Received: 11 July 2018 | Revised: 5 September 2018 | Accepted: 14 September 2018

DOI: 10.1002/prot.25608

WILEY PROTEINS

RESEARCH ARTICLE

DelPhiPKa: Including salt in the calculations and enabling polar residues to titrate

Swagata Pahari | Lexuan Sun | Sankar Basu | Emil Alexov

Department of Physics and Astronomy, Computational Biophysics and Bioinformatics, Clemson University, Clemson, South Carolina

Correspondence
Emil Alexov, Department of Physics and Astronomy, Computational Biophysics and Bioinformatics, Clemson University, Clemson, SC 29634.
Email: ealexov@clemson.edu

Funding information
National Institute of General Medical Sciences, Grant/Award Number: R01GM093937

Abstract

DelPhiPKa is a widely used and unique approach to compute pK_a's of ionizable groups that does not require molecular surface to be defined. Instead, it uses smooth Gaussian-based dielectric function to treat computational space via Poisson-Boltzmann equation (PBE). Here, we report an expansion of DelPhiPKa functionality to enable inclusion of salt in the modeling protocol. The method considers the salt mobile ions in solvent phase without defining solute-solvent boundary. Instead, the ions are penalized to enter solute interior via a desolvation penalty term in the Boltzmann factor in the framework of PBE. Hence, the concentration of ions near the protein is balanced by the desolvation penalty and electrostatic interactions. The study reveals that correlation between experimental and calculated pK_a's is improved significantly by taking into consideration the presence of salt. Furthermore, it is demonstrated that DelphiPKa reproduces the salt sensitivity of experimentally measured pK_a's. Another new development of DelPhiPKa allows for computing the pK_a's of polar residues such as cysteine, serine, threonine and tyrosine. With this regard, DelPhiPKa is benchmarked against experimentally measured cysteine and tyrosine pK_a's and for cysteine it is shown to outperform other existing methods (DelPhiPKa RMSD of 1.73 vs RMSD between 2.40 and 4.72 obtained by other existing pK_a prediction methods).

KEYWORDS

electrostatics, Gaussian-based dielectric function, pKa's, Poisson-Boltzmann equation, proteins, salt concentration

1 | INTRODUCTION

The acidic and basic amino acid residues of a protein may be ionized at a particular pH. The ionization state of these residues contribute to conformation, stability, solubility, and function of the protein as a function of pH.[1–3] Hence, ionizable residues of biological molecules play a significant role in protein-protein interaction, protein-ligand binding, enzymatic reaction, and so forth.[4,5] As pK_a determines the ionization state of the residues, it is important to know the pK_a values of titrable groups in proteins and how these values depend on the surrounding environment. However, accurate prediction of pK_a's of ionizable groups remains a challenge since it requires to determine the equilibrium distribution of ionization states of side chain titrable residues along with the conformational changes.

Several computational techniques have been explored for pK_a's calculations.[6–10] These methods can be broadly classified into three classes: (*a*) microscopic; (*b*) empirical; and (*c*) macroscopic methods. All these techniques have their weakness and strength in capturing the biophysical insights which in turn establish their ability to match the experimentally determined pK_a values. All these three classes of approaches are described in detail by Alexov et al.[6] Microscopic methods such as molecular dynamics (MD), quantum mechanics molecular mechanics (QM/MM) consider the atomic level of detail of the system while calculating the thermodynamics properties. Generally, these methods are more accurate as they incorporate the conformational change of protein while a residue changes from protonated to unprotonated form. However, convergence problem in performing the configurational sampling, large computational cost makes these methods unrealistic for most of the protein applications. On the other hand, empirical methods use knowledge-based terms and scoring functions to describe the effect of the environment on pK_a's of protein residues.[11–14] Since they do not consider conformational sampling explicitly, they are very fast and in some cases quite accurate. Indeed, PROPKA has recently been found to be a reliable protein pK_a

predictor,[14–17] with reported RMSD as low as 0.89. Other classes of methods are macroscopic methods based on continuum electrostatics (CE) using either Poisson Boltzmann (PB) or General Born (GB) equations. Traditional implementation of these methods requires that the computational space be divided into solute and solvent regions and no conformational changes to be considered. Several groups tried to improve the CE approaches by coupling ionization and conformation using Monte Carlo (MC) sampling technique, introducing side chain flexibility,[18–20] modifying hydrogen bond orientation,[21,22] and adding extensive side chain rotamer sampling.[22,23] All these additional implementations made the PB-based CE method more accurate. Recently, we reported a Gaussian-based method, where a smooth Gaussian-based dielectric function describes the dielectric properties of the solute and solvent on the same footage. This pK_a's prediction method, which has been already implemented in DelPhi,[24] uses Gaussian-based smooth dielectric function which in turn mimic the plausible conformational changes associated with changes of ionization states. In this approach, there is no need to define molecular surface while calculating the electrostatic energies and the corresponding pK_a's.

