

AMOEBA Force Field Trajectories Improve Predictions of Accurate pK_a Values of the GFP Fluorophore: The Importance of Polarizability and Water Interactions

Published as part of *The Journal of Physical Chemistry* virtual special issue "Biomolecular Electrostatic Phenomena".

Yu-Chun Lin, Pengyu Ren, and Lauren J. Webb*



Cite This: <https://doi.org/10.1021/acs.jpcb.2c03642>



Read Online

ACCESS |



Metrics & More

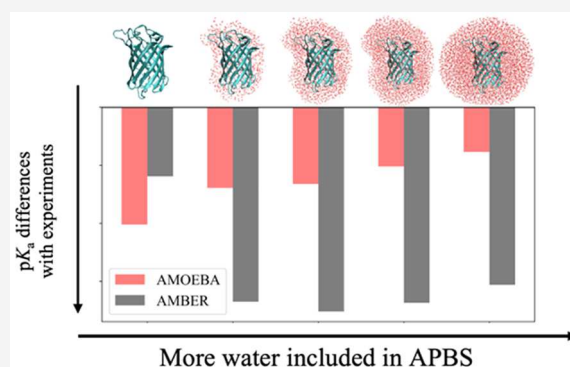


Article Recommendations



Supporting Information

ABSTRACT: Precisely quantifying the magnitude, direction, and biological functions of electric fields in proteins has long been an outstanding challenge in the field. The most widely implemented experimental method to measure such electric fields at a particular residue in a protein has been through changes in pK_a of titratable residues. While many computational strategies exist to predict these values, it has been difficult to do this accurately or connect predicted results to key structural or mechanistic features of the molecule. Here, we used experimentally determined pK_a values of the fluorophore in superfolder green fluorescent protein (GFP) with amino acid mutations made at position Thr 203 to evaluate the pK_a prediction ability of molecular dynamics (MD) simulations using a polarizable force field, AMOEBA. Structure ensembles from AMOEBA were used to calculate pK_a values of the GFP fluorophore. The calculated pK_a values were then compared to trajectories using a conventional fixed charge force field (Amber03 ff). We found that the position of water molecules included in the pK_a calculation had opposite effects on the pK_a values between the trajectories from AMOEBA and Amber03 force fields. In AMOEBA trajectories, the inclusion of water molecules within 35 Å of the fluorophore decreased the difference between the predicted and experimental values, resulting in calculated pK_a values that were within an average of 0.8 pK_a unit from the experimental results. On the other hand, in Amber03 trajectories, including water molecules that were more than 5 Å from the fluorophore increased the differences between the calculated and experimental pK_a values. The inaccuracy of pK_a predictions determined from Amber03 trajectories was caused by a significant stabilization of the deprotonated chromophore's free energy compared to the result in AMOEBA. We rationalize the cutoffs for explicit water molecules when calculating pK_a to better predict the electrostatic environment surrounding the fluorophore buried in GFP. We discuss how the results from this work will assist the prospective prediction of pK_a values or other electrostatic effects in a wide variety of folded proteins.



INTRODUCTION

Electric fields generated by the three-dimensional position of amino acid atomic charges within a protein have long been recognized as a fundamental driving force directing critical biomolecular properties including folding, dynamics, catalysis, ligand binding, protein–protein interactions, and other functions.^{1–6} Quantifying the magnitude and direction of electric fields within and around proteins is a complex task because these fields are generated by a large amount of partial charges interacting through space over short and long-range.^{7–9} It has been demonstrated that it is often necessary to consider Coulombic information from residues far from an enzyme active site in order to replicate experimentally observed effects such as reduction potential¹⁰ or enzyme kinetic efficiency.^{11–13} Taken together, long-standing work

from both experimental and computational perspectives have demonstrated that a measurable indicator of integrated protein electric field combined with a physical model that can support the experimental measurements is extremely complicated.

Although there are a number of experimental methods for measuring electric fields in proteins, including NMR spectroscopy^{14–16} and vibrational Stark effect spectroscopy,^{17,18} by far

Received: May 26, 2022

Revised: September 13, 2022

the most common method has been measuring pK_a shifts of titratable residues and interpreting those shifts in terms of local electric field around the proton.^{19–23} This is because the tendency of the residue to stay in its charged versus neutral state can be accurately measured through ^1H and ^{13}C chemical shifts by NMR or by absorption or emission energy changes of an internal spectroscopic probe.^{14,15,17,19,22} Moreover, the tendency of a residue to stay in its charged or neutral state in a protein is physically determined by the electrostatic environment generated from the positions of partial charges of all atoms in that protein, including associated water molecules.^{23–26} Thus, the development of the understanding on the pK_a values of residues at the active site or targeted location in a protein is the base of any protein application technology and the corresponding model theory. Although measured pK_a values are good indicators for local fields generated by the surrounding electrostatic environment, even with this information it is challenging to determine which specific interactions contribute to a pK_a value because it is a measurement of a change in free energy of two separate states (protonated and deprotonated), affected by the chemical environment over a range of distances from the position of the titrated proton. For example, the pK_a value of a lysine side chain in water is 10.5, but when lysine is buried in a low dielectric, hydrophobic protein interior, pK_a values as low as 5.3 have been measured.²⁰ Similarly, the side chain of aspartic acid has a pK_a of 1.99 in water, but this can increase to as high as 9.30 when the residue is buried in a more hydrophobic protein interior.²⁶ However, knowing that a hydrophobic environment stabilizes the neutral form of an acid or base does not provide the specific magnitude and direction of electric fields or indicate which atoms over which distance range are involved, a level of detail necessary for understanding the effect of electric field on all aspects of protein structure and function. To obtain this, many attempts have been made to theoretically predict the pK_a values of different targeted residues using quantum mechanics, combining quantum mechanics with classical molecular mechanics (QM/MM),^{27,28} different approaches of continuum electrostatics,^{22–25,29–34} and the more recent constant pH molecular dynamics (MD) simulations.^{35–37} Among these techniques, the more conventional and available electrostatic calculation methods could accurately predict the pK_a values of protein residues that are exposed to solvent, but the pK_a values of residues that are buried in a complicated environment such as a protein interior remain challenging.^{29,38} The pK_a values of buried residues could possibly be accurately calculated by constant pH MD simulations;^{36,37} however, this is not currently available with polarizable force fields, the focus of the current work.

One significant challenge in quantifying the electrostatic effect of a protein interior accurately with all relevant physical details is the computational cost. A common strategy to calculate pK_a values of buried residues is to use a thermodynamic cycle to calculate the differences between the electrostatic free energy of a titratable residue in water and in a protein embedded in continuum solvent by solving the linear Poisson–Boltzmann (PB) equation.^{39–42} While the direction of the pK_a shifts between water and protein for protein surface residues can usually be predicted accurately,^{23,35} the magnitude of the pK_a shifts can be off by several pK_a units, especially for buried residues.^{19,23,35} Combining the PB equation with different MD force fields allowed us to improve the reliability of pK_a calculations by optimizing the computational

procedures. In this work we explore the contributions of three common contributions to inaccurate predictions of pK_a values: (1) relegating all water molecules, no matter their distance from the protein surface, to part of the dielectric continuum that may neglect critical nonbonded interactions; (2) ignoring many-body polarization during calculations;^{19,29,38} and (3) inaccurate structural ensembles generated by MD simulations that will skew accurate representation of electrostatic effects.^{19,35,40,43}

To conduct MD simulations, it is necessary to choose a potential energy function to evaluate the forces on every atom, such as that shown in eq 1, where the total potential energy of the system is a sum of potential energies from bond vibration (U_{bond}), angle bending (U_{angle}), vibrational and bending coupling ($U_{\text{b}\theta}$), out-of-plane bending (U_{oop}), dihedral torsions (U_{tor}), van der Waals interactions (U_{vdw}), permanent electrostatic interactions ($U_{\text{ele}}^{\text{perm}}$), and many-body induction or polarization ($U_{\text{ele}}^{\text{ind}}$).⁴⁴

$$U = U_{\text{bond}} + U_{\text{angle}} + U_{\text{b}\theta} + U_{\text{oop}} + U_{\text{tor}} + U_{\text{vdw}} + U_{\text{ele}}^{\text{perm}} + U_{\text{ele}}^{\text{ind}} \quad (1)$$

