Assessing the conformational equilibrium of carboxylic acid via QM and MD studies on acetic acid

Victoria T. Lim,[1] Christopher I. Bayly,[2] Laszlo Fusti-Molnar,[3] and David L. Mobley[1,4,a)]
[1)] *Department of Chemistry, University of California, Irvine*
[2)] *OpenEye Scientific Software*
[3)] *QuantumFuture Scientific Software LLC*
[4)] *Department of Pharmaceutical Sciences, University of California, Irvine*

(Dated: November 21, 2018)

Accurate hydrogen placement in molecular modeling is crucial for studying the interactions and dynamics of biomolecular systems. It is difficult to locate hydrogen atoms from many experimental structural characterization approaches, such as due to the weak scattering of x-ray radiation. Hydrogen atoms are usually added and positioned *in silico* when preparing experimental structures for modeling and simulation. The carboxyl functional group is a prototypical example of a functional group that requires protonation during structure preparation. To our knowledge, when in their neutral form, carboxylic acids are typically protonated in the *syn* conformation by default in classical molecular modeling packages, with no consideration of alternative conformations, though we are not aware of any careful examination of this topic. Here, we investigate the general belief that carboxylic acids should always be protonated in the *syn* conformation. We calculate and compare the relative energetic stabilities of *syn* and *anti* acetic acid using *ab initio* quantum mechanical calculations and atomistic molecular dynamics simulations. We show that while the *syn* conformation is the preferred state, the *anti* state may in some cases also be present under normal NPT conditions in solution.

I. INTRODUCTION

The carboxyl functional group, –COOH, is widespread in nature and highly biochemically relevant. It is present in amino acids that compose proteins, fatty acids of cell membranes, and naturally occurring organic compounds (*e.g.*, niacin, citric acid, biotin). This group is very common in medicinal compounds, found in over 450 marketed drugs including nonsteroidal anti-inflammatory drugs (*e.g.*, aspirin, ibuprofen), antibiotics (*e.g.*, penicillin) and cholesterol-lowering statins (*e.g.,* atorvastatin (Lipitor)).[1,2] The presence of the hydrophilic carboxyl moiety on organic compounds can confer high solubility in water,[3–5] which can be important to consider when designing new chemical reactions or developing new medicinal compounds. This group can also have important implications for pharmaceutical drugs; for example, drugs with a carboxyl functional group can be more metabolically unstable[6] or have more difficulty diffusively crossing membranes.[1,6] Given the carboxyl group's ubiquitous presence in nature and its importance as a functional group, understanding its conformational preferences in various settings is fundamental for the design, modification, and property prediction of new and existing molecules.

The preferred orientation of hydroxyl in the carboxyl functional group in solution is a matter of some debate, even for acetic acid, an archetypal carboxylic acid. The two equilibrium conformations are denoted *syn* (Figure 1(a)), where the O=C–O–H dihedral angle is defined here to be 0°, and *anti* (Figure 1(b)), where the O=C–O–H dihedral angle is defined here as 180°. It is widely believed that the preferred conformation of carboxyl is the *syn* arrangement, from which there is a large energetic penalty to reach the *anti* arrangement. The reasoning behind this idea lies in the perceived extra stability of intramolecular hydrogen bonding that occurs in the *syn* structure. This belief is supported by a number of experimental and theoretical studies done in gas phase, and there is no doubt that this is the preferred conformation in the gas phase.[7–11]

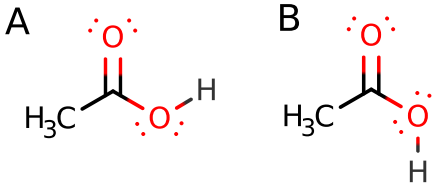

Figure 1: Lewis structure of acetic acid in (a) *syn* and (b) *anti* conformation.

The orientational preference of COOH is considerably more complex outside of gas phase. While some workers remain convinced that *syn* will be more stable, a variety of evidence indicates that this may not always be the case. A recent review article[12] discusses the competition between intramolecular and intermolecular hydrogen bonds in solution, stating that an intramolecular hydrogen bond may be disrupted in protic solution, such as water, when the increase in internal energy is offset by two or more solute-solvent intermolecular hydrogen bonds. Another study found that the carboxyl group has no strong preference, kinetically and thermodynamically, for the *syn* (or *anti*) conformation in proton trans-

[a)] Electronic mail: dmobley@mobleylab.org

fer catalysis.[13] The *anti* state may also be important to consider when calculating solvation free energies.[14] Finally, the *anti* conformation is not insignificantly represented in structures from the Cambridge Structural Database,[15] supported by related crystallographic and theoretical charge density studies.[16,17]

Past work investigating acetic acid in solvent predominantly considers the *syn* state such as in studies characterizing hydrogen-bonding interactions of acetic acid microhydrates using DFT-B3LYP calculations,[18] or assessing the dimer form in various stages of hydration theoretically[19,20] and experimentally.[21] One recent work examining solvent stabilization using DFT-ωB97X-D calculations[22] indicates that water may modulate the conformational preferences of acetic acid; however, to our knowledge there has not been a systematic investigation of the preferred conformational state of the carboxyl group in solution. We believe that this collection of evidence on the orientational preference of COOH in solvent lacks a clear, definitive answer on whether both conformational states of the carboxyl group may reasonably be populated in normal aqueous solution when this group is in its neutral protonation state.

In this work, we aim to understand the relative conformational stability and energetic barrier for carboxyl functional group interconversion in both gas and aqueous phases. We present our investigation on monomeric acetic acid using both *ab initio* quantum mechanical (QM) calculations and atomistic molecular dynamics (MD) simulations.

II. METHODS

Past gas phase QM studies clearly indicate a preference for the *syn* structure of the COOH group such as in acetic acid. However, classical all-atom MD simulations show that both are equally favorable in solution, at least with the energy model ("force field") employed.[14] This could be a real effect of water on the conformational preferences, or a limitation of the force field employed. Therefore, we need to examine a more intermediate region between gas phase QM and solution-phase atomistic MD to settle the issue more definitively. Specifically, we look at QM in solvent (implicit, and with explicit waters) as well as QM data of snapshots pulled from MD simulations, while on the MD end, we consider the effects of force field as well as solvation state.

