JCTC Journal of Chemical Theory and Computation

PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK_a Predictions

Mats H. M. Olsson,* Chresten R. Søndergaard, Michal Rostkowski, and
Jan H. Jensen*

*Department of Chemistry, University of Copenhagen, Universitetsparken 5,
Copenhagen, Denmark*

Abstract: In this study, we have revised the rules and parameters for one of the most commonly used empirical pK_a predictors, PROPKA, based on better physical description of the desolvation and dielectric response for the protein. We have introduced a new and consistent approach to interpolate the description between the previously distinct classifications into internal and surface residues, which otherwise is found to give rise to an erratic and discontinuous behavior. Since the goal of this study is to lay out the framework and validate the concept, it focuses on Asp and Glu residues where the protein pK_a values and structures are assumed to be more reliable. The new and improved implementation is evaluated and discussed; it is found to agree better with experiment than the previous implementation (in parentheses): rmsd = 0.79 (0.91) for Asp and Glu, 0.75 (0.97) for Tyr, 0.65 (0.72) for Lys, and 1.00 (1.37) for His residues. The most significant advance, however, is in reducing the number of outliers and removing unreasonable sensitivity to small structural changes that arise from classifying residues as either internal or surface.

Introduction

Ionizable residues in proteins play key roles in protein function; they are of profound importance for protein catalysis, protein stability, and ligand−protein and protein−protein interactions.[1-6] All properties that depend on these ionizable residues are therefore also pH-dependent, for example, proteins for acidophilic organisms are usually more stable toward lower pH values than normal and acid−base catalytic proteins have a pH optimum when the acid is protonated and the base is unprotonated, which can be traced to the protein pK_a values. Thus, understanding, predicting and modulating protein pK_a values has become an important feat for biochemistry and protein engineering in its own right. During the last decades, however, calculating pK_a values has also become a gauge for our ability to describe and predict the electrostatic interactions in proteins and therefore a first reliable validation of assessing the energetics in protein reactions.[5] Titrating a residue is a well-defined and from a

kinetic perspective uncomplicated event that can be measured with, for instance, NMR. From a computational perspective, on the other hand, a residue titration is still a major challenge that depends on the dielectric response of the entire protein + water environment surrounding the ionizable residue. It is in other words sensitive to all of its surroundings: water reorientation, protein reorganization, water penetration, and for the extreme cases, partial protein unfolding.

Over the last decades, there has been a major effort, and significant progress, in describing protein electrostatics; see references for reviews.[7-13] The majority of methods describe the environment using continuum electrostatics, such as Poisson−Boltzmann (PB) and generalized Born (GB) approaches, or a regular force field approach, for example, all-atom molecular dynamics simulations. Though the increasing number of pK_a predicting methods is encouraging, most methods require a significant computational effort with calculation times ranging from several minutes or hours to days. In response to this shortcoming, empirical methods provide an alternative in that they can calculate all pK_a values for a medium-sized protein within a few seconds. All these

* To whom correspondence should be addressed. E-mail: mats@kemi.ku.dk (M.H.M.O.); jhjensen@kemi.ku.dk.

methods have their strengths and weaknesses. Typically, the force-field based methods are the more rigorous since they include a full microscopic protein-dipole model (with or without electronic polarization) and can calculate the full thermodynamic cycle by free energy perturbation (FEP), thereby including explicit protein configurational sampling as the residue changes from the charged to uncharged form. However, severe convergence problems, a steep learning curve, and an exceptionally large computational effort make these approaches impractical for most real-life protein applications. On the other extreme, we have the empirical approaches that use scoring functions or effective potentials to describe the influence of the environment. These are easy-to-use and exceptionally fast, but in contrast to the other methods, empirical methods rely on representing the environment by undefined functions or descriptors. These are by definition nonrigorous, and since they are normally based on using single averaged structures obtained, for example, by X-ray crystallography, they cannot include configurational sampling in the conventional explicit sense. Empirical approaches are nevertheless very useful for the vast majority of protein pK_a values, and they are typically found to predict protein pK_a values as reliably as most more rigorous methods.[14–16] For PROPKA, which has recently been referred to as "the empirical method to beat",[14] the initially reported rmsd was found to be 0.79 (0.89 if pK_a values determined within an upper or lower bound is included).[17]

In a recent and independent study, where the more common and most easy-to-use programs were validated, PROPKA was found to be one of the most reliable protein pK_a predictors of the four tested.[15] Even though this is clearly encouraging, we have recently found physical inconsistencies in the previous and current versions of PROPKA (PROP-KA1[17] and PROPKA2[18]) that make some predictions behave erratic. The major problems arise from the discrete classification of residues into either surface or buried groups. Though such a distinction seem to have been very fruitful, it is clearly not a division made by nature; and undoubtedly, the correct physics is identical even though one type of residues might be more difficult to model than the other. The unfortunate either/or classification is further aggravated since coulomb interactions, which can be quite large, are only included for buried residues. Since a contact number defines the buried residues, that is, counting atoms within a radius cutoff, the residue classification can become very sensitive to the position of the residue or atoms close to a radius cutoff when the atom count is close to the junction between these classifications (see Results and Discussion for an illustrative example).

In this study, we have resolved the erratic behavior caused by the distinct classification into surface and internal residues by interpolating between these residue types to make the transition continuous. Since the surface versus buried classification of residues is central to calculating the charge−charge and hydrogen-bonding contributions to the pK_a shift and all contributions are interdependent, it has been necessary to reevaluate all aspects of the PROPKA theory. Thus, we have altered the functional forms of the charge−charge and desolvation contributions very significantly and PROPKA

has been reparameterized completely. This reassessment has rectified some significant outliers where the error in the predicted pK_a value has been several pH units, and concurrently we have removed a number of irregular interaction exceptions and reduced the number of parameters. The objective of this work, however, has not primarily been to increase the accuracy by the reparameterization, but to describe the interactions consistently and to call attention to a problem that might appear in other methods that utilize empirical information.

As we move to justifying and describing the model in the following section, we should keep in mind that we are not looking for an exact theory or rigorous treatment but for a simple and computationally fast approximation that has the most important features of our system and treats the dominant physical effects appropriately. In the following section, we introduce the overall concepts and contributions to calculating pK_a values. In the remaining section, we determine the model parameters, evaluate its expected accuracy by applying it to a set of proteins where the pK_a values have been determined experimentally, and discuss the performance and validity of both our approach and data test set. Finally we look at an illustrative example where the surface/internal classification makes the pK_a prediction exceedingly sensitive to a small change in structure for PROPKA2 and how this is resolved with the new approach.

Methods and Concepts

Since empirical pK_a predictors rely heavily on parametrization and calculating perturbations, and the relation between their contributions and conventional electrostatics is less obvious, we start by briefly scrutinizing the relevant thermodynamic cycle. The reaction we want to describe is the deprotonation of an ionizable residue in its protein environment and its change in free energy. Thus, we write a general deprotonation reaction as

$$AH \rightarrow A^- + H^+ \tag{1}$$

where A can be any ionizable group. The relevant free energy change we need to consider for protein pK_a calculations is best obtained by examining a thermodynamic cycle that considers a residue in its protein position and its reference water reaction as was previously introduced by Warshel and co-workers.[5] Thus, we write the free energy change for eq 1 in the protein as

$$\Delta G^{\text{Protein}}(AH \rightarrow A^- + H^+) = \Delta G^{\text{Water}}(AH \rightarrow A^- + H^+) + \Delta G^{\text{Water}\rightarrow\text{Protein}}_{\text{Solvation}}(A^-) - \Delta G^{\text{Water}\rightarrow\text{Protein}}_{\text{Solvation}}(AH) \tag{2}$$

Here, the $\Delta G^{\text{Protein}}$ and ΔG^{Water} terms are the definition of the pK_a value of AH in the protein and in water, in free-energy units, and the two last terms are the solvation free energies of moving the deprotonated and protonated form of the residue from water to its site in the protein. Since, we are ultimately interested in calculating protein pK_a values, we use the relationship between free energy and pK_a values

$$\Delta G = 2.30RT \cdot \Delta pK_a \tag{3}$$

define the effect of the protein on the reference water reaction as

$$\Delta p K_a^{\text{Water} \rightarrow \text{protein}} = \frac{1}{2.30RT} \cdot (\Delta G_{\text{Solvation}}^{\text{Water} \rightarrow \text{Protein}}(A^-) -$$

$$\Delta G_{\text{Solvation}}^{\text{Water} \rightarrow \text{Protein}}(AH)) \quad (4)$$

and rewrite eq 2 for residue i as

$$pK_{a,i}^{\text{Protein}} = pK_{a,i}^{\text{Water}} + \Delta p K_{a,i}^{\text{Water} \rightarrow \text{Protein}} \quad (5)$$

