JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

# *Pcetk*: A pDynamo-based Toolkit for Protonation State Calculations in Proteins

Mikolaj Feliks[†,‡,§] and Martin J. Field*[,†,‡,§]

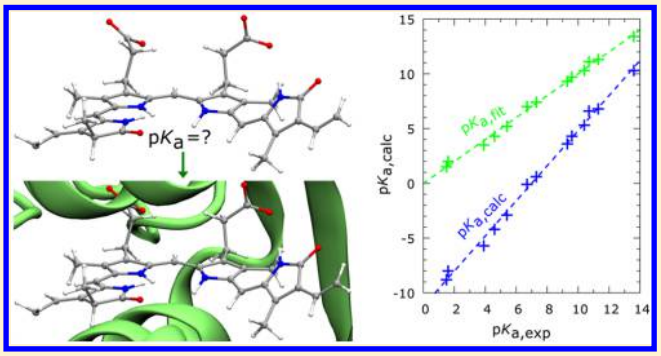
[†]Université Grenoble Alpes, IBS, F-38044 Grenoble, France

[‡]CNRS, IBS, F-38044 Grenoble, France

[§]CEA, IBS, F-38044 Grenoble, France

Supporting Information

**ABSTRACT:** *Pcetk* (a pDynamo-based continuum electrostatic toolkit) is an open-source, object-oriented toolkit for the calculation of proton binding energetics in proteins. The toolkit is a module of the pDynamo software library, combining the versatility of the Python scripting language and the efficiency of the compiled languages, C and Cython. In the toolkit, we have connected pDynamo to the external Poisson−Boltzmann solver, extended-MEAD. Our goal was to provide a modern and extensible environment for the calculation of protonation states, electrostatic energies, titration curves, and other electrostatic-dependent properties of proteins. *Pcetk* is freely available under the CeCILL license, which is compatible with the GNU General Public License. The toolkit can be found on the Web at the address http://github.com/mfx9/pcetk. The calculation of protonation states in proteins requires a knowledge of p$K_a$ values of protonatable groups in aqueous solution. However, for some groups, such as protonatable ligands bound to protein, the p$K_a^{aq}$ values are often difficult to obtain from experiment. As a complement to *Pcetk*, we revisit an earlier computational method for the estimation of p$K_a^{aq}$ values that has an accuracy of ±0.5 p$K_a$-units or better. Finally, we verify the *Pcetk* module and the method for estimating p$K_a^{aq}$ values with different model cases.




## INTRODUCTION

The prediction of protonation states in protein structures acquired from X-ray crystallography represents a nontrivial modeling problem.[1] Proteins usually contain a large number of protonatable groups that can bind or release protons. These groups include, for example, the side-chains of aspartate, glutamate, and histidine, as well as different types of protonatable ligands buried in the protein. Moreover, some of these groups or ligands are multiprotic and can bind or release more than one proton per group. The knowledge of the protonation states of these groups is crucial for understanding many biologically relevant phenomena, such as enzyme catalysis.[2]

The process of binding and release of protons is controlled by pH, and by the interactions of protonatable groups with their molecular environment.[1,3] A protein with $N$ protonatable groups will have $2^N$ possible protonation states, assuming that each group can exist in only two forms, protonated or deprotonated. For even a small protein, $2^N$ is a very large number, which makes it difficult to consider individually all possible protonation states and their relative energies.[4]

Many modeling approaches have been used to assign the positions of protons in crystal structures of proteins. These include assignment based on visual inspection of the crystal structure, chemical intuition or empirical databases or statistics. Some of these, such as the well-known PROPKA methods,[5,6] can give results that agree well with experiment. However, different case studies show that simple approaches may not always be reliable.

Computational methods based on the Poisson−Boltzmann equation are among the most accurate *ab initio* approaches for studying proton binding energetics in proteins.[1] In this paper, we describe an open-source software toolkit based on the pDynamo library[7] and the extended-MEAD package[8,9] that implements a continuum electrostatic model for protonation state calculations in proteins. The majority of software in the field of computational chemistry is traditionally written in Fortran or C. However, there has been a trend during the past decade or so to use higher-level languages, such as Python, and object-oriented design in scientific programming.[10,11] At the same time, the open-source approach to the licensing of scientific software has gained increased recognition, including in the chemical community.[12−14] Software packages following these new trends include, for example, pDynamo[7] for quantum chemical/molecular mechanical calculations, MMTK[15] for classical simulations, PyQuante for the development of quantum chemical methods, cclib[16] for the analysis of quantum

**Received:** May 7, 2015
**Published:** September 22, 2015





DOI: 10.1021/acs.jcim.5b00262
J. Chem. Inf. Model. 2015, 55, 2288−2296

chemical calculations, PyMOL[17] for molecular editing and visualization, Biopython[18] for biological computation, and many others.

Although electrostatic calculations based on the Poisson−Boltzmann equation have become an almost routine approach for studying the titration of proteins, we developed our toolkit because we felt that there was a lack of software to perform these calculations in a convenient manner. The toolkit is closely integrated with pDynamo, but it can also be coupled to other Python-based programs and tools. The use of the Python scripting language and an object-oriented programming model means that the toolkit provides a flexible and easily extensible framework for studying the titration of proteins. The input file to perform a calculation is simply a Python script, which enables the automation of many routinely performed tasks. However, the script itself does not have to compromise between the ease of use of Python and the efficiency of calculations, because all time-consuming routines of the toolkit have been written in C and Cython.

To calculate the protonation state of a titratable group in a protein, it is necessary to know, as a prerequisite, the p$K_a$ value of this group in aqueous solution, p$K_a^{aq}$. p$K_a^{aq}$ values of titratable groups are usually acquired from experiments and can be looked up in different databases or compilations. However, the p$K_a^{aq}$ values of unusual titratable groups, such as protein-bound ligands, are often not available, because they may be difficult to measure or otherwise determine experimentally. Alternatively, these unknown p$K_a^{aq}$ values can be estimated by using computational approaches.[19−21] In this paper, as a complement to our developmental work on the *Pcetk* toolkit, we expand on the recently proposed method of Muckerman and co-workers[20] for the calculation of absolute p$K_a^{aq}$ values. The method is able to predict the p$K_a^{aq}$ values of different organic compounds with a satisfactory accuracy of ±0.5 p$K_a$-unit. Together, the new pDynamo module and the extended method for determining p$K_a^{aq}$ values from a framework that allows for the treatment of both standard and unusual protonatable groups in proteins.