We present an overview of our approach then discuss the methods in further detail. A torsion drive was conducted on acetic acid over the aforementioned dihedral angle. We conduct restrained geometry optimizations using two different QM methods, each with and without the presence of implicit solvent. Then, we perform QM geometry optimizations on microhydrated configurations of the *syn* and *anti* structures of acetic acid with three explicit waters in implicit solvent. Our final work on the QM front involved a set of geometry optimizations on pentahydrated acetic acid with varied water configurations obtained from MD simulations. We compared these energies for both *syn* and *anti* structures.

On the MD side, we compute a series of free energy landscapes, also known as potentials of mean force (PMFs), from driving the relevant torsion in acetic acid. We evaluated the sensitivity of these one-dimensional free energy surfaces to the force field, partial charge assignment, and solvation state. We consider the force field because this factor is likely to vary among users running MD simulations. The partial charge set assigned to a solute depends on the initial conformation and is typically fixed throughout MD simulations, so we investigate potential implications of choosing one set or another. Finally, we compare the results of gaseous and aqueous phases to shed light on how reasonably the *syn* and *anti* states may be occupied in either scenario.

A. Ab initio torsion drive of acetic acid

Acetic acid configurations of the carboxyl O=C–O–H dihedral angle were generated and used as input for both QM torsion drives and MD umbrella sampling simulations. The dihedral angle was rotated using VMD[23] in 15° increments from 0° to 360°, yielding 24 total conformations.

The QM torsion drives were run using Turbomole version 7.1[24] with two different levels of theory: HF/6-31G* and TPSSh-D3BJ/def2-TZVP. The former method, using Hartree-Fock reference[25] with the Pople 6-31G* basis set,[26,27] was chosen for consistency with the methods often employed in parameterization of force fields used for molecular simulation.[28] This low level method also provides historical perspective contributing to the strong bias favoring the *syn* conformation of the carboxyl group. Taking a more rigorous approach, we employed the TPSSh hybrid functional[29] with Grimme's D3 dispersion correction[30] and Becke-Johnson damping,[30] in combination with the Karlsruhe triple-zeta basis set def2-TZVP.[31] We also run calculations in implicit solvent with each of the aforementioned methods using the conductor-like screening model (COSMO) with outlying charge corrections.[32]

B. Ab initio unrestrained geometry optimization of microhydrated acetic acid

In addition to the QM torsion drives, we evaluate relative energies of the *syn* and *anti* conformations of acetic acid in the presence of explicit waters. We first consider a microhydrated state with three explicit waters arranged to maximize hydrogen bonding to acetic acid. These water molecules were placed in separated positions surrounding acetic acid around each of the *syn* and *anti* conformers using Pymol.[33] Two water molecules were placed on either side of the carbonyl group, and one was placed

in line with the hydroxyl group (see supplementary information).

For each system, a quick MM optimization was performed via OpenEye's OEChem Python Toolkit[34] using the MMFF94S force field.[35] Subsequently, a two-stage gas phase optimization was completed with the Psi4[36] software package. Geometries obtained from the first level of theory, MP2/def2-SV(P),[31,37] were used as inputs for QM optimization on the subsequent level of theory, B3LYP-D3MBJ/def2-TZVP.[31,38] Finally, QM optimization with solvation was performed using the Turbomole version 7.1[24] software package with COSMO-TPSSh-D3BJ/def2-TZVP.[29–32]

C. Ab initio geometry optimizations from molecular dynamics configurations

We sample various configurations of water molecules around acetic acid by running separate MD simulations of the *syn* and *anti* conformations in a box of TIP3P water molecules. The structures were solvated using Antechamber[39] with in an cubic box with TIP3P waters,[40] such that the minimum distance between the solute and the edge of the periodic box was 12 Å. Dynamics were run using GROMACS version 5.0.4 with the leap-frog stochastic dynamics integrator and a 2 fs time step. We use a Langevin thermostat for the temperature at 298.15 K with a frictional constant of 2.0 ps^{-1}. The pressure was maintained at 1 atm using the Parrinello-Rahman pressure coupling scheme with a time constant of 10 ps^{-1} and an isothermal compressibility of 4.5×10^{-5} bar^{-1}. These systems were simulated with 2500 steps of steepest descent minimization, 50 ps constant volume and temperature (NVT) equilibration, 5 ns of constant pressure and temperature (NPT) equilibration, and then 5 ns of NPT production. Trajectory snapshots were extracted of the most similar configurations for acetic acid and its five closest waters using a root-mean-square deviation clustering of geometries with a 2 Å cutoff. This yielded 14 snapshots for the penta-hydrated *syn* conformation and 17 snapshots for the penta-hydrated *anti* conformation. Each snapshot was MM-optimized via OpenEye's OEChem Python Toolkit[34] using the MMFF94S force field,[35] then subsequently QM-optimized using Turbomole version 7.1[24] with COSMO-TPSSh-D3BJ/def2-TZVP.[29–32]

D. MD simulations with umbrella sampling along carboxyl dihedral angle

We used umbrella sampling[41] molecular dynamics to compute a potential of mean force (PMF) to analyze the free energy landscape projected onto this one-dimensional coordinate. We compared MD results with the GAFF[42] and GAFF2[43] classical all-atom force fields, with partial charges assigned by the AM1-BCC[44] approach. We consider effects of the solute partial charges in the MD simulations by carrying out MD simulations with AM1-BCC charges assigned from the *syn* configuration as well as charges assigned from the *anti* configuration. Energetics were examined in gas phase, then in solvent using explicit TIP3P water molecules.[40]

These simulations were run using GROMACS version 5.0.4.[45] Each acetic acid configuration generated in VMD was set with partial charges from the AM1-BCC charge model[44] on the *syn* (0°) conformation as implemented in OpenEye's Python toolkits.[34] The partial charges of the solute depend on initial configuration, so we also consider the *anti* (180°) conformation for computing partial charges. The O=C–O–H dihedral angle was restrained in both gas phase and aqueous MD simulations, using a harmonic force constant of 300 kJ/mol/(rad^2) (approximately 0.022 kcal/mol/(deg^2)).

For the gas phase simulations, the reference temperature of 298.15 K was maintained using Langevin dynamics with a frictional constant of 1.0 ps^{-1}. Maintaining the GROMACS parameters described earlier, the systems underwent steepest descent minimization over 2500 steps, NVT equilibration for 50 ps, and NVT production for 1 ns.

