JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

QM/MM Prediction of the Stark Shift in the Active Site of a Protein

Daniel J. Sandberg, Aleksandra N. Rudnitskaya, and José A. Gascón*

Department of Chemistry, University of Connecticut, Storrs, Connecticut 06269, United States

Supporting Information

ABSTRACT: Recent developments in the biophysical characterization of proteins have provided a means of directly measuring electrostatic fields by introducing a probe molecule to the system of interest and interpreting photon absorption in the context of the Stark effect. To fully account for this effect, the development of accurate atomistic models is of paramount importance. However, suitable computational protocols for evaluating Stark shifts in proteins are yet to be established. In this work, we present a comprehensive computational method to predict the change in absorption frequency of a probe functional group as a direct result of a perturbation in its surrounding electrostatic field created by a protein environment, i.e., the Stark shift. We apply the method to human aldose reductase, a key protein enzyme that catalyzes the reduction of monosaccharides. We develop a protocol based on a combination of molecular dynamics and moving-domain QM/MM methods, which achieves quantitative agreement with experiment. We outline the difficulties in predicting localized electrostatic field changes within a protein environment, and by extension the Stark shift, due to a protein site mutation. Furthermore, the combined use of Stark effect spectroscopy and computational modeling is used to predict the protonation state of ionizable residues in the vicinity of the electrostatic probe.

INTRODUCTION

Electrostatic interactions have long been recognized as the primary driving force in many biological processes, including catalysis,[1] ion channel gating and selectivity,[1−4] ligand binding,[5−7] and macromolecular signal translocation.[1,8−10] The importance of electrostatic interactions in biology drives the need for computational methods, capable of calculating electrostatic forces within a protein, that have been vetted by direct comparison to experimental data.[11−18] The dynamic nature of large macromolecules, such as proteins, suggests that any assessment of electrostatic effects should involve finite temperature simulations, generally employing molecular mechanics (MM) force fields. However, MM force fields typically used within Monte Carlo or molecular dynamic trajectories are often unable to reproduce reliable electrostatic potentials.[19−21] Attempts have been made over the past decade to extend standard MM to include improved treatments for polarization effects[22−30] and other quantum mechanical (QM) based events, such as hydrogen bond formation and breaking.[31] Emerging methods require increased levels of theory, but ultimately, the accuracy of the results can only be quantified by comparison to direct measurements of the electrostatic field, which is often difficult to isolate in biological systems.

Direct experimental evaluation of the internal electrostatic fields within a protein has been a challenge in the biophysical characterization of proteins. Internal electrostatic fields can be very large in magnitude (>10 MV/cm) and can fluctuate over small distances due to the presence of polar residues. Pioneering work has been done by Boxer and co-workers to measure shifts in the electrostatic field created by a protein, resulting from an amino acid mutation, by means of vibrational Stark effect (VSE) spectroscopy.[32−43] In this technique, a functional group non-native to the protein structure, e.g. a nitrile group, is introduced to the protein via chemical modification of a known ligand for the protein of interest.[42] Such a method has been extended to several proteins by alternatively introducing a non-native amino acid.[35] In the case of modifying a known ligand with a nitrile group, the response of the nitrile's vibrational energy to an electrostatic field is first calibrated by obtaining a series of Stark spectra at different, known external field strengths, $\mathbf{F}_{ext}$. In this first experimental step, $\mathbf{F}_{ext}$ is not the field produced by the protein moiety but rather an applied external field used only to calibrate the response of the nitrile absorption energy to the presence of an electrostatic field. The shift in the vibrational wavenumber, $\Delta\tilde{\nu}_{CN}$, measured by infrared spectroscopy, is proportional to the strength of the applied electrostatic field vector

$$\Delta E = hc\Delta\tilde{\nu}_{CN} = -\Delta\boldsymbol{\mu}_{CN}\cdot\mathbf{F}_{ext} \quad (1)$$

where h is Planck's constant, c is the speed of light, and $|\Delta\boldsymbol{\mu}_{CN}|$ is the Stark tuning rate in debyes. After this calibration step has been performed, the probe molecule binds to a protein and functions as an electrostatic field sensor, detecting changes in the electrostatic field generated by the protein after amino acid substitution or other perturbations in the protein structure. While in the calibration experiment the applied field is relative to any ambient field, in the protein experiment, the wild type (WT) protein is considered the baseline and the single vibrational absorption frequency of the nitrile is recorded. A site mutation to the protein is then made, affecting the field surrounding the nitrile and, by direct consequence, affecting the vibrational absorption frequency. Because the Stark tuning rate has been obtained from the calibration experiment, the shift in the vibrational frequency resulting from the mutation can be directly related to the shift in the electrostatic field projected along the nitrile probe. Equation 2 follows logically from eq 1

Received: May 21, 2012
Published: June 15, 2012

$$\Delta E = hc\Delta\tilde{\nu}_{CN} = -\Delta\boldsymbol{\mu}_{CN}\cdot\Delta\mathbf{F}_{protein} \quad (2)$$

where $\Delta\mathbf{F}_{protein}$ is the change in the electrostatic field of the protein due to a mutation. (Interested readers are referred to ref 42 and references therein for additional details regarding the experimental protocol.) This technique was successfully applied to the study of human aldose reductase (hALR2) using the nitrile probe shown in Figure 1, {5-chloro-2-[[(4-cyanobenzyl)-