In what follows, we first introduce the basic concepts and functionalities of the developed toolkit. We do not discuss all details of the employed computational methods, because they are well established and have been extensively reviewed.[1,2,22−26] Instead, we focus mainly on the practical side of the Poisson−Boltzmann model and its calculation. Next, we introduce a method for the calculation of p$K_a$ values of unusual protonatable groups in aqueous solution. Finally, we discuss two examples. The first is a basic study that features the calculation of protonation states in lysozyme, using the *Pcetk* toolkit. The second determines the protonation states of the IFP2.0 fluorescent protein that contains the biliverdin chromophore, which is a complex protonatable group.

## METHODS AND IMPLEMENTATION

The development of the toolkit follows a modern approach to software engineering,[14] and the code of the program is maintained under a version control system in a publicly accessible repository. The toolkit is a module of the pDynamo library, with different functionalities split into various submodules. These functionalities and their implementation are comprehensively discussed in section S1 in the Supporting Information. In the present section, we concentrate on the description of the used theoretical models.

**Calculation of the Electrostatic Potential.** Electrostatic interactions play a dominant role in proteins.[27−29] The electrostatic potential for a solvent-immersed protein is usually calculated by using the linearized form of the Poisson−Boltzmann equation. Details of this approch have been reviewed many times,[1,30] so we only recapitulate here a few important aspects. First, the system is divided into a low dielectric region, which is the protein phase, and a high dielectric region, which is the solvent phase. Although the choice of the dielectric constant for each region may not always be straightforward,[29] values of $\epsilon_p = 4$ and $\epsilon_s = 80$ are often used. Prior to the calculation, it is necessary to know the partial charges and radii of protein atoms. These are usually taken from a force field or, in the case of unusual protonatable groups, determined from *ab initio* calculations. The protein is mapped onto a grid by using a simple interpolation scheme.[30] Each node of the grid is assigned a dielectric constant according to the boundary between the protein and solvent. The boundary is determined by rolling a ball imitating a solvent molecule on the surface of the protein. Finally, the Poisson−Boltzmann equation is solved numerically by using the finite difference method. The values of the electrostatic potential are calculated on grid nodes. Usually, a series of grids of increasing resolution is used, which is known as focusing. During each focusing step, the initial values of the electrostatic potential are taken from the previously calculated values on the coarser grid. The starting grid is centered on the protein and the subsequent finer grids are centered on the protonatable group of interest.

*Pcetk* splits the protein model into different groups of charges, namely the protonatable residues and a nonprotonatable protein surrounding, and employs MEAD to calculate the electrostatic potential and interaction energies between them. Because of the flexible programming model that we have used, it is possible in future versions of the program that MEAD could be complemented or replaced by a different Poisson−Boltzmann solver.

**Basic Concepts of the Electrostatic Model.** Titratable groups in a protein, such as aspartates, glutamates, histidines, or protonatable ligands, will be referred to as titratable sites, or just sites. Each site has different protonation forms. Most have just two: protonated and deprotonated, although sites such as histidine or ligands with more than two protonatable positions exist in multiple forms; these are called multiprotic sites. Different forms of a site are called instances. In the presently described model, the instances of a site differ only in the values of their atomic charges, so changing the protonation state of a site is equivalent to changing its set of atomic charges. No atoms are physically added or removed during the protonation or deprotonation reactions. Instead, removable atoms, for example the carboxyl hydrogen during the deprotonation of glutamate, are simply assigned zero charges.

Each instance of a site is characterized by its label, a set of charges, the relative energy of the model compound, $G^{model}$, and the number of bound protons. We note that other electrostatic models include instances of sites that differ also in conformation; these are called rotameric instances,[31,32] although we do not treat these here. For each instance of each site, we perform calculations first in a model compound, which is the reference system, and second in the protein, which is the target system (see Figure S1).

Model compounds are constructed by taking all atoms of the residue to which the site belongs, with additional backbone fragments from the two neighboring residues. Thus, a model compound can be understood as a complete amino acid, for which a p$K_a^{aq}$ value has been experimentally determined, with


DOI: 10.1021/acs.jcim.5b00262
J. Chem. Inf. Model. 2015, 55, 2288−2296

some additional molecular environment. Model compounds of protein-bound ligands are handled in a similar way, but they do not include the additional molecular neighborhood. The protein model is divided into titratable sites and a nontitratable background. The charges of the background are constant and so are independent of the current protonation state of the protein.

*State Vector.* The protonation state of a protein with $N$ protonatable sites is represented by an $N$-dimensional vector. Each component of this so-called state vector represents a particular site. The value of the component determines which instance of the site is active in the current protonation state of the protein. In existing approaches,[4,33] the components of this vector often take the values of 0 or 1 only, depending on whether the site is deprotonated or protonated, respectively. In this work, however, the value of the component does not necessarily indicate the actual protonation state of a site, but rather contains an index that identifies the currently active instance of a site. For example, for histidine the indices 0, 1, 2 indicate the fully, $\delta$-, and $\epsilon$-protonated instances, respectively, whereas 3 indicates the fully deprotonated instance. For every protonation state of the protein, otherwise known as a microstate, described by the state vector, one can calculate the microstate energy.

*Energy Function.* In a naive application of the Poisson−Boltzmann formalism, one would solve the Poisson−Boltzmann equation for the energy of each protonation state separately. For a protein with only standard protonatable residues, this would be $2^N\ 4^{N_{\mathrm{His}}}$ calculations, where $N$ is the number of titratable sites with only two possible forms, protonated and deprotonated, and $N_{\mathrm{His}}$ is the number of histidines, which have four. Once the state energies are known, they can be compared to determine the protonation state of the protein with the lowest energy and, hence, the highest energy. The energy function used in the present model, however, enables one to solve the Poisson−Boltzmann equation only $2N + 4N_{\mathrm{His}}$ times, which is computationally far more feasible. This approach is possible, because the protein model has been partitioned into separate protonatable sites, whose energies can be precalculated, tabulated, and looked up during the later calculations for different combinations of sites and their protonations.[1,34]

The energy of a protonation state, $G^{\mathrm{micro}}$, for a given state vector, $\vec{\nu}$, and pH, can be expressed as

$$G^{\mathrm{micro}}(\vec{\nu}, \mathrm{pH}) = \sum_{i=1}^{N^{\mathrm{sites}}} G_i^{\mathrm{intr}} - \sum_{i=1}^{N^{\mathrm{sites}}} n_i(-RT\ln 10\mathrm{pH}) + \sum_{i=1}^{N^{\mathrm{sites}}} \sum_{j}^{j<i} W_{i,j} \quad (1)$$

The first two terms are a sum of intrinsic energies, $G_i^{\mathrm{intr}}$, minus the current number of protons bound to each site, $n_i$, running over all titratable sites of the model. The value of each intrinsic energy depends on the currently active instance of the site, hence the $(\vec{\nu})$-dependence in eq 1. The third term is a double sum that collects electrostatic interaction energies, $W_{i,j}$, between active instances of sites. Each element of the symmetric matrix, $W_{i,j}$, represents a particular interaction of an instance of a site with another instance of another site.[1] To reduce numerical error, the matrix is symmetrized by $W_{i,j} = 1/2\ (W_{i,j} + W_{j,i})$ before the titration calculations.[35] Diagonal elements representing self-interactions are set to zero. Both $G_i^{\mathrm{intr}}$ and $W_{i,j}$ are precalculated with MEAD. Note that these components depend parametrically on the current temperature and ionic strength, although for clarity we have omitted this dependence from the equation.