For the explicit solvent simulations, each of the 24 configurations was solvated using Antechamber[39] within an isometric box with TIP3P waters,[40] such that the minimum distance between the solute and the edge of the periodic box was 12 Å. Other MD simulation settings remained the same as previously described in the section "*Ab initio* geometry optimizations from molecular dynamics configurations." These systems were simulated with 2500 steps of steepest descent minimization, 50 ps NVT equilibration, 50 ps NPT equilibration, and 5 ns NPT production. The configurations with dihedral angle around 270° seemed not converged, so six conformations were extended 5 ns for a total of 10 ns each: 65°, 90°, 105°, 255°, 270°, 285°. However, there was little to no change in the resulting PMFs.

Analysis of all umbrella sampling simulations was completed with MBAR algorithm[46] to produce the potentials of mean force (PMFs) for rotation of the carboxyl dihedral angle.

III. RESULTS AND DISCUSSION

Results from both QM and MD approaches support the general understanding and former work that *syn* is favored in gas phase. They also indicate that the *anti* conformation may also be populated to a significant extent in water. We address our QM results first and then discuss MD results.

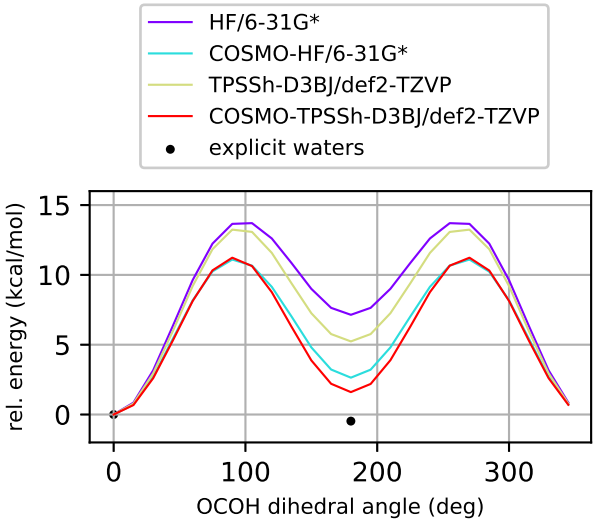

Figure 2: QM torsion drive of acetic acid carboxyl dihedral angle for HF and TPSSh methods. In each case, implicit solvation with COSMO reduces the energy barrier and the relative minima energy to 5-7 kcal/mol and 2-3 kcal/mol respectively. The black points are final energies from geometry optimizations of trihydrated acetic acid with COSMO-TPSSh-D3BJ/def2-TZVP.

A. Ab initio torsion drive of acetic acid

Our QM calculations in gas phase and implicit solvent show that *syn* is highly favored in the gas phase but the difference becomes less significant in solvent. From the torsion drive obtained via *ab initio* QM calculations, the *syn*-*anti* energy difference is 7.14 kcal/mol with the basic HF/6-31G* method and decreases to 5.24 kcal/mol with the higher level of theory using the TPSSh functional (Figure 2). With COSMO, a similar trend is seen in which the higher level of theory yields a smaller energy difference between the *syn* and *anti* structures. With either level of theory, adding implicit solvent significantly lowers the relative energy difference between *syn* and *anti* from 5-7 kcal/mol to 2-3 kcal/mol. A 5-7 kcal/mol difference is large enough that such configurations would occur only extremely rarely, whereas 2-3 kcal/mol is enough that such conformations will occur sporadically in solution (3-7% of the time) and could potentially easily be stabilized by interactions with a nearby receptor or other biomolecule with a strain energy no larger than that reported in many binding interactions,[47,48] making it potentially relevant functionally.

We now turn our focus to the energy barrier from the *syn* state to the *anti* state. This feature is not particularly critical in molecular simulation, as in most cases systems will be at equilibrium given sufficient relaxation time and sampling. That being said, the energy barrier has implications for interconversion between the two states. One conformation may be more structurally relevant than the other in certain scenarios, and a modeler may wish to achieve an accurate representation of the populations of both conformations. The barrier associated with the rotation of the carboxyl dihedral angle determines how easy it is to interconvert between and sample different conformations. From our QM results, we see a large energetic cost or barrier of 13-14 kcal/mol separating the *syn* form from the *anti* form in gas phase. Solvation with COSMO reduces this barrier height to around 11 kcal/mol.

Overall, the relative energy difference between the *syn* and *anti* conformations of acetic acid appears not very large, especially in the aqueous solutions relevant to biochemistry. These relative energy comparisons from the QM torsion drives are summarized in the top four lines of Table I. Note that, from our QM results, these are relative energies rather than relative free energies; with MD in the following section; we obtain relative free energies.

B. Ab initio unrestrained geometry optimization of microhydrated acetic acid

To rule out the possibility that stabilization of the *anti* form is due to implicit solvent model alone, and to determine whether explicit water might provide additional stabilization, we examined acetic acid with explicit water molecules. *Ab initio* QM calculations with COSMO on acetic acid with three explicit water molecules suggest an energetically preferred *anti* conformation in aqueous solution. Water molecules were placed in various positions around acetic acid in an attempt to maximize intermolecular hydrogen bonding interactions. Two arrangements of the water molecules were considered for each acetic acid conformer. In gas phase, the two *anti* configurations minimized to the same structure (Figure 3(b)), and the minimized *syn* configurations differed in energy by about 0.11 kcal/mol. We proceeded with the lower energy *syn* configuration (Figure 3(a)). These structures were optimized with inclusion of implicit solvent using COSMO and the TPSSh-D3BJ/def2-TZVP method. Although these acetic acid systems are surrounded by only three water molecules, observed optimized acetic acid geometries agree extremely well with past experimental and theoretical studies[7,9] of other solvated structures (see supplementary information). Specifically, bond lengths differed from past work by no more than 0.03 Å, and angles were within one degree. In the COSMO-optimized structures, we observe that the *anti* tetramer is the lower-energy conformation by 0.47 kcal/mol. Compared to the previously discussed torsion drive without explicit water molecules, the addition of explicit waters stabilized the *anti* form by 2 kcal/mol more than implicit solvent alone. These two energy values are shown as black points in Figure 2.