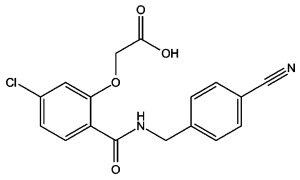

Figure 1. {5-Chloro-2-[[(4-cyanobenzyl)amino]carbonyl]phenoxy}-acetic acid (**1**).

amino]carbonyl]phenoxy}acetic acid (herein referred to as **1**). The absorption frequency of the nitrile functional group is distinct from any frequency of the apoprotein and is therefore easily distinguished from other absorption peaks in the spectrum. With this method, a test set of 10 mutants of the protein was generated to determine the tuning effect of the probe's environment. Experimental measurements of the shift in the electrostatic field resulting from a residue mutation provide a set of data that can be used to determine the accuracy of computational methods focused on predicting electrostatics.

Equation 2 implies that one can directly obtain the Stark shift by computing the change in the electrostatic field due to a residue mutation. Such calculation has proven to be a complex problem. Models that predict Stark shifts may yield accurate results for small systems[38,44,45] but provide only marginal agreement when applied to the study of the electrostatic fields in protein environments.[41−43,46−50] Computational modeling performed by Webb and Boxer pointed out the difficulty in predicting the Stark shift. Specifically, they noted that the conformational flexibility of amino acid side chains near the nitrile played an important role. Many of the mutations analyzed had multiple possible conformations, resulting in many possible Stark shift values varying in magnitude by as much as 20 MV/cm. Thus, single configuration evaluations cannot provide a meaningful estimate, unless the structure represents an average configuration for the system.[42] More recently, an attempt has been made at crystallizing the hALR2 protein with bound **1**. Unfortunately, the low binding affinity of the ligand precluded a crystal structure. To improve binding affinity, a nitro group was added at the meta position of the phenyl ring. Interestingly, this chemical modification resulted in two IR peaks in the Stark spectrum, attributed to a hydrogen-bonded and non-hydrogen-bonded population. A potential hydrogen bond between the nitrile and a neighboring side chain complicates the interpretation of the Stark shift due to the fact that (1) hydrogen bonding will affect the sensitivity of the nitrile to electrostatic environments and (2) hydrogen bonding will draw electron density from the lone pair on the nitrile nitrogen and induce a shift in the IR absorption frequency, an effect for which eq 2 does not account. Despite these challenges, Xu and co-workers were able to successfully use molecular dynamics simulations to accurately depict the proportion of the two microenvironments around the nitrile, H-bonded and non-H-bonded, though they noted that the prediction of the spectral shifts was incorrect due to the neglected effect of the H-bond. The authors noted the need for further theoretical work, possibly utilizing QM/MM methods.[43]

This study strives to outline a reliable method to predict electrostatic field changes, and thus Stark shifts, resulting from a protein site mutation, incorporating both finite temperature simulations and ab initio quality structures and charges. We focus on the data generated for the original ligand **1**,[42] which has a single Stark shift for each mutation and therefore a single microenvironment. It is shown that the combination of finite temperature simulation and QM/MM methods leads to an evaluation of the electrostatic field in very good quantitative agreement with experimental measurements. This has important implications for future interpretations of Stark effect spectroscopy in proteins with the aid of modeling. In particular, we show that this interplay between experiment and computation can be used to reliably infer the protonation state of the probe's neighboring amino acids, which were left uncertain in the original work.

COMPUTATIONAL METHODS

A model of hALR2 bound to **1** was constructed from available crystallographic data,[51] analyzed by molecular dynamics (MD) calculations, and refined, first by a QM/MM hybrid approach and then by moving-domain QM/MM (MOD-QM/MM). MOD-QM/MM is a self-polarization technique that is used to derive point-charges with ab initio quality for all atoms in the protein.[19,20]

Molecular Dynamics. A complete atomistic model of the human aldose reductase (hALR2) protein was constructed from the PDB entry 2PZN,[51] using the Maestro GUI included in the Schrödinger software suite.[52] The IDD393 ligand, {5-chloro-2-[[(3-nitrobenzyl)amino]carbonyl]phenoxy}acetic acid, was modified to the nitrile-containing analog, {5-chloro-2-[[(4-cyanobenzyl)amino]carbonyl]phenoxy}acetic acid (**1**)[42] (Figure 1). All residues were protonated based on a neutral pH. Tautomers for histidine residues were determined on the basis of the potential of making hydrogen bonds with neighboring groups. The MD simulations were carried out with the program Desmond,[53] which generated a cubic simulation box of 70 Å on each side, with periodic boundary conditions. The protein model was solubilized with a spherical point-charge (SPC) water model with neutralizing sodium and chloride ions for charged residues. Explicit Na^+ and Cl^- ions were randomly placed within the simulation box to mimic a 0.3 M NaCl solution. The WT model had a total of 39 473 atoms. The OPLSAA force field[54] was used for all atoms in the protein, the NADP cofactor, and the ligand. For the nitrile ligand, electrostatic potential (ESP) charges were computed by performing an in vacuo density functional (DFT) minimization using the program Jaguar.[55] This DFT calculation employed the Becke functional with Lee−Yang−Parr correction (B3LYP) and the 6-31g* basis set. OPLSAA was used for all other force field parameters in the ligand. During the MD, an additional 200 kcal/(mol Å^2) restraint potential was assigned to all backbone atoms to prevent unrealistic distortions from the crystal structure. The system was equilibrated in several steps. First a minimization via a steepest decent algorithm for 10 steps was performed with all heavy atoms under harmonic restraints, followed by an L-BFGS minimization algorithm, applied until the energy gradient was below 50 hartree. This initial step allowed for the solvent and salt ions to relax around the protein. With only the backbone restrained, the system was minimized again by steepest decent followed by an L-BFGS minimization