Here, $pK_{a,i}^{\text{Water}}$, which is usually referred to as the model value, is the pK_a value of the corresponding residue in water, and since these values are well-known (and collected in table S1 in Supporting Information), we are left with calculating the difference in pK_a between protein and water. This is the effect exerted by the protein on the pK_a value and can in principle be calculated correctly for instance with FEP approaches. In PROPKA, however, we emphasize computational speed and simplicity and see the protein as a small environmental perturbation to the water reference. Thus, we express the total environmental perturbation as a sum of effective perturbation contributions from protein groups. This is generally justified if the effect of the protein, $\Delta p K_a^{\text{Water} \rightarrow \text{Protein}}$, is small compared to the solvation energies involved in the reference reaction. This seems indeed to be the case: the vast majority of protein ionizable residues have pK_a values very similar to their corresponding water reference, and even for a significantly shifted residue the effect of the protein is more than an order of magnitude smaller than the absolute solvation energies involved (i.e., the solvation energy for acetic acid is close to 10 and 80 kcal/mol for the protonated and unprotonated form, respectively, whereas each pH-unit shift from the protein environment corresponds to 1.36 kcal/mol of perturbed solvation free energy). If the perturbation becomes too large or, more importantly, is associated with significant structural rearrangement, it is quite likely that this description will fail. A reaction where the solute is being charged, as in eq 1, is by nature also strongly coupled to the environment and will therefore also undergo non-negligible structural rearrangement. However, since we never calculate any single-configuration energies, but define effective pK_a contributions, $\Delta p K_a^{\text{Water} \rightarrow \text{Protein}}$ or $\Delta G^{\text{Water} \rightarrow \text{Protein}}$, we can fit the functions to include "average structural reorganization" implicitly. The challenge we face at this point is obviously that we do not know the functional form of these effective protein perturbations. Before we proceed to an atomistic detail, which is done in separate subsections below, we need to get an overview and define the types of perturbations we need to consider. Since we now have defined that we need to calculate "Water → Protein" perturbations, we omit the superscript and start by writing the protein perturbation, $\Delta p K_a$, for residue i as

$$\Delta p K_{a,i} = \Delta p K_{a,i}^{\text{Self}} + \Delta p K_{a,i}^{QQ} \quad (6)$$

Here, the second term is the Coulomb contribution because of the protein charge−charge interactions with all other charged or ionizable groups, and the first term, the self-energy or intrinsic contribution, is the remaining contribution

that is obtained when all other charged and ionizable groups are kept in their neutral form. The self-perturbation can in turn be divided into two major components: the desolvation and intrinsic electrostatic energy according to

$$\Delta p K_{a,i}^{\text{Self}} = \Delta p K_{a,i}^{\text{Desolv}} + \Delta p K_{a,i}^{Q\mu} \quad (7)$$

The desolvation term describes the desolvation penalty or the loss of solvation energy exerted by the protein as protein atoms replace ambient water, whereas the electrostatic term describes the substituting solvating effect from those atoms (i.e., interactions from nearby protein (di)polar groups such as NH and CO groups). For practical purpose, which is explained below, we approximate this term with the dominant hydrogen-bonding interactions and a usually much less important interaction representing unfavorable electrostatic interactions that typically cannot be assigned to hydrogen bonding according to

$$\Delta p K_{a,i}^{Q\mu} \simeq \Delta p K_{a,i}^{\text{HB}} + \Delta p K_{a,i}^{\text{RE}} \quad (8)$$

Coulomb. Even though Coulomb interactions were early seen as the major contribution to protein pK_a values,[19] it has since long been recognized that, in fact, the intrinsic term is usually more important.[20] In PROPKA3, we have, contrary to previous versions of PROPKA that use a linearized form of Coulombic interactions for buried residues and no Coulomb interactions for surface residues, decided to adopt a regular $1/r$ term and calculate the pK_a contribution to residue i from charge j as

$$\Delta p K_{a,i}^{QQ} = \sigma_{ij} \cdot \frac{244}{\varepsilon \cdot r_{ij}} \cdot w(r_{ij}) \quad (9)$$

Here, σ_{ij} is the sign of the function determining if the interaction shifts $pK_{a,i}$ up or down, 244 is the normal Coulomb's law coefficient converted into pK_a units, ε is the dielectric constant that screens the Coulomb interaction between two ionizable residues, r_{ij} is the distance between the residue charge-centers (defined by Table S1 in Supporting Information), and $w(r_{ij})$ is a distance-dependent weight function.

σ clearly depends on the residue types. For instance, the Coulomb interaction between the opposite charges of an acid and a base stabilizes the configuration where both residues are ionized; thus, the pK_a is shifted down for the acid ($\sigma = -1$) and shifted up for the base ($\sigma = +1$). For two acids, however, there will only be a Coulomb interaction for the residue with the higher pK_a value since this residue is in its neutral form when the residue with lower pK_a value titrates. The Coulomb interaction between two negative groups is unfavorable and results in raising the higher pK_a value (i.e., $\sigma = +1$), whereas the lower pK_a value is unaffected (i.e., $\sigma = 0$). In summary, σ can be defined as

$$\sigma_{ij} = \begin{cases} -1 & \text{if } i \in \text{acids and } j \in \text{bases} \\ & \text{or } i \in \text{bases and } pK_{a,i} < pK_{a,j} \\ +1 & \text{if } i \in \text{bases and } j \in \text{acids} \\ & \text{or } i \in \text{acids and } pK_{a,i} > pK_{a,j} \\ 0 & \text{otherwise} \end{cases} \quad (10)$$
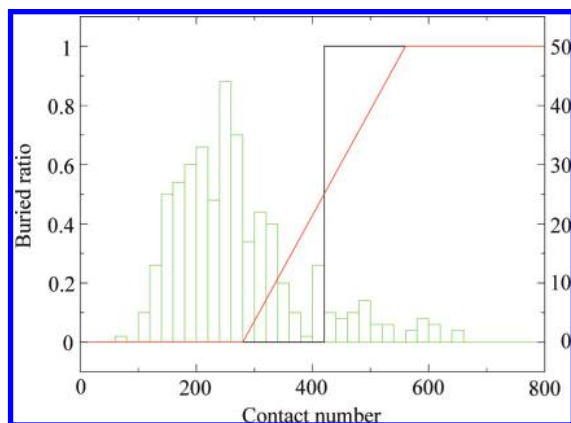
Figure 1. Problem with classifying residues as either surface or buried is resolved by interpolating between these extremes using a buried ratio. The figure shows the buried ratio (the single-residue based position-dependent weight function, $w(N)$) for PROPKA3 (red) and PROPKA2 (black) as a function of heavy-atom count (N) within a 15 Å radius of the charge center (left y-axis). The background histogram (green) shows the distribution of heavy-atom counts for the 363 Asp + Glu residues found in the 22 proteins used for validating Asp and Glu pK_a values (right y-axis).

The resulting coupled titrations are then resolved iteratively using a scheme akin to the Tanford−Roxby scheme.[21]

The value of the dielectric constant in proteins is a matter of much debate and frequently some confusion and is therefore maybe not so straightforward. In principle, each ion−ion pair interaction has its own dielectric constant depending on the two residues' position in the protein, their solvent exposure, and the ability of the surrounding environment to respond to changes in the electric field. Clearly, PROPKA must be pragmatic in this context, but it seems reasonable to use a comparatively large value, $\varepsilon_{surface}$ that is close to ε_{water}, if the residues are close to the protein surface, and a smaller value, ε_{buried}, if they are buried deeply in the interior of the protein (see table 2). Since it also seems reasonable to make a linear interpolation between these extreme cases, we use an effective dielectric constant obtained as

$$\varepsilon = \varepsilon_{surface} - (\varepsilon_{surface} - \varepsilon_{buried}) \cdot w_{pair}(N) \qquad (11)$$

where $w_{pair}(N)$ is a position-dependent weight function that depends on the location of the two residues in the protein, that is, to what degree the residues are buried in the protein (their buried ratios). A residue's buried ratio is obtained by its contact number, that is, by counting the number of protein heavy atoms, N, within a 15 Å sphere from its charge center, and defined according to

$$w(N) = \begin{cases} 0 & \text{if } N \leq N_{min} \\ \dfrac{N - N_{min}}{N_{max} - N_{min}} & \text{if } N_{min} < N < N_{max} \\ 1 & \text{if } N \geq N_{max} \end{cases} \qquad (12)$$

This function is depicted by the red line in Figure 1, and as can be seen in the figure, a residue is considered fully on the surface when $N \leq 280$ (N_{Min}) and fully buried when $N \geq 560$ (N_{Max}). Thus, a buried residue concurs with being surrounded by a large number of protein atoms, whereas a surface residue concurs with being surrounded by few protein atoms and therefore many solvent molecules. Between these extremes, a straight line interpolates the buried ratio. From the background (green) histogram one can also see that the majority of Asp and Glu residues are with these cutoff values as expected identified as surface residues (237 of 363) or only slightly buried, and only 11 of 363 residues are considered fully buried. This is reasonable since most proteins considered in this study are comparatively small and therefore have few truly buried residues. For Coulomb interactions it is more useful to consider the residue pair rather than the individual residues. Therefore $w_{pair}(N)$ in eq 11 is instead obtained by summing the interacting residues' contact numbers ($N_{Pair} = N_i + N_j$) and use $N_{Min} = 560$ and $N_{Max} = 1120$, that is, doubling the single-residue based values, when calculating the position dependent weight for the pair. Thus, for two surface residues, where the number of neighboring protein atoms is small, N_{Pair} is close to or smaller than N_{Min}, $w_{pair}(N) \approx 0$, and we obtain the anticipated $\varepsilon = \varepsilon_{surface}$ from eq 11. Similarly, for two buried residues, where the residues are surrounded by an abundance of protein atoms, N_{Pair} is close to or larger than N_{Max}, $w_{pair}(N) \approx 1$, thus, recapturing $\varepsilon = \varepsilon_{buried}$.