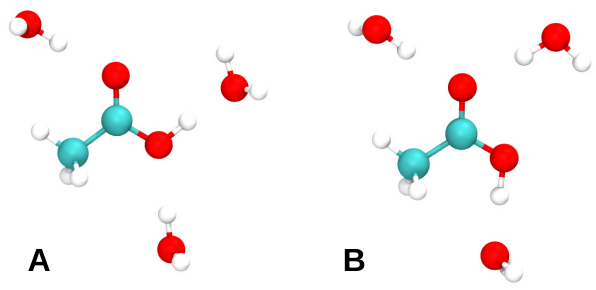

Figure 3: Acetic acid conformation with three water molecules for (a) *syn* and (b) *anti* forms.

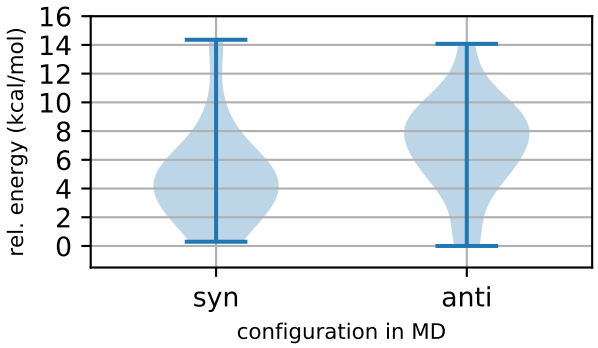

Figure 4: Violin plots for relative energy distributions of pentahydrated *syn* and *anti* conformations of acetic acid. The data represent COSMO-TPSSh-D3BJ/def2-TZVP energies of configurations taken from MD simulations of the *syn* form (14 snapshots) and the *anti* form (17 snapshots).

C. Ab initio geometry optimizations from molecular dynamics configurations

Recent work on the microhydration of acetic acid suggests that the particular arrangement of water molecules may be important when comparing energetic stabilities of acetic acid conformations.[22] The water molecules in the trihydrated clusters of the previous section were placed to maximize the hydrogen bonding network between acetic acid and water to promote solvent stabilization. However, other arrangements of water molecules could lead to lower energy configurations. This means that results could be artifacts of water placement. Given that we are interested in solution-phase behavior, the *actual* solution-phase geometry of water molecules around acetic acid then becomes very important. To address this point, we sample various conformations of water molecules around acetic acid by running molecular dynamics simulations for each of the *syn* and *anti* forms. Configurations of acetic acid with its five nearest waters were clustered by root-mean-square deviation of geometries. The most common arrangements were extracted for QM optimization in implicit solvent using the method COSMO-TPSSh-D3BJ/def2-TZVP. The violin plots in Figure 4 display the distributions for the relative energies of the *syn* (left side) and *anti* (right side) pentahydrated configurations of acetic acid. The distribution for the *syn* configurations skews toward lower energies compared to the *anti* configurations. However, the energy values of the extrema are quite similar, and the population of the *anti* form at low energies is nonnegligible.

D. MD simulations with umbrella sampling along carboxyl dihedral angle

Our above QM calculations study only conformational energies, not free energies, so we computed the one-dimensional free energy landscape (the potential of mean force, or PMF) of rotating the acetic acid dihedral angle with classical molecular dynamics. The MD results in gas phase and in explicit solvent are in qualitative agreement with our QM data and indicate that water substantially increases the stability of the *anti* conformation. We considered various force fields, partial charge sets, and solvation states for a total of eight PMFs. Atomic partial charges are held fixed within our simulations, as is typical in MD, but these charges are sensitive to the molecular conformation when assigning charges, so we assigned charges using both conformations. Hereafter we use the notation `SC` for acetic acid partial charges obtained from the *syn* conformation and `AC` for charges obtained from the *anti* conformation. Error bars on the PMFs are obtained from the MBAR estimator.[46] We discuss each of these three factors (force field, charge set, and solvation state) separately and present a comprehensive comparison in Figure 8 and in Table I.

Considering the GAFF and GAFF2 force fields, the PMFs are in good agreement with each other in both gas and aqueous phases as well as with either `SC` or `AC` (Figures 5, 8). We observe consistent relative free energies between the *syn* and *anti* minima. In gas phase, for the `SC` solute, the *syn* structure is favored in free energy by 6.2±0.2 kcal/mol with GAFF and 5.9±0.2 kcal/mol with GAFF2 (Figure 5, red vs. blue). In aqueous phase, the *anti* structure is favored in free energy by -0.7±0.1 kcal/mol with GAFF and -1.4±0.1 kcal/mol with GAFF2 (Figure 8, yellow vs. brown). These qualitative conclusions are the same when considering the `AC` solute. In this case, with gas phase, *syn* is favored in free energy by 3.4±0.2 kcal/mol for GAFF and 3.3±0.2 kcal/mol for GAFF2 (Figure 8, green vs. purple). Conversely, the `AC` aqueous phase structures show a preference for the *anti* state by -1.3±0.1 kcal/mol for GAFF and -1.6±0.1 kcal/mol for GAFF2 (Figure 8, pink vs. gray). Thus, GAFF and GAFF2 give very similar results for the conformational equilibrium of acetic acid which holds true regardless of the partial charge set. Overall these re-

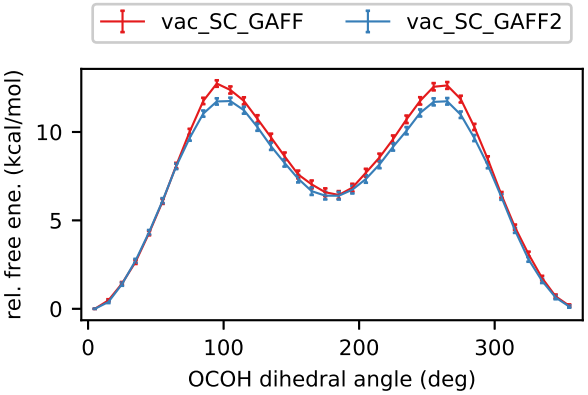

Figure 5: Comparison of GAFF and GAFF2 force fields in PMFs of rotating the acetic acid carboxyl dihedral angle. Both are in strong agreement with each other. The PMFs displayed in this figure came from gas phase simulations with *syn* charges. Similar conclusions were drawn for PMFs from aqueous simulations and from using *anti* charges (Figure 8).