within an energy convergence of 5 hartree. The system then underwent a period of equilibration MD for 12 ps at 10 K with harmonic restraints on all heavy atoms and then 12 ps with harmonic restraints only on the backbone atoms. This equilibration procedure was repeated at 300 K, followed by an additional 24 ps of equilibration at 300 K and then a 5 ns production trajectory was collected. The orientation of the side chains of key residues within the binding pocket converged on an average value within the first 1−2 ns of the production trajectory and we therefore consider a 5 ns trajectory sufficient sampling of the protein and ligand conformations (see the Supporting Information). For all steps, a 2 fs integration interval for closely interacting atoms (<9 Å) and 6 fs integration interval for far-interaction atoms was used. The simulation was performed in an *NPT* ensemble at 300 K and 1 atm using a reversible system propagation algorithm (RESPA), a Nosé-Hoover thermostat, and a Martyna−Tobias−Klein barostat. A total of 1042 snapshots were recorded in 4.8 ps intervals.

Ensemble Averaging. Equation 2 relates the dot product between the difference dipole moment and the external electrostatic field vector to the transition energy. Thus, effectively, the experiment measures $F_{\parallel}$, the projection of the electrostatic field along the probe's dipole moment. This quantity is computed according to the expression

$$F_{\parallel} = \frac{1}{4\pi\varepsilon_o\varepsilon_{\text{eff}}} \sum_i \sum_{j\in i} \frac{q_{ij}}{|\mathbf{r}_p - \mathbf{r}_{ij}|^3}(\mathbf{r}_p - \mathbf{r}_{ij})\cdot\mathbf{n}_p \tag{3}$$

where i goes over all residues in the protein including the cofactors, and j goes over all atoms in residue i. q_{ij} and $\mathbf{r}_{ij}$ are the charge and position vector of the jth atom in residue i, respectively, $\mathbf{r}_p$ is the position vector at the nitrile's midpoint, $\mathbf{n}_p$ is a unit vector pointing from the nitrile probe carbon to the nitrogen, ε_0 is the permittivity of a vacuum, and ε_{eff} is the effective dielectric constant. In the case of the MD simulation, all electrostatic interactions are explicitly included and therefore $\varepsilon_{\text{eff}} = 1$. To compute an appropriate average $\langle F_{\parallel}\rangle$ and its standard deviation for the molecular dynamics trajectory, the 1048 snapshots were divided into nine blocks of 100 snapshots and a 10th block of 148 snapshots. For each block a numerical average value was determined and then an average and standard deviation for the blocks were evaluated. This form of ensemble averaging was performed using the molecular mechanics methods that were used to generate the dynamics trajectory. At higher levels of theory, this type of averaging would be intractable, and we therefore switch to the single-snapshot approach described below.

In the analysis of the molecular dynamics simulations, a noteworthy observation is the large standard deviation (1−2 MV/cm) observed for the field projection for a single trajectory. This behavior is not unexpected. As previously noted by Suydam and co-workers,[41] the nitrile probe can be significantly more sensitive to field shifts caused by local, structural rearrangement than to the average field shift due to a mutation. Our own analysis of the standard deviation as a function of different parameters (e.g., time step, simulation length, size of periodic box) for the MD calculation reached the same conclusions, providing further evidence that such standard deviations are true representations of the inherent noise in the finite temperature MD averaging of electric fields in proteins.

Single-Snapshot Approach. The protocol that would yield the highest accuracy would involve, in principle, computing the Stark shift for a large trajectory with a high level of theory and then using ensemble averaging methods to obtain a value for the Stark shift. However, it is necessary to balance accuracy with tractability. Furthermore, the implementation of a high level theory over the entire trajectory, if feasible, would still be susceptible to the level of noise noted in the MD, overshadowing any improvement in accuracy. With this in mind, we devised a method to incorporate quantum chemical refinement in a tractable and efficient way. First, we observed that the Stark shift, computed with molecular mechanics methods, displayed a dependence on several residue side chains, with varying dihedral angle orientations and distances from the nitrile probe. Therefore, to obtain a representative conformation, we computed the average Stark shift from the molecular dynamics trajectory. Then, the snapshot with the closest Stark shift value to the average was selected for higher-level calculations (QM/MM minimization and single-point MOD-QM/MM). In all cases, the largest difference between the MD average and the selected snapshot was 0.02 MV/cm. Thus, by matching the computed average Stark shift we are implicitly including conformational sampling and at the same time we remove the inherent noise from the long-range electrostatic interactions observed in the MD. We emphasize the fact that this procedure is predictive, as we do not use any knowledge of the experimental Stark shift in the selection of the snapshot.