The electrostatic interactions within solute depend on the properties of the solvent. Typically charge residues are located on the surface of the solute, these are accessible to salts (NaCl, KCl, K_2SO_4, etc.) which are present in the solvent. Thus, the presence of salt affects the screening of electrostatic interactions. Furthermore, ions can bind to oppositely charged protein residues[25,26] or also can cause chaotropic effect in solvent structure. Therefore, protein stability can be changed in the presence of salt. It mostly depends on the charge distribution on protein surface and in turn on pH as indicated by the following experimental observations. It has been found that the stability of chymotrypsinogen increases with salt concentration below pH 1.5 but it decreases at higher pH. Surprisingly, no effect on stability of barnase has been revealed in presence of 600 mM KCl at pH 3.5 but it stabilizes the protein at higher and lower pH. Hence, the effect of salt on protein stability is not linear and upfront. Therefore, it is difficult to understand clearly the role of salt from investigation at a single pH. It is essential to study the titration curve at a wide range of pHs and salt concentration to reveal the interplay between ionization processes and salt contribution.

Typically, titratable groups are considered to be Asp, Glu, His, Lys, and Arg residues, since they are frequently ionized at physiological pH. However, there are other residues, typically refereed as polar residues, which may also titrate in physiological pH. The list includes Ser, Thr, Cys, and Tyr residues. These groups play an important role when present in catalytic site and the state of protonation is the key to their function. Therefore, it is essential to determine the pK_a for these groups to understand catalytic reaction and its pH dependence. For example, serine protease is an enzyme that catalyzes the hydrolysis of a peptide bond with an active site serine residue.[27] The serine residue in this case acts as a nucleophile during the catalysis. Furthermore, the catalytic cycles of cysteine protease and tyrosine phosphatase involve the deprotonation of cysteine or tyrosine.[28,29] Lastly, in cysteine protease family, thiolate is an essential intermediate which undergoes nucleophilic attack in the active site of protein.[30,31] Therefore, it is essential to know the ionization state of serine, cysteine, tyrosine or threonine to understand the pH dependence of catalysis.

In this work, we report two new functionalities of DelPhiPKa. The first one is the addition of salt in the modeling protocol. This is done without determining solute-solvent interface, that is, molecular surface free protocol. The upgraded DelphiPKa is benchmarked against experimentally measured pK_a's and is shown to deliver better results compared with the previous version. The second development allows for polar residues such as cysteine, serine, threonine and tyrosine to be treated as titratable residues and their pK_a's to be predicted. The DelPhiPKa predictions for pK_a's of cysteine residues in catalytic site are tested by comparing with experimental data and it is shown that the results are better than previously reported by other existing methods. DelPhiPKa is also shown to reproduce the experimental pK_a for tyrosine and to predict the pK_a shift based on the surrounding environment for serine.

2 | METHODS

We briefly describe the DelPhiPKa methods for predicting pK_a's of titrable residues here. Details of implementation of the method can be found elsewhere.[24] DelPhiPKa method calculates the probability of protonation of each titrable residue as a function of pH and determines the pK_a as pH at which probability of protonation is 50%. For doing so, the electrostatic free energy of the titrable residues in their protonated as well as deprotonated states is calculated using DelPhi built-in module. It is important to mention that the calculations are done with smooth Gaussian-based dielectric function and do not require determining solute-solvent interface.[32,33] Our previous works[32,33] have demonstrated that such an approach results in dielectric function inside the macromolecule that varies from low reference value of about 2 up to 20 and more depending on the atomic packing. At the van der Waals (vdW) surface, it increases smoothly and reaches the value of 80 in bulk water. Thus, Gaussian-based smooth dielectric function is designed to capture several effects: (*a*) dielectric inhomogeneity of macromolecules; (*b*) existence of water cavities inside, if any; (*c*) the fuzziness of the macromolecule-water region; and (*d*) the ability of ions to visit space close to vdW surface.

3 | ELECTROSTATIC FREE ENERGY

Smooth Gaussian based dielectric function is used for all the electrostatic energy calculations, the implementation of the model in DelPhi has been described in previous work.[34] The details of the calculation of electrostatic energy and its components (charge-charge pairwise interaction, $G^{pairwise}_{i,j(charged)}$; polar energy term, $G^{polar}_{i,charged}$; desolvation energy, $\Delta G^{\text{desol}}_{i,\text{charged}}$) were explained in earlier paper[24]. These three energy components are determined for *i*th residue in both charged ($G^{pairwise}_{i,j(charged)}$, $G^{polar}_{i,charged}$, $\Delta G^{\text{desol}}_{i,\text{charged}}$) as well as neutral state ($G^{paiwise}_{i,j,neutral}$, $G^{polar}_{i,neutral}$, $\Delta G^{desol}_{i,neutral}$). We obtain the change in pairwise interaction, polar energy term and delsovation energy due to change in protonation state as follows:

$$\Delta G_{i,j}^{pairwise} = G_{i,j,charged}^{pairwise} - G_{i,j,neutral}^{pairwise}, \quad (1)$$

$$\Delta G_{i}^{polar} = G_{i,charged}^{polar} - G_{i,neutral}^{polar}, \quad (2)$$

$$\Delta\Delta G_{i}^{desol} = \Delta G_{i,charged}^{desol} - \Delta G_{i,neutral}^{desol}. \quad (3)$$

Therefore, the total electrostatic energy of the protein at given protonation state is expressed as

$$G_i = \gamma(i)\left[2.3k_bT\left(pH - pk_{a_i}^{ref,solvent}\right)\right] + \left(\Delta G_i^{polar} + \Delta\Delta G_i^{desol}\right) + \sum_{j=1,j\neq i}^{N} \Delta G_{i,j}^{pairwise}. \quad (4)$$