Common classical force fields use fixed charges for every atom without considering the polarization effect ($U_{\text{ele}}^{\text{ind}}$)^{19,45,46} for computational speed. Although MD simulations using these fixed charge force fields can predict the secondary structures of peptides and proteins with a high degree of accuracy,^{47,48} they may fail to generate structures that can be used to accurately predict the pK_a values of buried residues.¹⁹ One possible reason for this is that the short-ranged polarization between atoms makes strong contributions to pK_a in addition to long-ranged permanent electrostatic interactions, suggested by several studies that compared the calculated pK_a values from both fixed charge and polarizable force fields.^{39,40,49,50} However, previous studies have either only included polarization in the energy minimization process or applied auxiliary particles to polarizable atoms rather than directly considering many body polarizabilities by applying atomic dipole induction. To investigate this, we have chosen to explore the polarizable atomic multipole optimized energetics for biomolecular applications (AMOEBA) force field that includes many-body polarization through atomic dipole induction. Moreover, the permanent electrostatic interaction term ($U_{\text{ele}}^{\text{perm}}$) in AMOEBA also considers permanent dipole and quadrupole interactions to generate a more detailed description of electrostatic fields generated within a protein that are from atomic charge only.^{39,44}

Previous work in our laboratory used the Amber03 fixed charge force field to perform MD simulations of structures of mutants of superfolder GFP (hereafter GFP).^{19,51} The wild type structure of this construct (taken from protein data bank (PDB) accession number 2b3p) is shown in Figure 1.⁵¹ It includes 11 strands arranged in a β -barrel, the fluorescent chromophore, multiple water molecules within the protein interior (two are shown in Figure 1), and a nearby residue T203 that was mutated in this work. The naturally occurring fluorophore located inside the GFP β -barrel can be either protonated (A state) or deprotonated (B state), shown in Figure 2. To quantify the effect of electric fields on the electrostatic environments around the fluorophore, mutations were made at a nearby residue, T203, to Cys, Phe, His, Asn, Ser, and Tyr. The experimentally measured pK_a values of the fluorophore in these mutants ranged from 6.48 to 7.95¹⁹ because of changes in the electrostatic environment near the

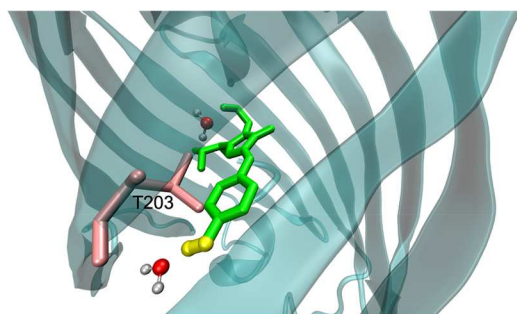


Figure 1. Representative structure of WT GFP (cyan) enclosing the TYG fluorophore (green) with the hydroxyl group (yellow) that is titrated in experiments and simulations. T203 (pink) and two nearby water molecules are also shown.

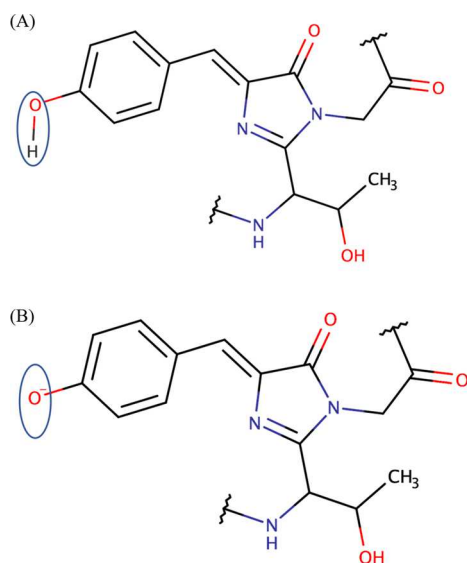


Figure 2. Structure of the chromophore that was used to parametrize the molecule for AMOEBA simulations showing the protonated (A) and deprotonated (B) states.

chromophore controlled by the chemical structure at position 203. The different hydrogen bonding network between nearby water molecules, protein residues around the fluorophore, and the phenol group of the fluorophore determined the pK_a values of the fluorophore in these mutants. Interactions between the fluorophore and the side chains of aromatic residues, such as tyrosine and phenylalanine, tended to stabilize the neutral form of the fluorophore. However, the computational predictions made using Amber03 trajectories with linear PB equations in our previous work were not able to reproduce the experimental pK_a values of the fluorophore.¹⁹

In this work, we used the AMOEBA polarizable force field in MD simulations to compare pK_a predictions directly to the Amber03 fixed charge force field. Equilibrium ensembles from AMOEBA MD simulations were used to calculate the electrostatic free energies of the A and B states of the fluorophore in GFP by solving the linear PB equation.^{39–42} In a typical application of the linear Poisson–Boltzmann equation, all of the water molecules are replaced by a dielectric continuum of 78.54.^{19,52} In this work, to better compare the pK_a values calculated from the trajectories from Amber03¹⁹ versus those from AMOEBA force fields, different amounts of water molecules were treated explicitly by PB. In this

treatment, every atom in the simulation (protein and water) for trajectories from both force fields was represented by the atomic charges and radii based on the corresponding values of the two force fields as water at increasingly large distances away from the fluorophore was included explicitly in the PB calculation. The pK_a values of the fluorophore generated by these two different force fields were compared with the previously published experimental pK_a data sets.¹⁹ We found that including more explicit water molecules in PB calculations lowered the accuracy of the predicted pK_a values from the Amber03 force field trajectories but increased the accuracy of predicted pK_a values from AMOEBA trajectories. Moreover, the greater discrepancy in pK_a values from Amber03 trajectories were not random but instead were caused by overstabilization (and subsequent decrease in free energy) of the deprotonated states of each of the mutants as more water molecules were included explicitly.

METHODS

Molecular Dynamics (MD) Simulation—AMOEBA Force Field. The parameters for the AMOEBA polarizable force field for the GFP fluorophore in both protonated (A) and deprotonated (B) states (Figure 2) were derived from the AMOEBA parameter generator program Poltype.⁵³ Gaussian 09 was used for the QM calculations in Poltype.⁵⁴ The initial GFP structures of WT and all mutants were taken from our previous work,¹⁹ which describes preparation of each structure and the considerations of the protonation states of the residues as they relate to previous studies.^{46,51,55} We conducted MD simulations of WT (T203), T203C, T203F, T203H, T203N, T203S, and T203Y with both protonated and deprotonated states of the fluorophore by using the AMOEBA force field implemented in Tinker version 8.7.⁵⁶ Packmol⁵⁷ was used to build the simulation box containing 5–6 Na^+ to neutralize the charge of the GFP and water molecules to fill a cubic box of $98 \text{ \AA}^3$.⁴⁴ After energy minimization of the system, the MD simulations started at constant volume heating from 0 to 298 K for 6.6 ns with a time step of 2 fs, followed by constant pressure MD at 298 K for 2 ns at a time step of 2 fs. The production MD simulations were done with the averaged constant volume for 90 ns by using a time step of 2 fs with frames recorded every 4 ps. From the energy minimization process to every phase of the MD simulations, the permanent charge, dipole, quadrupole, and many-body polarization were all considered. For all mutants, we calculated the root-mean-square deviation (RMSD) of protein backbone atoms between each MD structure and the initial energy minimized protein structure to determine when the AMOEBA structures became stable. These results are shown in Figure S1; we observed that the RMSD values became stable after a few ns. For the fluorophore itself, the RMSD values between the AMOEBA trajectories and the fluorophore of the initial energy minimized proteins were less than $1 \text{ \AA}$ throughout the simulation, shown in Figure S2. We took the trajectories, including the water molecules and every atom in the protein, that were generated from the last 40 ns, collected in 500 ps steps, to calculate pK_a values.

The corresponding RMSD values from Amber03 trajectories for the protein backbone atoms for each captured structure and the initial energy minimized protein are shown in Figure S3.¹⁹ The parameters of the protonated and the deprotonated fluorophore were generated by Nifosi et al.⁵⁸ The Amber03 structures included the MD simulations on the protein

structures by Amber03 force field^{59,60} in the GROMACS (5.0.4) software package and explicit solvation with TIP3P water.^{61,62} We observed that the RMSD values remained low (<2.7 Å) after 10 ns. As previously observed, the AMOEBA polarizable force field gave rise to larger fluctuations (measured by RMSD) during MD simulations of proteins than fixed charge force fields due to the differences in torsional parameters.⁶³ RMSD values for the fluorophore when compared to the initial energy minimized structure were less than 1 Å throughout the simulation, shown in Figure S4.¹⁹ We used trajectories from 10 to 50 ns generated from the previous work¹⁹ to further calculate pK_a values.

Combination of MD Trajectories with Electrostatic Free Energy Calculation in APBS. We treated both the Amber03 and the AMOEBA MD trajectories exactly the same way since they were the input trajectories in the APBS calculations. We referred this treatment to previous work that calculated the pK_a values of carboxylic residues in protein with the both fixed charge (OPLS) and the polarizable force field (PFF) energy minimized structures followed by an electrostatics calculation with PB. In that work, the authors concluded that the PFF structure combined with PB equation predicted pK_a values closer to the experimental results despite the fact that the PB calculation does not take advanced electrostatic terms into account.³⁹ The value of the dielectric constant of GFP in this work was determined by referring to the previous work,¹⁹ which suggested the screening effect throughout the dielectric constant between 2 and 8 were similar¹⁹ when the MD trajectories were generated by all-atom simulations for GFP. The value 6 was then determined based on the better correlation with experimental pK_a values of GFP fluorophore;¹⁹ the value 6 was then adopted in this work.

