Reliable and Accurate Prediction of Single Residue pKa Values Through Free Energy Perturbation Calculations

Dilek Coskun[†], Wei Chen[†], Anthony J. Clark[†], Chao Lu[†], Edward D. Harder[†], Lingle Wang[†], Richard A. Friesner[‖], Edward B. Miller[†,*]

†Schrödinger, Inc., 1540 Broadway, New York, New York 10036, United States

‖Department of Chemistry, Columbia University, New York, 3000 Broadway, MC 3110, New York 10036, United States

ABSTRACT: Accurate prediction of the pKa's of protein residues is crucial to many applications in biological simulation and drug discovery. Here we present the use of free energy perturbation (FEP) calculations for the prediction of single protein residue pKa values. We begin with an initial set of 191 residues with experimentally determined pKa values. To isolate sampling limitations from force field inaccuracies, we develop an algorithm to classify residues whose environments are significantly affected by crystal packing effects. We then report an approach to identify buried histidines that require significant sampling beyond what is achieved in typical FEP calculations. We therefore define a clean dataset not requiring algorithms capable of predicting major conformational changes on which other pKa prediction methods can be tested. On this data set, we report an RMSE of 0.76 pKa units for 35 ASP residues, 0.51 pKa units for 44 GLU residues, and 0.67 pKa units for 76 HIS residues.

I. Introduction.

Accurate calculation of the pKa's of protein residues is an essential molecular modeling task in many important applications in biological simulation and drug discovery[1,2]. Protein conformations, ligand binding affinities, and protein-protein interfaces can all exhibit a critical dependence upon the protonation state of key titratable residues. Use of the wrong protonation state can lead not only to quantitative errors, but to qualitatively incorrect conclusions.

However, obtaining high precision in pKa evaluations is a very challenging task. To date, a variety of empirical and continuum solvent-based approaches have been employed. For highly solvent exposed residues, the best of these methods performs reasonably well. As solvent exposure decreases, the ability to obtain results that correlate well with experiment diminishes. Explicit solvent-based MD simulations have also been explored, most often via the use of pH dependent simulations based on lambda dynamics[3]. Such approaches have shown some promise but have not yet been tested on the large data sets needed to make a statistical assessment of error.

There are a large number of published protonation prediction methods available that listing them all is beyond the scope of this paper. Therefore, we briefly highlight some of the more popular and often benchmarked methods to illustrate the range of reported accuracies and dataset sizes used. One of the most popular empirical pKa predictors is PROPKA3.[4] PROPKA3 reports a combined RMSE for aspartate and glutamate of 0.79 pKa units over a dataset of 201 residues and an RMSE of 1.0 pKa units for histidine for a dataset of 31 residues. Continuum solvent based approaches such as H++[5] and MCCE2[6] are often cited. H++ reports an RMSE of ~1.4 pKa units for 201 ionizable groups in 23 high-quality protein structures. MCCE2 reports an RMSE of 0.77 pKa units for 225 surface residues and 1.2 pKa units for 80 buried residues. Buried residues are presented separately as a significantly greater challenge than surface residues in MCCE2. Molecular dynamics-based method MD/GB/TI reports an RMSE of 1.4 pKa units for the pKa prediction of 80 different residues (20 each for aspartate, glutamate, lysine, and histidine)[7].

Constant pH molecular dynamics (CpHMD) approaches are an active area of work in multiple groups. These methods are able to handle coupled protonation states, a limitation of our approach here. However, the computational costs of these methods are non-trivial, often employing implicit solvent methods to cope, and the dataset sizes reported in the literature are often much smaller than the empirical methods, which provides challenges for statistical validation of these methods.

Swalis, York and Roitberg report a CpHMD method that combines explicit solvent MD with protonation state changes performed under implicit solvent at fixed intervals[8]. The method is benchmarked for 10 titratable residues in hen egg white lysozyme (7 ASP, 2 GLU, and 1 HIS residue) with an overall RMSE of 0.82 pKa units.

Radak et al report a CpHMD method that uses Monte Carlo moves consisting of short non-equilibrium MD trajectories with explicit solvent[9]. Their publication reported results for a set of titratable residues within Staphylococcal nuclease. Of the residues with precise experimental pKa values (this excludes experimental pKa values such as < 2.20), there are 6, 11, and 2 ASP, GLU and HIS residues, respectively. The overall RMSE for this set is 0.58 pKa units[10].

Harris and Shen report a GPU CpHMD approach that uses a GB continuum solvation model and reports an RMSE of 0.92, 0.61, and 1.04 pKa units for ASP, GLU and HIS datasets, respectively[11]. The number of titratable residues tested were 42,

54, and 18 residues for ASP, GLU and HIS, respectively, one of the largest datasets for CpHMD validation.

Recently, Harris et al produced a follow-up paper with the use of explicit solvent[12]. The dataset consisted of titratable residues in six proteins. RMSEs reported are 0.76, 0.69, and 0.92 pKa units for ASP, GLU, and HIS datasets, respectively, which showed improvements compared to the use of implicit solvent. The number of titratable residues tested were 28, 31, and 10 for ASP, GLU and HIS, respectively.

In several recent papers, we have demonstrated the ability of the Schrodinger FEP+ package, using recent versions of the OPLS force field, to predict the effects of mutation on protein-protein binding to an accuracy of better than 1 kcal/mole, including the challenging case of charge changing mutations[13–15]. These results provide a reason for optimism that pKa calculations can similarly be addressed successfully. In the present paper, we report an initial exploration of this hypothesis, using standard data sets from the literature as our initial test cases.

In order to appropriately assess the accuracy of rigorous physics-based pKa calculations, a key problem, heretofore not discussed in the literature, must be overcome. The data sets mentioned above are based on high resolution crystal structures of small proteins. These proteins have a high surface to volume ratio, which leads to a profusion of crystal packing interactions affecting a significant number of the histidines and carboxylate residues of the protein. Even if we assume that the secondary structure framework is minimally affected by the crystal environment, loops constrained by crystal packing in the crystal may well have a significantly different conformation in solution. And as the FEP simulations we perform are of relatively short duration, it is unlikely in such cases that the loop will have time to reorganize into the correct conformation. The challenge is therefore to disentangle errors caused by problematic starting conformations of the residue and its environs, as opposed to those due to the computational methodology. We note that a similar analysis would apply to all of the pKa prediction methods cited above.

To address this problem, we formulate a scoring function which evaluates the degree to which crystal packing affects each of the test cases in our data set. A cutoff value of the scoring is defined using a training set of residues, and performance of the FEP calculations on residues that score below the threshold (implying minimal crystal packing perturbation) and above it is computed. A significant improvement in the RMS error and correlation coefficient can be observed for the residues deemed to be minimally affected by crystal packing, which we interpret as reflecting the performance that one can expect if the correct solution structure is used as a starting point. We note that our scoring function can also be used to determine which loops are likely distorted by crystal packing in any crystal structure; if any such loops impact key functional interactions (such as ligand binding), refinement of these loop structures may be necessary to achieve accurate results in, for example, FEP simulations of protein-ligand binding.

Once crystal packing effects have been (approximately) removed, the remaining errors in pKa prediction must be due to one of three sources: experimental uncertainty, incomplete sampling, or errors in the force field model. Incomplete sampling is most likely to arise when the crystal structure is adapted to the neutral form of the residue, and a transformation to the charged species creates a highly unfavorable set of interactions (note that this is far more likely to be an issue for a histidine residue as opposed to a carboxylic acid, simply because the great majority of carboxylates are crystallized in the charged form). In some cases, the unfavorable set of interactions of the charged species induces a significant conformational change, which the limited FEP MD sampling is not able to fully access. We have developed an algorithm, specific to histidine, which can automatically identify such cases, based on both the crystal structure and the MD trajectory from the FEP simulation. This algorithm captures a high fraction of the largest outliers in the histidine data set, and explains why the raw histidine results are substantially worse than those for carboxylates (not only in our approach, but in others as well). We develop a simple, one parameter empirical correction for these buried histidine cases which in essence reverts their error distribution to a value comparable to that obtained for the overall data set.