In this way, total electrostatic energy of each microstate is calculated and using the distribution of this energy, probability of ionization of *i*th residue is calculated at a given pH. The probability of ionization of *i*th residue at a given pH is determined using the Boltzmann distribution of corresponding electrostatic energy

$$P_i = \frac{\sum_{m=1}^{M} \chi(i).e^{-G_m(pH)/kT}}{\sum_{m=1}^{M} e^{-G_m(pH)/kT}}, \quad (5)$$

where system has total M microstates and at *m*th microstate, energy is $G_m(pH)$ at a given pH. $\chi(i)$ is 1 when *i*th residue is ionized and 0 if it is neutral. The probability of ionization of each ionizable residue is calculated as a function of pH. The pH at which probability of ionization is 50% is defined as pK_a. To reduce computational cost, DelPhiPKa applies clustering approach to enumerate the Boltzmann distribution as described in the original paper.[32]

The treatment of mobile ions in the framework of Poisson-Boltzmann equation (PBE) by adding a desolvation penalty term is introduced by Jia et al.[35] The same method is used here for treating a given concentration of salt which is comprised of mobile ions. The desolvation penalty is calculated via Born equation

$$\Delta G_{solv} = \frac{N_A z^2 e^2}{8\pi\epsilon_0 r_0}\left(\frac{1}{\epsilon_r} - \frac{1}{\varepsilon_w}\right), \quad (6)$$

where N_A is the Avogadro constant, z is the valence, e is the charge, r_0 is the effective radius of the ion, ϵ_0 is the permittivity of the vacuum. ϵ_r and ε_w are the dielectric constant at a given location and in bulk water, respectively. Therefore, incorporating the desolvation penalty in PBE, the corresponding PBE is written as

$$\nabla.[\varepsilon(r)\nabla\varphi(r)] = -4\pi\left(\rho_{solute}(r) + \sum_{i=1}^{N} 2q_i c^{bulk}\left[\exp\left(\frac{-\Delta G_{solv}}{RT}\right)\right]\left(\frac{-q_i\varphi(r)}{RT}\right)\right). \quad (7)$$

The ΔG_{solv} term is nonzero only in the close proximity of the protein molecules, it is zero in bulk water.

4 | DATABASE AND PARAMETERS OF DELPHIPKA

In order to predict the accuracy of the method, we benchmarked the calculated pK_a's against an experimental protein pK_a's database (http://compbio.clemson.edu/databases/database_pKa.xlsx) compiled by us, which is a large dataset containing 82 proteins and 773 residues. Most of the residues of this database are surface exposed and accessible to the salt. The database includes protein pK_a database (http://pka.engr.ccny.cuny.edu/) and cases compiled from literature. Note that pK_a-cooperative database is not suitable for this work, since here we emphasize on the role of salt, while pKa-cooperative set is comprised of buried residues away from water phase (and thus not sensitive to salt). The structure files were taken from PDB databank. The missing atoms and residues were re-generated using PROFIX (a software module within the JACKAL package, https://honiglab.c2b2.columbia.edu/software/Jackal/Jackalmanual.htm#profix). We removed all the compounds, other than the amino acids, such as heme, SO_4, PO_4, all ions and water molecules. For the pK_a's calculations, we used sigma = 0.70 and internal dielectric = 8 in DelphiPKa. These parameters are optimized through benchmarking against experimental data.[24] It is shown in an earlier paper[24] that the Gaussian dielectric provides a significant improvement in pK_a's prediction compared to homogeneous model. Therefore, in this work, Gaussian smooth dielectric model is used. The atomic charges and radii are taken from Amber forcefield parameters for all the calculations.[36]

5 | RESULTS AND DISCUSSIONS

With regards with the new features of DelPhiPKa reported in this work, several investigations were carried out. First, we investigated the change of the performance of DelPhiPKa upon adding the salt option; this was done by benchmarking the predicted pK_a's against experimentally determined pK_a's (see Methods section) in presence and absence of salt. Second, we probed the performance of DelPhiPKa to predict experimentally measured pK_a shifts of various residues due to variation in salt concentrations. Finally, we checked DelPhiPKa performance against experimental data of pK_a's of Cys and Tyr residues. We also tested the ability of DelPhiPKa in predicting pK_a shift for Ser residue depending on the structural environment.

5.1 | Does the salt option improve pK_a's predictions?

The total RMSD and correlation coefficient for the predicted pK_a's of total 752 residues (Asp, Glu, His, and Lys) using DelphiPKa is 0.74 and 0.96, respectively, in presence of salt (I = 0.15 M). The plot of experimental vs calculated pK_a's is shown in Figure 1A. The slope and intercept of the linear fit is 0.94 and 0.41, respectively. The deviation of calculated from experimental pK_a, both for all the residues as well as for each residue type, are presented in Table 1. Among the pK_a values of 752 residues, for 487 residues (65%), the deviation of calculated values from experimental one is less than 0.5 pK unit. For 660 residues (88%), the deviation is less than 1 pK unit and only for 3% of the residues, the shift is greater than 2 pK unit. Most of the residues, which falls under these 3%, are buried inside the protein (Table 2). Thus, Table 2 indicates that buried residues (RSA < 10) have higher RMSDs irrespective of residue types. Better predictions can be made if larger variance of the Gaussian function is used as pointed in earlier work.[32] Focusing on different residue types separately, it has been found that