For the last 40 ns of both AMOEBA and Amber03 trajectories, we collected structures every 500 ps to calculate the free energy differences between the protonated and deprotonated fluorophore in GFP using the adaptive Poisson–Boltzmann Solver (APBS) to calculate protein free energy.⁶⁴ The goal was to calculate the free energy of the fluorophore in water and in GFP ($\Delta\Delta_a G$), as shown in eq 2:

$$\Delta\Delta_a G = -RT \ln(K_a) \quad (2)$$

using a convenient thermodynamic cycle illustrated in Figure 3.^{19,65–68} Here, the pK_a shift (ΔpK_a) between the fluorophore in water (8.2) and in protein was determined using eqs 3 and 4:

$$\begin{aligned} \Delta\Delta_a G &= \Delta_a G_{\text{free}} - \Delta_a G_{\text{protein}} = \Delta G_{\text{xferB}} - \Delta G_{\text{xferA}} \\ &= -RT \ln(K_a) \end{aligned} \quad (3)$$

$$\Delta pK_a = \frac{\Delta_a G_{\text{free}} - \Delta_a G_{\text{protein}}}{RT \ln 10} = \frac{\Delta G_{\text{xferB}} - \Delta G_{\text{xferA}}}{RT \ln 10} = \frac{\Delta\Delta_a G}{RT \ln 10} \quad (4)$$

where $\Delta\Delta_a G$ is the free energy difference between the fluorophore in an aqueous environment isolated from the protein ($\Delta_a G_{\text{free}}$) and the protein environment ($\Delta_a G_{\text{protein}}$). We calculated $\Delta\Delta_a G$ by determining $\Delta_a G_{\text{xferA}}$ and $\Delta_a G_{\text{xferB}}$, which represent the differences in electrostatic free energy of the fluorophore transferred from water to the protein environment for the A state and B state, respectively. These were calculated from eq 5:

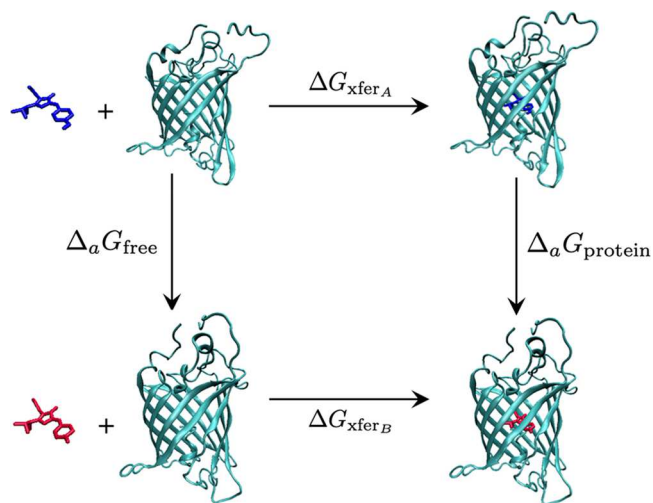


Figure 3. Free energy cycle used to calculate the electrostatic free energy differences between the A (blue) and B (red) states of the fluorophore. The GFP structure is shown in cyan.

$$\begin{aligned} \Delta G_{\text{xferA}} &= \frac{1}{2} \int \rho_A(\vec{r}) \Psi_A(\vec{r}) dV_A \\ \Delta G_{\text{xferB}} &= \frac{1}{2} \int \rho_B(\vec{r}) \Psi_B(\vec{r}) dV_B \end{aligned} \quad (5)$$

by integrating charge density $\rho(\vec{r})$ and electric potential $\Psi(\vec{r})$ over the total volume of the system. We used APBS version 1.4 to perform the free energy calculations^{64,69} by solving the linear Poisson–Boltzmann equation (eq 6):

$$\nabla \cdot \epsilon(\vec{r}) \nabla \Psi(\vec{r}) = \epsilon(\vec{r}) \bar{\kappa}^2 \Psi(\vec{r}) - 4\pi \rho(\vec{r}) \quad (6)$$

where $\epsilon(\vec{r})$ is dielectric constant as a function of position vector and was assigned 78.54 for implicit water¹⁹ and 6 for the protein based on previous work,¹⁹ $\Psi(\vec{r})$ is electric potential, $\bar{\kappa}^2$ is the ion accessibility coefficient, and $\rho(\vec{r})$ is the charge density with respect to the positions. For every atom in the protein, we included the atomic charge and radius explicitly with the corresponding values from either the Amber03 or AMOEBA force fields during the APBS free energy calculations. For water molecules, we included the partial charges and radii of water molecules with the corresponding values from the two force fields during the free energy calculations that were (1) separated from the fluorophore phenol oxygen by 5 Å, (2) separated from the fluorophore molecule by distances of 20 Å, and 35 Å, and (3) separated from the protein surface by 4, 8, and 12 Å. Any water not contained within these regions was treated as a continuum and assigned a dielectric constant 78.54. For the two sets of trajectories, the ions were not included in PB calculations due to negligible interactions with the low ion concentration.

RESULTS AND DISCUSSION

Comparing pK_a Values Generated from AMBER and AMOEBA. We compared the experimental pK_a values to those calculated from both Amber03ff¹⁹ and AMOEBA trajectories combined with APBS. While an accurate approximation of free energy calculations may be achieved by conducting APBS without every explicit water molecule for some compact globular proteins or the targeted residues located on the protein surface,^{52,70} it may not be true for the fluorophore in the β -barrel structure of GFP that contains several water

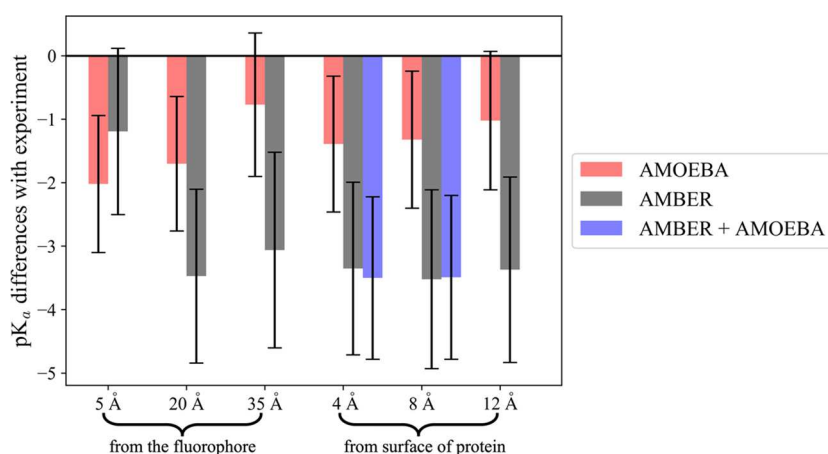


Figure 4. Average pK_a difference from experiment determined from simulations from AMOEBA (red), Amber03 (gray), and Amber03 trajectories analyzed with AMOEBA atomic charges and radii (blue) with different amounts of explicit water included in the APBS calculation. “5 Å” indicates all water within a radius of 5 Å from the fluorophore probe; “20 Å” indicates all water within a radius of 20 Å from the fluorophore; “35 Å” indicates all water within a radius of 35 Å from the fluorophore; “4 Å” indicates all water within a radius of 4 Å from the surface of the protein; “8 Å” indicates all water within a radius of 8 Å from the surface of the protein; “12 Å” indicates all water within a radius of 12 Å from the surface of the protein.

molecules that are very close to the titratable proton of the fluorophore, as shown in Figure 1.^{19,71} Therefore, we first calculated the pK_a values from both Amber03 and AMOEBA trajectories with APBS free energy calculations that included explicit water molecule within 5 Å of the fluorophore phenol oxygen. We calculated the average of the differences between the experimentally measured and calculated pK_a for WT and all mutants of interest for both Amber03 and AMOEBA trajectories; these results are shown in Figure 4. Although the major conformational change did not occur in both Amber03 and AMOEBA trajectories as we can see from the RMSD values in Figure S1 and S3, the calculated pK_a values showed different trends in the trajectories from the two force fields in Figure 4. PB calculations using trajectories generated from both Amber03 and AMOEBA force fields predicted lower pK_a values than the experimental results. The error in calculated pK_a when water molecules within 5 Å of the fluorophore phenol oxygen were included explicitly were larger for AMOEBA (-2.0 ± 1.1) than for Amber03 (-1.2 ± 1.3). We further investigated how the cutoff distance of explicitly included water molecules affect free energy (and thus pK_a) calculations by testing a variety of different treatments of explicit versus implicit water around the fluorophore in the APBS pK_a calculations for both AMOEBA and Amber03 trajectories by changing cutoff distance of explicit water molecules included in the pK_a calculations. The average differences between experimental pK_a values and results of these two force fields when considering different quantities of water are shown in Figure 4. First, we treated water molecules that were within 20 Å of fluorophore explicitly in APBS pK_a calculations. We again observed that both trajectories resulted in lower pK_a values than experimental results; however, the treatment of the two force fields resulted in very different predictions from the 5 Å water shell. In this case, the pK_a differences between the experimental results and the Amber03 predictions were larger (-3.5 ± 1.4) than the pK_a differences between experimental results and the AMOEBA trajectories (-1.7 ± 1.1) with the same explicit water cutoff. These data indicate that including more water molecules explicitly in APBS calculations would cause pK_a predictions from