**Calculation Protocol.** The aim of the calculation is to determine, at a given pH, ionic strength and temperature, the protonation states of protonatable groups in the protein. To achieve this aim, the *Pcetk* module follows a procedure consisting of two separate steps. During the first step, the protein model is analyzed and split into titratable sites and a nontitratable background. Water molecules, if present in the model, are removed, since they will be replaced by a continuum representation of the solvent. For each instance of each site, electrostatic energy terms are calculated using MEAD. The way in which pDynamo employs MEAD to calculate electrostatic energies is similar to how it interfaces with the program ORCA[36] to calculate quantum chemical energies and properties. The communication between the user script and MEAD is handled automatically by the module. Finally, the calculated electrostatic energies are used to obtain the intrinsic and interaction energies, $G_i^{\mathrm{intr}}$ and $W_{i,j}$, respectively. These energies are tabulated and the first calculation step is complete.

In the second step, the values of $G_i^{\mathrm{intr}}$ and $W_{i,j}$ are combined to calculate microstate energies and protonation state probabilities for the complete protein model. If the studied protein is reasonably small, the probabilities of protonation states can be calculated directly from the statistical mechanical partition function. However, this analytic treatment is not feasible for proteins with more than about 25 titratable sites, because of the combinatorial explosion of the number of possible protonation states. For larger proteins, the probabilities can be estimated by Monte Carlo sampling.[37] This can be done with the in-house routines implemented in the module or by employing the external sampling program, GMCT.[38] The application of the Monte Carlo method for the determination of protein protonation states has been described comprehensively elsewhere.[1,37]

The probabilities of protonation states can be calculated for a particular pH, for example pH = 7, or for a range of pH values, usually from 0 to 14, to give titration curves. After the lowest energy protonation state of the protein has been determined, it is possible to define a subset of titratable sites and calculate changes of the state energy depending on the protonation states of these sites. A comparison of these so-called substate energies can be used, for example, to determine how much energy is needed to protonate or deprotonate a titratable group of special interest in the protein, such as a catalytically important residue in the active site of an enzyme.[39,40]

*Calculation of the Electrostatic Energy Terms.* Because of the energy model that we employ, energy terms describing individual sites can be precalculated separately. This approach enables simple, coarse-grained parallelization that scales linearly with the number of CPUs. For each instance of a site, two electrostatic energy terms are calculated, namely the Born energy, $G^{\mathrm{Born}}$, and the background energy, $G^{\mathrm{back}}$. The former term is defined as the electrostatic interaction of a set of charges with its own reaction field, whereas the latter term describes the interaction of the set with other surrounding charges. The calculations of $G^{\mathrm{Born}}$ and $G^{\mathrm{back}}$ are performed separately for each instance of each site embedded first in the model compound and second in the protein (Figure S1). Additionally, for a site embedded in the protein, interaction energies, $W_{I,j}$, are


DOI: 10.1021/acs.jcim.5b00262
J. Chem. Inf. Model. 2015, 55, 2288−2296

calculated between the current instance of the site and all possible instances of other sites.

The calculations in the model compound and protein are performed by the extended-MEAD programs my_2diel_solver and my_3diel_solver, respectively. The continuum electrostic model consists of two phases characterized by different dielectric constants, usually $\epsilon_p = 4$ for the model compound/protein phase and $\epsilon_s = 80$ for the solvent phase. For the calculations discussed in this work, the solvent was modeled as water with an ionic strength of $I = 100$ mM and a temperature of $T = 300$ K. The molecular volume of the model compound/protein was defined by an ion exclusion layer of 2.0 Å and a solvent probe radius of 1.4 Å. The volume is determined automatically by MEAD. The Poisson−Boltzmann equation was solved on a starting grid of $121^3$ nodes at a resolution of 2.0 Å, followed by three focusing steps at resolutions of 1.0, 0.5, and 0.25 Å.

In the next step, the calculated self-and background energies can be combined into heterogeneous transfer energies:

$$\Delta G^{\text{heterotrans}} = (G^{\text{Born}}_{\text{protein}} + G^{\text{back}}_{\text{protein}}) - (G^{\text{Born}}_{\text{model}} + G^{\text{back}}_{\text{model}}) \quad (2)$$

For each instance of each site, the transfer energy is added to the model energy of the instance in aqueous solution, $G^{\text{model}}$, which gives the intrinsic energy in the protein, $G^{\text{intr}}$. The intrinsic energy of a protein-buried instance of a site can be interpreted as the energy that the instance would have if there were no interactions between the sites in the protein. For one of the instances of a site, $G^{\text{model}}$ is assumed to be zero, whereas the model energies of the other instances are equal to the free energies of the corresponding deprotonation reactions. The relation between $G^{\text{model}}$ and $pK_a^{\text{aq}}$ is simply:

$$G^{\text{model}} = RT \ln 10 pK_a^{\text{aq}} \quad (3)$$

*Calculation of the Probabilities of Protonation States.* As mentioned previously, the probabilities of protonation states can be calculated analytically or estimated by Monte Carlo sampling. In the $i$-th protonation state of the protein, the probability of occurrence of instance $\mu$ can be calculated as

$$\langle x_\mu \rangle = \frac{1}{Z} \sum_{i=1}^{N^{\text{states}}} \mu^i e^{-(G_i^{\text{micro}}/RT)},\ Z = \sum_{i=1}^{N^{\text{states}}} e^{-(G_i^{\text{micro}}/RT)} \quad (4)$$

The probabilities are simply observables of the partition function. The partition function, $Z$, requires the calculation of the state energies, $G^{\text{micro}}$, of all possible protonation states of the protein. The $\mu^i$ factor adopts the values of 1 or 0, depending on whether the instance is or is not active in the $i$-th protonation state of the protein. One of the methods of dealing with too many protonation states is via Metropolis Monte Carlo sampling.