Before we continue showing the resulting pK_a Coulomb contribution, we note that the way we use the contact number is different compared to what is commonly used in structural bioinformatics since we use a comparatively large sphere radius, use the charge center for the titrating group as center, and count all heavy atoms rather than only the Cα or Cβ atoms. Though these modifications clearly evade the concept of "contact", since an atom 15 Å away is hardly in contact with the center, our modification is clearly more suitable for pK_a calculations since they are centered on the site of interest for titration and more sensitive to report on the local environment. Using atom-based contact numbers rather than residue-based, which is the case when you count Cα or Cβ atoms, makes our approach less sensitive to the "all or nothing" behavior because of the radius cutoff and can also account for fractions of residues. More importantly, Coulomb interactions and solvation effects are usually considered long-ranged and goes beyond the first coordination sphere.

The distance-dependent weight function, $w(r_{ij})$ in eq 9, is formally defined by

$$w(r_{ij}) = \begin{cases} \dfrac{r_{ij}}{r_{min}} & \text{if } r_{ij} \leq r_{min} \\ \dfrac{r_{ij} - r_{min}}{r_{max} - r_{min}} & \text{if } r_{min} < r_{ij} < r_{max} \\ 1 & \text{if } r_{ij} \geq r_{max} \end{cases} \qquad (13)$$

where r_{Max} and r_{Min} are cutoff values that defines two extreme points: the distance where the Coulomb contribution attains an upper limiting value and the distance cutoff for the interaction. The first makes PROPKA predictions more robust by enforcing the contribution to stay constant at distances shorter than r_{Min} (4 Å), see also Figure 2. Omitting this cap increases the sensitivity for short-distance contributions that are more likely to be due to problems in geometry
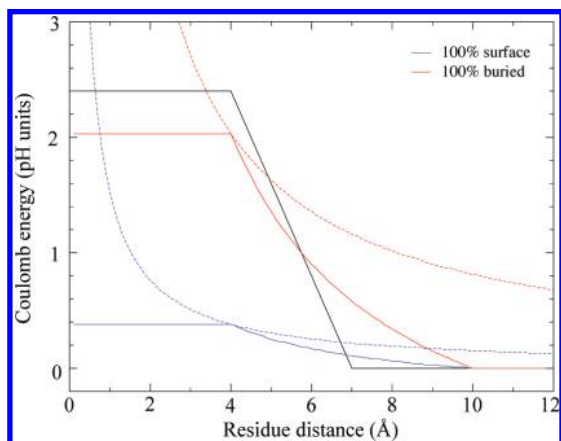
PROPKA3

J. Chem. Theory Comput., Vol. 7, No. 2, 2011 **529**

Figure 2. Coulomb interaction in PROPKA3 depends on how buried the interacting pair is. The interaction between two surface residues is screened by a dielectric constant similar to water (blue), whereas two buried residues are screened significantly less (red). The corresponding dashed lines correspond to a true Coulomb energy where $w(r) = 1$ is constant. The Coulomb contribution used for buried interactions in previous PROPKA versions is included in black for comparison; note that Coulomb contribution is zero for surface residues in previous versions.

for poor-resolution X-ray structures rather than real effects. At distances between r_{Min} and r_{Max}, $w(r_{ij})$ decreases linearly from 1 to 0 so that the Coulomb contribution vanishes smoothly at r_{Max} (10 Å). The important property in this range, unlike for a regular cutoff, is that we avoid discontinuous jumps. By using this weight function we effectively also obtain a distance-dependent dielectric response of the environment since the effective dielectric at longer distances is much larger than at short distances. However, though this is an important and nontrivial topic, see for instance the introduction of ref 22 for an overview, we see our approach as rather pragmatic but effective and gives a well-defined interaction cutoff.

The resulting Coulomb contribution is depicted for two extreme cases in Figure 2: the Coulomb interaction between two surface residues (solid blue line) and between two buried residues (solid red line); for comparison we have also included the linearized Coulomb contribution that was used for buried interactions in PROPKA2 (solid black line). We find that reasonable values for the dielectric constants are $\varepsilon_{surface} = 160$ and $\varepsilon_{buried} = 30$. These seemingly counterintuitive values are actually quite well founded. First, since PROPKA is parametrized using pdbfiles determined by X-ray crystallography, we have to accept that $\varepsilon_{surface} \neq \varepsilon_{water}$ since the surface residues are heavily influenced by neighboring protein cells by, for example, making hydrogen bonds to neighboring proteins and do not reflect the "correct" residue geometry in water. Presumably, the lack of solvation also makes charged residues seek alternative positions to fully solvated configurations. Therefore, ion-pairs between a surface acid and base are much more common in crystals than in a fully solvated protein in water, and in order to compensate for these artificial interactions it is necessary to "over-screen" the Coulomb interaction. If structures, however, are obtained from long MD simulations with correct

boundary conditions, or possibly also from NMR constraints, it is probably a better strategy to use $\varepsilon_{surface} = 80$. Without going much further into what the more commonly used protein dielectric constants signifies, we conclude that the popular $\varepsilon_p = 4$, which is associated with the self-energy, is inappropriate for charge−charge screening.[23] Instead, one needs to consider the dielectric response of the entire protein + water system with electronic polarization, protein+water dipolar reorganization, and water penetration; and eventually find that an appropriate value is probably in the range 20 to 80, which is in agreement with our $\varepsilon_{buried} = 30$. We also note that the previous version of PROPKA effectively used the corresponding $\varepsilon_{surface} = \infty$ and $\varepsilon_{buried} \approx 25$, but without interpolating between the two types. In fact, our $\varepsilon_{surface}$ and ε_{buried} should instead be compared to the ε_{app} used by Warshel and co-workers who have used $\varepsilon_{app} = 40$ with success in the PDLD/S-LRA model.[24] We also find significant similarities in the screening of Coulomb interactions between our approach and the modulated screened Coulomb potential (SCP) approach by Mehler et al.,[25] which uses distance-dependent sigmoidal screening functions, modulated by the protein microenvironment, to screen electrostatic interactions. The dielectric constant of these screening functions approach 80 as the distance increases and attains 40 already at a distance of 8 Å. For comparison, we have also included the regular Coulomb's law contribution, which can be obtained by using $w(r_{ij}) = 1$, as dashed blue (surface residues) and dashed red (buried residues) lines in Figure 2. Thus, the effect of $w(r_{ij})$ can be seen by comparing the corresponding solid and dashed lines. Finally, it should be noted that though the PROPKA2 (black line) and PROPKA3 (red solid line) contributions seem to be quite similar for fully buried residues, for example, the max pK_a contribution is capped at similar 2.4 and 2.0 pH units for 4 Å, the most important advance lies in interpolating those residues that are only slightly buried.

Desolvation. The desolvation contribution to the ΔpK_a values corresponds in principle to creating the cavity in the solvent that is occupied by the protein surrounding the ionizable residue. Thus, it is an energy penalty for excluding parts of the water solvation that disfavors the charged form and raises the pK_a value of acids and lowers the pK_a value of bases. The full details of the model we use in PROPKA3 will be published elsewhere, but in essence it depends on the solvent volume that has been excluded and its distance from the ionizable residue's charge center. The total desolvation to residue i is calculated as

$$\Delta pK_{a,i}^{Desolv} = c \cdot \sum_j^N \frac{V_j}{r_{ij}^4} \tag{14}$$

where the sum runs over all non-hydrogen atoms (N), c is an empirically determined constant whose value depend on the solvation properties of water, V_j is the effective volume occupied by a nearby atom j, and r_{ij} is the distance from the charge center of residue i to the center of atom j. Each protein atom thus increases the desolvation contribution to the pK_a, but large atoms (with larger volume) change the pK_a more than smaller atoms and atoms very close to the residue change the pK_a values much more than more distant atoms.