AMOEBA trajectories to improve but Amber03 predictions to worsen and deviate further from experimental values. We next treated every water molecule that was within 35 Å of the fluorophore explicitly in the APBS calculation for both AMOEBA and Amber03 simulations. As shown in Figure 4, we observed that this further improved the agreement between experiment and AMOEBA-derived pK_a (-0.8 ± 1.1) but the agreement for Amber03 was essentially unaffected (-3.1 ± 1.5). We therefore observe a clear pattern in which including more explicit water decreases the accuracy of pK_a predictions in Amber03, but increases it in AMOEBA. The result from Amber03 in particular is surprising as one would expect that explicitly treating water molecules improve the physical representation and thus prediction of the pK_a shift.

To determine the amount of water needed to be treated explicitly to predict the pK_a values with the greatest accuracy and understand the reasons for these observations, we compared these results to a variety of selections for number and distance of water to be included in the electrostatics calculation for both Amber03 and AMOEBA trajectories. For example, water molecules that were within 35 Å of the fluorophore formed a thick hydration shell around the protein surface, shown in Figure 5A. We calculated the radial distribution function $g(r)$ between the fluorophore and water molecules for all mutants, including the protonated and deprotonated state. This is the probability of finding water molecules within the distances of the protonated and deprotonated fluorophores for all mutants; a representative result for the WT A state is shown in Figure 5B. Essentially, the radial distribution functions of the water molecules surrounding the protein (and thus the fluorophore) were identical with the two force fields out to a distance of >30 Å, and water molecules that were farther than 30 Å from the fluorophore were too far to have direct hydrogen bonding and other nonbonding interactions with the GFP surface residues. Thus, to understand how the water molecules interact with protein residues and how the interactions affect the pK_a values in Amber03 and AMOEBA trajectories, we calculated the pK_a values in both trajectories by including the water molecules that were within a set of different distances of the GFP surface.

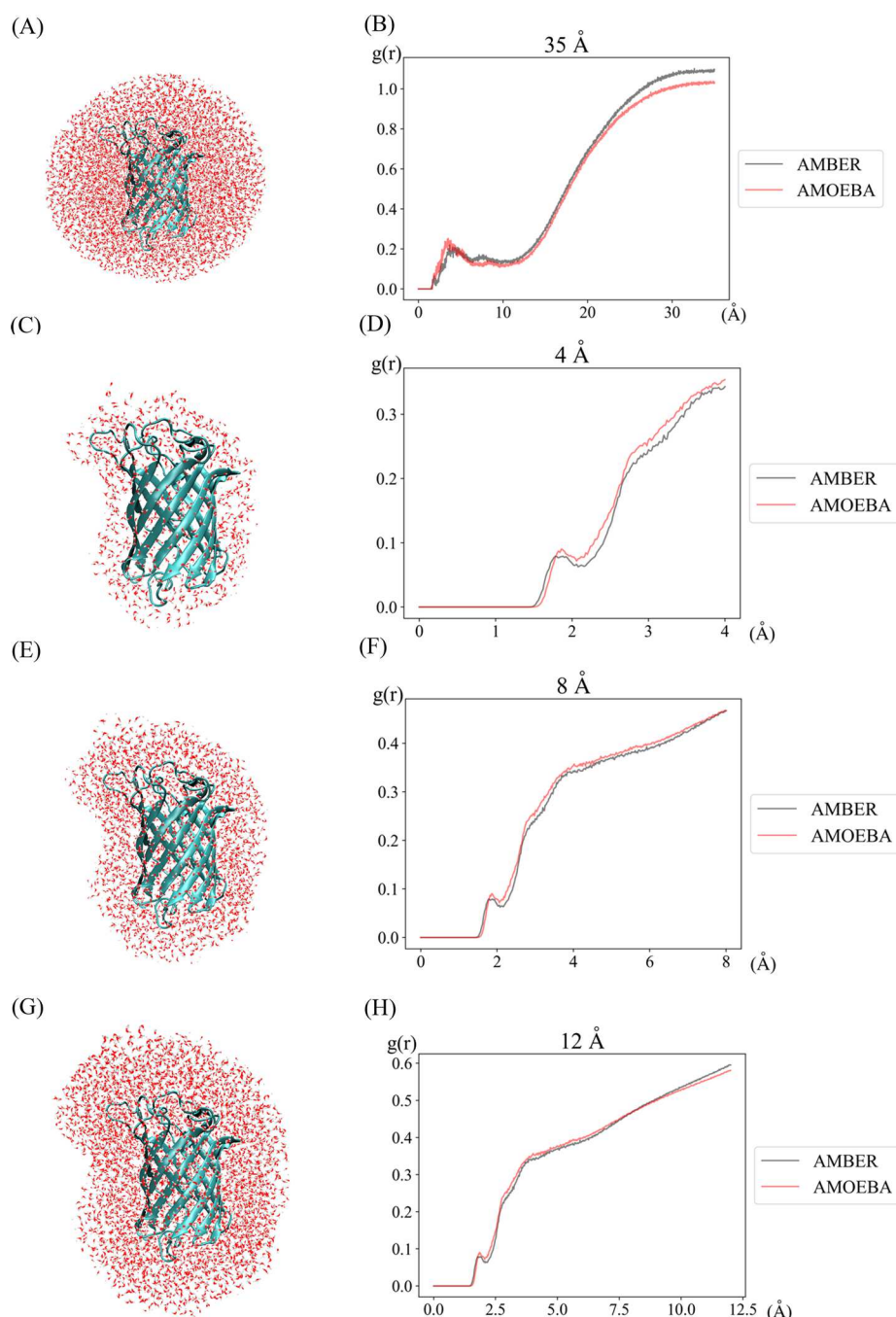


Figure 5. Representative structures of GFP (cyan) and water molecules (red and white) included in APBS calculations. Left column: water molecules that are within 35 Å of the fluorophore (A), or 4 Å (C), 8 Å (E), and 12 Å (G) from the surface of the protein. Right column: representative radial distribution function of the WT A state calculated from these structures with AMOEBA (red) and Amber03 (gray).

In a nonspherical protein like GFP, choosing water distances from the buried fluorophore might not accurately represent the full extent of water influencing all protein atoms in generating the overall electric field impinging on the fluorophore. We therefore also examined selecting water as a function of distance for the entire protein surface. In Figure 5C,D, we show the structure and radial distribution function, respectively, of water molecules within 4 Å of every surface residue of the GFP. In Figure 5D, the peak at 2 Å and a shoulder peak at around 3 Å indicate that this strategy includes slightly more than two hydrating monolayers, significantly less than occurs when counting every water molecule within 35 Å of

fluorophore. In this case, pK_a values predicted from AMOEBA trajectories (-1.4 ± 1.1) were significantly closer to the experimental results than the Amber03 predictions (-3.4 ± 1.4), shown in Figure 4. To explore this further, we added additional water in increasing distances away from the protein surface, 8 Å (Figure 5E,F) and 12 Å (Figure 5G,H). This resulted in 4 and 5, respectively, layers of water molecules surrounding the entire protein, and pK_a predictions that were more accurate from AMOEBA trajectories (-1.3 ± 1.1 and -1.0 ± 1.1 , respectively) than that of Amber03 (-3.5 ± 1.4 and -3.4 ± 1.5 , respectively). Both the radial distribution function from Figure 5H and the pK_a values indicate that the