We can then estimate force field error/experimental error (which cannot be disentangled easily) by examining a data set from which both the suspected crystal packing and incomplete sampling outliers (as defined above) have been removed. The resulting RMS error is consistent with that obtained in our prior work on FEP binding affinity calculations in protein-protein complexes [13]. In these latter efforts, we have discovered a few unusual structures (carboxylic acid making an exceptional number of hydrogen bonds) where the errors (estimated indirectly based on binding affinity data) are much larger, which we attribute to the failure to explicitly treat polarization in the OPLS4 force field (testing of the performance of an explicitly polarizable model for these systems is currently in progress). However, our assessment based on the residue environments for the current data set is that none of the cases discussed in the current paper fall into this category.

The paper is organized as follows. In section II, we describe the literature data set that we have assembled, comprising 105 Asp/Glu and 86 His cases in total. Section III summarizes our FEP methodology for evaluating pKa's of the residues in the data set, including force field improvements motivated by comparing calculated pKa's with experiment. Section IV compares computed and experimental pKa values for all residues in the data set, without considering crystal packing effects. In section V, we then develop a model to identify residues that are strongly affected by crystal packing and train the parameters of the model using the experiment/theory comparisons obtained in Section IV (using standard statistical techniques to assess overfitting). In section VI, we present attempts to resolve errors due to limited sampling by using additional simulation time, and then identifying cases via an automated algorithm which are suspected to suffer from sampling limitations that are presumed to require simulation time well beyond the scope of this work. In section VII the discussion, we compare the results obtained when crystal packing and limited sampling effects are excluded with those obtained from the unfiltered data sets and draw conclusions with regard to the accuracy and reliability of our explicit solvent based pKa prediction methodology. In section VIII, the conclusion, we summarize our results and outline future directions.

II. Literature Data Set for pKa Calculations

We use a data set of 191 experimentally obtained pKa values for individual protein residues from 44 separate proteins. Of these 191 residues, 157 residues are from crystal structures and the remaining 34 from NMR structures. This dataset has overlap

with the PKAD[16], although for some proteins we refer to more recent experimental data. For example, for human hemoglobin, PDB ID 1HHO, we use a 2010 reference[17] for HIS pKa values rather than a publication from 1980[18] referenced by the PKAD. Supplemental Table S9 lists the PDB IDs of each protein and the associated reference for single residue pKa values. The limited size of the data set is due primarily to the paucity of reliable experimental measurements that are available in the literature. For example, while the PKAD reports experimental pKa values for 225 histidine residues, many of them are duplicates of the same protein with different PDB IDs, for example, PDBIDs 1A6K and 1A6M.

III. FEP Methodology for pKa Prediction

The pK_a of a titratable group within a protein, $pK_{a,protein}$, is determined by its intrinsic pK_a, also referred to as model pK_a, $pK_{a,model}$ and the perturbation created by the protein environment, ΔpK_a.

$$pK_{a,protein} = pK_{a,model} + \Delta pK_a$$

Equation 1. Calculation of single titratable group pK_a

The ΔpK_a, also referred to as the pK_a shift, is calculated using free-energy perturbation[15] to arrive at a ΔΔG of deprotonating a single residue in the protein versus solution environment. Equation 2 shows the approach for calculating ΔpK_a. Calculations are run with a default of 20 ns using the OPLS4 force-field[19] in SPC explicit water in the 2022-1 Schrödinger release[20]. This release includes additional refinements to the HIS side chain parameters, although no fitting was done to pKa results. Details about the generation of improved HIS side chain parameters are in Appendix B in the supporting information. Sodium and chloride ions are added to match experimental salt concentration, when available, otherwise we default to 0.15 M. Supplemental Table S9 lists the salt concentration used for each titratable residue. Complete details of the protein FEP methodology are discussed in reference [15].

$$\Delta pK_a = \frac{\Delta\Delta G}{\ln(10)\,RT} = \frac{1}{\ln(10)\,RT}\left(\Delta G_{protein} - \Delta G_{solution}\right)$$

Equation 2. Calculation of the pK_a shift.

In Equation 2, $\Delta G_{solution}$ is calculated using just the titratable residue capped with ACE and NMA, run in explicit solvent.

For the imidazole ring of histidine, there are two distinct microscopic pK_a values for the delta and epsilon nitrogens. We protonate in place the target histidine to HIP, the charged histidine form, and then perform FEP calculations mutating the HIP to HIE and separately from HIP to HID. The free energy difference, ΔΔG, is calculated by considering both mutations as shown:

$$\Delta\Delta G = \Delta\Delta G_{HIP\to HIE} - RTln\left(1 + 10^{pK_{a,model(HIE)} - pK_{a,model(HID)}} exp\frac{\Delta\Delta G_{HIP\to HIE} - \Delta\Delta G_{HIP\to HID}}{RT}\right)$$

Equation 3. Calculation of the free energy difference for histidine.

The model HIS micro-pKa values in Equation 3 are taken from the literature[21] and are shown in Table 1.

Table 1. Microscopic pKa values for histidine tautomers

HIS Tautomer	Microscopic pKa
HID	6.92
HIE	6.53

In summary for the prediction of ASP/GLU pKa shift, we directly apply Equation 2 simulating with FEP just the perturbation from the charged to neutral state. For histidine pKa prediction, we perform FEP twice, mutating from HIP to either HID or HIE, calculate an overall ΔΔG using Equation 3, insert the result into Equation 2, and then use that result in Equation 1 with the $pK_{a,model}$ being the HIE microscopic pKa.

We utilize the OPLS4 force field for all FEP+ pKa simulations that follow. As is discussed in detail in ref. [19], parameters for ionic species were reoptimized in OPLS4, fitting initially to quantum mechanical calculations of hydrated dimer and cluster (ion plus multiple water) energies for a number of important ions, including the charged states of the carboxylate and histidine side chains. First van der Waals parameters were refit for the ions, and then torsions were optimized appropriately in response to the van der Waals changes.

Refinement of the torsional parameters was then carried out for the carboxylate side chains (ASP and GLU by fitting to pKa data, using the same data set as we discuss herein). This led to a significant improvement in the RMS error as compared to the older set of torsional parameters and was essential for achieving the quality of results discussed below. Our objective in the present paper with regard to the carboxylate data set is to partition the remaining error into crystal packing, incomplete sampling, and the remaining experimental/force field component, as discussed in the Introduction above.

No explicit fitting to the histidine pKa results was carried out in the development of OPLS4, or subsequently, so this data set constitutes an independent test of the accuracy of the model. Achieving reliable predictive capabilities for HIS has been very challenging in previous work on pKa prediction; for example, in ref. [11], which uses a continuum solvent based simulation approach, a correlation coefficient of nearly zero was obtained for a histidine data set that is a subset of the one used in the present paper. As in the case of the carboxylates, we can obtain an assessment of the force field accuracy, to within the limits of the experimental error bars, by identifying and filtering out outlier cases due to crystal packing and incomplete sampling.

As the work here is focused on single residue pKa prediction, we set the protonation state of the surrounding residues as if the system was at the pH of the experimental pKa of the target residue. If the surrounding residue's experimental pKa or model

pKa (for cases where the experimental pKa is unknown) is within 1 pKa unit of the titratable residue's pKa, then we kept the protonation state of the neighboring residue as assigned by Schrödinger's Protein Prep Wizard[22,23]. For example, for a target ASP with pKa of 4, a surrounding ASP with pKa of 5.5 is set to be protonated. If the surrounding ASP also had a pKa of 4, the protonation state was assigned by Schrödinger's Protein Prep Wizard[22,23].

A final point is that there are a small number of cases where it appears as though the pKa's of two or more residues are coupled, and a more powerful simulation approach than single residue pKa prediction (fixing the remaining residues at either their experimental values, when known, or otherwise at estimated values) is required. In such a situation, improved results can be obtained via constant pH simulations utilizing lambda dynamics to simultaneously optimize all of the relevant protonation states. Our constant pH simulation methodology is currently under development and will be described in more detail in a subsequent publication. The approach is similar to published work from other groups[11,12,24–26], with the exception of the use of the OPLS4 force field with explicit solvent as opposed to alternatives such as CHARMM or AMBER.

IV. Initial FEP Results for pKa prediction

In Figure 1 we show the initial performance of protein FEP for pKa prediction on the complete data set of 191 residues.