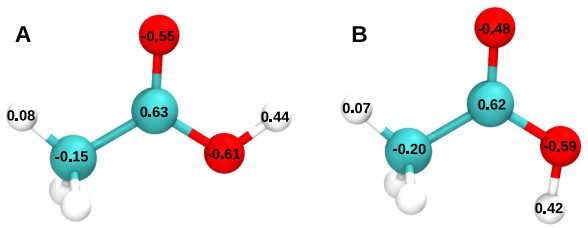

Figure 6: AM1-BCC charges generated for (a) *syn* and (b) *anti* configurations of acetic acid.

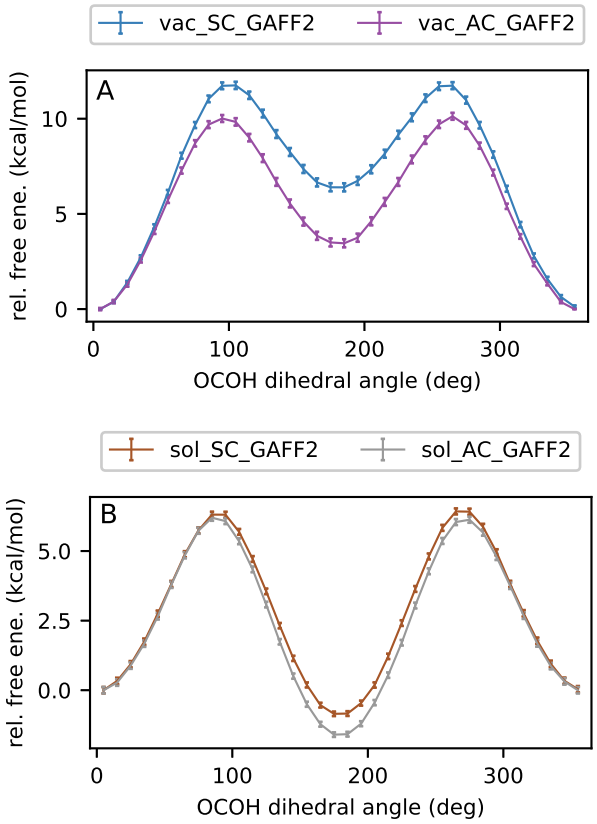

Figure 7: Comparison of *syn* and *anti* solute charges in PMFs of rotating the acetic acid carboxyl dihedral angle. In each situation with *anti* charges (top) and *syn* charges (bottom), the `AC` set more strongly stabilizes the *anti* conformation than the `SC` set.

sults, at least within the classical framework, indicate that explicit solvent provides approximately 5-8 kcal/mol of stabilization of the *anti* conformation relative to the *syn* conformation. This trend is in the same direction as that provided by COSMO implicit solvent, but provides further stabilization.

We also compare the two force fields in terms of the conformational transition barriers. We note that the GAFF barrier height is higher than the GAFF2 barrier in each pairwise combination of the two force fields with various solvent and charge models. The barrier height differences are 0.9±0.4 kcal/mol in gas phase (compare barrier heights in Figure 8 for red vs. blue and for green vs. purple). The rotational barriers differ by 0.6±0.2 kcal/mol in aqueous phase (compare barrier heights in Figure 8 for yellow vs. brown and for pink vs. gray). For both gaseous and aqueous states, the PMFs from the same force field match each other to *lesser* degree than PMFs with matching charges (`SC` or `AC`). For example, in Figure 8, the red and green curves are more distinct from each other, while the red and blue curves are more similar. Since the partial charges of the solute may affect the PMFs more so than the force field, as shown here, one should carefully consider other likely conformations when assigning partial charges. Next we further investigate the solute partial charge sets.

There is a pronounced difference in the PMFs depending on the conformation used to charge acetic acid (Figure 6). Charges are typically fixed throughout a molecular dynamics simulation, meaning that initial charge assignment is important for capturing correct energetics throughout a simulation. The free energy difference between the *syn* and *anti* structures is notably larger in gas phase than in water. When we use the *syn* form to obtain AM1-BCC charges (`SC`), the gas phase PMFs are higher in energy for both the barrier height and the two minima (Figure 7 (a)) compared to using the *anti* form to obtain AM1-BCC charges (`AC`). Qualitatively, the `SC` set is slightly stronger in magnitude than the `AC` set, which is consistent with the intramolecular hydrogen bonding aspect of the *syn* conformation. The stronger `SC` partial charges contribute to increased stabilization of the lower-energy *syn* structure in gas phase, which results in a greater free energy difference and barrier height com-

pared to AC. On the other hand, in water, (Figure 7 (b)), *syn* and *anti* are closer in relative free energy for SC than for AC. In this setting, *syn* is higher in energy than *anti*. Once again, the stronger SC partial charges contribute to increased stabilization of *syn*, in this case via more stabilizing interactions with the solvent. Here, the two minima are closer in free energy. Therefore we see again that the relative free energies at the minima are governed more strongly by solute charges than by force field.

We take a final look at the MD PMFs in the lens of gaseous versus aqueous phases. These results are in harmony with earlier work on ibuprofen (a carboxylic acid) which found that the *syn* conformation was favorable in vacuum but the *anti* conformation was slightly preferred in water.[14] The major takeaway from the aqueous phase PMFs is that the *anti* conformation of acetic acid is the lower free energy state in solution due to an increased ability to form stabilizing interactions with the solvent. This conclusion qualitatively parallels the result obtained with COSMO-QM calculations on microhydrated acetic acid which showed that the *anti* conformation is lower in energy than the *syn* conformation by about 1.6 kcal/mol, at least for certain arrangements of water molecules.

Overall, the MD results are qualitatively consistent with QM calculations in determination of relative energy differences of the minima and energy barriers for conformational interconversion. The SC charge set seems better than the AC set in reproducing the relative energy differences obtained with QM DFT in gas phase and in implicit solvent, consistent with our previous practice of considering this conformation more important when assigning charges.