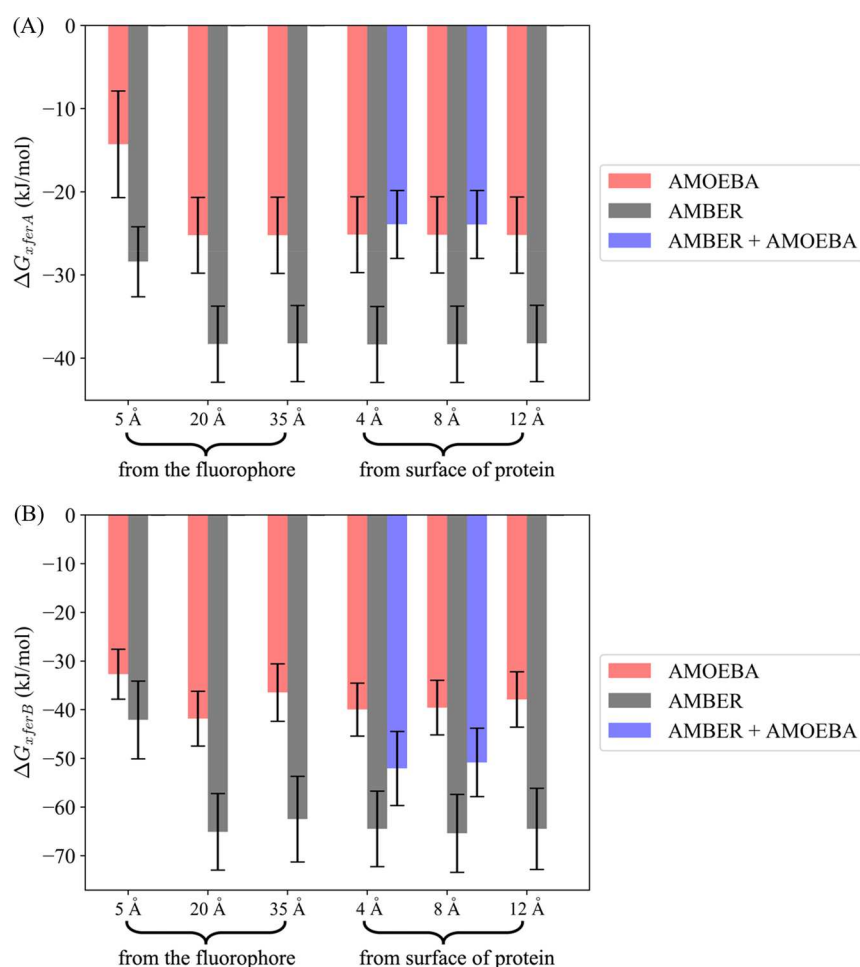


Figure 6. Water molecules included in APBS calculation with respect to the average free energy of the A (A) and B (B) states calculated through AMOEBA (red) or Amber03 (gray) trajectories and Amber03 trajectories with AMOEBA atomic charges and radii (blue). “5 Å” indicates all water within a radius of 5 Å from the fluorophore probe; “20 Å” indicates all water within a radius of 20 Å from the fluorophore; “35 Å” indicates all water within a radius of 35 Å from the fluorophore; “4 Å” indicates all water within a radius of 4 Å from the surface of the protein; “8 Å” indicates all water within a radius of 8 Å from the surface of the protein; “12 Å” indicates all water within a radius of 12 Å from the surface of the protein.

water molecules within 12 Å away from the protein surface may be very close to, or already at the limit of bulk water. The water molecules that were farther than 12 Å from every residue were unlikely to interact with any residue either inside and outside of the β -barrel GFP, so the water molecules more likely performed as bulk water molecules. We conclude that the water molecules farther than 12 Å away from the protein surface could be considered as continuum solvent in the fluorophore pK_a calculations. In summary, we found that either (1) including larger radii of water away from the GFP fluorophore explicitly or (2) including multiple water layers (smaller radii) of every protein surface residue explicitly make the predictions from AMOEBA trajectories more accurate than the predictions from Amber03 trajectories. The only scenario in which the AMOEBA prediction was less accurate than the Amber03 prediction occurred when we only included 5 Å of water molecules within the fluorophore phenol oxygen.

Comparing Free Energy Predictions between Force Fields. To account for the reason for the decrease in accuracy of predicted pK_a values from Amber03 trajectories when more water molecules were treated explicitly in the APBS calculation, we investigated the average solvation free energy differences in both A and B states (ΔG_{xferA} and ΔG_{xferB} , respectively) for both force fields. These results are shown in

Figure 6. Because the pK_a values were calculated from the differences between ΔG_{xferA} and ΔG_{xferB} , examining these two terms separately is needed to disentangle the contributions of the decrease in accuracy of Amber03 predicted pK_a values as more water was included explicitly. The value of ΔG_{xferA} in Amber03 trajectories that included water molecules within 5 Å of fluorophore pK_a probe was -28.4 ± 4.2 kJ mol $^{-1}$, which was approximately 10 kJ mol $^{-1}$ less negative than the ΔG_{xferA} with more water being treated explicitly; all other water inclusion strategies resulted in ΔG_{xferA} of approximately -38 kJ mol $^{-1}$ (see Figure 6A). A similar trend was found in AMOEBA protonated state trajectories; the value of ΔG_{xferA} in AMOEBA trajectories that includes water molecules within 5 Å of fluorophore pK_a probe was -14.3 ± 6.4 kJ mol $^{-1}$, which was approximately 11 kJ mol $^{-1}$ less negative than the ΔG_{xferA} with more water being treated explicitly (approximately -25 kJ mol $^{-1}$ for every other water inclusion strategy, see Figure 6A). We did the same comparison for ΔG_{xferB} in the two force fields, shown in Figure 6B. The value of ΔG_{xferB} in Amber03 trajectories that included water molecules within 5 Å of the pK_a probe was -42.1 ± 8.0 kJ mol $^{-1}$, which was around 23 kJ mol $^{-1}$ more negative than the ΔG_{xferB} with more water being treated explicitly (approximately -65 kJ mol $^{-1}$ for every other water inclusion strategy; see Figure 6B). However, for

AMOEBA trajectories of the deprotonated state, the differences between $\Delta G_{\text{xfer B}}$ with water molecules within 5 Å of the fluorophore probe included and the $\Delta G_{\text{xfer B}}$ with more water molecules included was much smaller, around 6.6 kJ mol⁻¹, compared to that of Amber03, 23 kJ mol⁻¹. Therefore, the reason for the lower accuracy in pK_a predictions from Amber03 trajectories when more water molecules were treated explicitly in APBS calculations was because Amber03 created structures which overpredicted the stabilization of the deprotonated (B) state, leading to more negative values of $\Delta G_{\text{xfer B}}$ in Amber03 than in AMOEBA. Because pK_a is a calculation of the difference between $\Delta G_{\text{xfer A}}$ and $\Delta G_{\text{xfer B}}$ (eq 4), this will lead directly to incorrectly large values of pK_a for Amber03 but not AMOEBA.

One possible explanation for this observation is that in a fixed charge force field like Amber03, Coulombic forces cause water molecules that are farther away from the negatively charged B state to stabilize significantly more than the polarizable multipole-based AMOEBA model. The electrostatic interactions in the two models exhibit very different distance dependence since higher multipole moments and polarization decay faster than monopoles. In addition, many body polarization provides electronic dielectric screening that is missing in fixed charge models. Such dielectric screening was shown essential for protein-ion recognitions.^{72,73} The fixed charge based electrostatic interactions in Amber03 over-stabilizes the negative charge on the deprotonated fluorophore, when individual water molecules which are farther away from the B state fluorophore were included. On the other hand, because AMOEBA considers dipole, quadrupole, and induced dipole moments of each atom, the balance of short and long-range Coulombic forces of the explicit water is more realistic. The water molecules that are farther away from the fluorophore in AMOEBA trajectories can be affected by both the interactions with other water molecules and protein residues and less of that from the net charge on the fluorophore, including the negative charge on the B state. Therefore, with farther water molecules being assigned partial charges and radii in APBS free energy calculations, the longer-range Coulombic forces from these explicit water molecules in the B state of Amber03 trajectories would over stabilize the electrostatic free energy, which in turn causes the two force fields to have the opposite trends of the predictions in the pK_a values when more water is included explicitly. Previous work comparing calculated pK_a values between fixed charged and polarizable force field MD trajectories found that the pK_a values of some acidic residues were lower in the fixed charge treatment than in the polarizable Drude force field treatment when the same amount of water was included explicitly in PB-based free energy calculations.⁴⁰ This suggests that including polarizability explicitly in MD simulations for protein heterogeneity is crucial for the free energy calculations.⁴⁰ The work reported here confirms this and highlights that this effect is especially important for the charged state (which in this case is deprotonated). AMOEBA simulations of protein–ligand binding have also suggested that ignoring polarization overpredicts protein–ligand and protein-ion electrostatic attractions, a similar effect to what we observe here.^{73,74}

Applying AMBER Trajectories with AMOEBA Atomic Charges and Radii in APBS Calculations. The results in Figure 4 suggest that using AMOEBA and more water molecules leads not only to substantially greater accuracy in pK_a prediction but also unfortunately to increasing computa-

tional expense. We therefore tested the utility of a shortcut: determining the structure with Amber03, and pK_a values with atomic charges and radii from AMOEBA. To do this, we took Amber03 trajectories of our WT and mutant GFPs containing either water molecules 4 or 8 Å away from every residue of the GFP (Figure 5C,E), then implemented partial charges and radii from AMOEBA for every atom for every explicit atom included in the APBS calculation. Results are shown in Figure 4 (blue column). The average difference in the pK_a values between the calculated and the experimental results were both -3.5 ± 1.3 for 4 and 8 Å of water, respectively, essentially identical to the results from using Amber03 for both MD and PB calculations.