Having selected the snapshot from each MD simulation according to the protocol specified above, the solvent molecules were deleted and an effective dielectric constant was employed to appropriately scale the electrostatic field shift. Early and perhaps overly simplified treatments of protein electrostatics in the context of continuum models employed an effective dielectric of ∼2−4 for a protein interior depending on the degree of hydration.[56] Subsequent developments in the study of protein electrostatics indicate that appropriate values range from 2 to 20[57,58] and that the basis for a protein dielectric is a measure of electrostatic interactions not explicitly included in the model.[58] Because our model ultimately treats polarization effects within the system on a quantum mechanical level, it is reasonable to expect a dielectric constant in the lower end of this range. In addition, as we note below, at no time in our simulation do solvent molecules enter near the cavity. Considering these observations, we chose $\varepsilon_{\text{eff}} = 2.0$.

Moving-Domain QM/MM. A moving-domain QM/MM partitioning scheme[19] was employed to replace the OPLSAA charges with ESP charges for all atoms in the system. Prior to the implementation of moving-domain QM/MM, the ligand in each of the selected snapshots was minimized at the QM/MM level using the program QSite[59] with the same level of DFT specified above. The 37 atoms of the ligand were designated as the QM region, while atoms outside the QM region were treated according to the OPLSAA force field constraining all protein coordinates. In the MOD-QM/MM procedure each protein amino acid, the NADP cofactor, and the ligand were taken as individual domains. MOD-QM/MM consists of partitioning the protein in residues (domains) and performing successive QM/MM calculations on each domain followed by ESP charge computation. Because the QM region interacts with MM charges at every step, the procedure incorporates polarization effects and produces a self-consistent set of ab initio quality charges for the entire system. Typically, convergence is achieved in three iterations. The in-house program modQ3M was used for this procedure. It is worth

noting that implementation of MOD-QM/MM, which provides a new set of charges for the atoms in the system, affects the partial charges on the carbon and nitrogen atoms of the nitrile. Although we did observe small changes in the partial charges on these atoms, they were omitted from the summation in eq 3, congruent to the protocol used by Webb and Boxer.[42] By omitting these atoms, we focus on the shift in the electrostatic field resulting from the changes in the protein and the rest of the ligand atoms.

RESULTS AND DISCUSSION

Stark Shift Prediction. For each mutant model, the $\Delta\langle F_{\|}\rangle$ value was computed, indicating the difference between the MD ensemble average electrostatic field projection onto the nitrile for the WT protein and that of a mutant protein:

$$\Delta\langle F_{\|}\rangle = \langle F_{\|}\rangle_{\text{Mutant}} - \langle F_{\|}\rangle_{\text{WT}} \quad (4)$$

The hALR2 protein forms a tight binding pocket around the nitrile-bearing ligand comprised of both hydrophobic and polar residues (see Figure 2). The structure and composition of the

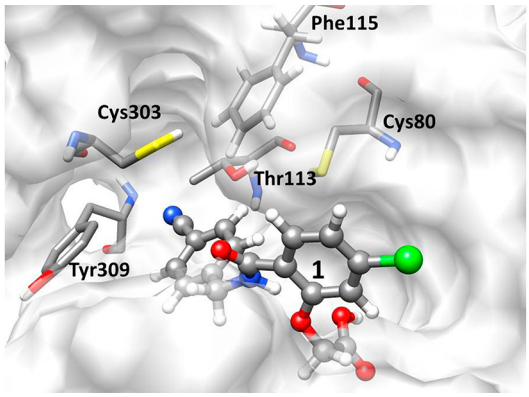

Figure 2. Cavity of the WT model of hALR2. The nitrile-containing probe molecule is shown in ball-and-stick representation. The residues in the binding pocket, which were targeted for mutation or significantly affected the $\Delta F_{\|}$ value, are shown in tube representation.

binding pocket prevent water molecules and salt ions from entering the pocket, and therefore, no significant contribution from the solvent to the Stark shift is expected. In fact, at no time during the MD trajectory did solvent or ion atoms come closer than 5 Å to the nitrile midpoint, resulting in a negligible solvent contribution ($\Delta\langle F_{\|}\rangle_{\text{Solvent}} \approx 0$). The MD ensemble average contribution to $\langle F_{\|}\rangle$ from the nitrile ligand structure was also consistent for all protein mutants and the wild type, also resulting in $\Delta\langle F_{\|}\rangle_{\text{Ligand}} \approx 0$.

Analysis of the protein contributions to the Stark shift revealed its pronounced dependency on the side chain orientation of key residues. Consider, for example, the dihedral angle ϕ, formed by the C_α–C_β–O–H group of Thr113 in WT. According to the MD simulation, this dihedral can adopt two possible conformations; in one (Figure 3A) $\phi \sim 52°$ and in the other (Figure 3B) $\phi \sim -90°$. The electrostatic field contour maps, plotted behind each respective structure, indicate profound differences in the electrostatic interactions at the probe midpoint. The computed difference in the $F_{\|}$ for the two conformers was 24.6 MV/cm. An investigation of other side chains near the binding pocket, even with a slightly polar functionality, i.e., C_α–C_β–S–H for Cys80, C_α–C_β–S–H for Cys303, and C_ε–C_γ–O–H for Tyr309, revealed a similar