As it turns out, our pK_a value predictions are greatly improved if we use the same scheme as for the Coulomb contribution and define one constant, $c_{surface}$, for surface residues and one, c_{buried}, for buried residues (see Table 2) and interpolate between them to get

$$c = c_{surface} - (c_{surface} - c_{buried}) \cdot w(N) \quad (15)$$

Here, $w(N)$ is again the buried ratio defined in eq 12 for the Coulomb interaction, but note that we use the single-charge-center based contact number N and N_{Min} and N_{Max} is half the value of those in the Coulomb contribution. The reason for the improved accuracy with two c values could be several, but probably again it is associated with the dielectric response and problems with the structure; surface residues have an artificial tendency to aggregate in ion-pairs and seek alternative interactions on the protein surface because of crystal packing rather than extend into the solvent. Thus, its value is on the average smaller to compensate for artificial interactions. Other plausible reasons are that we have interaction cutoff distances that effectively truncate the protein, a simplified term for the electrostatic self-energy, thereby omitting protein solvation, or simply by using different Coulomb contributions for surface and buried interactions. Another, and maybe more likely, reason is that we are using a static average structure for the protein and are therefore missing protein reorganization and water penetration due to changes in the local structure around the ionizable residue as it changes protonation state. Regardless of the true origin, all these effects make the effective desolvation notably smaller on the surface than in the interior of the protein.

Intrinsic Electrostatics. In principle, the intrinsic electrostatics includes all interactions between the ionizable residue and the remaining protein, apart from the Coulomb energy, that affects the deprotonation energy of the residue. In a classical electrostatic sense this would include protein permanent dipoles (and higher multipole terms) and van der Waals interactions from the entire protein. To include all these interactions, however, would not be feasible for PROPKA if we want to apply it to larger systems with a reasonable computational effort. Instead, we partition the contribution into short-distance and long-distance contributions based on a hypothetical cutoff value, say 6 Å, and assume that the contribution from groups further than this cutoff tend to contribute to the residue deprotonation energy, on the average, with similar magnitude as ambient water and therefore do not contribute to ΔpK_a. Thus, we use a minimal model[26] of the protein and focus on the short-distance contributions for the water → protein perturbation. This is of course not strictly true, as can be seen for protein solvation free-energy simulations that typically converges with 16 Å simulation radii,[27] but it is probably a good approximation for surface and moderately buried residues. Nevertheless, this approximation is probably case-dependent and rely to a great degree on contribution cancelations for more buried residues. The dominant short-distance contributions come from close polar interactions, for example, hydrogen bonds of the type $COO^{\ominus} \cdots HN$ for carboxylic acids or $NH^{\oplus} \cdots OC$ for bases. Since, presumably, a large part of the calculation time is spend on these evaluations, we use the simplest possible function to describe this interaction, that is,

$$\Delta pK_{a,i}^{HB} = \begin{cases} c^{HB} \cdot w(r_{ij}) \cdot \cos\theta & \text{if } \theta \geq 90^{\circ} \\ 0 & \text{if } \theta < 90^{\circ} \end{cases} \quad (16)$$

Here, c^{HB} is a constant parameter determined in this study (which contains both dipole and van der Waals contributions), $w(r_{ij})$ is a distance dependent weight function, and θ is the angle formed by the hydrogen bond and hydrogen-bond acceptor (with the hydrogen atom as the apical center). This way, we have a simple function that describes the most important features of a hydrogen bond: a strong interaction for close distances where the hydrogen is aligned pointing directly toward the ionizable group or hydrogen-bond acceptor, and a vanishing contribution for distant interactions or where the hydrogen does not point toward the acceptor.

The distance dependence for these hydrogen-bonding interactions is a compromise between what is reasonable and what is computationally convenient. In principle, the interaction is akin to a charge-dipole interaction and should therefore depend on $1/r^2$. However, as it represents the effective ΔpK_a contribution, which also contains reorganization and other compensating effects from the environment, we cannot assume such a form apriori. Instead, we use the simpler linearized form that was also previously employed in PROPKA2[17]

$$w(r_{ij}) = \begin{cases} 1 & \text{if } r_{ij} \leq r_{min} \\ \dfrac{r_{ij} - r_{min}}{r_{max} - r_{min}} & \text{if } r_{min} < r_{ij} < r_{max} \\ 0 & \text{if } r_{ij} \geq r_{max} \end{cases} \quad (17)$$

where r_{ij} is the hydrogen-bond distance (e.g., the shortest distance between H and OD1 or OD2 for an Asp backbone hydrogen bond), r_{Min} is a short cutoff where the contribution attains a plateau, and r_{Max} is the distance cutoff where the interaction vanishes. Just as for the Coulomb contribution, the contribution cap for short distances helps making the predictions insensitive to exceptionally short contacts, which probably arise from uncertainties in structure rather than from physical strong interactions. In PROPKA2, the value of these cutoffs were set to 2 and 3 Å, respectively, and then adjusted for some interaction types to accommodate experimental points. In the new PROPKA3 parameter set, however, we have only fitted the c^{HB} parameter, and instead obtain the cutoffs by considering histograms with observed bonding distances in X-ray pdb structures (an example of such a histogram, the resulting distances, and further discussion can be found in Figure S1 and Table S2 in the Supporting Information). Clearly, it would be desirable to fit also these values, but since we have a limited data set to fit against and an approximate model, we believe the structure-derived values are more appropriate.

For freely rotating hydrogen bonds, for example, the ROH group of Ser, we assume that the hydrogen bond is flexible enough to reorient its direction to its optimal orientation, i.e. directly point toward its hydrogen-bond acceptor, which is equivalent to setting $\cos\theta = 1$ in eq 16 and gives

PROPKA3

J. Chem. Theory Comput., Vol. 7, No. 2, 2011   **531**

$$\Delta pK_{a,i}^{HB} = \begin{cases} c^{HB} & \text{if } r_{ij} \leq r_{min} \\ c^{HB} \cdot \dfrac{r_{ij} - r_{min}}{r_{max} - r_{min}} & \text{if } r_{min} < r_{ij} < r_{max} \quad (18) \\ 0 & \text{if } r_{ij} \geq r_{max} \end{cases}$$

For such an interaction, we do not need to predict the actual position of the hydrogen atom but use the heavy-atom distance as a measure of hydrogen-bond strength, and thereby avoid a well-known problem in single-structure modeling of proteins (the values of r_{Min} and r_{Max} will be ~1 Å larger for heavy-atom distances on the account of not using the hydrogen position). For the remaining sp^2-hybridized hydrogen atoms, however, we have to create its position explicitly since we cannot assume that the hydrogen atom can adopt such an optimal interaction. This is done in a simplified way described elsewhere.[18] Note that PROPKA3 uses eq 15 with explicit hydrogen atoms for all sp^2-hybridized atoms, whereas PROPKA2 used eq 18 for all hydrogen bonds except for interactions with the peptide backbone. This alteration was not found to give any significant difference for the training set, but was adopted to make the description more consistent. As was found in this study, however, it is necessary to include these hydrogen-bonded interactions consistently for all types of ionizable residues. Failing to include this interaction term, as in PROPKA2 where it was included for acids but not for bases, overturns the important desolvation/resolvation balance and results in unphysical pK_a predictions (as can be seen for Lys in the following section).

Another short-distance term that is included in the intrinsic electrostatic contribution, ΔpK_a^{RE}, has the opposite effect to the previous hydrogen bonding in that it presents a possible hydrogen-bond acceptor for acids or hydrogen-bond donor for bases. The effect of this "reversed hydrogen bond", however, is presumably better described as an unfavorable dipole interaction of the type $COO^{\ominus} \cdots OC$ for acids, raising the pK_a value, and $NH^{\oplus} \cdots HN$ for bases, lowering the pK_a value. These interactions are far less common since it is destabilizing in nature, effectively raising the energy of the ionized form of the residue and thereby contributes to making the local protein structure less stable. However, these types of interactions *are* found in proteins, and can in some cases be used by the protein to tune pK_a values for catalytic function; for instance acid−base catalyzed reactions require a protonated acid with an elevated pK_a value for the catalytic activity. Since it is unfavorable in nature, it rarely interacts with a specific atom but with the charge of the ionizable residue, and thus, we model it with the same functional form as eqs 16 and 17, but instead of using the shortest atom-to-atom distance we use the atom-to-charge-center distance. These interactions are not considered freely rotatable and therefore always angular dependent. This term was not included in previous versions of PROPKA since it is not a hydrogen bond, but it does make sense from an electrostatic point of view and contribute to some pK_a values if given similar parameters as for the hydrogen bonding interactions. Unfortunately, there are too few instances of this interaction

in the training set to further probe its effect and to obtain reliable parameters; thus, the interaction remains rather ad hoc.

Results and Discussion

The main goal of this study is to probe and correct the inconsistent treatment of internal and surface residues. In this section, we parametrize and validate our new implementation of PROPKA that uses the interpolation scheme outlined in the previous section. For this purpose, we primarily consider Asp and Glu residues since we believe the protein pK_a values and structures are more reliable for these than for Tyr, Lys, and His residues.