Briefly, a Monte Carlo procedure starts from the generation of a random state vector, followed by a series of scans. During a scan, the state vector undergoes a number of random changes or moves. Single moves randomly choose a site and change its active instance. Double moves[37] choose a random pair of sites and simultaneously change both their instances. The purpose of double moves is to improve the sampling of strongly interacting sites in situations where the transition between two low-energy states would be hindered by high-energy states. The moves generate random protonation states of the protein, which can be accepted or rejected, depending on their calculated energies and the Metropolis criterion. If the energy of a new state is lower or equal to the energy of the old state, i.e., $\Delta G^{\text{micro}} = G^{\text{micro}}_{\text{new}} - G^{\text{micro}}_{\text{old}} \geq 0$, the new state is accepted. However, the new state can still be accepted, even with a higher energy, if the Metropolis criterion is fulfilled:

$$rand(0, 1) \leq e^{-\Delta G^{\text{micro}}/RT} \quad (5)$$

where $rand(0,1)$ is a randomly generated number between 0 and 1. Rejection of the new state means that the state vector reverts back to the old state.

The goal of each Monte Carlo scan is to generate a state vector representing a low-energy, statistically relevant protonation state of the protein. This low-energy state is then added to the final ensemble of states. The number of moves during the scan is proportional to the number of sites, so that each site has a chance to be sampled. Observables, such as probabilities of occurrence of individual instances, are calculated after the completion of all scans by counting the number of states where each instance is active, and averaging the counts over all accumulated states. For the generation of random numbers during the Monte Carlo sampling, *Pcetk* uses the Mersenne Twister generator,[41] as implemented in pDynamo.

We used the following Monte Carlo parameters for the calculations described in this work. Each MC calculation consisted of 500 equlibration scans followed by 20 000 production scans. Double moves were used for the pairs of sites whose absolute interaction energy was calculated to be 2.0 kcal/mol or greater. The frequency of double moves was proportional to the number of strongly interacting pairs. Titration curves were calculated at a resolution of 0.5 pH-units for a range of pHs between 0 and 14.

**Calculating $pK_a^{\text{aq}}$ Values of Nonstandard Protonatable Groups.** To calculate the protonation state of a titratable group in the protein, the $pK_a$ value of this group in aqueous solution has to be known. Protein-bound protonatable ligands and cofactors often represent a modeling problem, because their $pK_a^{\text{aq}}$ values are difficult to determine experimentally. A number of computational strategies have been proposed for the reliable prediction of unknown $pK_a^{\text{aq}}$ values.[42] With quantum chemical composite methods, such as Gaussian-$n$ or CBS, it is often possible to calculate reaction free energies in the gas-phase that are accurate to within 1 kcal/mol. However, the calculation of reaction free energies in solution remains difficult, mainly due to the deficiencies of the available solvation models, in particular when dealing with ionic species. These models, for example PCM[43] or COSMO,[44] often over- or underestimate solvation energies, because they do not take into account interactions at the solute−solvent interface, such as hydrogen bonding.[45]

In the present work, we modified and applied the recenly proposed method by Muckerman and co-workers[20] to the calculation of $pK_a^{\text{aq}}$ values of a multiprotic protonatable chromophore, biliverdin. Briefly, our modification relies on a more accurate calculation of the electronic energy contribution to the reaction free energy in vacuum, using the MP2/CBS level of theory (see section S2 in the Supporting Information for details). In a first step, a training set was prepared that consisted of 13 cationic acids with known $pK_a^{\text{aq}}$ values, as listed in Table 2. The acids were selected in such a way as to cover a wide range of $pK_a^{\text{aq}}$ values. For each acid, a straightforward calculation of its $pK_a^{\text{aq}}$ value was performed by using the protocol described in section S2. A comparison between these


DOI: 10.1021/acs.jcim.5b00262
J. Chem. Inf. Model. 2015, 55, 2288−2296

calculated and the experimental values shows a poor agreement, as seen on Figure 2 (blue lines). According to the previous study, the error in such calculations arises mainly from inaccurate differential solvation free energies of the acid and its conjugate base.[20] However, the error is systematic and so the training set of acids can be employed to calibrate a correction function. Linear fitting of the calculated values to the experimental ones was performed to obtain the parameters of the correction function (green lines on Figure 2). The fitting was done by using the least-squares method, as implemented in NumPy.[46] In the second step, the straightforward $pK_a^{aq}$ calculation was performed for each protonation form of a free biliverdin. The obtained initial $pK_a^{aq}$ values were then corrected by using the function obtained from our training set of acids.

## RESULTS AND DISCUSSIONS

The *Pcetk* toolkit is provided with a number of test and benchmark cases. Two of these examples will be discussed in the following section. For the preparation of the protein models, see section S3 in the Supporting Information.

**Predicting Protonation States in Lysozyme.** Lysozyme is a well-studied protein, which is frequently used as a system for the testing and benchmarking of electrostatic calculations. The protein contains only standard protonatable residues. Section S4 in the Supporting Information provides a detailed description of the Python script used to perform the calculations on lysozyme.

The calculations were done on two alternative structures of lysozyme.[47,48] For most protonatable groups, the accuracy of our calculations for the prediction of $pK_{1/2}$ values is similar to that of previous studies (see Table 1). The discrepancies that result are due to the slightly different parameters used during the calculations, such as, for example, the grid resolutions, the atomic charges and radii, the number of MC steps taken and the criteria for double moves during the MC sampling. Knowledge of the protonation states of Glu35 and Asp52 is particularly important, because they make up the enzymatic active site. For Glu35, the $pK_{1/2}$ values were calculated to be around 5.5, regardless of the protein structure used, which is in a good agreement with the experimentally determined $pK_a$ of 6.1. The $pK_{1/2}$ values of Asp52 in the original structure of lysozyme[47] (resolution 2.50 Å) was calculated to be 6.3, whereas the experimental $pK_a$ value for this group is known to be 3.6. In earlier works, the overestimation of the $pK_{1/2}$ value of Asp52 was even more pronounced. However, the Asp52 side-chain shows a slightly different conformation in the updated structure of lysozyme[48] (resolution 1.97 Å). Using the updated structure, we calculated the $pK_{1/2}$ of Asp52 to be 2.8, which is considerably closer to the experimentally observed value. A noticeable deviation from the experiment still remains for the calculated $pK_{1/2}$ of the N-terminus. However, this deviation can be explained by the high flexibility of protein termini that is commonly observed, which complicates the $pK_a$ calculations based on static protein structures. Overall, the results generated by the toolkit show a reasonable agreement with the experimental $pK_a$ values and the previously calculated $pK_{1/2}$ values.