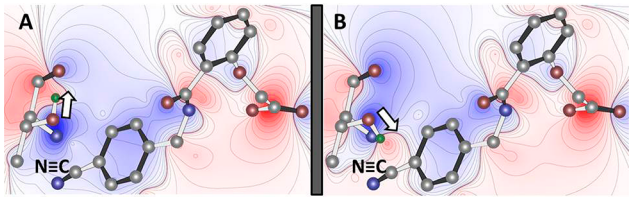

Figure 3. Electrostatic potential contour plot for two conformations of the WT hALR2 protein. Panel A shows the electrostatic potential for the ligand in the presence of Thr113 with a C_α–C_β–O–H dihedral angle of 52° and a clearly negative potential at the nitrile midpoint. Panel B demonstrates that a rotation of the Thr113 dihedral to −90° for these same atoms causes a significant change in the electrostatic field at the nitrile probe (25 MV/cm lower).

behavior. In general, a large range of values and signs for $F_{\|}$ was observed during the 5 ns production run, as much as 30 MV/cm, due to the sensitivity of $F_{\|}$ to the side chain orientations. Such fluctuations are, in turn, responsible for the somewhat large standard deviations in the values of $\langle F_{\|}\rangle_{\text{Total}}$ (1–2 MV/cm), which in many cases are larger than or equal to the measured shifts. Figure 3 clearly demonstrates such effect.

Table 1 shows the average MD Stark shift values for all single and double mutants considered. The MD ensemble average

Table 1. Summary of Computed Electrostatic Field Projection Values for Different Levels of Theory Compared to Experiment[a]

model	exp[b]	⟨MD⟩[c,d]	MM[c,e]	modQ3M[c,e]
wild type		−18.76 ± 1.49	−0.31	−11.17
ΔC80A	−1.0	+0.5	+2.1	−5.4
ΔC80N	+0.3	+4.5	+3.6	+1.8
ΔC303N	−3.2	−0.3	+0.0	−1.5
ΔC303DH	−6.5	+2.3	+1.9	−1.2
Δ*F*115Y	−5.7	−0.6	−3.5	−5.0
Δ*T*113A	−8.6	−12.7	−5.8	−5.4
Δ*T*113 V	−4.5	−11.7	−6.4	−7.6
Δ*T*113A/C303N	−10.4	−8.7	−3.3	−4.9
Δ*T*113A/C303DH	−7.3	−13.9	−5.8	−6.3
Δ*S*302R$^+$/C303D$^-$	−4.7	−17.8	−2.5	−3.8
rms Error		6.5	4.2	3.2
Mean Signed Error		−0.7	+3.2	−1.2

[a]All values are reported in MV/cm. [b]Experimental values were obtained from Webb and Boxer.[42] [c]The true value for the calculated electrostatic field in the WT model is reported for reference but subtracted from the true value for all other mutants to generate the value of the shift. [d]The MD ensemble average was computed as outlined in the methods section. [e]An effective dielectric of 2.0 was employed to scale electrostatic effects appropriately, accounting for the fact that calculations were done in vacuo.[56,60]

yielded an overall rms error of 6.5 MV/cm. For two of the 10 mutant models, the incorrect direction of the Stark shift was predicted and errors that were 1 order of magnitude larger than the measured shift were observed for three other models. We conclude then that despite the proper introduction of conformational sampling, the magnitude of the fluctuations are such that the agreement with the experimental shifts is marginal and unreliable. As pointed out in the methods section, we explored the possibility of using a single-snapshot approach that would allow us to incorporate both an average conformation and a higher level of theory. Thus, for each model trajectory a single snapshot was selected (according to

the criteria explained in the previous section) and the solvent molecules were removed. Using the single snapshot, $F_{||}$ was computed with eq 3, first using OPLSAA charges on all atoms in the system (referred as MM in Table 1) and then again using MOD-QM/MM charges (modQ3M in Table 1). In both calculations, an effective dielectric $\varepsilon_{eff} = 2.0$ was implemented in eq 3. The error dropped slightly for the set evaluated with MM charges, but the mean signed error increased as a result of consistently positively shifted values of $\Delta F_{||}$ compared to experiment. Further improvement of the model was accomplished by incorporation of polarization effects for all atoms in the system via the MOD-QM/MM approach, which essentially improves upon the MM method by incorporating polarization.[19] With this approach, a ∼50% reduction of the rms error was observed compared to MD results, and, perhaps more importantly the direction of the Stark shift was correctly predicted for every mutant model. The improvement observed using the QM protocol is therefore the result of a better treatment of the cavity's electrostatic effect onto the nitrile probe.

An interesting question to address in more detail is about how $\Delta F_{||}$ changes upon changing the level of theory and how the residues contribute to this shift. Consider, for instance, the mutant C303N. While the MD treatment produces a $\Delta F_{||}$ that is incorrect by over an order of magnitude and the MM treatment produces no $\Delta F_{||}$ whatsoever, the MOD-QM/MM treatment results in the correct direction of the shift (−1.5 MV/cm) in good agreement with experiment (−3.2 MV/cm). For this particular mutant, Figure 4 shows a residue screening