We then examined the average solvation free energy differences in the protonated and deprotonated states ($\Delta G_{\text{xfer B}}$ and $\Delta G_{\text{xfer A}}$) for the Amber03 trajectories with AMOEBA atomic charges and radii being applied. The results are shown in Figure 6. For $\Delta G_{\text{xfer A}}$, we found that applying AMOEBA atomic charges and radii to Amber03 trajectories dramatically reduced the free energy stabilization of the protonated A state, resulting in values of $\Delta G_{\text{xfer A}}$ that were similar to the trajectories generated with AMOEBA alone. However, for $\Delta G_{\text{xfer B}}$, although applying the AMOEBA atomic charges and radii to Amber03 trajectories did decrease the free energy of stabilization somewhat, it still overestimated the stabilization of the B state by the included explicit water molecules. This uneven shift in the electrostatic free energy between the A and B states resulted in similar pK_a predictions to a treatment entirely with Amber03. It therefore appears that subtle structural differences caused by the fixed charge versus polarizable force fields, both of which lead to correct and stable structures based on the RMSD results in Figure S3 and S1 for this model protein and the amount of surrounding water molecules from the radial distribution function in Figure 5B,D,F,H, still cause significant errors in electrostatic parameters such as pK_a .

In this work, we found that both Amber03 and AMOEBA trajectories generated pK_a predictions that were lower than the experimental results. This contrasts with other studies of pK_a calculations that found that there was no consistency in whether predicted pK_a values would be higher^{20,39,40,75} or lower^{39,40} than the experimental results. For example, previous work has found that calculated pK_a values are usually higher than the experimental results for basic residues such as cysteine⁷⁵ or lysine.^{20,40} However, other investigations of acidic residues such as aspartate and glutamate have resulted in predicted pK_a values that were higher than the experimental results when using a fixed charged force field, but lower than experimental results using a polarizable force field.³⁹ Others have reported that both fixed charge and polarizable trajectories generated lower pK_a values than the experimental results.⁴⁰ In our approach, both Amber03 and AMOEBA trajectories calculated average pK_a values less than the experimental values, overpredicting the tendency of the fluorophore to act as an acid. The lower predictions in pK_a values for both trajectories may be due to the observation that the APBS free energy calculations overstabilize the deprotonated state of the fluorophore. We hypothesized because the APBS electrostatic free energy calculations only considered charge–charge interactions, the stabilization of the negatively charged deprotonated fluorophore was more significant than the neutral protonated fluorophore. The other possible reason for the underestimation of pK_a values from both Amber03 and

AMOEBA trajectories could be due to the systematic error of the electric fields near the charged residues. Previous work on electric field calculations from both fixed charge general AMBER force field (GAFF)^{76,77} and AMOEBA polarizable force fields trajectories compared with the vibrational Stark effect (VSE) spectroscopy experiments reported that both trajectories could overestimate the electric field in the vicinity of a charged group in penicillin G ranging from 25 to 30 MV cm⁻¹ depending on the conformation of the molecule.⁷⁸ Although the measure of electric field between this report and the results presented here (VSE spectroscopy versus pK_a measurements) are totally different, based on the conversion of 1 MV cm⁻¹ $\cong$ 0.6 kJ mol⁻¹, the results of these two works are similar both in the direction of the error and the order of magnitude. The structures generated by the electric fields around the charged deprotonated fluorophore in both trajectories may also cause unrealistically low solvation free energy results from APBS.

The other major finding in this work is that including more explicit water molecules in APBS calculations improves predicted pK_a values from AMOEBA trajectories but worsens those from Amber03. Previous investigators have compared pK_a values between structures from fixed charge and other polarizable force fields, but often only included the explicit water molecules around the target residue, which does not reveal the role that explicit water molecules play in free energy calculations.^{39,40} More accurate predictions from AMOEBA-derived trajectories were found as more explicit water molecules were included in the APBS calculation. A possible reason is that the explicit polarizability between the protein and water interface is needed to accurately reflect electrostatic fields generated by such complex many-body interactions. This has been previously suggested when comparing AMOEBA and fixed charge models for calculating electric fields of organic solute in both hydrophobic and hydrophilic solvents, which indicated that explicit polarizability is most critically needed at the interface between a polar and a nonpolar molecule.⁷⁹ Because the surface of a protein is the site of many molecular interactions of varying amino acid side chain hydrophobicity and water, more accurate inclusion of the polarizability at the interface of the surface of the protein and water molecules from AMOEBA could be beneficial. Conversely, because Amber03 trajectories do not consider permanent dipole, quadrupole, and polarizability for positions of all atoms, including water molecules around surface residues would therefore not be predicted to improve pK_a calculations when compared to experiment. Therefore, the AMOEBA MD trajectories that considered many-body polarization could better account for the interactions between protein residues and water molecules and thus generate the more accurate free energy predictions when water molecules were treated explicitly. Moreover, by applying the AMOEBA atomic charges and radii to Amber03 trajectories in APBS calculations, accuracies from the computationally cheaper Amber03 fixed charged model cannot be improved. This result strengthens our conclusion that including the permanent dipole, permanent quadrupole, and polarizability throughout the MD simulations improves the prediction of the fluorophore pK_a in GFP mutants. Work is ongoing in our laboratory to investigate whether this applies only to this unusual protein (with its hollow water-containing interior) or is generalizable to other proteins of interest. Overall, we have demonstrated a viable method for calculating protein electrostatic fields that

strikes a balance between the computational time and the considerations of the relevant physical details.

CONCLUSIONS

Here we have compared two common computational strategies for generating biomolecular structures for pK_a calculations directly against a robust experimental data set in an effort to understand the importance of including polarizability in accurate predictions of protein electric field through pK_a. In particular, we have shown that the amount of explicit water molecules included in APBS calculations do affect the pK_a values of the fluorophore in GFP for both AMOEBA and Amber03 trajectories. We found that although the major conformational change did not occur in both Amber03 and AMOEBA trajectories, the calculated pK_a values showed the opposite trends from the two sets of trajectories. Including more water molecules in APBS calculations decreases the accuracy of pK_a values of the fluorophore in GFP in the ensemble of structures predicted by Amber03 fixed charge force field; conversely, pK_a predictions increase in accuracy when ensembles of the same structures are assembled using the polarizable AMOEBA force field. We propose that because long-range Coulombic interactions dominate in the intermolecular interactions in Amber03 trajectories, there is an obvious increase in the error of the free energy stabilization of the deprotonated B state in Amber03 trajectories compared to those in AMOEBA when more water is treated explicitly in APBS calculations. The larger negative value of the B state free energy in Amber03 trajectories results in the lower pK_a values with respect to experimental results than AMOEBA predictions. These results provide a useful methodology to investigate other pK_a or free energy related protein electrostatic calculations for applications across biophysical chemistry and are being expanded to other proteins in our laboratory.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcb.2c03642>.

RMSD values of AMOEBA trajectories for the protonated and deprotonated fluorophore and the GFP; RMSD values of AMBER trajectories for the protonated and deprotonated fluorophore and the GFP; AMOEBA force field parameter derivations of protonated and deprotonated fluorophore (PDF)

AUTHOR INFORMATION

Corresponding Author

Lauren J. Webb — Department of Chemistry and Texas Materials Institute, and Interdisciplinary Life Sciences Program, The University of Texas at Austin, Austin, Texas 78712-1224, United States; orcid.org/0000-0001-9999-5500; Email: lwebb@cm.utexas.edu

Authors

Yu-Chun Lin — Department of Chemistry, The University of Texas at Austin, Austin, Texas 78712-1224, United States; orcid.org/0000-0002-4839-8551

Pengyu Ren — Department of Biomedical Engineering, The University of Texas at Austin, Austin, Texas 78712, United States

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jpcb.2c03642>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by NSF (grant number MCB-1714555) and The Welch Foundation (grant number F-1722). P.R. acknowledges support by NIH (R01GM114237). We are also grateful to the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing high performance computing resources for all the APBS related calculations.