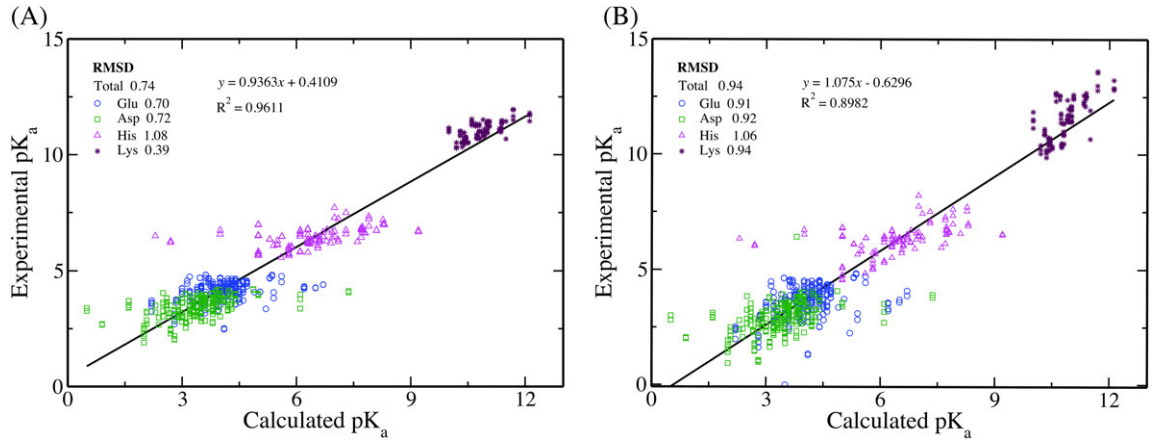

FIGURE 1 Benchmark of calculated pK_a against protein pK_a database for each individual residue type at (A) I = 0.15 M and (B) I = 0

the RMSD for Glu and Asp is similar (0.70 and 0.72, respectively) [Figure 1A]. However, for Lys, the RMSD is significantly lower (0.39) and for HIS it is the largest (1.08). Hence, the overall correlation of the calculated pK_a's with experiment seems exceptionally good for such a large dataset.

The agreement between calculated pK_a's and the experimental one drops significantly when salt concentration is not taken into account. The total RMSD in this case becomes 0.94. Therefore, presence of salt plays a crucial role in determining the pK_a of ionizable residues. The plot of experimental pK_a vs calculated pK_a for all the 752 residues in absence of salt is shown in Figure 1B. The slope and intercept for the linear fit are 1.075 and 0.63, respectively. Figure 1B also displays deviation of calculated pK_a with respect to experimental one for each individual residue types.

The RMSDs for Glu, Asp, and Lys in absence of salt are much higher (0.91, 0.92, and 0.94, respectively) compared to RMSDs in presence of salt (0.70, 0.72, and 0.39, respectively). For His, it is practically the same (1.06 without salt and 1.08 in presence of salt). These results indicate that salt effect is playing a vital role in determining ionization state of titratable groups.

It should be pointed out that the improvement of the accuracy of pK_a's predictions is not only due to inclusion of the salt, but the treatment of the system via a smooth Gaussian-based dielectric function. If one uses the traditional protocol that considers solute as a low dielectric cavity with a sharp border between solute and solvent, with the same parameters and salt concentration reported above, the corresponding RMSD is 1.24 (Figure S1), compared with 0.74 above.

For accessing the quality of the results, we compared the calculated data with the null model predictions, that is setting all protein pK_a shifts to 0 and so, all the pK_a values are equal to the intrinsic pK_a of the corresponding residues. In Table 3, the RMSD of all the residue types are shown in presence of salt (I = 0.15 M), absence of salt (I = 0) and null model. Table 3 reflects that DelPhiPKa outperforms the null model by an average of 0.19 pK unit. While this is a relatively small improvement, it should be pointed out that previous works were unable to improve the null model.[16,37] One of the reasons for such relatively small improvement is that the experimental shifts of pK_a values are only between 0.5 and 1.5 pK units.

TABLE 1 Statistics of the deviation of calculated pK_a from the experimental dataset

$\lvert pKa_{exp}\text{-}pKa_{cal}\rvert$	Total (752)	Glu (258)	Asp (275)	His (111)	Lys (109)
>0.5	35%	37%	32%	54%	20%
>1.0	12%	13%	11%	24%	3%
>1.5	6%	6%	6%	11%	...
>2.0	3%	2%	4%	6%	...
>2.5	...	...	2%	5%	...
>3.0	...	...	1%	3%	...