Consistent Treatment of Internal and Surface Residues. The best way to exemplify the inconsistent treatment of internal and surface residues is to consider the three protein configurations of barnase, that is, models, in the crystallographic asymmetric unit provided by pdbfile 1A2P. For most residues, the pK_a values are very similar for these copies, but for Glu 73 and Lys 27, PROPKA2 predicts pK_a values to be 1.5, 1.3, and 3.4 and 11.6, 11.3, and 10.2, respectively, for the three models (see Figure S2 and Table S3 in Supporting Information). The conspicuous deviation for the third model is because the charge center of Lys 27, the NZ atom, has moved 0.3 Å toward the solvent, which results in the heavy-atom count, 412, 412, and 381, going under the critical cutoff value 400 and therefore reclassifies the residue from being buried to being on the surface. This reclassification leaves out the charge-stabilizing Coulomb interaction, worth 1.9 pH units for each residue, giving a higher pK_a value for Glu 73 and lower pK_a value for Lys 27. This is clearly an artifact of the model, where in extreme cases, a miniscule change such as moving one or a few atoms from being 15.6 Å from the residue to 15.4 Å can abruptly "switch on" Coulomb interactions worth up to 2.4 pK_a units each. In this particular case, the problem is not so problematic since we would average the pK_a values of the three configurations to obtain an apparent pK_a value. PROPKA3, on the other hand, finds Glu 73 to be 71%, 72%, and 72% buried and Lys 27 to be 32, 33, and 23% buried for each model structure, and includes the Coulomb interaction with a pair-weight of 52%, 53%, and 48% for each model structure and gives pK_a values of 5.2, 5.2, and 5.3 and 10.9, 10.9, and 10.6, respectively.

A more common and more severe manifestation, however, is when making mutations or adding/removing ligands close to ionizable residues. This can easily change the heavy-atom count within a fixed radius for residues in the vicinity of the mutation; for example, a single mutation $G \rightarrow R$ adds 7 heavy atoms. It should also be noted that this inconsistent treatment of surface and internal residues would most likely give severe problems when improving the geometry by using a rotational library or Monte Carlo configuration searches since it creates discontinuous jumps in the scoring function.

Parameterization. Part of the new parametrization strategy is to fit the PROPKA parameters to fewer, in most cases uncomplicated, and more trusted experimental pK_a values. To avoid the more difficult ionizable groups in the para-

Table 1. Nonadjustable Parameters and Descriptors

interaction	type/use	parameter	value
Coulomb	buried ratio	$R_C{}^a$	15.0
	buried ratio	N_{Min}	280
	buried ratio	N_{Max}	560
	$w(r)$	r_{Min}	4.0
	$w(r)$	r_{Max}	10.0
desolvation	VDW volume	V_C	20.58
	VDW volume	V_{C4}	38.79
	VDW volume	V_N	15.60
	VDW volume	V_O	14.71
	VDW volume	V_S	24.43

a Contact radius.

Table 2. Fitted PROPKA3 Parameters

interaction	parameter	value
Coulomb	$\varepsilon_{surface}$	30
	ε_{buried}	160
desolvation	$c_{surface}$	3.375
	c_{buried}	13.5
intrinsic electrostatics	c^{HB}	0.85
	c^{RE}	0.80

metrization might seem a bit counterintuitive since improving these would also improve our agreement with experiment the most (lower the rmsd), but if the ionization process involves interaction changes that we do not include in our model, we could introduce significant inconsistencies and cover up problems. This seems to have been the case with previous versions of PROPKA. Such cases could include parametrizing against coupled residues where it is even difficult to assign experimental pK_a values to individual residues or using poor structures with significant errors (including crystal effects). Thus, we avoid ionizable residues with seemingly unorthodox interactions, residues whose titration is strongly coupled with nearby residues, multichain proteins, and proteins where we cannot find a reasonable X-ray structure (we avoid NMR structures since they generally seem to give poorer agreement with experiments[28]). Likewise, pK_a values that are known by upper or lower bounds seem less reliable since in fact we do not know the pK_a values of these and do not know what error we are making for them. Our goal is to make as general interactions as possible and to avoid exceptions that in effect only impact very few buried residues and therefore can be seen as adjusting individual points. It is quite conceivable that such exceptions and overparameterization could in fact be contributing to the low rmsd initially reported for PROPKA and in part behind some of the problematic residues reported.[17]

The effective perturbations described in the previous section use two sets of parameters: nonadjustable descriptors that describe a property or cutoff value of the system (compiled in Table 1), and adjustable parameters that directly determine the contribution to the pK_a values (compiled in Table 2). The latter fitted parameters are then obtained by modifying their value to give a reasonable representation of the training set consisting of 85 ionizable Asp and Glu residues from the compilation of Forsyth et al.[29] and Song et al.[30] (compiled in tables S4 and S5 in Supporting Information) and a low rmsd compared to experiment. Thus, we have not automatically minimized the rmsd since we find
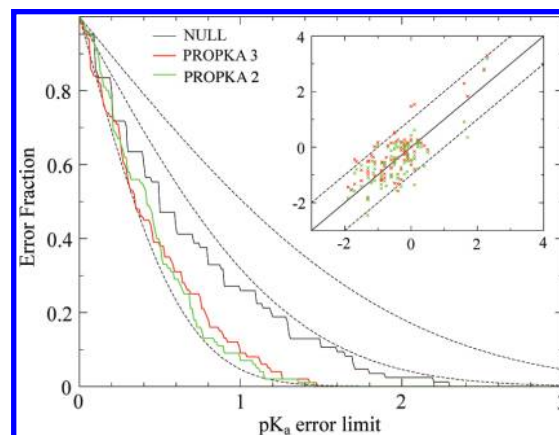

Figure 3. Error-fraction plot shows that predictions with PROPKA3 are comparable to PROPKA2 for the training set of 85 Asp and Glu pK_a values. The plot shows the error fraction for a given error limit (red: PROPKA3 and green: PROPKA2); for example, it can be seen that 9% of the points have an error >1 pH unit. For comparison the black solid line shows the predictions using the Null-model, and the dashed black lines show the error fraction corresponding to a normal distribution with standard deviations of 0.5, 1.0, and 1.5 (starting from bottom left corner). The inset shows a traditional scatter plot with the black solid line representing pK_a^{calcd} = pK_a^{exp} and black dashed lines pK_a^{calcd} = p$K_a^{exp} \pm 1$.

that doing so give outliers unreasonably large weight and return seemingly untenable parameters, which probably comes from having too few pK_a values with reliable X-ray structure. In addition to these 85 training residues, we acknowledge that our participation in the pK_a cooperative[31] and especially 20 Asp and Glu residues from what we term the Telluride data set has willingly or unwillingly affected the way we have defined and obtained the perturbations.

Aspartic and Glutamic Acids. Even though the training set cannot tell us what accuracy to expect from our new approach, it does provide a first indication how well PROPKA3 predicts Glu and Asp pK_a values compare to PROPKA2. As it turns out, the overall performance is very similar when compared for the 85-residue training set: the rmsd is 0.59 and 0.56 respectively. This correspondence with experiment seems exceptionally good, but we should keep in mind that we have chosen this data set to be uncomplicated. Never the less, the rmsd for the Null model, that is, setting all protein pK_a shifts to 0 and therefore all pK_a values to its model value (3.8 for Asp and 4.5 for Glu), provides a useful metric for assessing the quality of a particular data set. In this case, we obtain 0.89, which seems to be moderate, and find 26% of the residues being shifted by more than 1 pH unit. To get a better validation we need to consider another test set, but before we do that, we find it useful to assess the result closer.

A more detailed analysis of the outcome is provided by Figure 3, which shows the error distribution (as the error fraction of a given error limit) for the pK_a values in a convenient way. The dashed black lines depict the expected error fractions for a set of pK_a values whose error is given by a normal distribution with a standard deviation of 0.5, 1.0, and 1.5 (from left to right). By inspection, we can see that the red and green lines follow reasonably well the dashed

line corresponding to standard deviation 0.5 as expected from rmsd values of 0.59 and 0.56. Moreover, we can get the expected fraction of points with an error worse than any given acceptable limit. For instance, from the red line, we see that 9% of the pK_a values has an error larger than 1 pH unit for PROPKA3, whereas the corresponding value for PROPKA2 is 7%. From the inset scatter plot, it can also be verified in a traditional way that the two distributions indeed seem very similar. Obviously, we have not learnt more about pK_a values by comparing our result with normal distributions; however, the graph provides a more consistent and unbiased view on the result. For instance, frequently results are presented in terms of the percentage of points having an error <1 pH unit, for example, in a recent study PROPKA was found to have 88% within 1 pK_a unit error.[15] This number is clearly useful, but there is a reasonable risk that considering only one (or a few) limits could give a misleading view on one method over another; for example, if one method has more outliers than another but does better for the majority of points, choosing a relatively small limit favors that method whereas choosing a relatively large acceptable error disfavors it. As can be seen from this plot in Figure 3, choosing the limit close to 0.4 pH units gives 55% and 46% for PROPKA3 and PROPKA2, respectively, whereas choosing a limit of 0.75 pH units gives 76% and 84% respectively. In our view, there is no significant difference in accuracy based on this data set between these two versions. For the limit 1 pH unit, there is no problem, since we get 91% and 93%, respectively, which correctly reflect the similarity, but chances may be that a biased behavior occurs at this point, and clearly routinely choosing a value might lead to misleading percentages. Regardless of the outcome, presenting the error as in Figure 3 gives the entire picture and the expected error fraction can easily be read for any limit. Judging from the rmsd and the error-fraction curves, both PROPKA versions perform significantly better than the Null model.