**Predicting Protonation States in the IFP2.0 Fluorescent Protein.** Plants and microbial organisms use proteins, such as the phytochromes, to detect light. Phytochromes are interesting because they can be genetically engineered to gain fluorescent properties and so become infrared fluorescent proteins[49,50] (IFPs). Central to the phytochrome is a bilin chromophore, for example biliverdin, which has four protonatable positions (Figure 1). The knowledge of the $pK_a$ values of these positions is key to understanding the photochemistry of IFPs. However, there seems to be a lack of consent in the literature as to the protonation behavior of protein-bound biliverdin.[51,52] In addition, to the best of our knowledge, no electrostatic calculations have been done to elucidate the protonation states of this chromophore.

**Table 1. Comparison of Experimental[58−60] $pK_a$ and Calculated $pK_{1/2}$ Values for Different Protonatable Groups in Lysozyme from Chicken[a]**

| site | $pK_a^{exper}$ | $pK_{1/2}^{a}$ | $pK_{1/2}^{b}$ | $pK_{1/2}^{c}$ | $pK_{1/2}^{d}$ |
|---|---|---|---|---|---|
| N-terminus | 7.9 | 6.4 | 5.1 | 11.7 | 11.2 |
| His 15 | 5.8 | 4.0 | 2.4 | 4.7 | 4.5 |
| Glu 7 | 2.6 | 2.1 | 3.2 | 2.9 | 3.2 |
| Glu 35 | 6.1 | 6.3 | 5.7 | 5.5 | 5.4 |
| Asp 18 | 2.9 | 3.1 | 1.6 | 1.7 | 1.9 |
| Asp 48 | 4.3 | 1.0 | 2.5 | <1 | <1 |
| Asp 52 | 3.6 | 7.0 | 7.4 | 6.3 | 2.8 |
| Asp 66 | 2.0 | 1.7 | 1.5 | 0.3 | 0.3 |
| Asp 87 | 3.6 | 1.2 | 1.9 | <1 | 1.1 |
| Asp 101 | 4.1 | 7.9 | 4.3 | 4.0 | 4.2 |
| Asp 119 | 2.5 | 3.2 | 3.6 | 3.7 | 4.0 |
| Tyr 20 | 10.3 | 14.0 | 12.7 | 10.4 | 13.6 |
| Tyr 23 | 9.8 | 11.7 | 9.5 | 11.1 | 10.9 |
| Tyr 53 | 12.1 | 20.8 | >16 | >16 | >16 |
| Lys 1 | 10.8 | 9.6 | 11.2 | 9.7 | 8.7 |
| Lys 13 | 10.5 | 11.6 | 12.9 | 10.4 | 10.6 |
| Lys 33 | 10.6 | 9.6 | 10.0 | 8.3 | 9.4 |
| Lys 96 | 10.8 | 10.4 | 10.7 | 8.1 | 9.4 |
| Lys 97 | 10.3 | 10.6 | 10.9 | 10.4 | 10.8 |
| Lys 116 | 10.4 | 9.9 | 10.3 | 8.5 | 9.2 |
| C-terminus | 2.8 | 2.3 | 2.7 | 2.0 | 2.3 |

[a]A $pK_{1/2}$ is defined as the pH at which the probability of protonation of a site equals 1/2. Columns denoted by a and b represent earlier calculations performed by Bashford[59] and Miteva,[61] respectively, using the original structure of lysozyme[47] (PDB code 7LYZ). The last two columns contain the results of our calculations done on the original (c) and updated (d) structures of lysozyme[48] (PDB code 2LZT).

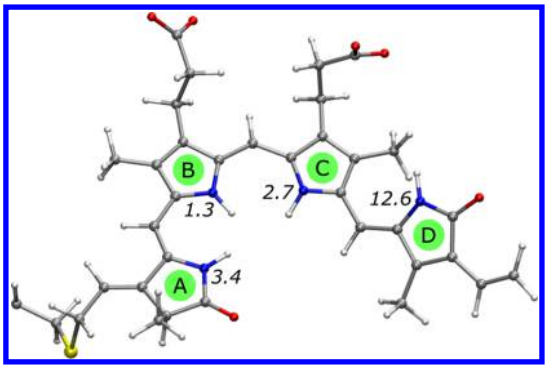



**Figure 1.** Biliverdin chromophore in its fully protonated form as seen in the crystal structure of IFP2.0.[50] The four pyrrole rings are highlighted in green and labeled by capital letters. The small italic numbers indicate the calculated aqueous solution $pK_a$ values, $pK_a^{aq}$ values, of each of the rings. The $pK_a^{aq}$ values were calculated by using a modified version of the method of Muckerman and co-workers,[20] as described in the text. The considerable acidity of rings A, B, and C is probably due to the repulsion between the inner pyrrole protons.



DOI: 10.1021/acs.jcim.5b00262
J. Chem. Inf. Model. 2015, 55, 2288−2296

Table 2. Cationic Acids with Experimentally Known $pK_a^{aq}$ Values Used as a Training Set for the Correction Function[a]

| | | (a) with MP2/CBS correction | | | (b) without MP2/CBS correction | | |
|---|---|---|---|---|---|---|---|
| | $pK_{a,exp}$ | $pK_{a,calc}$ | $pK_{a,fit}$ | $\Delta pK_{a,fit-exp}$ | $pK_{a,calc}$ | $pK_{a,fit}$ | $\Delta pK_{a,fit-exp}$ |
| 2,5-dichloro-anilinium | 1.5 | −8.8 | 1.5 | −0.0 | −8.5 | 1.3 | −0.2 |
| 4-cyano-anilinium | 1.6 | −8.0 | 2.0 | 0.4 | −8.4 | 1.4 | −0.2 |
| 4-bromo-anilinium | 3.9 | −5.7 | 3.5 | −0.4 | −5.2 | 3.3 | −0.6 |
| anilinium | 4.6 | −4.2 | 4.3 | −0.3 | −3.8 | 4.1 | −0.5 |
| p-anisidinium | 5.4 | −2.9 | 5.2 | −0.2 | −1.3 | 5.6 | 0.3 |
| 2,6-dimethyl-pyridinium | 6.7 | −0.1 | 7.0 | 0.3 | 1.9 | 7.5 | 0.8 |
| 2,4,6-collidinium | 7.3 | 0.6 | 7.4 | 0.0 | 3.6 | 8.5 | 1.2 |
| benzylammonium | 9.3 | 3.6 | 9.3 | −0.0 | 4.7 | 9.2 | −0.1 |
| DMAP | 9.6 | 4.3 | 9.7 | 0.1 | 6.8 | 10.4 | 0.8 |
| lysine | 10.4 | 5.3 | 10.3 | −0.1 | 6.1 | 10.1 | −0.3 |
| triethylammonium | 10.7 | 6.6 | 11.1 | 0.4 | 6.5 | 10.2 | −0.5 |
| pyrrolidinium | 11.3 | 6.8 | 11.3 | −0.0 | 7.4 | 10.8 | −0.4 |
| guanidinium | 13.6 | 10.3 | 13.4 | −0.2 | 11.6 | 13.3 | −0.3 |