TABLE 2 Statistics of RMSD based on the residue positions

	Total		Glu		Asp		His		Lys	
Relative surface accessible area	%	RMSD	%	RMSD	%	RMSD	%	RMSD	%	RMSD
RSA < 10	0.07	1.75	0.06	1.41	0.08	1.66	0.16	2.07	...	...
10 < RSA < 20	0.06	0.96	0.04	1.27	0.08	0.49	0.11	1.31	0.01	0.17
20 < RSA < 50	34	0.64	34	0.59	27	0.72	45	0.72	37	0.44
50 < RSA	53	0.53	57	0.59	57	0.51	30	0.65	63	0.36

TABLE 3 RMSD for each residue type

	GLU	ASP	HIS	LYS
Total number	258	275	111	109
Calc. pK_a at I = 0.15	0.70	0.75	1.08	0.39
Calc. pK_a at I = 0	0.91	0.92	1.06	0.94
Null-model	1.04	0.93	1.21	0.49

5.2 | Benchmarking DelPhiPKa against experimental data of salt dependence of pK_a's

Here, we investigate whether the predicted pK_a shifts due to change in salt concentration correlate with the experimental data. We collected four datasets from several studies[38–40] for four different proteins where pK_a's were determined experimentally as a function of salt concentration (salt concentration was varied in the range of 0.01-1.5 M). Lee et al.[38] performed acid/base titrations of SNase by using H NMR spectroscopy at different concentrations of KCl. It is reported that with increase in the salt concentration from 0.01 M to 1.5 M, pK_a values of His-8, His-46, His-121, and His-124 increase by 0.92, 0.44, 1.05, and 0.93 pK units, respectively. We calculated the pK_a values of these four His residues using DelPhiPKa. The salt dependencies of experimental and calculated pK_a values for SNase are compared in supporting information (Table S1). The correlation between the calculated and experimental pK_a values and the variations of both experimental and calculated pKa's as a function of salt concentration for these four His residues are shown in Figure S2 and Figure S3, respectively. The correlation coefficient (*R*) of 0.95, 0.96, 0.98, and 0.99 for His-8, His-46, His-121, and His-124, respectively, indicates that DelPhiPKa methods can well reproduce the salt dependence of measured pK_a values.

Kao et al.[39] investigated the salt dependence of His pK_a values of sperm whale myglobin (Pc-Mb). The experimentally measured pK_a values indicate that despite of having difference in their solvent accessibility and nature of the surrounding residues, most of the His residues exhibit similar rise of their pK_a values (~0.3 pK unit) on changing the salt concentration from 0.02 M to 1.5 M (Table S2). The correlation of experimental vs calculated pK_a's and the change in experimental as well as predicted pKa's with variation of salt concentration for all the His residues are presented in Figures S4 and S5, respectively. In Table S2, for His-36, measured experimental value indicates that pK_a's are insensitive to salt concentration between 0.02 M and 1.5 M. DelPhiPKa also predicts that there is a negligible increase (0.02 pK unit) in pK_a's by increasing the salt concentration from 0.02 to 0.2 M and a slight change of 0.05 pK unit going from 0.5 M to 1.5 M of salt. This observation reveals that the DelPhiPKa model is reasonably accurate in predicting the effect of salt. We estimated the correlation coefficient of the calculated pK_a's for all the His residues except His-36 and plotted in Figure S4. We obtained the correlation coefficient of 0.98, 0.91, 0.85, 0.98, 0.63, and 0.99 for His-12, His-48, His-81, His-113, His-116, and His-119, respectively, for Pc-Mb. In the same paper,[39] authors also described the salt sensitivity of His pK_a for horse heart Mb (Eq-Mb). The calculated pK_a's are compared with the experimentally measured pK_a's in Table S3. Similarly to Pc-Mb, in Eq-Mb, for His-36, the pK_a values are found to be not influenced by salt concentration both in experiment and model calculation. Correlation between experimental and calculated pK_a's and variation of these pKa's with salt concentration are plotted in Figure S6 and S7 respectively for the 6 His residues in Eq-Mb excluding His-36. We can see in Figure S6 that DelPhiPKa predicted pK_a's behave the same way as experimental pK_a's, as we change the salt concentration.

Abe et al.[40] evaluated the influence of salt concentration on the pK_a values of acidic residues in hen egg white lysozyme. The predicted pK_a values along with experimental pK_a's are shown in Table S4 which reveals that except for Glu35 and Asp87, DelPhiPKa can capture the salt sensitivity of the pKa for other residues similar way that in experiment. Glu35 is located in the active site of lysozyme and it is not surface exposed. Therefore, the influence of the presence of salt is negligible in this case. We can observe in Table S4 that the pK_a of Asp87 decreases with increase of salt concentration from 0.005 M to 0.1 M but increases from 0.1 M and 0.4 M. Such an unusual behavior may be caused by the fact that Asp87 is located at the N-terminal of α-helix (residues 88-98) and the conformation may be affected by the salt concentration. Since in our calculations, the structure is kept rigid, this may explain why we were unable to reproduce the abnormal change of experimentally measured pK_a. The correlation between predicted and experimental pK_a's is estimated and plotted for six acidic residues (except Glu35 and Asp87) in Figure S8. The correlation coefficient for Glu7, Asp18, Asp48, Asp52, Asp101, and Asp119 is determined to be 0.95, 0.99, 0.88, 0.94, 0.93, and 0.99, respectively. In Figure S9, the variation of both experimental and calculated pKa with varying salt concentration is plotted.

Lastly, we compared all the calculated pK_a's with the corresponding experimental data coming from 78 residues and four proteins and plotted them together in Figure 2 (excluding His-36 in both Pc-Mb and Eq-Mb, Glu35 and Asp87 in lysozyme). The correlation coefficient of 0.92 and RMSD of 0.54 indicate that DelPhiPKa method can accurately predict the effects of salt concentration.