A better test set can be conceived by combining the previous test sets compiled in studies by Forsyth et al.,[29] Stanton and Houk,[16] and Song et al.[30] and again exclude those titrations that are determined by upper or lower bounds, those deemed ambiguous assignment (in Song et al.), and proteins where we only find NMR structures but include those that were found to have unorthodox interactions or otherwise difficult. Also the 85 pK_a values in the training set are included in this data set, which, though it does not follow traditional test-set philosophy, we think is reasonable when comparing to PROPKA2 since PROPKA2 was parametrized using 314 pK_a values, and the study did not make any distinction between training set and test set. Thus, setting up a fictitious "test set" that has been included when parametrizing one version but not the other will clearly bias the evaluation. For the resulting 201 Asp and Glu residues, we obtain an rmsd of 0.79, 0.91, and 1.06 for PROPKA3, PROPKA2, and the Null-model respectively, which show that PROPKA3 provides a moderate but significant improvement over PROPKA2. The rmsd values are summarized in Table 3.

Even though obtaining an rmsd below 0.8 (for PROPKA3) is overall pretty good, it seems from the error-fraction plot

Table 3. rmsd Summarized for Each Residue Type

	COO	ASP	GLU	TYR	LYS	HIS
pK_a values	201	101	100	11	51	30
PROPKA 3	0.79	0.77	0.80	0.75	0.65	1.00
PROPKA 2	0.91	0.94	0.87	0.97	0.72	1.37
Null-model	1.06	1.23	0.86	0.70	1.01	0.93

in Figure 4 and the rmsd values in Table 3 that PROPKA outperforms the Null-model by disappointingly little; the solid thin black error-fraction curve representing the Null-model prediction is much closer to the red and green PROPKA curves than it was for the training set. It seems obvious that all pK_a-predicting methods should outperform that model. However, it rather reflects a common problem with pK_a benchmarks: a comparatively poor test set. In this test set, similar to most others, the vast majority of experimental pK_a values is only shifted by -1.5 to 0.5 pH units and therefore reflects the model pK_a value rather than the $\Delta pK_a^{Water \rightarrow Protein}$ perturbation.

We also find by comparing the colored error-fraction curves with the black dashed curves that the error distributions do not follow the expected errors from a normal distribution. Instead, there is a larger fraction of residues that have a larger error. Even though PROPKA generally predicts pK_a values as reliably as other methods, it reveals a major problem for empirical pK_a predictors; they probably rely more than the more rigorous methods on the quality of the data set, and not surprisingly, it has been found that PROPKA did especially well for surface residues.[15]

We also find that much of the improvement comes from reducing the number of outliers; looking at the scatter-plot inset to the error fraction plot presented in Figure 4, we find that PROPKA2 has two extreme outliers, Asp 73 in barnase and Glu 178 in *Bacillus agaradhaerens* xylanase, whereas the outliers for PROPKA3 are much less pronounced. This is indeed also verified by the rmsd 0.73, 0.79, and 1.03 we instead get if we reduce the data set with four residues (the two worst for each version). However, from the error fraction in figure 4 we can see that PROPKA3 (red curve) is slightly better than PROPKA2 (green curve) for the entire range of
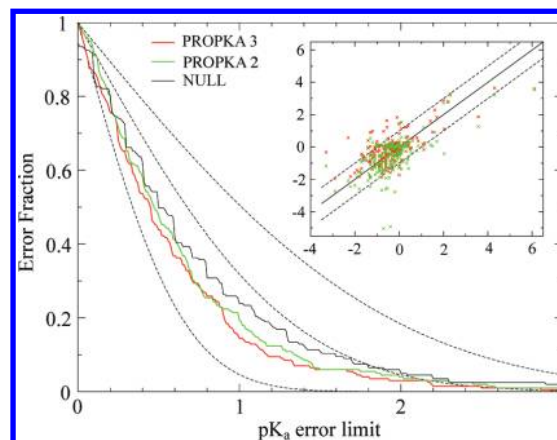

Figure 4. Error-fraction plot shows that PROPKA3 (red) is an overall improvement compared to PROPKA2 (green) using all 201 Asp+Glu residues considered in this study; this is also confirmed by the rmsd of 0.79 and 0.91, respectively.
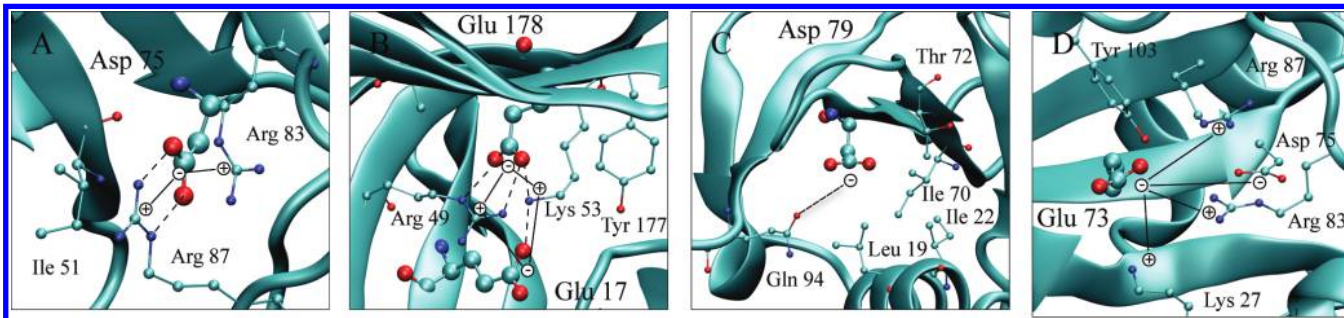
Figure 5. The figure shows the most important interactions identified by PROPKA3 determining the pK_a value of (A) Glu 75 in Barnase, (B) Glu 17 and Glu 178 in *Bacillus agaradhaerens* xylanase, (C) Asp 79 in RNase SA, and (D) Glu 73 in Barnase as discussed in the text. Hydrogen bonds are depicted by dashed lines and Coulomb interactions are indicated by solid lines. A: Hydrogen bonds with Arg 83 lowers the pK_a value of Glu 75 by 1.0 pH unit and Coulomb interactions with Arg 83 and Arg 87 by 0.9 pH units each. B: The Coulomb interaction with Arg 49 lowers the pK_a value of Glu 17 by 1.7 pH units, and the interaction with Lys 53 lowers it by 1.9 pH units (0.9 from hydrogen bond and 1.0 from Coulomb). Interactions with Arg 49 lowers the pK_a value of Glu 178 with 2.9 pH units (0.9 and 2.0) and Lys 53 lowers it with 2.5 pH units (0.8 and 1.7). C: Desolvation from nearby hydrophobic residues raises the pK_a value of Asp 79; an unfavorable interaction with Gln 94 probably raises the pK_a further making it more similar to the experimental value. D: A hydrogen bond from Tyr 103 lowers the pK_a value of Glu 73 by 0.9 pH units, and Coulomb interactions with Lys 27, Arg 83, and Arg 87 lowers it further by 0.5, 0.4, and 0.4 pH units.

error limits, which indicates that the improvement cannot necessarily be attributed to fewer outliers.

If we instead of using all 201 pK_a values follow more traditional test-set philosophy and remove the 85-point training set from the test set, the rmsd increases to 0.86, 1.08, and 1.17 for PROPKA3, PROPKA2, and the Null-model. This seems to be only a moderate increase for PROPKA3, whereas it is more significant for PROPKA2. This seems quite counterintuitive since many of these residues have been included in parametrizing PROPKA2 but not PROPKA3. However, this reduces the size of the test set significantly and is therefore uncertain how statistically significant these number dissections are, which is why we prefer relying on the extended test set. Never the less, obtaining clearly better rmsd for PROPKA3 for a test set depleted of uncomplicated residues and enriched with difficult, and generally biased toward PROPKA2, again underlines that PROPKA3 is really a significant improvement over PROPKA2.

In this study, we have not made an effort separating between Asp and Glu residues since we believe the systematic error between the two is probably much smaller than the error for the individual Asp or Glu residues, but we note that the rmsd is 0.77 (101 residues) for Asp and 0.80 (100 residues) for Glu residues using PROPKA3. This corroborates that the accuracy is not significantly different between these residues as suggested by the studies of Stanton et al.[16] and Song et al.[30] (we obtain 0.78 (89) for Asp and 0.82 (94) for Glu using the data set in Song et al). Instead, the corresponding rmsd for the Null-model, 1.23 and 0.86 (1.28 and 0.84 for the Song et al data set), indicates that the data set for Asp contains far more shifted residues than that of Glu and is therefore inherently more difficult to predict. This further illustrates the importance of having comparable data sets when making validation comparisons.