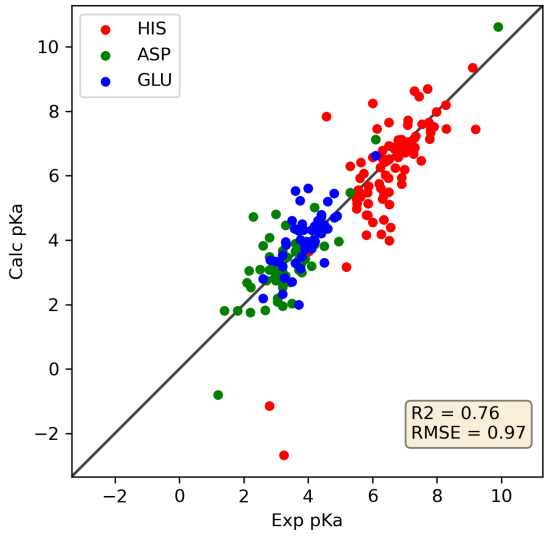

Figure 1. Correlation between experimental pKa and calculated pKa for the full data set of 191 residues.

The plot shows three large outliers. These outliers are 3EBX A:HIS6, 3SSI A:HIS43, and 2LZT A:ASP66. These outliers all shift the FEP calculated pKa to be lower than experiment, preferring the deprotonated form of the residue.

As we will discuss in more detail below, we attribute the presence of these large outliers (as well as a considerable number of less extreme outliers in Figure 1) to the difficulty we have using relatively short FEP simulations (20 ns) to sample large conformational changes in the protein structure. There are two major factors potentially leading to the requirement for substantial conformational modification of the starting experimental structure. Firstly, crystal packing interactions can lead to significant distortions of loop geometries, as compared to what would be observed in solution (which is where the experimental pKa measurements are carried out). Secondly, the experimental structure is obtained at a specific pH, which mandates one particular protonation state of the residue of interest. However, the FEP simulation to predict the pKa of the residue requires sampling both protonation states equally well. In a nontrivial number of cases, a change of protonation state from that present in the experimental structure leads to a very substantial conformational rearrangement of the protein in response. In such cases, the calculated altered protonation state can be significantly destabilized as compared to the lowest free energy basin that is accessible experimentally.

An illustrative example of the effects of crystal packing is presented in Table 2: the results for the prediction of HIS 52 in ovomucoid inhibitor across both several crystal structures and an NMR structure. For the NMR structure, 1OMU, the first NMR model was arbitrarily used. Notably, the pKa prediction for the NMR structure is significantly closer to experiment than either crystal structure and is of course free from crystal contact artifacts.

Table 2. Prediction of the pKa of HIS52 starting from three experimental structures[a]

	1PPF (Xtal)	2OVO (Xtal)	1OMU (NMR)	Experiment
Predicted pKa	5.50	5.76	6.47	7.50
Error	-2.00	-1.74	-1.03	N/A

[a.] PDB IDs 1OMU and 1PPF are both of Turkey ovomucoid inhibitor, as is the experimental result. PDB ID 2OVO is from Silver Pheasant, 98% sequence identical to Turkey ovomucoid inhibitor.

Figure 2 shows the difference between crystal structure 2OVO and NMR structure 1OMU. The titratable histidine, HIS52, is shown near the center of the image. As shown, the histidine is in nearly an identical position with only a minor change to its loop. The only significant change is in the position of the N-term of the protein which in the NMR structure folds towards HIS52 while in the crystal structure is pointing away from HIS52, packed against other crystal mates (not shown). Here the influencing crystal contacts are not directly in contact with the titratable residue but rather interact with residues forming the environment around the titratable residue. In the next section we demonstrate an empirical model to classify titratable residues as being affected by crystal contacts, either directly or through contacts to environment residues.

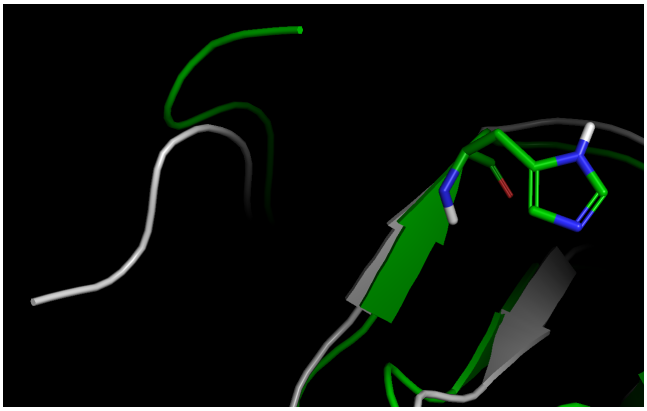

Figure 2. Comparison of the protein environment around HIS52 for crystal structure 2OVO (white) versus NMR structure 1OMU model 1 (green).

An illustrative example of a residue which is in the neutral state in the crystal structure and is then poorly sampled when converted to the charged state, is 3SSI HIS43, the largest FEP outlier in Figure 1. The environment of HIS43 in the crystal structure is shown in Figure 3. It can be seen that the residue is completely buried. When a hydrogen is added to yield a charged histidine, the effect is to create a buried charge with no solvent exposure. The FEP simulation relieves the worst of the electrostatic clashes by flipping the backbone residue, but cannot, even in a 100 ns trajectory (Table 4), enable significant solvation of the histidine (or formation of a stable salt bridge, which would be an acceptable substitute). The large underestimation of the pKa as compared to experiment (in essence asserting that the charged histidine state is massively destabilized) is the result.

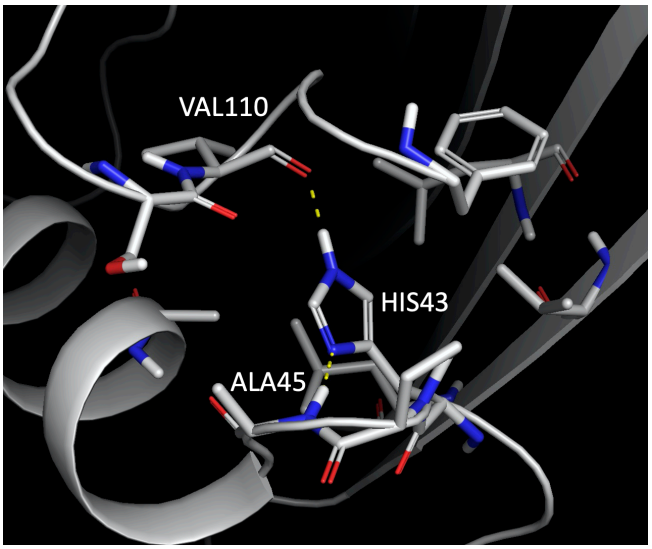

Figure 3. The environment around 3SSI A:HIS43. The histidine in the neutral form is an acceptor for a buried NH group, the backbone NH of A:ALA45. The histidine is completely buried with both imidazole nitrogens acting as either a donor or acceptor to the protein backbone.

In order to statistically analyze the errors across our data set, we need automated approaches to identify the systematic problems with crystal packing and incomplete sampling of an alternative charge state that have been discussed above. We present initial versions of such approaches in the next two sections. Larger and more diverse experimental data sets will enable refinement of these tools going forward.

V. Empirical Model for Identifying Residues Substantially Impacted by Crystal Packing

In Supplemental Table S9, we list the proteins from which the data sets discussed in Section II are drawn. The proteins are quite small, typically with an average of 163 residues. This leads to a large number of crystal contacts present in the structures that we use as a starting point for FEP simulations, due to the high surface to volume ratio of proteins of this size. As noted above, significant crystal packing effects on loops can lead to a distorted starting point for a solution phase simulation (which is where the pKa experiments are carried out) and hence erroneous prediction results if the solution phase conformation cannot be accessed by the FEP simulation in the allotted time frame.

We define two qualitatively different types of crystal packing effects which our model is then aimed at identifying. In the first, the target residue (the residue for which a pKa calculation is to be performed via FEP) is itself in a loop, and that loop has a sufficient number of crystal contacts in proximity to the residue to make it likely that in solution, a different loop conformation would be adopted. It is of course possible that this loop conformation would, perhaps accidentally, yield a similar pKa to that of the loop in the crystal structure, but there is no guarantee that this is the case.

The second situation is one in which the target residue can be located on any type of protein segment (loop, helix, or strand), but is in close contact with a neighboring loop which has a substantial number of crystal contacts. Such a loop can significantly impact the pKa of the target residue in two different directions. Firstly, it can block access of solvent to the side chain moiety, thus destabilizing the charged state relative to the uncharged state. Secondly, it can facilitate salt bridge formation by creating a favorable geometrical situation for this to occur, in this case stabilizing the charged state relative to the uncharged state.