To summarize our PMF results, we considered the effects of force field, charge set, and solvation state on the relative minima free energies as well as on the transition barriers between the two minima. The force fields GAFF and GAFF2 yielded generally similar results to each other. The PMFs in both gas phase and aqueous phase revealed strong dependence on solute charges, especially at the minima. More specifically, the set of partial charges assigned to acetic acid is sensitive to the orientation of O–H in the carboxyl group, leading to several kcal/mol variations in the free energy differences between the *syn* and *anti* structures. Lastly, the dihedral rotation free energy barriers between the *syn*-*anti* conformations are more dependent on the charge set than the force field in gas phase simulations, while they are more influenced by the force field in aqueous phase simulations. All eight PMFs, obtained from permutation of the force field, solute charges, and solvation state are summarized in Figure 8 and Table I.

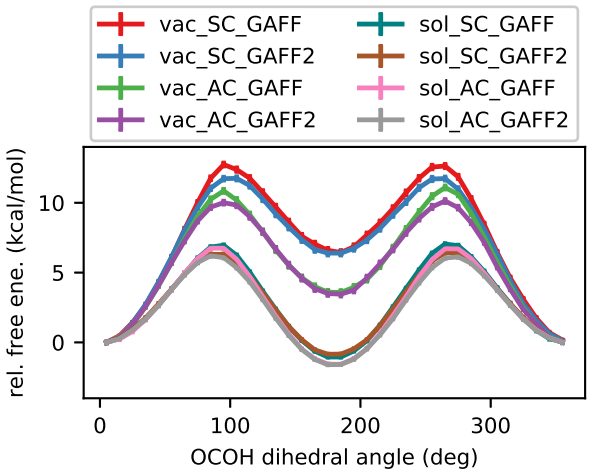

Figure 8: PMFs of rotating the acetic acid carboxyl dihedral angle. We consider variations on the force field (GAFF, GAFF2), solute AM1-BCC partial charges (starting from *syn* or *anti*), and solvation state (gas phase, explicit TIP3P waters).

Table I: Summary of relative energy differences between *syn* and *anti* conformations of acetic acid as well as free energy barriers of interconversion. The first four lines are results from QM torsion drives, the fifth from unrestrained QM optimizations, and the last eight from umbrella sampling via atomistic molecular dynamics simulations. Energies are listed in units of kcal/mol.

Method	Solvation	minima[a]	barrier
HF/6-31G*	gas	7.142	13.706
HF/6-31G*	COSMO	2.761	11.158
TPSSh/def2-TZVP[b]	gas	5.235	13.242
TPSSh/def2-TZVP	COSMO	1.626	11.230
3 explicit waters[c]	COSMO	-0.473	n/a
vac_SC_GAFF	gas	6.2±0.2	12.7±0.3
vac_SC_GAFF2	gas	5.9±0.2	11.7±0.3
vac_AC_GAFF	gas	3.4±0.2	11.0±0.3
vac_AC_GAFF2	gas	3.3±0.2	10.1±0.3
sol_SC_GAFF	TIP3P	-0.8±0.1	7.0±0.2
sol_SC_GAFF2	TIP3P	-0.7±0.1	6.4±0.2
sol_AC_GAFF	TIP3P	-1.3±0.1	6.7±0.2
sol_AC_GAFF2	TIP3P	-1.4±0.1	6.1±0.2

[a] All relative energy differences are taken with respect to acetic acid's *syn* conformation.
[b] Dispersion corrections added with all TPSSh calculations in this work. See details in text.
[c] QM optimization completed with TPSSh-D3BJ/def2-TZVP level of theory.

IV. CONCLUSIONS

Our results call into question the conventional wisdom that carboxylic acids will almost always be in the "more stable" *syn* conformation in biomolecular systems. Typically, the increased stability of the *syn* form is understood to be from the stabilizing intramolecular interaction between the hydrogen atom in the hydroxyl group and the carbonyl oxygen. This idea is in tune with gas phase results we present in this work. However, in aqueous phase, we conclude that the *anti* state may nearly be as

populated as the *syn* state due to stabilizing interactions from the solvent. Thus, for MD studies that involve a carboxylic acid or other functional group with possible intramolecular hydrogen bonds, it may be necessary to ensure sufficient sampling of all potentially relevant conformations in solution. This can be challenging given the particularly large barrier associated with rotation of the carboxylic acid torsion.

Our findings also have implications for partial charge calculations for parameter assignment for MD simulations. Carboxylic acids are a case in which neither partial charge set adequately represents the electrostatics of the solute as it samples various conformations. When generating an empirical force field, such as for a small molecule ligand, charges are typically computed for a particular given conformation. These fixed charges are then used for scenarios involving conformational change. In this work, we observe that different solute charges may lead to deviations in relative free energies to as large as 3 kcal/mol. Interconversion is not expected to be frequent, given that the torsional barrier is at least 6 kcal/mol. For that reason, one may wish to treat *syn* and *anti* conformation charges individually, though this could present difficulties in cases that interconversion is needed for convergence (*e.g.,* a carboxylic acid in a binding site where one conformation forms better contacts than the other). As an alternative approach, the use of polarizable charges may provide a more holistic picture of the carboxyl group's variable nature.

The carboxyl conformational equilibrium has implications for several other types of studies. Hydration free energy calculations may lead to results which depend substantially on the starting conformation. For example, kinetic trapping into one particular conformation can lead to computed hydration free energies which are sensitive to starting conformation and vary by more than 2 kcal/mol because of large torsional barriers.[14] An accurate insight into the preferred aqueous phase structure of the carboxyl group is important for catalysis, with impacts in atmospheric science and industrial processes.[49] Further impact may be in crystal engineering and drug co-crystallization, in which the carboxyl group is often used to promote aqueous solubility.[5] Theoretical studies on proton transfer such as on solvated acetic acid[50] or on green fluorescent protein[51,52] typically employ the *syn* conformation due to its expected energetic preference; however, it is worth investigating possible adaptations of carboxyl groups to their local environments. Being aware of the carboxyl moiety's nuanced conformational preferences in different environments may thus lead to better insight for calculated properties, reactivity, and molecular design.

ACKNOWLEDGMENTS

The authors thank Prof. Filipp Furche and Matthew Agee for helpful discussions on QM methods and for support in using the Turbomole software package, respectively. VTL acknowledges funding the National Science Foundation Graduate Research Fellowship Program. DLM appreciates financial support from the National Institutes of Health (1R01GM108889-01) and the National Science Foundation (CHE 1352608), and computing support from the UCI GreenPlanet cluster, supported in part by NSF Grant CHE-0840513.