5.3 | Benchmarking Cys pK_a's

The pK_a of a Cys residue, not involved in disulfide bridge, can vary significantly compared to their intrinsic pK_a. It was shown experimentally that pK_a's of Cys residues in non-catalytic sites are usually in the range

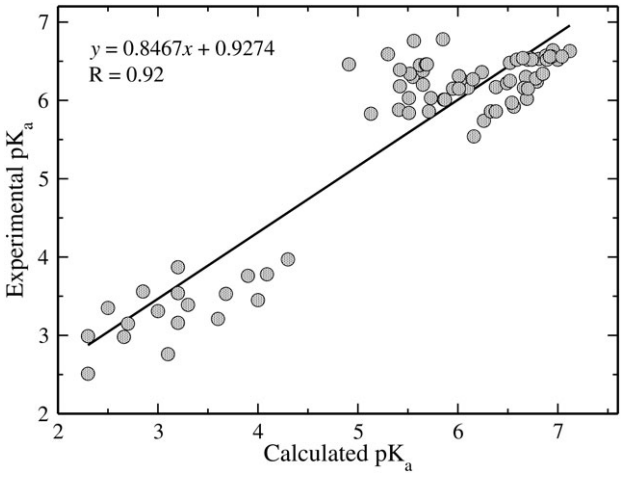

FIGURE 2 Overall correlation between experimental and calculated pK_a's at various salt concentrations

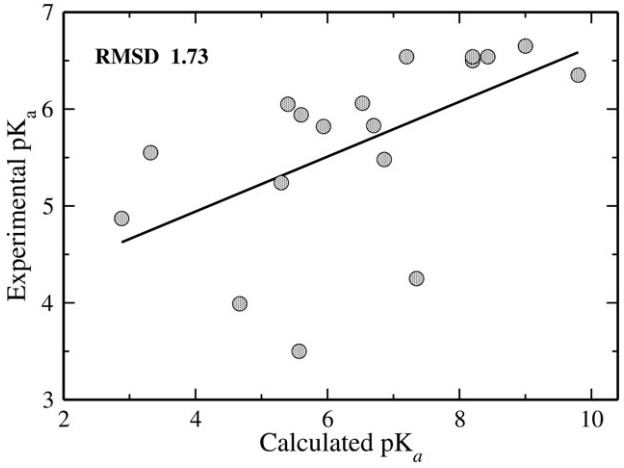

FIGURE 3 Correlation between experimental and calculated cysteine pK$_a$'s using the new DelPhiPKa

of 7-9 whereas in the catalytic site they can be as low as 2.88.[41–43] Thus, the pK$_a$ value of Cys residue depends on the environment that protein provides. Awoonor-Williams and Rowley[44] generated a Cys pK$_a$'s experimental data set comprised of 18 Cys residues in 12 proteins including both elevated and lowered pK$_a$'s. The same work calculated the pK$_a$ of these 18 residues with three methods using implicit solvent model (MCCE, PROPKA, and H++) as well as explicit solvent models with CHARMM36 and AMBER force field. The article concluded that the methods using implicit solvent model is not reliable in predicting Cys pK$_a$'s with RMSD between 3.41 and 4.72 whereas the RMSD obtained using explicit solvent models varied between 2.40 and 3.20. It was suggested by the authors that the methods still need to be improved for predicting Cys pK$_a$ since the null-model predicts the RMSD of 2.74 which is close to the best RMSD achieved by the explicit model. Here, we benchmark the new DelPhiPKa against the same dataset. The calculated pK$_a$'s are compared with the experimentally measured data and plotted in Figure 3. The RMSD obtained is 1.73 which is the lowest among all the five methods including implicit as well as explicit solvent models. Thus, the new version of DelPhiPKa outperforms all other existing methods (implicit and explicit water models) and also the null-model in predicting the Cys pK$_a$'s.

5.4 | Benchmarking Tyr pK$_a$'s

There are only few experimental pK$_a$ values for Tyr residues reported in the literature[45] among which structure files are available for three of them.[46,47] Hence, we calculated and compared the pK$_a$'s only for these three residues belonging to two different proteins. The predicted pK$_a$ for Tyr30 and Tyr49 in ribonuclease Sa (RNase Sa) is 10.42 and 10.60 respectively whereas the experimentally measured pK$_a$'s are 11.3 and 10.63, respectively.[45] Khare et al.[47] measured the pK$_a$ of Tyr33 in the B1 immunoglobulin G- (IgG-) binding domains of protein G, the value of which is 11 whereas DelPhiPKa predicts the pKa as 10.72. These results indicate that DelPhiPKa is successful in predicting pKa's for Tyr residues as well (an example of Ser pKa is provided in supplementary material, Figure S10).