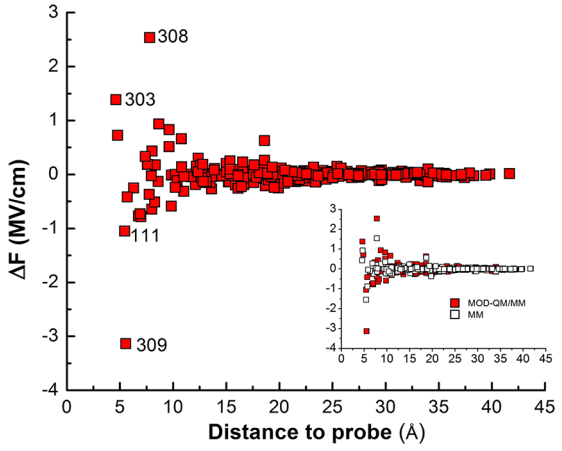

Figure 4. Difference in electrostatic field projection as a function of the distance between a residue (averaged over all atoms within the residue) and the probe for the MOD-QM/MM C303N model ($\Delta F_{||}^{C303N}$). Residues with $|\Delta F_{||}^{C303N}| > 1.0$ MV/cm are labeled. The inset figure shows a comparison between the polarized MOD-QM/MM (solid squares) model and the unpolarized MM model (open squares).

to determine the contribution of each amino acid to the shift and their average distance to the probe. Specifically, it shows that Tyr309 and Asp308 have a predominant effect on the overall shift. What causes the mutation at 303 large contributions from other residues, in particular 309 and 308? The $C_\varepsilon-C_\gamma-O-H$ dihedral angle Tyr309 rotates from −10.6° in the WT model to −30.2° in the C303N model. That alone resulted in a contribution to the predicted field shift of approximately −1.5 MV/cm when OPLSAA charges were assigned to atoms. However, that same geometry yielded a shift of approximately −3.0 MV/cm when charges were refined with MOD-QM/MM. Thus, the mutation introduces a new net force on the side chain of Tyr309, resulting in the 40° rotation and a consequent shift in the electrostatic field. Asp308 responds to the new force as well by rotating the charged carboxylic acid side chain −30°, from ∼104° in WT to ∼74° in C303N. These residues are all facing directly into the binding pocket and undergo rearrangement with nearly no perturbation to the other residues located around the ligand and the ligand itself. These structural rearrangements do not seem to affect any purported hydrogen-bonding interactions but rather subtly shifts the minimum energy conformations. In this new conformation, polarization effects further enhance the electrostatic contribution coming from these residues, resulting in a predicted negative shift of the projected electrostatic field along the nitrile bond, in agreement with experimental results. The overall effect of polarization is shown in Figure 4 (inset), which demonstrates that the redistribution of charges in the two models results in a more substantial contribution from many more amino acids. Both MM and MOD-QM/MM treatments are, nevertheless, consistent in determining the residues that contribute significantly to $\Delta F_{||}$. While we note that the agreement with experiment is not perfect, we find encouraging that the incorporation of ab initio quality structures and electrostatics performed in such a practical manner leads to a favorable comparison. Thus, we assert that any error introduced by the single-snapshot approach is outweighed by the improvement in the results from the higher-level of theory.

Protonation State Prediction via Combined Stark Effect Spectroscopy and Computational Analysis. Three mutant models examined in this study, C303D, T113A/C303D, and S302R/C303D, substitute a titratable residue near the binding pocket of the nitrile probe. At near neutral pH, the protonation state of mutations involving aspartic acid is ambiguous (a priori). This study shows, for the first time, that Stark effect spectroscopy in combination with adequate computational modeling can be used to predict the protonation state of a residue. The data in Table 2 indicate the computationally predicted $\Delta F_{||}$ values for the different protonation states of the potentially charged species introduced by the mutation. The notation C303DH indicates the neutral, protonated form of aspartic acid, whereas $C303D^-$ indicates a charged, deprotonated form. Due to the large pK_a of the arginine side chain, this is assumed positively charged. In both the C303D and T113A/C303D models, C303DH demonstrated a lower error, with good quantitative agreement,

Table 2. Electrostatic Field Projection Data Compared to WT for Titratable Mutations[a]

model	exp[b]	modQ3M[c]
$\Delta C303D^-$	−6.5	−17.80
ΔC303DH	−6.5	−1.2
$\Delta T113A/C303D^-$	−7.3	−18.9
ΔT113A/C303DH	−7.3	−6.3
$\Delta S302R^+/C303D^-$	−4.7	−3.8
$\Delta S302R^+$/C303DH	−4.7	−1.5

[a]All values are reported in MV/cm. [b]Experimental values were obtained from Webb and Boxer.[42] [c]An effective dielectric of 2.0 was employed to scale electrostatic effects appropriately, accounting for the fact that calculations were done in vacuo.[56,60]

compared to C303D$^-$. This result indicates that the environment surrounding position 303 increases the pK_a of the carboxylic acid side chain. In the case of the double mutant S302R/C303D, the high pK_a of the arginine and consequential positive charge lowers the pK_a of the carboxylic acid side chain of Asp303. Therefore, a model in which Arg302 and Asp303 are both charged and form a salt bridge yields a predicted $\Delta F_{\|}$ that is within 0.9 MV/cm of experiment. Part of this high accuracy results from the fact that the salt bridge stabilizes the geometry of the binding pocket, minimizing the error of selecting a single geometry to represent the system for the higher-level theory calculations. Thus, we emphasize that the combination of Stark effect spectroscopy and adequate computer modeling can be used to resolve the protonation state of titratable residues in any WT protein by introducing the nitrile probe, mutating the residue of questionable protonation state to an aliphatic residue, and then using the shift in absorption frequency to determine the shift in the electrostatic field projection. The experimental value is likely to match closely the modeling results of one protonation state while deviating significantly for the other.