Authors will release all atomic coordinates upon article publication.

REFERENCES

[1] C. Ballatore, D. M. Huryn, and A. B. Smith, "Carboxylic Acid (Bio)Isosteres in Drug Design," ChemMedChem **8**, 385–395 (2013).

[2] Y. Lou and J. Zhu, "Carboxylic Acid Nonsteroidal Anti-Inflammatory Drugs (NSAIDs)," in *Bioactive Carboxylic Compound Classes*, edited by C. Lamberth and J. Dinges (Wiley-VCH Verlag GmbH & Co. KGaA, 2016) pp. 221–236.

[3] S. E. David, P. Timmins, and B. R. Conway, "Impact of the counterion on the solubility and physicochemical properties of salts of carboxylic acid drugs," Drug Development and Industrial Pharmacy **38**, 93–103 (2012).

[4] D. P. McNamara, S. L. Childs, J. Giordano, A. Iarriccio, J. Cassidy, M. S. Shet, R. Mannion, E. O'Donnell, and A. Park, "Use of a Glutaric Acid Cocrystal to Improve Oral Bioavailability of a Low Solubility API," Pharmaceutical Research **23**, 1888–1897 (2006).

[5] G. R. Desiraju, "Crystal Engineering: From Molecule to Crystal," Journal of the American Chemical Society **135**, 9952–9967 (2013).

[6] T. Lassila, J. Hokkanen, S.-M. Aatsinki, S. Mattila, M. Turpeinen, and A. Tolonen, "Toxicity of Carboxylic Acid-Containing Drugs: The Role of Acyl Migration and CoA Conjugation Investigated," Chemical Research in Toxicology **28**, 2292–2303 (2015).

[7] P. I. Nagy, "The syn–anti equilibrium for the COOH group reinvestigated. Theoretical conformation analysis for acetic acid in the gas phase and in solution," Computational and Theoretical Chemistry **1022**, 59–69 (2013).

[8] P. I. Nagy, D. A. Smith, G. Alagona, and C. Ghio, "Ab initio studies of free and monohydrated carboxylic acids in the gas phase," The Journal of Physical Chemistry **98**, 486–493 (1994).

[9] H. Sato and F. Hirata, "The syn-/anti-conformational equilibrium of acetic acid in water studied by the RISM-SCF/MCSCF method1," Journal of Molecular Structure: THEOCHEM **461–462**, 113–120 (1999).

[10] K. B. Wiberg and K. E. Laidig, "Barriers to rotation adjacent to double bonds. 3. The carbon-oxygen barrier in formic acid, methyl formate, acetic acid, and methyl acetate. The origin of ester and amide resonance," Journal of the American Chemical Society **109**, 5935–5943 (1987).

[11] J. L. Derissen, "A reinvestigation of the molecular structure of acetic acid monomer and dimer by gas electron diffraction," Journal of Molecular Structure **7**, 67–80 (1971).

[12] P. I. Nagy, "Competing Intramolecular vs. Intermolecular Hydrogen Bonds in Solution," International Journal of Molecular Sciences **15**, 19562–19633 (2014).

[13] T. A. Montzka, S. Swaminathan, and R. A. Firestone, "Reversal of Syn-Anti Preference for Carboxylic Acids along the Reaction Coordinate for Proton Transfer. Implications for Intramolecular Catalysis," The Journal of Physical Chemistry **98**, 13171–13176 (1994).

[14] P. V. Klimovich and D. L. Mobley, "Predicting hydration free energies using all-atom molecular dynamics simulations and multi-

ple starting conformations," Journal of Computer-Aided Molecular Design **24**, 307–316 (2010).

[15]L. D'Ascenzo and P. Auffinger, "A comprehensive classification and nomenclature of carboxyl–carboxyl(ate) supramolecular motifs and related catemers: Implications for biomolecular systems," Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials **71**, 164–175 (2015).

[16]R. Pal, M. B. M. Reddy, B. Dinesh, M. A. Venkatesha, S. Grabowsky, C. Jelsch, and T. N. Guru Row, "Syn vs Anti Carboxylic Acids in Hybrid Peptides: Experimental and Theoretical Charge Density and Chemical Bonding Analysis," J. Phys. Chem. A **122**, 3665–3679 (2018).

[17]M. G. Medvedev, I. S. Bushmarinov, and K. A. Lyssenko, "Z-effect reversal in carboxylic acid associates," Chem. Commun. **52**, 6593–6596 (2016).

[18]Q. Gao and K. T. Leung, "Hydrogen-bonding interactions in acetic acid monohydrates and dihydrates by density-functional theory calculations," The Journal of Chemical Physics **123**, 074325 (2005).

[19]H. Pašalić, D. Tunega, A. J. A. Aquino, G. Haberhauer, M. H. Gerzabek, and H. Lischka, "The stability of the acetic acid dimer in microhydrated environments and in aqueous solution," Physical Chemistry Chemical Physics **14**, 4162–4170 (2012).

[20]J. Chocholoušová, J. Vacek, and P. Hobza, "Acetic Acid Dimer in the Gas Phase, Nonpolar Solvent, Microhydrated Environment, and Dilute and Concentrated Acetic Acid: Ab Initio Quantum Chemical and Molecular Dynamics Simulations," The Journal of Physical Chemistry A **107**, 3086–3092 (2003).

[21]B. Ouyang and B. J. Howard, "The monohydrate and dihydrate of acetic acid : A high-resolution microwave spectroscopic study," Physical Chemistry Chemical Physics **11**, 366–373 (2009).

[22]P. Krishnakumar and D. K. Maity, "Microhydration of Neutral and Charged Acetic Acid," The Journal of Physical Chemistry A **121**, 493–504 (2017).

[23]W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," Journal of Molecular Graphics **14**, 33–38 (1996).

[24]"TURBOMOLE V7.1 2016, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from `http://www.turbomole.com`.".

[25]M. Häser and R. Ahlrichs, "Improvements on the direct SCF method," Journal of Computational Chemistry **10**, 104–111 (1989).

[26]W. J. Hehre, R. Ditchfield, and J. A. Pople, "Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules," The Journal of Chemical Physics **56**, 2257–2261 (1972).

[27]P. C. Hariharan and J. A. Pople, "The influence of polarization functions on molecular orbital hydrogenation energies," Theoretica chimica acta **28**, 213–222 (1973).