6 | CONCLUSIONS

A new upgraded version of DelPhiPKa is reported in this work that enables salt concentration to be included in the surface-free protocol. In this way, DelPhiPKa remains a unique continuum electrostatic approach to calculate pK$_a$'s without determining the boundary between solute and solvent. The inclusion of the salt in the calculations is shown to deliver better results compared to the cases without salt. This confirms previous observation made in case of salt dependent protein-protein binding.[35] It should be pointed out that the improvement in pK$_a$'s predictions is significant (about 0.2 pK units) and it is achieved on larger dataset (larger than the original pK$_a$ dataset, see Methods section). Furthermore, the protocol was tested against experimentally determined changes of pK$_a$'s upon change of the salt concentration and very good correlations were obtained. It should be mentioned that both experimental and computed pK$_a$ changes are very small (less than half of pK$_a$ unit), which makes the comparison quite difficult. Lastly, DelPhiPKa was enabled to calculate pK$_a$'s of groups typically referred as polar groups. It was shown that the method outperforms all existing method and lower the RMSD even compared to explicit water models. Thus, here we report important enrichments of DelPhiPKa capabilities and significant improvement of its prediction accuracy. The improvement of prediction accuracy is attributed to both, the Gaussian-based smooth dielectric function and the novel treatment of ions.

ACKNOWLEDGMENTS

We thank Neocles Basil for pointing to us the necessity of allowing pK$_a$'s predictions for polar groups. The work was supported by a grant from , grant number .

ORCID

Emil Alexov https://orcid.org/0000-0001-5346-0156

REFERENCES

1. Honig B, Nicholls A. Classical electrostatics in biology and chemistry. *Science*. 1995;268(5214):1144-1149.
2. Warshel A, Russell ST. Calculations of electrostatic interactions in biological systems and in solutions. *Q Rev Biophys*. 1984;17(3):283-422.
3. Onufriev AV, Alexov E. Protonation and pK changes in protein-ligand binding. *Q Rev Biophys*. 2013;46(2):181-209.
4. Cherny VV, Murphy R, Sokolov V, Levis RA, DeCoursey TE. Properties of single voltage-gated proton channels in human eosinophils estimated by noise analysis and by direct measurement. *J Gen Physiol*. 2003;121(6):615-628.
5. Bashford D. Macroscopic Electrostatic Models for Protonation States in Proteins. *Front Biosci*. 2004;9:1082-1099.
6. Alexov E, Mehler Ernest L, Baker N, et al. Progress in the prediction of pKa values in proteins. *Proteins: Struct Funct Bioinform*. 2011;79(12):3260-3275.
7. Nicholls A, Honig B. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J Comput Chem*. 1991;12(4):435-445.
8. Bashford D. An object-oriented programming suite for electrostatic effects in biological molecules. An experience report on the MEAD project. Paper presented at: Scientific Computing in Object-Oriented Parallel Environments; 1997; Berlin, Heidelberg.