[a]Two sets of results are presented depending on how the electronic energy contribution to the reaction free energy was calculated. In set "a", the $E_{el}$ values were calculated at the MP2/CBS level of theory, whereas in set "b" the same level of theory was used as for the other calculations, namely BP/TZV(2d). The standard deviations of the obtained $pK_a^{aq}$ values were calculated to be 0.24 and 0.55 for sets "a" and "b", respectively. To train the correction function, set "a" was selected. The final correction function based on set "a" takes the following form: $pK_{a,fit} = 0.62\ pK_{a,calc} + 6.99$.

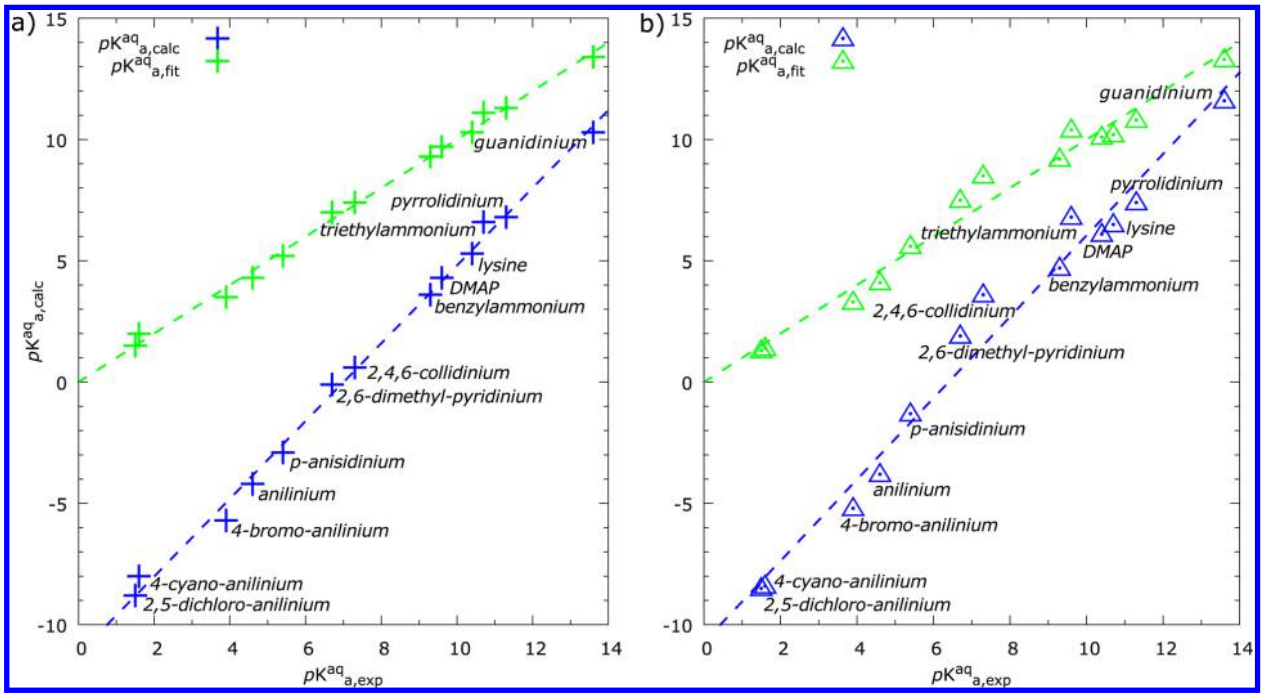



**Figure 2.** Calculated vs experimental $pK_a^{aq}$ values for the training set of 13 cationic acids (Table 2). Blue symbols indicate $pK_a^{aq}$ values calculated by using the method described in the Supporting Information. Green symbols indicate $pK_a^{aq}$ values after fitting to the experimental values. The green dashed lines indicate the exact agreement between the calculation and experiment. Two cases are considered depending on how the electronic energy contribution to the free energy of deprotonation was obtained. In case "a", $E_{el}$ was calculated at the MP2/CBS level of theory. In case "b", $E_{el}$ was obtained at the BP/TZV(2d) level, i.e., the same as for the calculation of geometries, frequencies and solvation energies. The standard deviations were calculated to be 0.24 and 0.55 for cases "a" and "b", respectively, which underlines the importance of an accurate electronic energy component in the calculation of $pK_a^{aq}$. Interestingly, without the MP2/CBS energy correction ("b"), the most pronounced discrepancies between the experimental and calculated $pK_a^{aq}$ values occur in the physiological range of pH, whereas the agreement is somewhat better for the more extreme pH values.

*Calculating $pK_a^{aq}$ Values of the Biliverdin Chromophore.* The aqueous solution $pK_a$ values, $pK_a^{aq}$, of biliverdin were estimated by using the modified method of Muckerman, as described in the Methods and Implementation section. The atomic coordinates of the biliverdin chromophore were extracted from the crystal structure of IFP2.0[50] (PDB code 4CQH). Each pyrrole ring of biliverdin can bind or release a proton, which leads to 16 possible protonation forms of the chromophore. However, we assumed that only one proton at a time can dissociate from biliverdin, because the multiply deprotonated forms would be energetically too unfavorable under physiological conditions. This assumption gives the A-, B-, C-, and D-deprotonated forms, where each letter indicates a pyrrole ring, and the fully protonated form of biliverdin. The four deprotonated forms have a net charge of zero, whereas the fully protonated form has a charge of +1.