[28]W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules," Journal of the American Chemical Society **117**, 5179–5197 (1995).

[29]V. N. Staroverov, G. E. Scuseria, J. Tao, and J. P. Perdew, "Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes," The Journal of Chemical Physics **119**, 12129–12137 (2003); "Erratum: "Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes" [J. Chem. Phys. 119, 12129 (2003)]," **121**, 11507–11507 (2004).

[30]S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu," The Journal of Chemical Physics **132**, 154104 (2010); S. Grimme, S. Ehrlich, and L. Goerigk, "Effect of the damping function in dispersion corrected density functional theory," Journal of Computational Chemistry **32**, 1456–1465 (2011).

[31]F. Weigend and R. Ahlrichs, "Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy," Physical Chemistry Chemical Physics **7**, 3297–3305 (2005).

[32]A. Klamt and G. Schüürmann, "COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient," Journal of the Chemical Society, Perkin Transactions 2 , 799–805 (1993); A. Klamt, V. Jonas, T. Bürger, and J. C. W. Lohrenz, "Refinement and Parametrization of COSMO-RS," The Journal of Physical Chemistry A **102**, 5074–5085 (1998); F. Eckert and A. Klamt, "Fast solvent screening via quantum chemistry: COSMO-RS approach," AIChE Journal **48**, 369–385 (2002); S. Sinnecker, A. Rajendran, A. Klamt, M. Diedenhofen, and F. Neese, "Calculation of Solvent Shifts on Electronic g-Tensors with the Conductor-Like Screening Model (COSMO) and Its Self-Consistent Generalization to Real Solvents (Direct COSMO-RS)," The Journal of Physical Chemistry A **110**, 2235–2245 (2006).

[33]Schrödinger, LLC, "The PyMOL molecular graphics system, version 1.7," (2015).

[34]OEChem, version 2.1.1, OpenEye Scientific Software Inc.: Santa Fe, NM, USA.

[35]T. A. Halgren, "Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94," Journal of Computational Chemistry **17**, 490–519 (1996); "Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions," **17**, 520–552 (1996); "Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94," **17**, 553–586 (1996); T. A. Halgren and R. B. Nachbar, "Merck molecular force field. IV. conformational energies and geometries for MMFF94," **17**, 587–615 (1996); T. A. Halgren, "Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules," **17**, 616–641 (1996); "MMFF VI. MMFF94s option for energy minimization studies," **20**, 720–729 (1999).

[36]J. M. Turney, A. C. Simmonett, R. M. Parrish, E. G. Hohenstein, F. A. Evangelista, J. T. Fermann, B. J. Mintz, L. A. Burns, J. J. Wilke, M. L. Abrams, N. J. Russ, M. L. Leininger, C. L. Janssen, E. T. Seidl, W. D. Allen, H. F. Schaefer, R. A. King, E. F. Valeev, C. D. Sherrill, and T. D. Crawford, "Psi4: An open-source ab initio electronic structure program," Wiley Interdisciplinary Reviews: Computational Molecular Science **2**, 556–565 (2012).

[37]C. Møller and M. S. Plesset, "Note on an Approximation Treatment for Many-Electron Systems," Physical Review **46**, 618–622 (1934).

[38]A. D. Becke, "Density-functional thermochemistry. III. The role of exact exchange," The Journal of Chemical Physics **98**, 5648–5652 (1993); C. Lee, W. Yang, and R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density," Physical Review B **37**, 785–789 (1988); S. H. Vosko, L. Wilk, and M. Nusair, "Accurate spin-dependent electron liquid correlation energies for local spin density calculations: A critical analysis," Canadian Journal of Physics **58**, 1200–1211 (1980); P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, "Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields," The Journal of Physical Chemistry **98**, 11623–11627 (1994).

[39]J. Wang, W. Wang, P. A. Kollman, and D. A. Case, "Automatic atom type and bond type perception in molecular mechanical calculations," Journal of Molecular Graphics and Modelling **25**, 247–260 (2006).

[40]W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," The Journal of Chemical Physics **79**, 926–935 (1983).

[41]G. M. Torrie and J. P. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling," Journal of Computational Physics **23**, 187–199 (1977).

[42]J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," Journal of Computational Chemistry **25**, 1157–1174 (2004).

[43]D.A. Case, R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao and P.A. Kollman, AMBER 2016, University of California, San Francisco (2016).

[44]A. Jakalian, B. L. Bush, D. B. Jack, and C. I. Bayly, "Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method," Journal of Computational Chemistry **21**, 132–146 (2000); A. Jakalian, D. B. Jack, and C. I. Bayly, "Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation," **23**, 1623–1641 (2002).

[45]M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," SoftwareX **1–2**, 19–25 (2015).

[46]M. R. Shirts and J. D. Chodera, "Statistically optimal analysis of samples from multiple equilibrium states," The Journal of Chemical Physics **129**, 124105 (2008).

[47]D. L. Mobley and K. A. Dill, "Binding of Small-Molecule Ligands to Proteins: "What You See" Is Not Always "What You Get"," Structure **17**, 489–498 (2009).

[48]M. A. Sahai and P. C. Biggin, "Quantifying Water-Mediated Protein–Ligand Interactions in a Glutamate Receptor: A DFT Study," J. Phys. Chem. B **115**, 7085–7096 (2011).

[49]M. Kumar, D. H. Busch, B. Subramaniam, and W. H. Thompson, "Organic Acids Tunably Catalyze Carbonic Acid Decomposition," The Journal of Physical Chemistry A **118**, 5020–5028 (2014).

[50]W. Gu, T. Frigato, T. Straatsma, and V. Helms, "Dynamic Protonation Equilibrium of Solvated Acetic Acid," Angewandte Chemie International Edition **46**, 2939–2943 (2007).

[51]C. Scharnagl, R. Raupp-Kossmann, and S. F. Fischer, "Molecular Basis for pH Sensitivity and Proton Transfer in Green Fluorescent Protein: Protonation and Conformational Substates from Electrostatic Calculations," Biophysical Journal **77**, 1839–1857 (1999).

[52]M. A. Lill and V. Helms, "Proton shuttle in green fluorescent protein studied by dynamic simulations," Proceedings of the National Academy of Sciences **99**, 2778–2781 (2002).