9. Mehler EL. The Lorentz-Debye-Sack Theory and Dielectric Screening of Electrostatic Effects in Proteins and Nucleic Acids. *Theoretical and Computational Chemistry*. 1996;3:371-405.
10. Schutz Claudia N, Warshel A. What are the dielectric "constants" of proteins and how to validate electrostatic models? *Proteins: Struct Funct Bioinform*. 2001;44(4):400-417.
11. Godoy-Ruiz R, Perez-Jimenez R, Garcia-Mira MM, Plaza del Pino IM, Sanchez-Ruiz JM. Empirical parametrization of pK values for carboxylic acids in proteins using a genetic algorithm. *Biophys Chem*. 2005;115(2):263-266.
12. Spassov VZ, Karshikov AD, Atanasov BP. Electrostatic interactions in proteins. A theoretical analysis of lysozyme ionization. *Biochim Biophys Acta (BBA) - Protein Struct Mol Enzymol*. 1989;999(1):1-6.
13. Krieger E, Nielsen JE, Spronk CAEM, Vriend G. Fast empirical pKa prediction by Ewald summation. *J Mol Graph Model*. 2006;25(4):481-486.
14. Li H, Robertson Andrew D, Jensen JH. Very fast empirical prediction and rationalization of protein pKa values. *Proteins: Struct Funct Bioinform*. 2005;61(4):704-721.
15. Bas Delphine C, Rogers David M, Jensen JH. Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins: Struct Funct Bioinform*. 2008;73(3):765-783.
16. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J Chem Theory Comput*. 2011;7(2):525-537.
17. Søndergaard CR, Olsson MHM, Rostkowski M, Jensen JH. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. *J Chem Theory Comput*. 2011;7(7):2284-2295.
18. You TJ, Bashford D. Conformation and hydrogen ion titration of proteins: a continuum electrostatic model with conformational flexibility. *Biophys J*. 1995;69(5):1721-1733.
19. Beroza P, Case DA. Including side chain flexibility in continuum electrostatic calculations of protein titration. *J Phys Chem*. 1996;100(51):20156-20163.
20. Warwicker J. Improved pK(a) calculations through flexibility based sampling of a water-dominated interaction scheme. *Protein Sci*. 2004;13(10):2793-2805.
21. Nielsen JE, Andersen KV, Honig B, et al. Improving macromolecular electrostatics calculations. *Protein Eng Des Sel*. 1999;12(8):657-662.
22. Nielsen Jens E, Vriend G. Optimizing the hydrogen-bond network in Poisson-Boltzmann equation-based pKa calculations. *Proteins: Struct Funct Bioinform*. 2001;43(4):403-412.
23. Georgescu RE, Alexov EG, Gunner MR. Combining conformational flexibility and continuum electrostatics for calculating pK(a)s in proteins. *Biophys J*. 2002;83(4):1731-1748.
24. Wang L, Li L, Alexov E. pKa predictions for proteins, RNAs, and DNAs with the Gaussian dielectric function using DelPhi pKa. *Proteins: Struct Funct Bioinform*. 2015;83(12):2186-2197.
25. Petukh M, Kimmet T, Alexov E. BION web server: predicting non-specifically bound surface ions. *Bioinformatics*. 2013;29(6):805-806.
26. Petukh M, Zhenirovskyy M, Li C, Li L, Wang L, Alexov E. Predicting nonspecific ion binding using DelPhi. *Biophys J*. 2012;102(12):2885-2893.
27. Sprang S, Standing T, Fletterick RJ, et al. The Three-Dimensional Structure of Asn102 Mutant of Trypsin: Role of Asp102 in Serine Protease Catalysis. *Science*. 1987;237:905-909.
28. Denu JM, Dixon JE. Protein tyrosine phosphatases: mechanisms of catalysis and regulation. *Curr Opin Chem Biol*. 1998;2(5):633-641.
29. Kolmodin K, Åqvist J. The catalytic mechanism of protein tyrosine phosphatases revisited. *FEBS Lett*. 2001;498(2):208-213.
30. Otto H-H, Schirmeister T. Cysteine proteases and their inhibitors. *Chem Rev*. 1997;97(1):133-172.
31. Chapman HA, Riese RJ, Shi G-P. Emerging roles for cysteine proteases in human biology. *Annu Rev Physiol*. 1997;59(1):63-88.
32. Wang L, Li L, Alexov E. pKa predictions for proteins, RNAs, and DNAs with the Gaussian dielectric function using DelPhi pKa. *Proteins*. 2015;83(12):2186-2197.
33. Li L, Li C, Zhang Z, Alexov E. On the dielectric "constant" of proteins: smooth dielectric function for macromolecular modeling and its implementation in DelPhi. *J Chem Theory Comput*. 2013;9(4):2126-2136.
34. Li L, Li C, Zhang Z, Alexov E. On the dielectric "constant" of proteins: smooth dielectric function for macromolecular modeling and its implementation in DelPhi. *J Chem Theory Comput*. 2013;9(4):2126-2136.
35. Jia Z, Li L, Chakravorty A, Alexov E. Treating ion distribution with Gaussian-based smooth dielectric function in DelPhi. *J Comput Chem*. 2017;38(22):1974-1979.
36. Aliev AE, Kulke M, Khaneja HS, Chudasama V, Sheppard TD, Lanigan RM. Motional timescale predictions by molecular dynamics simulations: Case study using proline and hydroxyproline sidechain dynamics. *Proteins*. 2014;82(2):195-215.
37. Antosiewicz J, McCammon JA, Gilson MK. The determinants of pKas in proteins. *Biochemistry*. 1996;35(24):7819-7833.
38. Lee KK, Fitch CA, Lecomte JTJ, García-Moreno EB. Electrostatic effects in highly charged proteins: salt sensitivity of pKa values of histidines in staphylococcal nuclease. *Biochemistry*. 2002;41(17):5656-5667.
39. Kao Y-H, Fitch CA, Bhattacharya S, Sarkisian CJ, Lecomte JTJ, García-Moreno EB. Salt effects on ionization equilibria of histidines in myoglobin. *Biophys J*. 2000;79(3):1637-1654.
40. Abe Y, Ueda T, Iwashita H, et al. Effect of salt concentration on the pK_a of acidic residues in lysozyme. *J Biochem*. 1995;118(5):946-952.
41. Pinitglang S, Watts AB, Patel M, et al. A classical enzyme active center motif lacks catalytic competence until modulated electrostatically. *Biochemistry*. 1997;36(33):9968-9982.
42. Zhang ZY, Dixon JE. Active site labeling of the Yersinia protein tyrosine phosphatase: the determination of the pKa of the active site cysteine and the function of the conserved histidine 402. *Biochemistry*. 1993;32(36):9340-9345.
43. Lohse DL, Denu JM, Santoro N, Dixon JE. Roles of aspartic acid-181 and serine-222 in intermediate formation and hydrolysis of the mammalian protein-tyrosine-phosphatase PTP1. *Biochemistry*. 1997;36(15):4568-4575.
44. Awoonor-Williams E, Rowley CN. Evaluation of methods for the calculation of the pKa of cysteine residues in proteins. *J Chem Theory Comput*. 2016;12(9):4662-4673.
45. Grimsley Gerald R, Scholtz JM, Pace CN. A summary of the measured pK values of the ionizable groups in folded proteins. *Protein Sci*. 2008;18(1):247-251.
46. Laurents DV, Huyghues-Despointes BMP, Bruix M, et al. Charge-charge interactions are key determinants of the pK values of ionizable groups in ribonuclease Sa (pI=3.5) and a basic variant (pI=10.2). *J Mol Biol*. 2003;325(5):1077-1092.
47. Khare D, Alexander P, Antosiewicz J, Bryan P, Gilson M, Orban J. pKa measurements from nuclear magnetic resonance for the B1 and B2 immunoglobulin G-binding domains of protein G: comparison with calculated values for nuclear magnetic resonance and X-ray structures. *Biochemistry*. 1997;36(12):3580-3589.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Pahari S, Sun L, Basu S, Alexov E. DelPhiPKa: Including salt in the calculations and enabling polar residues to titrate. *Proteins*. 2018;86:1277–1283. https://doi.org/10.1002/prot.25608