The goal of our scoring function is not to unambiguously identify residues whose pKa's would be definitively shifted by structural reorganization in the absence of crystal packing. Coming to such a conclusion requires a lot more computation; after all, just because a loop has many crystal contacts does not mean that the crystal conformation would not be the lowest free energy conformation in solution. Rather, it is to delineate the subset of residues from our data set for which a significant pKa shift upon such reorganization would have a strong possibility of occurring. We can then assess the accuracy of pKa prediction when these cases are removed from the database. While the cases remaining in the database may themselves be subject to some structural reorganization, the magnitude will presumably be significantly smaller, and in some cases may be remedied by the MD simulation carried out during the FEP calculations. Note that following the above argument, we expect to remove some test cases for which the prediction starting from the crystal structure has good agreement with experiment. This is unproblematic as long as a substantial fraction of the initial cases re-

main, as the primary objective is to create a clean data set allowing a better assessment of MD sampling and force field problems.

The scoring function is constructed via a two-part process:

(1) Identify any loops whose conformations would likely have a significant impact on the target residue pKa. As outlined above, the list of candidate loops includes both the loop on which the target residue resides (assuming it is on a loop) and also any loops in close contact with the target residue. We define a surrounding residue to be in close contact if there is a hydrogen bond to the titratable residue, including to its backbone, or the surrounding residue has a backbone atom within 5 Å from the titratable residue. If this surrounding residue is part of a loop, we consider the crystal contacts to that loop as well.

(2) For each relevant loop obtained from (1), enumerate a list of relevant crystal contacts (those sufficiently close to the target residue) and calculate an overall score to which each contact contributes, with the magnitude of the contribution dependent upon the contact distance, type of interaction, and whether that interaction involves backbone or side chain atoms. We optimize the scores of the various types of contacts by fitting to a subset of the experimental test cases as described below but impose physical constraints on the parameter values; for example, the impact of a backbone hydrogen bond should be larger than a simple pair of heavy atom contacts, shorter contacts should be more important than more distant ones, and backbone contacts should have more impact than those between a pair of side chains. A detailed description of the scoring function, and the approach used to optimize the parameters, is provided in Appendix A in the supporting information.

We have designed the scoring function to be a binary classifier: that is, the target residue is either considered to be substantially impacted by crystal packing (in which case we will be removing it from the data set to produce a "clean" subset of cases, hopefully able to achieve good solution pKa prediction starting from the crystal structure), or minimally impacted by crystal packing (in which case the residue will be retained in the clean data set). In order to check for overfitting, we adopt leave one out (LOO) calculations in which the scoring function is built without including data for each target residue.

To train the scoring function, we designate 15 titratable cases as true positive crystal contacts out of the 157 residues solved with x-ray crystallography. Each of the 15 cases are excluded one at a time and the scoring function re-optimized with the remaining 14 true positive cases. The one excluded case is then scored to determine if it would have been classified as impacted by crystal contacts.

Of the 15 cases, one case could not be recovered during leave-one-out optimization, 2RN2 A:ASP70, so this case is not excluded from the final statistics despite what we feel are significant crystal contacts. The scoring function weights various contacts (backbone-backbone, backbone-sidechain) between residues in the environment of the titratable residue and residues in crystal mates. Evidently, the crystal contacts to the environment residues around 2RN2 A:ASP70 are not similarly observed in any of the other 14 true positive cases leading to the failure during LOO optimization. This exposes a limitation of our dataset size. However, the purpose of the work here is to construct an unbiased, quantitative method to exclude cases suspected of being influenced by crystal packing which we feel is accomplished. What remains is a cleaner and unbiased dataset to evaluate pKa prediction accuracy.

For purposes of evaluating the accuracy of our pKa predictions, we define the clean data set from the LOO results. A total of 36 target residues are eliminated from the data set, leaving 121 residues from crystal structures plus the 34 residues from NMR, a total of 155 titratable residues to be incorporated into the pKa statistics. A complete list of the classification of each of the target residues is provided in Supplemental Table S3. We note that the fraction of residues that remain is sufficiently large for both carboxylates (79) and histidines (76) to constitute a critical mass of data. Furthermore, the distribution of pKa shifts that remains is similar (although not identical) to that in the original data set, so that there are a sufficient number of large pKa shifts to enable a meaningful assessment of the methodology (an objective that would be difficult to achieve if the cases were overwhelmingly composed of small shifts). Note that this data set can be used to more accurately evaluate the performance of other approaches to pKa prediction.

Finally, in order to further explore the validity of our hypothesis that crystal contacts with a key loop are responsible for many of the outliers, we select an example outlier, repredict the conformation of a proximate loop suspected to be influenced by a large number of crystal contacts, and carry out pKa calculations for the target residue using the top ranked structure obtained from loop prediction. The loop prediction is performed using the latest version of the Schrodinger Prime loop prediction methodology[27].

Briefly, the loop prediction algorithm proceeds as follows. We previously constructed a scoring function of cbeta to cbeta contacts, similar to a united residue approach[28]. As the backbone of the loop is being constructed, we score nascent loops with this scoring function to triage unlikely cbeta-cbeta contacts. Separately, we bin backbone positions by spatial position and the formation of backbone-to-backbone hydrogen bonds to consider diversity of loops. Once the backbone is placed, side chains are optimized on the loop and the surrounding environment, and the loop plus surrounding environment is minimized[29]. Side chain optimization and minimization is done with the OPLS4 forcefield[19] and the VSGB2 implicit solvent model[30].

Table 3. pKa prediction of ASP121 in PDB 7RSA with and without contacting loop prediction[a]

Structure	ASP121 pKa
Experimental pKa Measurement[31]	3.0
7RSA	4.84

7RSA Loop A:64-A:71 Prediction, ASP121	4.46
7RSA Loop A:64-A:71 Prediction, ASH121	3.63

[a] Loop prediction of residues A:64-A:71 is done twice, with the target residue, ASP121, in either the charged or neutral ASH state. The lowest energy loop prediction for each ASP state is reported.

Table 3 lists the results of pKa prediction when a loop interacting with the target residue is in an alternative conformation compared to the experimental coordinates. This case, 7RSA ASP121, is one of the cases that is considered a true positive for the purposes of building a crystal contact severity scoring function, as previously discussed in Section V. The target residue, ASP121, its interaction with loop A:64-A:71, and that loop's interaction with a crystal mate is shown in Figure 4. Here ASP121 is shown to have limited solvent exposure.

The pKa prediction using the deposited coordinates for 7RSA results in a pKa prediction of 4.84 compared to the experimental value of 3.0; the charged form of ASP is being overly penalized, presumably due to the very limited solvent accessibility caused by ASP121 packing against the A:64-A:71 loop.

As shown in Table 3, prediction of the loop with ASP121 in the protonated ASH state was necessary to yield pKa results in close agreement with experiment. This loop prediction shifted the backbone away from ASP121, moving the THR70 NH shown in Figure 4, by 2.93 Å from its native position into solution and away from the ASP. Evidently, this produces a loop conformation that permits proper sampling of both the ASP and ASH states during FEP, despite the fact that in the crystal structure, which was solved at pH 5.3, the ASP will still be charged[32].

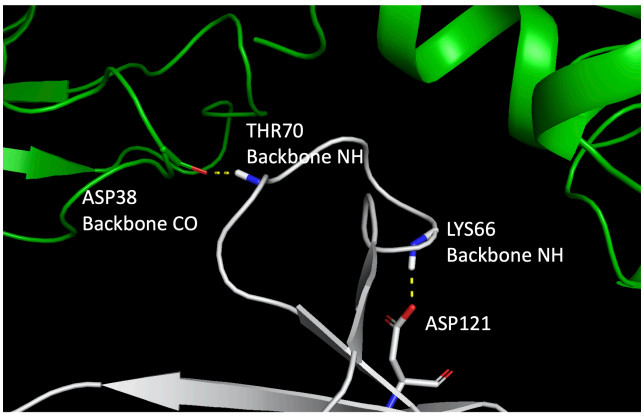

Figure 4. Interactions of ASP121 with loop A:64-A:71 and that loop's interaction with a crystal mate in PDB ID 7RSA. The asymmetric unit protein is shown in white with the crystal mate in green. ASP121 forms a hydrogen bond to LYS66 backbone NH. The interacting loop is closely surrounded by crystal mates and form a backbone to backbone hydrogen bond between THR70 NH and the ASP38 carbonyl in the crystal mate.