CONCLUSIONS

This study utilized MD, MM, and moving-domain QM/MM methods for the calculation of the electrostatic field within a protein. The perturbation in the protein structure caused by a mutation required molecular dynamics ensemble averaging to generate an appropriate geometry. Using a geometry predicted by MD averaging as a starting point, we implemented a QM/MM geometry minimization of the ligand followed by a polarization technique on the entire protein, obtaining very good agreement with experiment in the Stark shift for a number of single and double mutations. In particular, we obtained complete agreement between experiment and theory in the direction of the Stark shift and resolved the questionable protonation state of residues in three mutant models. This work demonstrated the use of experimental data to validate reliable computational-based protocols for the prediction of complex behavior in proteins, such as the substantial change in the electrostatic field resulting from a residue mutation. This study further underscored the importance of polarization effects in the electrostatic properties of biomolecules. Finally, the combination of Stark effect spectroscopy and computational modeling was used to determine the protonation states of ionizable residues. This work does, however, outline the difficulty of selecting an appropriate geometry for high-level theory calculations and demonstrates that models may deviate from reality if a good geometry is not selected. We offer these results toward the development of a computational method to predict the $\Delta F_{\|}$ resulting from perturbation to a biological system with the goal of expanding the method to improve overall accuracy as well as address the more complicated case of hydrogen-bonding interactions between the nitrile and neighboring residues. Directions of future study will include the use of a QM/MM driven-MD or Monte Carlo simulation (with semiempirical potentials as the QM part) or the use of an expanded QM region that would permit a better account of the effect of hydrogen bonds involving the probe.

ASSOCIATED CONTENT

Supporting Information

Additional information as noted in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Tel (860) 486-0591. Fax (860) 486-2981. E-mail: jose.gascon@uconn.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

J.A.G. is thankful for financial support from the Camille and Henry Dreyfus foundation and from NSF (CAREER Award CHE-0847340).

ABBREVIATIONS:

B3LYP, Becke functional with Lee−Yang−Parr correction; DFT, density functional theory; EE, electrically embedded; ESP, electrostatic potential; hALR2, human aldose reductase; L-BFGS, limited-memory Broyden−Fletcher−Goldfarb−Shanno; MD, molecular dynamics; MM, molecular mechanical; MV/cm, megavolts per centimeter; modQ3M, moving-domain quantum mechanics/molecular mechanics; PES, potential energy surface; QM, quantum mechanical; rms, root-mean-square; SPC, spherical point-charge; VSE, vibrational Stark effect; WT, wild type

REFERENCES

(1) Benkovic, S. J.; Hammes-Schiffer, S. *Science* **2003**, *301*, 1196.
(2) Stora, T.; Lakey, J. H.; Vogel, H. *Angew. Chem., Int. Ed.* **1999**, *111*, 399.
(3) Doyle, D. A.; Cabral, J. M.; Pfuetzner, R. A.; Kuo, A.; Gulbis, J. M.; Cohen, S. L.; Chait, B. T.; MacKinnon, R. *Science* **1998**, *280*, 69.
(4) Hong, H.; Szabo, G.; Tamm, L. K. *Nature* **2006**, *2*, 627.
(5) Simonson, T.; Archontis, G.; Karplus, M. *J. Phys. Chem. B* **1997**, *101*, 8349.
(6) Simonson, T.; Archontis, G.; Karplus, M. *J. Phys. Chem. B* **1999**, *103*, 6142.
(7) Vasilyev, V.; Bliznyuk, A. *Theor. Chem. Acc.* **2004**, *112*, 313.
(8) Gilles-Gonzalez, M.; Gonzalez, G. *J. Appl. Physiol.* **2004**, *96*, 774.
(9) Merrill, A. H., Jr.; Sullards, M. C.; Wang, E.; Voss, K. A.; Riley, R. T. *Environ. Health Perspect.* **2001**, *109*, 283.
(10) Resh, M. D. *Cell . Signal.* **1996**, *8*, 403.
(11) Exner, T. E.; Mezey, P. G. *J. Phys. Chem. A* **2002**, *106*, 11791.
(12) Exner, T. E.; Mezey, P. G. *J. Phys. Chem. A* **2004**, *108*, 4301.
(13) Gao, A. M.; Zhang, D. W.; Zhang, J. Z.; Zhang, Y. *Chem. Phys. Lett.* **2004**, *394*, 293.
(14) Li, S.; Li, W.; Fang, T. *J. Am. Chem. Soc.* **2004**, *127*, 7215.
(15) Mei, Y.; Zhang, D. W.; Zhang, J. Z. H. *J. Phys. Chem. A* **2005**, *109*, 2.
(16) Nakano, T.; Kaminuma, T.; Sato, T.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2000**, *318*, 614.
(17) Nakano, T.; Kaminuma, T.; Sato, T.; Fukuzawa, K.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2002**, *351*, 475.
(18) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27.
(19) Gascon, J. A.; Leung, S. S. F.; Batista, E. R.; Batista, V. S. *J. Chem. Theor. Comput.* **2006**, *2*, 175.
(20) Menikarachchi, L. C.; Gascón, J. A. *J. Mol. Model.* **2008**, 1.
(21) Sproviero, E. M.; Newcomer, M. B.; Gascón, J. A.; Batista, E. R.; Brudvig, G. W.; Batista, V. S. *Photosynth. Res.* **2009**, *102*, 455.
(22) Anisimov, V. M.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D. *J. Chem. Theor. Comput.* **2007**, *3*, 1927.
(23) Banks, J. L.; Kaminski, G. A.; Zhou, R.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **1999**, *110*, 741.
(24) Dehez, F.; Angyán, J. G.; Gutiérrez, I. S.; Luque, F. J.; Schulten, K.; Chipot, C. *J. Chem. Theor. Comput.* **2007**, *3*, 1914.
(25) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J. P. *J. Chem. Theor. Comput.* **2007**, *3*, 1960.