REFERENCES

- (1) Hayes, D. M.; Kollman, P. A. Role of Electrostatics in a Possible Catalytic Mechanism for Carboxypeptidase A. *J. Am. Chem. Soc.* **1976**, *98*, 7811–7816.
- (2) Warshel, A. Energetics of Enzyme Catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **1978**, *75*, S250–S254.
- (3) Honig, B.; Nicholls, A. Classical Electrostatics in Biology and Chemistry. *Science*. **1995**, *268*, 1144–1149.
- (4) Gunner, M. R.; Nicholls, A.; Honig, B. Electrostatic Potentials in Rhodospseudomonas Viridis Reaction Centers: Implications for the Driving Force and Directionality of Electron Transfer. *J. Phys. Chem.* **1996**, *100*, 4277–4291.
- (5) Warshel, A.; Papazyan, A. Electrostatic Effects in Macromolecules: Fundamental Concepts and Practical Modeling. *Curr. Opin. Struct. Biol.* **1998**, *8*, 211–217.
- (6) Romei, M. G.; Lin, C.; Mathews, I. I.; Boxer, S. G. Electrostatic Control of Photoisomerization Pathways in Proteins. *Science*. **2020**, *367*, 76–79.
- (7) Schreiber, G.; Fersht, A. R. Rapid, Electrostatically Assisted Association of Proteins. *J. Dairy Res.* **1979**, *3* (5), 427–431.
- (8) Li, L.; Alper, J.; Alexov, E. Cytoplasmic Dynein Binding, Run Length, and Velocity Are Guided by Long-Range Electrostatic Interactions. *Sci. Rep.* **2016**, *6*, 31523.
- (9) Fadrná, E.; Hladečková, K.; Koča, J. Long-Range Electrostatic Interactions in Molecular Dynamics: An Endothelin-1 Case Study. *J. Biomol. Struct. Dyn.* **2005**, *23*, 151–162.
- (10) Hosseinzadeh, P.; Marshall, N. M.; Chacón, K. N.; Yu, Y.; Nilges, M. J.; New, S. Y.; Tashkov, S. A.; Blackburn, N. J.; Lu, Y. Design of a Single Protein That Spans the Entire 2-V Range of Physiological Redox Potentials. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 262–267.
- (11) Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; et al. Biocatalytic Asymmetric Synthesis of Sitagliptin Manufacture. *Science*. **2010**, *329*, 305–310.
- (12) Jiménez-Osés, G.; Osuna, S.; Gao, X.; Sawaya, M. R.; Gilson, L.; Collier, S. J.; Huisman, G. W.; Yeates, T. O.; Tang, Y.; Houk, K. N. The Role of Distant Mutations and Allosteric Regulation on LovD Active Site Dynamics. *Nat. Chem. Biol.* **2014**, *10*, 431–436.
- (13) Osuna, S.; Jiménez-Osés, G.; Noey, E. L.; Houk, K. N. Molecular Dynamics Explorations of Active Site Structure in Designed and Evolved Enzymes. *Acc. Chem. Res.* **2015**, *48*, 1080–1089.
- (14) Markley, J. L. Observation of Histidine Residues in Proteins by Means of Nuclear Magnetic Resonance Spectroscopy. *Acc. Chem. Res.* **1975**, *8*, 70–80.
- (15) Bradbury, J. H.; Scheraga, H. A. Structural Studies of Ribonuclease. XXIV. The Application of Nuclear Magnetic Resonance Spectroscopy to Distinguish between the Histidine Residues of Ribonuclease. *J. Am. Chem. Soc.* **1966**, *88*, 4240–4246.
- (16) Yu, B.; Pletka, C. C.; Montgomery Pettitt, B.; Iwahara, J. De Novo Determination of Near-Surface Electrostatic Potentials by NMR. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118* (25), e2104020118.
- (17) First, J. T.; Slocum, J. D.; Webb, L. J. Quantifying the Effects of Hydrogen Bonding on Nitrile Frequencies in GFP: Beyond Solvent Exposure. *J. Phys. Chem. B* **2018**, *122*, 6733–6743.
- (18) Blasiak, B.; Ritchie, A. W.; Webb, L. J.; Cho, M. Vibrational Solvatochromism of Nitrile Infrared Probes: Beyond the Vibrational Stark Dipole Approach. *Phys. Chem. Chem. Phys.* **2016**, *18*, 18094–18111.
- (19) Slocum, J. D.; First, J. T.; Webb, L. J. Orthogonal Electric Field Measurements near the Green Fluorescent Protein Fluorophore through Stark Effect Spectroscopy and pK_a Shifts Provide a Unique Benchmark for Electrostatics Models. *J. Phys. Chem. B* **2017**, *121*, 6799–6812.
- (20) Isom, D. G.; Castañeda, C. A.; Cannon, B. R.; García-Moreno, B. E. Large Shifts in pK_a Values of Lysine Residues Buried inside a Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (13), S260–S265.
- (21) Langsetmo, K.; Fuchs, J. A.; Woodward, C. The Conserved, Buried Aspartic Acid in Oxidized Escherichia Coli Thioredoxin Has a pK_a of 7.5. Its Titration Produces a Related Shift in Global Stability. *Biochemistry* **1991**, *30*, 7603–7609.
- (22) Antosiewicz, J.; Mccammon, J. A.; Gilson, M. K. The Determinants of pK_as in Proteins. *Biochemistry* **1996**, *35*, 7819–7833.
- (23) Nielsen, J. E.; Gunner, M. R.; García-Moreno, E. B. The pK_a Cooperative: A Collaborative Effort to Advance Structure-Based Calculations of pK_a Values and Electrostatic Effects in Proteins. *Proteins Struct. Funct. Bioinforma.* **2011**, *79*, 3249–3259.
- (24) Forsyth, W. R.; Antosiewicz, J. M.; Robertson, A. D. Empirical Relationships between Protein Structure and Carboxyl pK_a Values in Proteins. *Proteins Struct. Funct. Genet.* **2002**, *48*, 388–403.
- (25) Gibas, C. J.; Subramaniam, S. Explicit Solvent Models in Protein pK_a Calculations. *Biophys. J.* **1996**, *71*, 138–147.
- (26) Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. A Summary of the Measured pK_a Values of the Ionizable Groups in Folded Proteins. *Protein Sci.* **2009**, *18* (1), 247–251.
- (27) Jensen, J. H.; Li, H.; Robertson, A. D.; Molina, P. A. Prediction and Rationalization of Protein pK_a Values Using QM and QM/MM Methods. *J. Phys. Chem. A* **2005**, *109*, 6634–6643.
- (28) Riccardi, D.; Schaefer, P.; Cui, Q. pK_a Calculations in Solution and Proteins with QM/MM Free Energy Perturbation Simulations: A Quantitative Test of QM/MM Protocols. *J. Phys. Chem. B* **2005**, *109*, 17715–17733.
- (29) Schutz, C. N.; Warshel, A. What Are the Dielectric “Constants” of Proteins and How to Validate Electrostatic Models? *Proteins Struct. Funct. Genet.* **2001**, *44*, 400–417.
- (30) Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. Modeling Electrostatic Effects in Proteins. *Biochim. Biophys. Acta, Proteins Proteomics* **2006**, *1764*, 1647–1676.
- (31) Li, H.; Robertson, A. D.; Jensen, J. H. Very Fast Empirical Prediction and Rationalization of Protein pK_a Values. *Proteins Struct. Funct. Genet.* **2005**, *61*, 704–721.
- (32) Olsson, M. H. M.; SØndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK_a Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.
- (33) Reis, P. B. P. S.; Vila-Vicosa, D.; Rocchia, W.; Machuqueiro, M. PypKa: A Flexible Python Module for Poisson-Boltzmann-Based pK_a Calculations. *J. Chem. Inf. Model.* **2020**, *60*, 4442–4448.
- (34) Khaniya, U.; Mao, J.; Wei, R. J.; Gunner, M. R. Characterizing Protein Protonation Microstates Using Monte Carlo Sampling. *J. Phys. Chem. B* **2022**, *126*, 2476–2485.
- (35) Alexov, E.; Mehler, E. L.; Baker, N.; Baptista, A. M.; Huang, Y.; Milletti, F.; Erik Nielsen, J.; Farrell, D.; Carstensen, T.; Olsson, M. H. M.; Shen, J. K.; Warwicker, J.; Williams, S.; Word, M. J. Progress in the Prediction of pK_a Values in Proteins. *Proteins Struct. Funct. Genet.* **2011**, *79*, 3260–3275.
- (36) Wallace, J. A.; Wang, Y.; Shi, C.; Pastoor, K. J.; Nguyen, B. L.; Xia, K.; Shen, J. K. Toward Accurate Prediction of pK_a Values for Internal Protein Residues: The Importance of Conformational Relaxation and Desolvation Energy. *Proteins Struct. Funct. Bioinforma.* **2011**, *79*, 3364–3373.