Figure 5 shows the performance of pKa prediction following the exclusion of 36 cases selected by our empirical model as likely to be affected by crystal packing. Excluded now are two of the three severe outliers that were previously mentioned in Section IV and shown in Figure 1, as well as cases which were not outliers but must be excluded for consistent application of empirical model of crystal packing severity. The overall R^2 improved from 0.76 to 0.79 and the RMSE was reduced from 0.97 to 0.88 pKa units. The one remaining exceptionally large outlier is 3SSI A:HIS43 which is addressed in the next section.

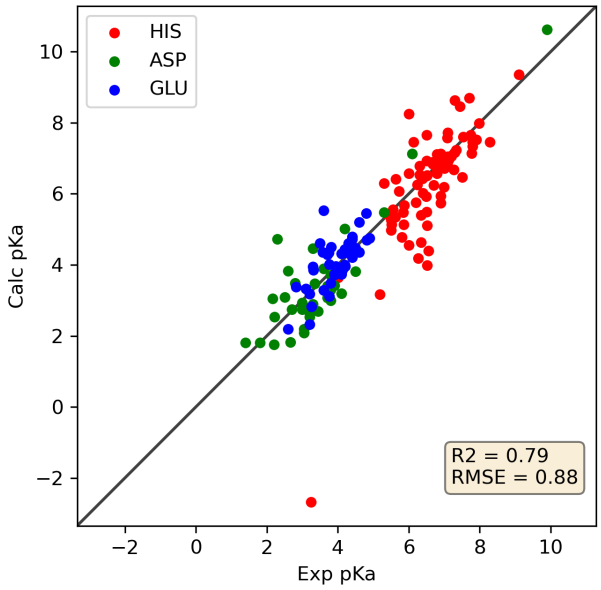

Figure 5. Correlation between experimental pKa and calculated pKa for the data set of 155 residues which excludes 36 residues that were classified as being potentially affected by crystal packing.

VI. Detailed Analysis of Buried Histidine Outliers

Shown in Figure 3 is the local crystal structure of the remaining large outlier, 3SSI HIS43. Based on the experimental pKa of this residue as compared to the pH at which crystallization was carried out, we can confidently assert that the crystal structure incorporates the neutral form of the HIS residue. This histidine is entirely buried with zero solvent exposed surface area. The delta nitrogen is acting as an acceptor for the backbone NH of ALA45 while the epsilon nitrogen is a hydrogen bond donor to the backbone carbonyl of VAL110. If a proton is added to the delta nitrogen, a severe electrostatic clash will be introduced with the ALA 45 backbone NH; furthermore, a net charge will be buried in a highly hydrophobic region. Understandably, the experimental[33] pKa of this histidine is severely shifted down to 3.25, favoring the neutral form, relative to free histidine which has a pKa of 6.38[21].

While it is trivial to qualitatively predict that this histidine will prefer the neutral state, a quantitative prediction of the pKa of this group is challenging. The FEP calculation considers the charged state to be even more unfavorable relative to the neutral state than experiment, yielding a pKa of -2.68. The error here is presumed to be due to the necessity of a major structural reorganization of the environment for the protonated form of the histidine which is not correctly sampled over our default 20 ns simulation.

The 3SSI HIS43 test case is an extreme example of a phenomenon that can be seen more generally in our HIS data set. Histidine residues that are partially or fully buried, and manifest a substantial number of hydrophobic contacts with the protein, are invariably in the neutral state in the experimental structure unless they form a salt bridge with a carboxylate group. Such structures may be initially highly unfavorable for the charged

form of HIS, and hence require extensive conformational sampling to achieve their lowest free energy state. A short molecular dynamics simulation, as is carried out in the FEP+ pKa prediction protocol, may not be sufficient to achieve the requisite conformational change, thus leading the charged state to be calculated as too unfavorable as compared to experiment. Major problems of this type are much less likely to arise for carboxylic acid residues because they generally occur in the charged form in a crystal structure (although incomplete sampling for a carboxylate case is certainly possible, and likely accounts for some of the discrepancies between theory and experiment seen in Figure 5). Consequently, in the present paper, we carry out a detailed analysis of the buried HIS cases in our data set, as discussed below.

The first step in this analysis is to identify the subset of HIS cases in which at least one of the two histidine nitrogens is buried, and for which the environment provides significant hydrophobic contacts (if there are few hydrophobic contacts, the molecular dynamics will most likely be able to reorganize the charged HIS state without a problem). A key parameter is the minimum number of hydrophobic contacts, which we set at 6. How we precisely count hydrophobic contacts is explained below. Table 4 presents a list of all of the HIS cases which satisfy these criteria. It can be seen that many (although not all) of the entries in Table 4 are significant outliers; indeed, they constitute a high percentage of the large outliers in the HIS data set.

The next question is whether any cases can be accurately corrected by extending the FEP simulations to a longer (but still practically accessible) timescale. Table 4 shows the results obtained when the FEP simulations are extended to 100 ns rather than the default 20 ns runs. Most cases are altered only minimally, but 3RN3 A:48 improves significantly, and 3SSI A:43 more modestly.

An important feature of the simulation results, the persistence of salt bridge formation, is also displayed in Table 4. We have already excluded from this table cases where the HIS is hydrogen bonded to a carboxylate in the crystal structure (thus forming a salt bridge in the charge state). However, in a fair number of cases in Table 4, a salt bridge, not present originally, forms during the course of the FEP simulation of the charged state. There is a sharp divide between cases where salt bridge persistence is greater than or less than 70%. Strikingly, all of the cases in the latter category have reasonably good agreement with experiment if a 100 ns simulation is used in FEP. In the low salt bridge persistence category, with the exception of 3RN3: A48 (discussed immediately below), the errors in pKa prediction are always in the direction of insufficiently stabilizing the charged state, consistent with the hypothesis that the simulation is unable to carry out a conformational change needed to optimize the free energy of that state, with the error being greater than 1 pKa unit.

3RN3:48 has a highly unusual environment which leads to adequate solvation of the charged state despite the low level of salt bridge persistence. Specifically, we find that in our FEP simulations, on average the charged HIS side chain makes 5.73 close interactions with acceptors, a qualitatively larger number than is observed for any other residue in Table 4, and substantially more than one would expect to see for a HIS monomer in bulk solution.

We define an acceptor interaction to be either a hydrogen bond or an aromatic CH interaction. The donor here is coming from the protonated histidine sidechain, which has no acceptors. The donor-acceptor distance must be less than 2.85 Å, and the donor angle better than 90º. When the acceptor is a carbonyl, we measure the distance from the donor to the acceptor oxygen and separately the distance to the carbonyl carbon attached to the acceptor oxygen. The unfavorable hydrogen-carbon distance must be 0.5 Å or larger than the favorable hydrogen-oxygen distance. In essence we require good hydrogen bonding or aromatic CH---O geometry.

It is reasonable to expect that such an environment does not require substantial reorganization to achieve the correct low free energy basin of the charged HIS species. A snapshot of the average structure observed from the molecular dynamics simulation of 3RN3:A48 HIP is shown in Figure 6, with the HIP/acceptor interactions indicated.

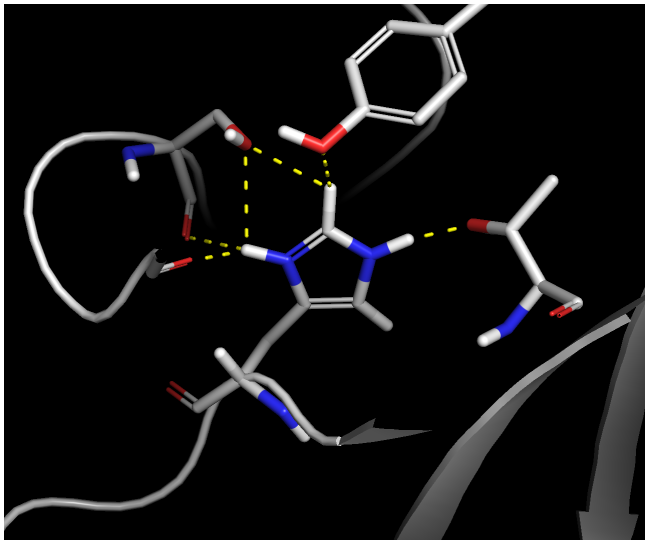

Figure 6. An individual MD frame during the 100 ns FEP simulation of 3RN3 A:48. This frame comes from the replica with lambda=0, the HIP state is physically present. This frame shows 6 acceptor interactions, including hydrogen bond and aromatic CH interactions.