DOI: 10.1021/acs.jcim.5b00262
J. Chem. Inf. Model. 2015, 55, 2288−2296

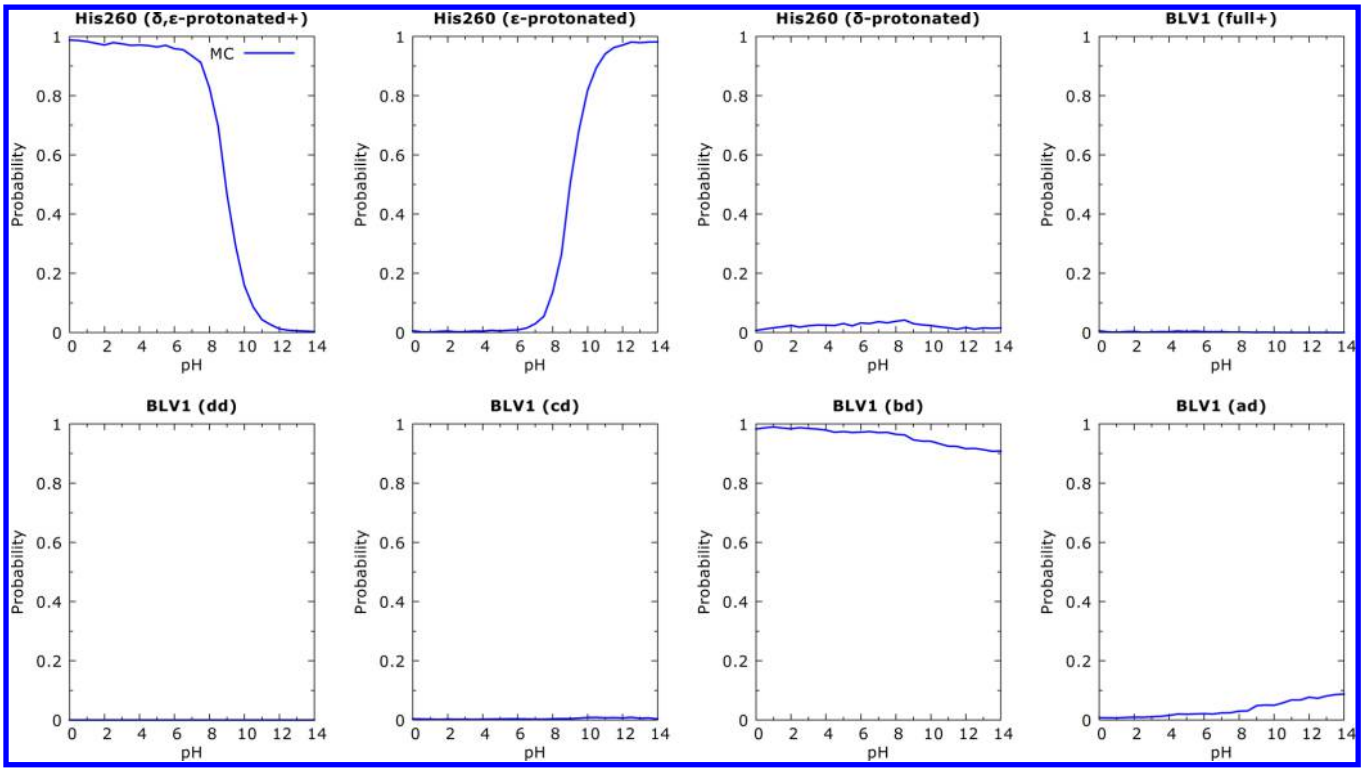



**Figure 3.** Titration curves of different protonation forms of the biliverdin chromophore (BLV1) and His260 in IFP2.0. The curves were calculated at a resolution of 0.5 pH-unit by using the in-house Monte Carlo routines of *Pcetk*. Labels *ad*, *bd*, *cd*, and *dd* indicate the chromophore deprotonated at pyrrole ring A, B, C, and D, respectively, whereas *full*(+) indicates the fully protonated chromophore. For clarity, the fully deprotonated form of His260 is omitted, since its probability of occurrence was calculated to be always zero. The plots representing the two acetate groups of biliverdin are also not shown, because these groups were found to be always deprotonated. The calculations indicate that in IFP2.0 the chromophore is predominantly present in its B-deprotonated form. His260 is fully protonated at neutral pH, but at basic pH, the $\delta$-proton of His260 is transferred to the solvent, rather than to ring B of the chromophore.

On the basis of the training set of 13 cationic acids (Table 2a and Figure 2a), the gradient and intercept parameters of the linear correction function were calculated to be 0.62 and 6.99, respectively. We estimate the final, corrected $pK_a^{aq}$ values of the four pyrrole rings of biliverdin to be (values in parentheses indicate the initial, uncorrected $pK_a^{aq}$ values): 3.4 (−5.7), 1.3 (−9.2), 2.7 (−6.9), and 12.6 (8.9) for rings A, B, C, and D, respectively. The lowest $pK_a^{aq}$ value of 1.3 was found for the B pyrrole ring, and so ring B has the highest probability of being deprotonated in aqueous solution. The calculated $pK_a^{aq}$ values indicate that rings A, B, and C are all considerably acidic, which can be explained based on the conformation adopted by biliverdin in the protein. The repulsion between the strongly interacting inner protons of the first three pyrrole rings can destabilize the fully protonated form of biliverdin. The $pK_a^{aq}$ value of ring D is significantly higher, because the proton of this ring is not in close contact with the other protons.

For the same training set of cationic acids, the replacement of the MP2/CBS electronic energy component by the component calculated at the lower level of theory was found to worsen significantly agreement with experiment (Table 2b and Figure 2b). Thus, the standard deviation of the predicted versus experimental $pK_a^{aq}$ values was calculated to be 0.24 pH-units with the MP2/CBS correction, included and 0.55 without. This observation highlights the importance of an accurate electronic energy component in $pK_a^{aq}$ calculations.[19] Additionally, we considered geometries, frequencies, and solvation energies calculated by using the B3LYP density functional and the 6-311G+(d,p) 5d basis set. The use of a different density functional and basis set was found to somewhat affect the predicted $pK_a^{aq}$ values, since now the standard deviation was calculated to be 0.36 (Table S1a), or 0.59 without the accurate electronic energy component (Table S1b). Overall, the best agreement with experiment for the calculated $pK_a^{aq}$ values was achieved by combining the BP/TZV(2d) (geometries, frequencies, solvation energies) and MP2/CBS (electronic energies) levels of theory.

*Calculating Protonation States of Biliverdin in the Protein.* The protonation states of the IFP2.0 protein, including the protein-bound biliverdin chromophore, were calculated by using the script described in section S5 in the Supporting Information. In the lowest energy protonation state, the biliverdin chromophore is deprotonated at pyrrole ring B, both acetate groups at rings B and C are deprotonated, and His260 is doubly protonated (Table S2). The state representing the situation where pyrrole ring B has been protonated from His260 (state 5) is 3.4 kcal/mol higher in energy than the most stable state. From our calculations, we conclude that the protein environment of IFP2.0 does not stabilize the fully protonated form of the biliverdin chromophore, which remains B-deprotonated regardless of the current pH, as seen in Figure 3.