(26) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A. *J. Comput. Chem.* 2002, *23*.
(27) Lamoureux, G.; Roux, B. *J. Phys. Chem. B* **2006**, *110*, 3308.
(28) Nakagawa, S.; Mark, P.; Ágren, H. *J. Chem. Theor. Comput.* **2007**, *3*, 1947.
(29) Soteras, I.; Curutchet, C.; Bidon-Chanal, A.; Dehez, F.; Ángyán, J. G.; Orozco, M.; Chipot, C.; Luque, F. J. *J. Chem. Theor. Comput.* **2007**, *3*, 1901.
(30) Xie, W.; Gao, J. *J. Chem. Theor. Comput.* **2007**, *3*, 1890.
(31) Jorgensen, W. L.; Jensen, K. P.; Alexandrova, A. N. *J. Chem. Theor. Comput.* **2007**, *3*, 1987.
(32) Andrews, S. S.; Boxer, S. G. *J. Phys. Chem. A* **2000**, *104*, 11853.
(33) Andrews, S. S.; Boxer, S. G. *J. Phys. Chem. A* **2002**, *106*, 469.
(34) Boxer, S. G. *J. Phys. Chem. B* **2009**, *113*, 2972.
(35) Fafarman, A. T.; Webb, L. J.; Chuang, J. I.; Boxer, S. G. *J. Am. Chem. Soc.* **2006**, *128*, 13356.
(36) Levinson, N. M.; Bolte, E. E.; Miller, C. S.; Corcelli, S. A.; Boxer, S. G. *J. Am. Chem. Soc.* **2011**, *133*, 13236.
(37) Park, E. S.; Andrews, S. S.; Hu, R. B.; Boxer, S. G. *J. Phys. Chem. B* **1999**, *103*, 9813.
(38) Saggu, M.; Levinson, N. M.; Boxer, S. G. *J. Am. Chem. Soc.* **2011**, *133*, 17414.
(39) Sigala, P. A.; Fafarman, A. T.; Bogard, P. E.; Boxer, S. G.; Herschlag, D. *J. Am. Chem. Soc.* **2007**, *129*, 12104.
(40) Suydam, I. T.; Boxer, S. G. *Biochemistry* **2003**, *42*, 12050.
(41) Suydam, I. T.; Snow, C. D.; Pande, V. S.; Boxer, S. G. *Science* **2006**, *313*, 200.
(42) Webb, L. J.; Boxer, S. G. *Biochemistry* **2008**, *47*, 1588.
(43) Xu, L.; Cohen, A. E.; Boxer, S. G. *Biochemistry* **2011**, *50*, 8311.
(44) Le, H. Q.; Zayhowski, J. J.; Goodhue, W. D. *Appl. Phys. Lett.* **1987**, *50*, 1518.
(45) Schuda, F.; Stroud, C. R.; Hercher, M. *J. Phys. B.* **1974**, *7*, L198.
(46) Ash, W. L.; Zlomislic, M. R.; Oloo, E. O.; Tieleman, P. *Biochim. Biophys. Acta* **2004**, *1666*, 158.
(47) Sheinerman, F. B.; Norel, R.; Honig, B. *Curr. Opin. Struct. Biol.* **2000**, *10*, 153.
(48) Simonson, T. *Curr. Opin. Struct. Biol.* **2001**, *11*, 243.
(49) Warshel, A.; Papazyan, A. *Curr. Opin. Struct. Biol.* **1998**, *8*, 211.
(50) Lösche, M.; Feher, G.; Okamura, M. Y. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 7537.
(51) Ruiz, F.; Hazemann, I.; Mitschler, A.; Andrzejj, J.; Schneider, T.; Karplus, M.; Podjarny, A. *Acta Crystallogr. Sect. D* **2004**, *60*, 1347.
(52) *Maestro v. 9.2*; Schrödinger, Inc.: New York, 2010.
(53) *Desmond Multisim 3.8.3.36*; D. E. Shaw Research: New York, 2010.
(54) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.
(55) *Jaguar v. 7.6*; Schrödinger, Inc.: New York, 2010.
(56) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144.
(57) Alexov, E.; Mehler, E. L.; Baker, N.; M. Baptista, A.; Huang, Y.; Milletti, F.; Erik Nielsen, J.; Farrell, D.; Carstensen, T.; Olsson, M. H. M.; Shen, J. K.; Warwicker, J.; Williams, S.; Word, J. M. *Proteins: Struct., Funct., Genet.* **2011**, *79*, 3260.
(58) Warshel, A.; Dryga, A. *Proteins: Struct., Funct., Genet.* **2011**, *79*, 3469.
(59) *QSite v. 5.7*; Schrödinger, Inc.: New York, 2010.
(60) Schultz, C. N.; Warshel, A. *Proteins: Struct., Funct., Genet.* **2001**, *40*, 400.