In Figure 7 we highlight another case, 1YPH C:HIS40, a buried histidine that is able to achieve close agreement with experiment (predicted pKa is 0.14 pKa units from experiment). We have examined the FEP molecular dynamics trajectory, and we observe that the protonated histidine residue is able to form a stable salt bridge with the carboxylate residue C:GLU70 over the course of the simulation. Figure 7 shows the initial structure and a snapshot from the FEP simulation in which the salt bridge has been formed. The persistence of the salt bridge after formation is documented in this figure which plots the closest histidine – glutamate distance as a function of simulation time. In contrast with that is 3SSI A:HIS43, shown in Figure 3, which scarcely forms a salt-bridge over the course of the simulation and is the largest outlier in the dataset with the predicted pKa in error by -5.93 pKa units, severely over penalizing the protonated state.

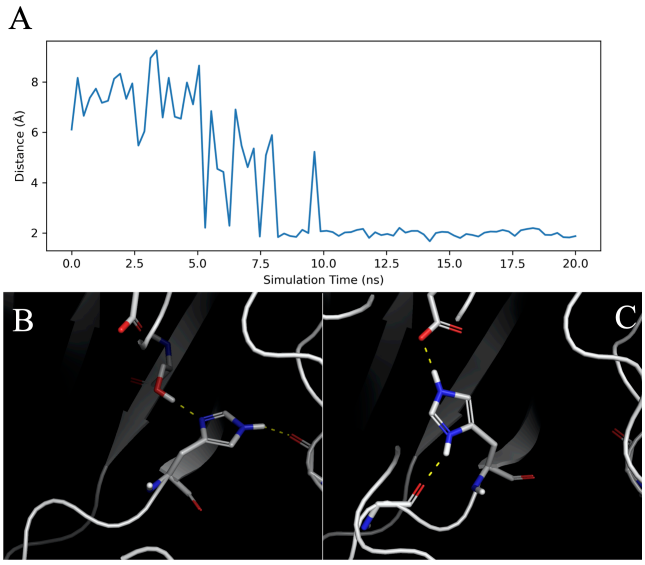

Figure 7. Formation of a new salt-bridge for 1YPH C:HIS40 when the histidine is protonated. (A**)** A plot of the distance in ångstroms between C:GLU70 versus simulation time. Reported is the closest distance between either HD1 or HE2 and carboxylate oxygen on GLU70. As can be seen, a stable salt-bridge forms at 10 ns. (B) The deposited coordinates for 1YPH C:HIS40 showing the neutral histidine is 5.88 Å from GLU70. (C) The final frame of the simulation at 20 ns where the histidine has flipped over to form the salt-bridge with GLU70.

The above results suggest a systematic approach to treating buried HIS cases in prospective applications of FEP for pKa prediction (i.e. when the experimental answer is not known in advance):

(1) Compute the solvent exposed surface area of both histidine nitrogens and count the number of hydrophobic atoms within 5 Å of the histidine sidechain heavy atoms. We define a hydrophobic atom to be any carbon atom not bonded to a heteroatom, except for the carbon atoms in the aromatic rings of TYR and TRP which we also consider hydrophobic. If at least one histidine nitrogen is buried and there are 6 or more hydrophobic contacts nearby, without a preexisting hydrogen bond to a negatively charged group (e.g. ASP or GLU), then a further investigation of the detailed behavior of the residue in an extended FEP simulation is required, as described in the steps below.

(2) Run a 100 ns FEP simulation for each case identified using the criteria of (1) above, and track the salt bridge persistence and average number of acceptor interactions. If it is above 70% or there are on average 5 or more acceptor interactions, accept the result as is. Otherwise, proceed to step 3 below.

(3) Apply an empirical correction to the FEP calculated pKa to take into account the estimated sampling error. Fitting the 9 cases which have between 6 and 10 hydrophobic contacts yields a value for the correction of 1.5 pKa units. 3SSI A:43 has a much more hydrophobic environment (15 hydrophobic contacts) and hence it makes sense that the error is much larger (5.07 pKa units). There is not enough data at present to build an interpolation function to reliably handle cases in this regime (one would want to interpolate between the correction of 1.5 and the 5.07 value obtained for 3SSI), but at the very least, one would be alerted to the presence of a challenging problem, and could explore more extensive conformational sampling options if it was important to obtain an accurate pKa value for the residue in question (as opposed to classifying it as having a very large pKa shift, which might be sufficient for most practical applications).

Table 4. Histidine residues where at least one nitrogen is buried, which makes a significant number of protein hydrophobic contacts, and which are not initially forming a salt bridge in the experimental structure[a]

Case	Exp pKa	Pred pKa (20ns)	Pred pKa (100 ns)	Salt-Bridge Persistence (100 ns)	Average Acceptor Interactions for HIP (100 ns)	Num. Hydrophobic Contacts	Buried N Atoms	pKa Correction	Pred pKa w/ Correction (100 ns)
1A6K A:119	6.26	4.18	4.52	60.05%	1.21	10	NE	+1.5	6.02
1DE3 A:137	5.80	4.78	4.79	1.20%	1.01	6	NE, ND	+1.5	6.29
1DE3 A:50	7.70	8.69	8.67	91.63%	1.86	6	ND	0	8.67
1DWR A:119	6.56	4.39	4.59	54.31%	1.04	8	ND	+1.5	6.09
1EY0 A:46	5.86	5.14	4.93	48.56%	1.19	9	ND	+1.5	6.43
1H4G B:11	6.52	3.98	4.00	38.04%	0.41	7	ND	+1.5	5.5
1HHO A:112	7.53	7.60	7.66	98.80%	1.72	13	ND	0	7.66
1PTD A:227	6.90	5.94	5.48	61.48%	2.67	8	NE	+1.5	6.98

1YPH C:40	7.20	7.06	7.68	79.19%	2.99	8	NE,ND	0	7.68
2CPL A:126	6.34	4.63	5.11	0.0%	2.26	10	ND	+1.5	6.61
3RN3 A:48	6.00	8.25	7.11	31.10%	5.73	8	NE, ND	0	7.11
3SSI A:43	3.25	-2.68	-1.82	0.48%	3.38	15	NE, ND	+5.0	3.18
3SSI A:106	6.00	4.55	4.99	3.59%	1.90	10	ND	+1.5	6.49
6GST A:83	5.18	3.17	3.16	0.24%	1.01	8	ND	+1.5	4.66

[a.] A histidine is considered buried if either the epsilon or delta nitrogen has no solvent exposed surface area and is forming a hydrogen bond. Salt-bridge persistence refers to the percent of the simulation time (100 ns) in which a salt bridge has formed between the protonated histidine and a carboxylate (GLU, ASP, etc). This salt-bridge was not present in the PDB deposited experimental structure. A protein-HIS side chain hydrophobic contact is defined by a hydrophobic atom of the protein approaching within 5A of a heavy atom in the HIS imidazole ring. We have estimated the number of contacts required for a critical mass as 6. An acceptor interaction is a hydrogen bond or aromatic CH bond where the donor is either a proton or aromatic CH on the protonated histidine.

VII. Discussion

Table 5. Final pKa prediction results across the various datasets

Amino Acid	Full Data Set			Crystal Packing Corrected Data Set			Packing Corrected + Buried N Removed	Packing Corrected + Empirical Correction
	ASP	GLU	HIS	ASP	GLU	HIS	HIS	HIS
Number of Cases	48	57	86	35	44	76	66	76
R^2	0.71	0.49	0.59	0.79	0.50	0.61	0.65	0.64
RMSE (pKa units)	0.85	0.62	1.20	0.76	0.51	1.08	0.59	0.67
RMSE (kcal/mol)	1.17	0.86	1.65	1.05	0.71	1.49	0.81	0.92

We are now in a position to determine the RMS errors and correlation coefficients obtained from FEP based pKa prediction for the three titratable residues under study (ASP, GLU, and HIS) as we systematically modify the data sets (Table 5). We consider four data sets: (1) the original, unfiltered data sets for all three residues; (2) data sets with crystal packing cases, as identified by the algorithm described in Section V, removed; (3) a HIS data set with both crystal packing cases and buried N cases removed that fail to form a salt bridge with a minimum salt-bridge persistence of 70% of the simulation or with an average of 5 acceptor interactions, as described in Section VI and Table 4. This is excluding the cases where we apply an empirical correction; (4) A HIS data set with the empirical corrections applied as listed in Table 4 to the remaining buried cases (as determined by salt bridge persistence and the buried NH descriptor). For each data set/protocol, we present the number of cases included, the correlation coefficient R^2, and the RMSE in pKa units and kcal/mol.