- (37) Huang, Y.; Chen, W.; Wallace, J. A.; Shen, J. All-Atom Continuous Constant pH Molecular Dynamics with Particle Mesh Ewald and Titratable Water. *J. Chem. Theory Comput.* **2016**, *12*, 5411–5421.
- (38) Warshel, A.; Dryga, A. Simulating Electrostatic Energies in Proteins: Perspectives and Some Recent Studies of pK_a s, Redox, and Other Crucial Functional Properties. *Proteins Struct. Funct. Bioinforma.* **2011**, *79*, 3469–3484.
- (39) MacDermaid, C. M.; Kaminski, G. A. Electrostatic Polarization Is Crucial for Reproducing pK_a Shifts of Carboxylic Residues in Turkey Ovomucoid Third Domain. *J. Phys. Chem. B* **2007**, *111*, 9036–9044.
- (40) Aleksandrov, A.; Roux, B.; Mackerell, A. D. pK_a Calculations with the Polarizable Drude Force Field and Poisson-Boltzmann Solvation Model. *J. Chem. Theory Comput.* **2020**, *16*, 4655–4668.
- (41) Fogolari, F.; Brigo, A.; Molinari, H. The Poisson-Boltzmann Equation for Biomolecular Electrostatics: A Tool for Structural Biology. *J. Mol. Recognit.* **2002**, *15*, 377–392.
- (42) Meyer, T.; Knapp, E. W. pK_a Values in Proteins Determined by Electrostatics Applied to Molecular Dynamics Trajectories. *J. Chem. Theory Comput.* **2015**, *11*, 2827–2840.
- (43) Li, L.; Li, C.; Zhang, Z.; Alexov, E. On the Dielectric “Constant” of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi. *J. Chem. Theory Comput.* **2013**, *9*, 2126–2136.
- (44) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; Distasio, R. A.; et al. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.
- (45) Reuter, N.; Lin, H.; Thiel, W. Green Fluorescent Proteins: Empirical Force Field for the Neutral and Deprotonated Forms of the Chromophore. Molecular Dynamics Simulations of the Wild Type and S65T Mutant. *J. Phys. Chem. B* **2002**, *106*, 6310–6321.
- (46) Nifosi, R.; Tozzini, V. Molecular Dynamics Simulations of Enhanced Green Fluorescent Proteins: Effects of F64L, S65T and T203Y Mutations on the Ground-State Proton Equilibria. *Proteins Struct. Funct. Genet.* **2003**, *51*, 378–389.
- (47) First, J. T.; Webb, L. J. Agreement between Experimental and Simulated Circular Dichroic Spectra of a Positively Charged Peptide in Aqueous Solution and on Self-Assembled Monolayers. *J. Phys. Chem. B* **2019**, *123*, 4512–4526.
- (48) Collier, G.; Vellere, N. A.; Yancey, J. A.; Stuart, S. J.; Latour, R. A. Comparison between Empirical Protein Force Fields for the Simulation of the Adsorption Behavior of Structured LK Peptides on Functionalized Surfaces. *Biointerphases* **2012**, *7*, 24.
- (49) Click, T. H.; Kaminski, G. A. Reproducing Basic pK_a Values for Turkey Ovomucoid Third Domain Using a Polarizable Force Field. *J. Phys. Chem. B* **2009**, *113*, 7844–7850.
- (50) Timothy, H. C.; Sergei, Y.; Ponomarev, G. A. K. Importance of Electrostatic Polarizability in Calculating Cysteine Acidity Constants and Copper(I) Binding Energy of *Bacillus Subtilis* Cop. *J. Comput. Chem.* **2012**, *33* (11), 1142–1151.
- (51) Pédelacq, J. D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. Engineering and Characterization of a Superfolder Green Fluorescent Protein. *Nat. Biotechnol.* **2006**, *24*, 79–88.
- (52) Simonson, T.; Brooks, C. L. Charge Screening and the Dielectric Constant of Proteins: Insights from Molecular Dynamics. *J. Am. Chem. Soc.* **1996**, *118*, 8452–8458.
- (53) Wu, J. C.; Chatterjee, G.; Ren, P. Automation of AMOEBA Polarizable Force Field Parameterization for Small Molecules. *Theor. Chem. Acc.* **2012**, *131*, 1138.
- (54) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09, Revision A.1*, Revision A.1; Gaussian, Inc.: Wallingford, CT, 2009.
- (55) Elsliger, M. A.; Wachter, R. M.; Hanson, G. T.; Kallio, K.; Remington, S. J. Structural and Spectral Response of Green Fluorescent Protein Variants to Changes in pH. *Biochemistry* **1999**, *38*, 5296–5301.
- (56) Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J. P.; Ren, P.; Ponder, J. W. Tinker 8: Software Tools for Molecular Design. *J. Chem. Theory Comput.* **2018**, *14*, 5273–5289.
- (57) Martínez, L.; Andrade, R.; Birgin, E. G. M. J. M. Packmol: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *J. Comput. Chem.* **2009**, *30*, 2157–2164.
- (58) Nifosi, R.; Tozzini, V. Molecular Dynamics Simulations of Enhanced Green Fluorescent Proteins: Effects of F64L, S65T and T203Y Mutations on the Ground-State Proton Equilibria. *Proteins Struct., Funct., Genet.* **2003**, *51*, 378–389.
- (59) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
- (60) Sorin, E. J.; Pande, V. S. Exploring the Helix-Coil Transition via All-Atom Equilibrium Ensemble Simulations. *Biophys. J.* **2005**, *88*, 2472–2493.
- (61) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
- (62) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.
- (63) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 4046–4063.
- (64) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. PDB2PQR: Expanding and Upgrading Automated Preparation of Biomolecular Structures for Molecular Simulations. *Nucleic Acids Res.* **2007**, *35*, 522–525.
- (65) Gorham, R. D., Jr.; Kieslich, C. A.; Nichols, A.; Sausman, N. U.; Foronda, M.; Morikis, D. An Evaluation of Poisson-Boltzmann Electrostatic Free Energy Calculations through Comparison with Experimental Mutagenesis Data. *Biopolymers* **2011**, *95* (11), 746–754.
- (66) Baker, N. A. Improving Implicit Solvent Simulations: A Poisson-Centric View. *Curr. Opin. Struct. Biol.* **2005**, *15*, 137–143.
- (67) Baker, N. A. Biomolecular Applications of Poisson-Boltzmann Methods. *Rev. Comput. Chem.* **2005**, *21*, 349–379.
- (68) Wagoner, J.; Baker, N. A. Solvation Forces on Biomolecular Structures: A Comparison of Explicit Solvent and Poisson-Boltzmann Models. *J. Comput. Chem.* **2004**, *25*, 1623–1629.
- (69) Unni, S.; Huang, Y.; Hanson, R.; Tobias, M.; Krishnan, S.; Li, W.; Nielsen, J. E.; Baker, N. A. W. Web Servers and Services for Electrostatics Calculations with APBS and PDB2PQR. *J. Comput. Chem.* **2011**, *32*, 1488–1491.
- (70) Kukic, P.; Farrell, D.; McIntosh, L. P.; García-Moreno, E. B.; Jensen, K. S.; Toleikis, Z.; Teilum, K.; Nielsen, J. E. Protein Dielectric Constants Determined from NMR Chemical Shift Perturbations. *J. Am. Chem. Soc.* **2013**, *135* (45), 16968–16976.
- (71) Lin, J. H.; Baker, N. A.; Andrew McCammon, J. Bridging Implicit and Explicit Solvent Approaches for Membrane Electrostatics. *Biophys. J.* **2002**, *83* (3), 1374–1379.
- (72) Jing, Z.; Rackers, J. A.; Pratt, L. R.; Liu, C.; Rempe, S. B.; Ren, P. Thermodynamics of Ion Binding and Occupancy in Potassium Channels. *Chem. Sci.* **2021**, *12*, 8920–8930.
- (73) Jing, Z.; Liu, C.; Qi, R.; Ren, P. Many-Body Effect Determines the Selectivity for Ca^{2+} and Mg^{2+} in Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (32), 7495–7501.
- (74) Jiao, D.; Golubkov, P. A.; Darden, T. A.; Ren, P. Calculation of Protein-Ligand Binding Free Energy by Using a Polarizable Potential. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 6290–6295.
- (75) Awoonor-Williams, E.; Rowley, C. N. Evaluation of Methods for the Calculation of the pK_a of Cysteine Residues in Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 4662–4673.

- (76) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (77) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247–260.
- (78) Kozuch, J.; Schneider, S. H.; Zheng, C.; Ji, Z.; Bradshaw, R. T.; Boxer, S. G. Testing the Limitations of MD-Based Local Electric Fields Using the Vibrational Stark Effect in Solution: Penicillin G as a Test Case. *J. Phys. Chem. B* **2021**, *125*, 4415–4427.
- (79) Fried, S. D.; Wang, L. P.; Boxer, S. G.; Ren, P.; Pande, V. S. Calculations of the Electric Fields in Liquid Solutions. *J. Phys. Chem. B* **2013**, *117*, 16236–16248.