760 (2008).
[23] J. Huang and A. D. MacKerell, Jr., Biophys. J. **107**, 991–997 (2014).
[24] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141–151 (1999).
[25] D. S. Cerutti, W. C. Swope, J. E. Rice, and D. A. Case, J. Phys. Chem. B **117**, 2328–2338 (2013).
[26] M. D. Smith, J. S. Rao, E. Segelken, and L. Cruz, J. Chem. Inf. Model. **55**, 2587–2595 (2015).
[27] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, and T. Lee, J. Comput. Chem. **24**, 1999–2012 (2003).
[28] D. A. Case, C. Scheurer, and R. Brüschweiler, J. Am. Chem. Soc. **122**, 10390–10397 (2000).