When the buried HIS outliers are either removed or corrected, the results for all three residue types are now comparable to those obtained for OPLS4 for an extensive series of small molecule binding affinity predictions[34,35], as well as for residue mutation free energy changes at protein-protein interfaces[13,15], taking into account for the latter that changing the charge of a buried residue could produce intractable sampling problems (similar to those seen here for the buried HIS residues discussed in Section VI). The RMS error range of approximately 0.7-1 kcal/mol has been difficult to reduce further, unsurprisingly since estimates of experimental error for both binding and pKa measurements are on the order of 0.5 kcal/mol RMS error. In the present case, this level of accuracy is suitable for assigning protonation states as a function of pH when studying challenging problems such as the prediction of pH dependent binding of antibodies, a key functionality for proteins such as the neonatal receptor FcRn. The accuracy is significantly better than results reported in previous work[4–7,11,12], although those papers did not remove crystal packing or remove or empirically correct extreme sampling cases as we have done here.

It is interesting to note that identification and removal of the ten buried HIS cases that fail to form a persistent salt-bridge or multiple acceptor interactions during FEP brings the HIS RMS error down to that observed for the ASP and GLU calculations. This is encouraging with regard to the performance of the OPLS4 force field model, and the approach used therein to optimize parameters for charged systems. As noted above, the HIS data set represents an independent test set for the protocol, with

no direct fitting of HIS parameters to pKa measurements. The results shown here argue that the force field is well balanced with regard to its treatment of both positive and negative ions.

It is clear from the above results that crystal packing effects can lead to a major perturbation of the predicted pKa's of protein titratable residues in solution, due to biasing of the initial conformation towards that seen in the crystal. Truly converged sampling would of course overcome any such difficulties. However, when standard FEP simulation protocols are used (even when the simulation time is extended to 100 ns), the trajectories are apparently not long enough to reach the correct conformational basin in a significant fraction of cases. The (admittedly anecdotal) success reported above in improving pKa predictions in such cases, via loop predictions in the solution environment, support this interpretation. Investigation of a larger set of analogous test cases would provide useful validation of our Prime based loop prediction algorithms. Furthermore, the algorithm we have developed to detect a high degree of crystal packing effects could be applied routinely at the beginning of any structure-based drug discovery project, and the affected loops repredicted. The resulting protein structures could then be tested for improvements in calculations other than pKa prediction, such as FEP computations of ligand binding affinity.

The results in the present paper provide a road map as to how to achieve accurate and robust predictions in prospective applications. When embarking on a project using a crystal structure as a starting point, the crystal packing classification algorithm is initially used to identify loops which are likely to have a modified conformation in solution. If the pKa of one or more residues impacted by these loops is deemed important, it will be necessary to apply structural refinement methods (e.g. long time MD simulations and/or loop predictions) to obtain a more accurate structure prior to running pKa calculations. Future work will focus on understanding in detail what sorts of refinement protocols are necessary for such an effort to yield useful pKa predictions.

As noted above, the fixed charge force field optimized in OPLS4 can break down when the environment of a carboxylate deviates in an extreme way from the normal solution environment, either via a large number (5-6) of persistent hydrogen bonds or an unusually small number of hydrogen bonds. In these cases, use of an explicitly polarizable model appears to be necessary to properly model the changes in effective hydrogen bond and salt bridge strength across different carboxylate environments if a first principles simulation approach is employed. Work in this direction is ongoing in our group.

VIII. Conclusion

In the present paper, we have described an FEP based algorithm for predicting pKa's of protein residues, using the OPLS4 force field. We show that a naïve assessment using standard experimental pKa data sets fail to properly evaluate the intrinsic accuracy of the force field model, due to the presence of significant crystal packing effects and also buried HIS cases which require a major conformational change that is not accessible via a 100 ns molecular dynamics simulation. We have developed automated algorithms to identify residues likely to be impacted by crystal packing, and buried neutral histidine residue which are likely to require major conformational change when converted to the charged HIP state. A simple one parameter empirical correction scheme has been developed which, at present for a limited data set, achieves significant improvement in the agreement with experiment for the relevant test cases.

In addition to providing confidence in prospective pKa prediction via FEP+/OPLS4 for new systems, our results in the present paper can serve a number of other functions. Firstly, they define a clean data set, not requiring algorithms capable of predicting major conformational changes, on which other pKa prediction methods (for example those employing continuum solvent models, which require much less computational effort than explicit solvent based FEP) can be tested. Secondly, a complementary data set is defined in which structural refinement of the protein crystal structure is required in order to achieve the appropriate lowest free energy structures for various residue protonation states. The accuracy of structural refinement approaches can then be tested by comparing the pKa's predicted by such approaches with experiment. One could also further validate these structural refinement predictions experimentally, e.g. via NMR or cryo-EM data at sufficient resolution.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Detailed description of the construction of the crystal packing severity scoring function, table of defined crystal packing true positives, table of final weights for the crystal packing scoring function, detailed description of the histidine force field parameterization and final set of force field parameters, table of all titratable residues, their experimental and predicted pKa values, and the experimental reference (PDF)

AUTHOR INFORMATION

Corresponding Author

* Edward B. Miller – Email: Ed.Miller@schrodinger.com

Author Contributions

The manuscript was written through contributions of all authors. / All authors have given approval to the final version of the manuscript.

Notes

The authors declare the following competing financial interest(s): R.A.F. has a significant financial stake in Schrödinger, Inc., is a consultant to Schrödinger, Inc., and is on the Scientific Advisory Board of Schrödinger, Inc.

REFERENCES

(1) Nielsen, J. E.; Gunner, M. R.; García-Moreno, B. E. The PKa Cooperative: A Collaborative Effort to Advance Structure-Based Calculations of PKa Values and Electrostatic Effects in Proteins. *Proteins* **2011**, *79* (12), 3249–3259. https://doi.org/10.1002/prot.23194.

(2) Navo, C. D.; Jiménez-Osés, G. Computer Prediction of PKa Values in Small Molecules and Proteins. *ACS Med. Chem. Lett.* **2021**, *12* (11), 1624–1628. https://doi.org/10.1021/acsmedchemlett.1c00435.

(3) Knight, J. L.; Brooks, C. L. 3rd. Lambda-Dynamics Free Energy

Simulation Methods. *J. Comput. Chem.* **2009**, *30* (11), 1692–1700. https://doi.org/10.1002/jcc.21295.

(4) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical PKa Predictions. *J. Chem. Theory Comput.* **2011**, *7* (2), 525–537. https://doi.org/10.1021/ct100578z.

(5) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: Automating PK Prediction and the Preparation of Biomolecular Structures for Atomistic Molecular Modeling and Simulations. *Nucleic Acids Res.* **2012**, *40* (W1), W537–W541. https://doi.org/10.1093/nar/gks375.

(6) Song, Y.; Mao, J.; Gunner, M. R. MCCE2: Improving Protein PKa Calculations with Extensive Side Chain Rotamer Sampling. *J. Comput. Chem.* **2009**, *30* (14), 2231–2247. https://doi.org/10.1002/jcc.21222.

(7) Stanton, C. L.; Houk, K. N. Benchmarking PKa Prediction Methods for Residues in Proteins. *J. Chem. Theory Comput.* **2008**, *4* (6), 951–966. https://doi.org/10.1021/ct8000014.

(8) Swails, J. M.; York, D. M.; Roitberg, A. E. Constant PH Replica Exchange Molecular Dynamics in Explicit Solvent Using Discrete Protonation States: Implementation, Testing, and Validation. *J. Chem. Theory Comput.* **2014**, *10* (3), 1341–1352. https://doi.org/10.1021/ct401042b.