Specific Residues. *Asp 75 in Barnase:* As was discussed previously,[17] PROPKA2 predicts Asp 75 in barnase to have a pK_a value of −1.2, giving an error of 4.3 pH units compared to the experimental 3.1. This was attributed to a possible miss-assignment with Asp 54. PROPKA3, however,

predicts this pK_a value to be close to 4.8, which is in much better agreement with experiment, but instead overshooting it with 1.7 pH units. The difference between these extremes can be traced to the larger, and more realistic, desolvation contribution from the close Arg 83, Arg 87 and Ile 51 residues. In PROPKA2, the residue is identified as buried, and experiences the full lowering of pK_a from the Arg residues (2 × 2.4 pH units), but only a small part of the desolvation penalty, whereas these contributions are better balanced in PROPKA3 (see figure 5A). This is a good example of the challenge to get the balance between electrostatic solvation and the desolvation penalty correctly (this residue was not included in the PROPKA3 training set).

Glu 178 and Glu 17 in Bacillus agaradhaerens Xylanase: Glu 178 was also found to have an extreme pK_a value of −0.4, giving it an error of 4.5 pH units compared to the observed 4.1. In this case, the error was attributed to local structural distortions by a ligand sugar. However, PROPKA3 predicts the pK_a value to be close to 3.6, which again is in much better agreement with experiment. Again, the discrepancy can be traced to the desolvation contribution by nearby Lys 53, Arg 49, Tyr 177, and Ala 180 residues (see Figure 5B). Also here PROPKA2 identifies the residue as buried, and it experiences the full lowering of the close Lys and Arg residues but only a small part of the desolvation penalty. The same improved behavior is seen for the nearby Glu 17, where the pK_a value is predicted to be 1.7 with PROPKA2, whereas the 4.2 obtained with PROPKA3 is much closer to the experimental 4.3. Again, the same Lys and Arg residues lowers the pK_a value for PROPKA2 but is not compensated for correctly by the desolvation penalty.

Asp 76 in RNase T1: Asp 76 in RNase T1 is predicted to be 3.5, which makes it the residue with the largest error for PROPKA3 in this study (compared to 0.5 for the experiment). This large error is difficult to rationalize using pdbfiles 1I0V and 1RGA, but we note that there is a nearby disulfide bridge, and since the titration experiment has been conducted in a low-pH environment (0.5), it seems reasonable that the large discrepancy could come from using a neutral-pH

PROPKA3

J. Chem. Theory Comput., Vol. 7, No. 2, 2011 **535**

structure when the experiment titration represents a low-pH environment without an SS-bond. A broken SS-bond would probably result in additional stabilization of the ionized form of the acid and lower its pK_a values. We also note that the PROPKA2 prediction is only slightly better, 2.8, and MCCE2 obtains similar unshifted pK_a values as we do, 4.1.[30]

Asp 79 in RNase SA: Asp 79 in RNase SA is predicted to have a pK_a value of 5.7, which is a 1.6 pH-unit underestimate compared to the experimental 7.3. The significant upshift in pK_a comes predominantly from the desolvation contribution from the local hydrophobic region formed by Leu 19, Thr 72, Ile 70, and Ile 22. However, inspecting the pdbfile one finds also an unfavorable side-chain amide CO interaction from Gln 94 pointing directly toward the acid (see Figure 5C). These types of interactions are presently not included in the PROPKA rules and undoubtedly this interaction make up parts of the pK_a discrepancy (we get about 0.5 pH units using the parameters for the backbone ΔpK_a^{RE} term).

Glu 73 in Barnase: Three close bases surround Glu 73 in barnase: Lys 27, Arg 87, and Arg 83 (at 4.6, 6.3, and 6.4 Å distance, see Figure 5D). Intuitively, it seems that a residue in this environment would have a significantly downshifted pK_a value because of the charge–charge stabilization, which is confirmed by an experimental pK_a of 2.2. However, PROPKA3 seem to underestimate these close Coulomb interactions (or possibly overestimate the desolvation) and we obtain 5.1; both MCCE2 and PROPKA2 predict the pK_a to be 2.1.

Glu 94 and Glu 184 in Bacillus agaradhaerens Xylanase: PROPKA3 and PROPKA2 predict the pK_a value of these residues to be 6.1 and 7.4 and 4.9 and 7.2, whereas it is found to be 3.9 and 6.5 experimentally. Even though both versions identify the catalytic residues as nucleophile and acid correctly, PROPKA2 is closer to their experimental values. However, the differences in determinants are not as significant as the pK_a values might imply; overall they are similar, but the desolvation is 0.6 pH units larger and the hydrogen bonds from Tyr 85, Tyr 96 and Gln 143 is together 0.7 pH units smaller. A possible source of error in PROPKA3 could be the Coulomb interactions with two nearby bases, that is, Arg 49 and Arg 129. Considering that the active-site region is buried and contains two tyrosine, two tryptophan, and at least one phenylalanine residue, it seems also plausible that the effective dielectric constant should be smaller than what we use (30) and the Coulomb interaction between the acids and bases therefore larger. This would lower the pK_a values of both catalytic acids.

As we have gone through the Asp and Glu acids that give PROPKA3 the biggest errors, it seems that much of the problems might be attributed to a too small Coulomb interaction with nearby bases, which typically leads to an over estimation of down-shifted pK_a values. The simplest remedy to this would be to reduce the dielectric constant. However, it was found that reducing ε to 20 and 80 resulted in worse rmsd for the training set, and as can be seen from Figure 4 most residues are well accounted for by $\varepsilon_{\text{buried}} = 30$ and $\varepsilon_{\text{surface}} = 160$; the exception is a number of down-shifted pK_a values. Including some of these residues in the

training set and refitting ε would clearly improve our model, however, it would also reduce the value of our test set.

Tyrosine, Lysine, and Histidine. Even though the focus in this study is on carboxylic acids, we also briefly assess the expected accuracy for predicting Tyr, Lys, and His pK_a values. If we use the experimental compilation of Song et al[30] and Stanton et al[16] and remove pK_a values that are determined to an upper and lower-limit, those deemed ambiguous assignment, and those where we only find NMR structures, we obtain 11 Tyr residues, 51 Lys residues, and 31 His residues. The error distribution for these data sets can be found in figures S4–S6 in the Supporting Information. For Tyr, the 11 residues result in an rmsd of 0.75, 0.97, and 0.70 for PROPKA3, PROPKA2, and the Null-model respectively. It may seem unfortunate that the Null-model has the lowest rmsd. However, in this case it reflects a rather poor test set whose residues are predominantly unshifted from their model values. For such a data set even the most accurate prediction methods would have problems out performing the Null-model. The only conclusion that seems reasonable at this point is that PROPKA3 represents a significant improvement compared with PROPKA2.

For the 51 Lys residues, we obtain an rmsd of 0.65, 0.72, and 1.01 for PROPKA3, PROPKA2, and the Null-model respectively. In this case we have significantly more points that covers a larger range, especially with the two mutations V66K and M102K (in the mutant C54T/C97A/M102K) in staphylococcal nuclease and RNase T1 where the residues have been buried in hydrophobic patches in the protein and therefore have 4 pH-unit down-shifted pK_a values. As expected, the rmsd of the Null-model is much worse when including these significantly shifted residues. Removing these points result in an rmsd of 0.65, 0.68, and 0.64, showing that much of the difficulty comes from these two residues. Never the less, we can also for Lys see an improvement for PROPKA3 compared to PROPKA2. From the inset scatter plot in figure S7, it can be seen that PROPKA2 predicts all Lys pK_a shifts to be negative. This comes from omitting the hydrogen-bond interactions with neighboring protein residues, thus, only including desolvation and Coulomb contributions. Since Lys is typically found close to the protein surface and PROPKA2 does not include surface Coulomb contributions, the pK_a shift is effectively determined by the desolvation penalty alone. Thus, the important balance between desolvation and protein resolvation is overturned resulting in these unrealistic predictions. In PROPKA3, however, all terms are included consistently, albeit parametrized for carboxylic acids, which results in an overall better physical description and a slightly better rmsd; part of this should probably also be attributed to the improved desolvation model.

For the 31 His residues, we obtain an overall rmsd of 1.00, 1.35, and 0.92 for PROPKA3, PROPKA2, and the Null-model, respectively. Before looking closer at His, however, it should be noted that we have in addition to the above criteria also excluded the residues coming from myoglobin from this data set since the large heme-ligand can presumably have significant influence on the pK_a value or structure and PROPKA3 cannot treat ligands yet. Much of the difference

in rmsd between PROPKA3 and PROPKA2 can be traced to four slightly downshifted pK_a values (experimental pK_a shift -0.5 to -0.2) that PROPKA2 greatly overestimate (calculated pK_a-shift -4.6 to -2.0). These are all buried residues that in PROPKA2 use the special extended radius for calculating the local desolvation, which seem to be greatly overestimated. The improved desolvation model in PROPKA3, on the other hand, seems to treat the desolvation penalty more balanced and predict the pK_a values much better. PROPKA3, on the other hand, seem to underestimate stabilizing interactions for three residues: the pK_a values of His 12 in RNase A, His 31 in Lysozyme T4, and His 53 in RNase SA are all underestimated by $2-2.5$ pH units. Judging from these 31 residues, it seems that PROPKA3 is a significant improvement over PROPKA2. However, considering the better rmsd for the Null-model and the generally very spread out scatter plot, we can only conclude that His residues continue to be a challenge for PROPKA and their rules needs to be revised at some point. Saying this, one should also keep in mind that these parameters have been obtained for carboxylic acids, and predicting His pK_a values is probably inherently much more difficult since the charge is much more delocalized and there are two potential titration sites.