(9) Radak, B. K.; Chipot, C.; Suh, D.; Jo, S.; Jiang, W.; Phillips, J. C.; Schulten, K.; Roux, B. Constant-PH Molecular Dynamics Simulations for Large Biomolecular Systems. *J. Chem. Theory Comput.* **2017**, *13* (12), 5933–5944. https://doi.org/10.1021/acs.jctc.7b00875.

(10) Radak, B. K.; Chipot, C.; Suh, D.; Jo, S.; Jiang, W.; Phillips, J. C.; Schulten, K.; Roux, B. Correction to Constant-PH Molecular Dynamics Simulations for Large Biomolecular Systems. *J. Chem. Theory Comput.* **2018**, *14* (12), 6748–6749. https://doi.org/10.1021/acs.jctc.8b01075.

(11) Harris, R. C.; Shen, J. GPU-Accelerated Implementation of Continuous Constant PH Molecular Dynamics in Amber: PK(a) Predictions with Single-PH Simulations. *J. Chem. Inf. Model.* **2019**, *59* (11), 4821–4832. https://doi.org/10.1021/acs.jcim.9b00754.

(12) Harris, J. A.; Liu, R.; Martins de Oliveira, V.; Vaquez Montelongo, E.; Henderson, J. A.; Shen, J. GPU-Accelerated All-Atom Particle-Mesh Ewald Continuous Constant PH Molecular Dynamics in Amber. *bioRxiv* **2022**, 2022.06.04.494833. https://doi.org/10.1101/2022.06.04.494833.

(13) Clark, A. J.; Negron, C.; Hauser, K.; Sun, M.; Wang, L.; Abel, R.; Friesner, R. A. Relative Binding Affinity Prediction of Charge-Changing Sequence Mutations with FEP in Protein–Protein Interfaces. *J. Mol. Biol.* **2019**, *431* (7), 1481–1493. https://doi.org/10.1016/j.jmb.2019.02.003.

(14) Clark, A. J.; Gindin, T.; Zhang, B.; Wang, L.; Abel, R.; Murret, C. S.; Xu, F.; Bao, A.; Lu, N. J.; Zhou, T.; Kwong, P. D.; Shapiro, L.; Honig, B.; Friesner, R. A. Free Energy Perturbation Calculation of Relative Binding Free Energy between Broadly Neutralizing Antibodies and the Gp120 Glycoprotein of HIV-1. *J. Mol. Biol.* **2017**, *429* (7), 930–947. https://doi.org/10.1016/j.jmb.2016.11.021.

(15) Steinbrecher, T.; Zhu, C.; Wang, L.; Abel, R.; Negron, C.; Pearlman, D.; Feyfant, E.; Duan, J.; Sherman, W. Predicting the Effect of Amino Acid Single-Point Mutations on Protein Stability—Large-Scale Validation of MD-Based Relative Free Energy Calculations. *J. Mol. Biol.* **2017**, *429* (7), 948–963. https://doi.org/https://doi.org/10.1016/j.jmb.2016.12.007.

(16) Pahari, S.; Sun, L.; Alexov, E. PKAD: A Database of Experimentally Measured PKa Values of Ionizable Groups in Proteins. *Database* **2019**, *2019*, baz024. https://doi.org/10.1093/database/baz024.

(17) Ho, C.; Yuan, Y. Haemoglobin: Cooperativity in Protein–Ligand Interactions. *eLS*. April 19, 2010. https://doi.org/https://doi.org/10.1002/9780470015902.a0001345.pub2.

(18) Ohe, M.; Kajita, A. Changes in PKa Values of Individual Histidine Residues of Human Hemoglobin upon Reaction with Carbon Monoxide. *Biochemistry* **1980**, *19* (19), 4443–4450. https://doi.org/10.1021/bi00560a010.

(19) Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G. A.; Dahlgren, M. K.; Russell, E.; Von Bargen, C. D.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *J. Chem. Theory Comput.* **2021**, *17* (7), 4291–4300. https://doi.org/10.1021/acs.jctc.1c00302.

(20) Schrödinger Release 2022-1: FEP+. Schrodinger, Inc: New York, NY 2021.

(21) Tanokura, M. 1H-NMR Study on the Tautomerism of the Imidazole Ring of Histidine Residues: I. Microscopic PK Values and Molar Ratios of Tautomers in Histidine-Containing Peptides. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1983**, *742* (3), 576–585. https://doi.org/https://doi.org/10.1016/0167-4838(83)90276-5.

(22) Schrödinger Release 2022-1: Protein Preparation Wizard. Schrodinger, Inc: New York, NY 2021.

(23) Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput. Aided. Mol. Des.* **2013**, *27* (3), 221–234. https://doi.org/10.1007/s10822-013-9644-8.

(24) Hayes, R. L.; Buckner, J.; Brooks, C. L. 3rd. BLaDE: A Basic Lambda Dynamics Engine for GPU-Accelerated Molecular Dynamics Free Energy Calculations. *J. Chem. Theory Comput.* **2021**, *17* (11), 6799–6807. https://doi.org/10.1021/acs.jctc.1c00833.

(25) Arthur, E. J.; Brooks, C. L. 3rd. Efficient Implementation of Constant PH Molecular Dynamics on Modern Graphics Processors. *J. Comput. Chem.* **2016**, *37* (24), 2171–2180. https://doi.org/10.1002/jcc.24435.

(26) Donnini, S.; Tegeler, F.; Groenhof, G.; Grubmüller, H. Constant PH Molecular Dynamics in Explicit Solvent with λ-Dynamics. *J. Chem. Theory Comput.* **2011**, *7* (6), 1962–1978. https://doi.org/10.1021/ct200061r.

(27) Miller, E. Advancing Loop Prediction to Ultra-High Resolution Sampling, Columbia University, 2014. https://doi.org/10.7916/D8R78CJ1.

(28) Liwo, A.; Pillardy, J.; Czaplewski, C.; Lee, J.; Ripoll, D. R.; Groth, M.; Rodziewicz-Motowidlo, S.; Kamierkiewicz, R.; Wawak, R. J.; Oldziej, S. UNRES: A United-Residue Force Field for Energy-Based Prediction of Protein Structure—Orgin and Significance of Multibody Terms. In *Proceedings of the fourth annual international conference on Computational molecular biology*; 2000; pp 193–200.

(29) Miller, E. B.; Murrett, C. S.; Zhu, K.; Zhao, S.; Goldfeld, D. A.; Bylund, J. H.; Friesner, R. A. Prediction of Long Loops with Embedded Secondary Structure Using the Protein Local Optimization Program. *J. Chem. Theory Comput.* **2013**, *9* (3), 1846–1864. https://doi.org/10.1021/ct301083q.

(30) Li, J.; Abel, R.; Zhu, K.; Cao, Y.; Zhao, S.; Friesner, R. A. The VSGB 2.0 Model: A next Generation Energy Model for High Resolution Protein Structure Modeling. *Proteins Struct. Funct. Bioinforma.* **2011**, *79* (10), 2794–2812. https://doi.org/10.1002/prot.23106.

(31) Baker, W. R.; Kintanar, A. Characterization of the PH Titration Shifts of Ribonuclease A by One- and Two-Dimensional Nuclear Magnetic Resonance Spectroscopy. *Arch. Biochem. Biophys.* **1996**, *327* (1), 189–199. https://doi.org/https://doi.org/10.1006/abbi.1996.0108.

(32) Wlodawer, A.; Svensson, L. A.; Sjoelin, L.; Gilliland, G. L. Structure of Phosphate-Free Ribonuclease A Refined at 1.26 .ANG. *Biochemistry* **1988**, *27* (8), 2705–2717. https://doi.org/10.1021/bi00408a010.

(33) FUJII, S.; AKASAKA, K.; HATANO, H. Acid Denaturation Steps of *Streptomyces* Subtilisin Inhibitor. *J. Biochem.* **1980**, *88* (3), 789–796.

(34) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137* (7), 2695–2703. https://doi.org/10.1021/ja512751q.

(35) Abel, R.; Wang, L.; Harder, E. D.; Berne, B. J.; Friesner, R. A. Advancing Drug Discovery through Enhanced Free Energy Calculations. *Acc. Chem. Res.* **2017**, *50* (7), 1625–1632. https://doi.org/10.1021/acs.accounts.7b00083.