Comparing to Other pK_a Predictors. Clearly, it is not straightforward to compare different pK_a predictors and validate them relative each other since it strictly requires us to compare the rmsd using identical data sets, but two of the data sets we have included in this study can give us an indication. If we reevaluate the rmsd values using the 183 Asp and Glu residues we have included from the Song et al. compilation,[30] we obtain rmsd values of 0.80 (PROPKA3), 0.91 (PROPKA2), 0.85 (MCCE2), and 1.08 (Null-model), and find that the performance is quite similar. If we instead use the 40 Asp and Glu residues we have included from the more stringent data set used in Stanton and Houk,[16] we obtain rmsd values of 0.96 (PROPKA3), 1.06 (PROPKA2), 0.74 (microenvironment SCP[32]), 0.97 (geometry-dependent dielectric[33]), 1.15 (MD/GB/TI[34]), 1.18 (EGAD[35]), 1.31 (MCCE[36]), and 1.37 (Null-model). Since this study does not include all pK_a values for all methods when obtaining the rmsd, we avoid judging the methods further, but conclude that PROPKA3 is a viable option even when compared to more rigorous approaches.

Conclusions

The overall goal of this study has been to clean up the ad hoc parametrization of previous versions of PROPKA and to treat the physics healthier. This includes using a more appropriate form of Coulomb interactions and a better model for desolvation penalties, but the most important achievement has been to treat all residues consistently without any discontinuous jumps and generalizing the effective potentials by minimizing the number of exceptions. The resulting new version of PROPKA has, not counting model pK_a values, six adjustable parameters (two for Coulomb interactions, two for desolvation penalties, and two for intrinsic electrostatics) and was parametrized against a subset of 85 unproblematic Asp and Glu residues with experimentally known pK_a values

and reasonable crystal structures. PROPKA2 had officially ten corresponding adjustable parameters (one for Coulomb interactions, two for desolvation penalties, and seven for hydrogen bonding), and was parametrized against 314 experimental pK_a values (many were even included in the test sets used to determine the accuracy in this study). However, effectively many of the radii and cutoffs in PROPKA2 should also be considered as parameters since they generally are an intricate part of determining the pK_a values and very interaction specific, and were frequently adjusted based on experimental pK_a values. In principle, there is nothing wrong with using a larger set of parameters; on the contrary, it is clearly better if the observable requires it, but in this case many of these seem to have compensated for an inappropriate model. For the new version, however, we have improved the model and based these radii on structure observations and unbiased descriptors rather than values adjusted to improve experimental data points.

In an ideal world where we have a "correct" ensemble of protein structures in their protein + water solvent (i.e., not from a crystal), we could most likely reduce these to three (one for each interaction type). If we indeed had such a situation, though, we would on the other hand considered it justified to parametrize the hydrogen bonding interactions individually, for example, for the ROH, CONH, NH, and SH interactions, since hydrogen bonding abilities are clearly different. When we arrive at that, we will probably see an increase in accuracy, but as long as we do not have better structures and more experimental values, there seem to be little use in diversifying the parameters. In fact, obtaining an rmsd close to 0.6 for the training set and 0.79 for the largest test set should in many aspects be considered better than expected; approximations such as using a minimum protein model and not including protein reorganization explicitly seems intuitively to be rather crude. The results are nevertheless encouraging. It should also in this context be noted that much of the success for PROPKA as a pK_a predictor tool does not lie in its functions being better or more accurate. Instead, it lies in that it tries to predict the bare minimum, is comparatively insensitive, and utilizes cancelation of errors efficiently. In fact, most other pK_a predictors are clearly more rigorous and can therefore be seen as having better rules. However, as these methods try to predict things rigorously, the larger are the effects involved and the bigger is the potential to over or underestimate their values. Finally, it should also be recognized that the most important parameter in virtually all current pK_a predicting programs is actually the reference model value pK_a^{Water}.

Supporting Information Available: Several additional tables and figures. This information is available free of charge via the Internet at http://pubs.acs.org/.

PROPKA3

J. Chem. Theory Comput., Vol. 7, No. 2, 2011 **537**

Note Added after ASAP Publication. This article was published ASAP on January 6, 2011. Reference 31 has been modified. The correct version was published on January 13, 2011.

References

(1) Raquet, X.; Lounnas, V.; LamotteBrasseur, J.; Frere, J. M.; Wade, R. C. *Biophys. J.* **1997**, *73*, 2416–2426.

(2) Nielsen, J. E.; Mccammon, J. A. *Protein Sci.* **2003**, *12*, 1894–1901.

(3) Lamotte-Brasseur, J.; Lounnas, V.; Raquet, X.; Wade, R. C. *Protein Sci.* **1999**, *8*, 404–409.

(4) Lamotte-Brasseur, J.; Dubus, A.; Wade, R. C. *Proteins* **2000**, *40*, 23–28.

(5) Warshel, A. *Biochemistry* **1981**, *20*, 3167–3177.

(6) Warshel, A. *Acc. Chem. Res.* **1981**, *14*, 284–290.

(7) Gunner, M. R.; Mao, J. J.; Song, Y. F.; Kim, J. *Biochim. Biophys. Acta, Bioenerg.* **2006**, *1757*, 942–968.

(8) Gunner, M. R.; Alexov, E. *Biochim. Biophys. Acta, Bioenerg.* **2000**, *1458*, 63–87.

(9) Beroza, P.; Case, D. A. *Energ. Biol. Macromol., Part B* **1998**, *295*, 170–189.

(10) Warshel, A.; Papazyan, A. *Curr. Opin. Struct. Biol.* **1998**, *8*, 211–217.

(11) Ullmann, G. M.; Knapp, E. W. *Eur. Biophys. J. Biophys.* **1999**, *28*, 533–551.

(12) Nielsen, J. E.; McCammon, J. A. *Protein Sci.* **2003**, *12*, 313–326.

(13) Mongan, J.; Case, D. A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 157–163.

(14) Lee, A. C.; Crippen, G. M. *J. Chem. Inf. Model.* **2009**, *49*, 2013–2033.

(15) Davies, M. N.; Toseland, C. P.; Moss, D. S.; Flower, D. R. *BMC Biochem.* **2006**, *7*.

(16) Stanton, C. L.; Houk, K. N. *J. Chem. Theory. Comput.* **2008**, *4*, 951–966.

(17) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins* **2005**, *61*, 704–721.

(18) Bas, D. C.; Rogers, D. M.; Jensen, J. H. *Proteins* **2008**, *73*, 765–783.

(19) Tanford, C.; Kirkwood, J. G. *J. Am. Chem. Soc.* **1957**, *79*, 5333–5339.

(20) Warshel, A.; Russell, S. T.; Churg, A. K. *Proc. Natl. Acad. Sci., India, Sect. B* **1984**, *81*, 4785–4789.

(21) Tanford, C.; Roxby, R. *Biochemistry* **1972**, *11*, 2192-&.

(22) Mallik, B.; Masunov, A.; Lazaridis, T. *J. Comput. Chem.* **2002**, *23*, 1090–1099.

(23) Schutz, C. N.; Warshel, A. *Proteins* **2001**, *44*, 400–417.

(24) Sham, Y. Y.; Chu, Z. T.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 4458–4472.

(25) Mehler, E. L.; Guarnieri, F. *Biophys. J.* **1999**, *77*, 3–22.

(26) Jensen, J. H.; Li, H.; Robertson, A. D.; Molina, P. A. *J. Phys. Chem. A* **2005**, *109*, 6634–6643.

(27) Olsson, M. H. M.; Hong, G. Y.; Warshel, A. *J. Am. Chem. Soc.* **2003**, *125*, 5025–5039.

(28) Powers, N.; Jensen, J. H. *J. Biomol. NMR* **2006**, *35*, 39–51.

(29) Forsyth, W. R.; Antosiewiez, J. M.; Robertson, A. D. *Proteins* **2002**, *48*, 388–403.

(30) Song, Y. F.; Mao, J. J.; Gunner, M. R. *J. Comput. Chem.* **2009**, *30*, 2231–2247.

(31) pKacoop. http://amylase.ucd.ie/pKacoop (accessed Jan 1, 2011).

(32) Hassan, S. A.; Guarnieri, F.; Mehler, E. L. *J. Phys. Chem. B* **2000**, *104*, 6490–6498.

(33) Wisz, M. S.; Hellinga, H. W. *Proteins* **2003**, *51*, 360–377.

(34) Simonson, T.; Carlsson, J.; Case, D. A. *J. Am. Chem. Soc.* **2004**, *126*, 4167–4180.

(35) Pokala, N.; Handel, T. M. *Protein Sci.* **2004**, *13*, 925–936.

(36) Georgescu, R. E.; Alexov, E. G.; Gunner, M. R. *Biophys. J.* **2002**, *83*, 1731–1748.