## ■ CONCLUSIONS

In this paper, we have described *Pcetk*, a software toolkit for protonation state calculations in proteins. The toolkit is designed as a module of the pDynamo software library and uses the external Poisson−Boltzmann solver, MEAD, for the


DOI: 10.1021/acs.jcim.5b00262
J. Chem. Inf. Model. 2015, 55, 2288−2296

calculation of electrostatic energy terms. The key features of the toolkit include the flexible environment based on the Python scripting language, in-house Monte Carlo routines for efficient sampling of protonation state energies, and automation of frequently used procedures. These features allow for writing very compact, easy to understand and robust scripts to perform and analyze electrostatic calculations. Moreover, *Pcetk* works well in a Python interactive shell, such as IPython,[53] which allows for quick testing or writing "one-liner" scripts.

At the present stage of development, the toolkit already provides a wide range of functionalities. Nonetheless, the used programming model and object-oriented design allow for easy extensions of the toolkit. Future developments may include, for example, the inclusion of a custom, GPU-accelerated[54] solver of the Poisson−Boltzmann equation to replace MEAD. Integration of *Pcetk* is planned with packages such as matplotlib[55] to automatize the plotting of titration curves or GTKDynamo,[56] the recent graphical front-end for pDynamo, to enable visualization options. Some extensions of the Monte Carlo routines are already underway, such as the implementation of triple moves to improve further the quality of sampling. Finally, the toolkit may enable the handling of new types of simulation in pDynamo, for example constant pH simulations.[57] Because the source code of *Pcetk* is stored in a version-controlled repository, other developers can easily make contributions in the form of source code patches or alternative branches. *Pcetk* is freely available at the Web address http://github.com/mfx9/pcetk. As of this writing, the current version of the toolkit is 1.0. An up-to-date user manual, installation instructions, and a set of test cases are included in the repository. The toolkit is distributed under the CeCILL license, which is a French equivalent of the GNU General Public License.

We have also revisited an earlier method[20] for the calculation of aqueous solution p$K_a$ values of unusual protonatable groups, such as the biliverdin chromophore. This method, which relies on a training set of chemically similar compounds with known p$K_a^{aq}$ values, was demonstrated to predict p$K_a^{aq}$ values with a reasonable accuracy of ±0.5 p$K_a$-units.[20] On the basis of our training set, we note that the quality of the p$K_a^{aq}$ prediction can be further improved by accurate calculation of the electronic energy contributions, using the MP2/CBS level of theory. Electrostatic calculations performed for biliverdin bound to IFP2.0 show that the chromophore is deprotonated at pyrrole ring B for the whole spectrum of pH. We propose that the fully protonated form of biliverdin is destabilized by the repulsion interactions of the inner pyrrole protons. The protein environment of IFP2.0 does not seem to stabilize the fully protonated form of biliverdin.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.5b00262.

Organization of the *Pcetk* toolkit, description of the calculation of the p$K_a^{aq}$ values of biliverdin and cationic acids, details of the setup of the protein model test cases, and summaries of the lysozyme and IFP2.0 *Pcetk* scripts and results. A table summarizing the p$K_a^{aq}$ calculations performed with the B3LYP density functional, and EST files containing the biliverdin parameters used during the electostatic calculations in IFP2.0 (PDF).

## ■ AUTHOR INFORMATION


### Corresponding Author

*M. Field. E-mail: e-mail: martin.field@ibs.fr.


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

M.F. and M.J.F. acknowledge financial support from the French National Research Agency (grant number ANR-11-BSV5-0012).

## ■ REFERENCES


(1) Ullmann, G. M.; Knapp, E. W. *Eur. Biophys. J.* **1999**, *28*, 533−551.
(2) Nielsen, J. E.; McCammon, J. A. *Protein Sci.* **2003**, *12*, 1894−1901.
(3) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *Biochemistry* **1996**, *35*, 7819−7833.
(4) Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 9556−9561.
(5) Sondergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7*, 2284−2295.
(6) Olsson, M. H. M.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7*, 525−537.
(7) Field, M. J. *J. Chem. Theory Comput.* **2008**, *4*, 1151−1161.
(8) Bashford, D. *Lect. Notes in Computer Science* **1997**, *1343*, 233−240.
(9) Essigke, T. *A Continuum Electrostatic Approach for Calculating the Binding Energetics of Multiple Ligands.* Ph.D. thesis, University of Bayreuth, Bayreuth, Germany, 2008. https://epub.uni-bayreuth.de/655/.
(10) Oliphant, T. E. *Comput. Sci. Eng.* **2007**, *9*, 10−20.
(11) Millman, K. J.; Aivazis, M. *Comput. Sci. Eng.* **2011**, *13*, 9−12.
(12) Guha, R.; Howard, M. T.; Hutchison, G. R.; Murray-Rust, P.; Rzepa, H.; Steinbeck, C.; Wegner, J.; Willighagen, E. L. *J. Chem. Inf. Model.* **2006**, *46*, 991−998.
(13) O'Boyle, N. M.; Guha, R.; Willighagen, E. L.; Adams, S. E.; Alvarsson, J.; Bradley, J. C.; Filippov, I. V.; Hanson, R. M.; Hanwell, M. D.; Hutchison, G. R.; James, C. A.; Jeliazkova, N.; Lang, A. S.; Langner, K. M.; Lonie, D. C.; Lowe, D. M.; Pansanel, J.; Pavlov, D.; Spjuth, O.; Steinbeck, C.; Tenderholt, A. L.; Theisen, K. J.; Murray-Rust, P. *J. Cheminf.* **2011**, *3*, 37.
(14) Gezelter, J. *J. Phys. Chem. Lett.* **2015**, *6*, 1168−1169.
(15) Hinsen, K. *J. Comput. Chem.* **2000**, *21*, 79−85.
(16) O'Boyle, N. M.; Tenderholt, A. L.; Langner, K. M. *J. Comput. Chem.* **2008**, *29*, 839−845.
(17) Schrödinger, LLC. *The PyMOL Molecular Graphics System*, Version 1.3r1; Schrödinger, LLC: New York, 2010.
(18) Cock, P. J. A.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; de Hoon, M. J. L. *Bioinformatics* **2009**, *25*, 1422.
(19) Liptak, M. D.; Shields, G. C. *J. Am. Chem. Soc.* **2001**, *123*, 7314−7319.
(20) Muckerman, J. T.; Skone, J. H.; Ning, M.; Wasada-Tsutsui, Y. *Biochim. Biophys. Acta, Bioenerg.* **2013**, *1827*, 882−891.
(21) Uddin, N.; Choi, T. H.; Choi, C. H. *J. Phys. Chem. B* **2013**, *117*, 6269−6275.
(22) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144−1149.
(23) Gunner, M. R.; Alexov, E. *Biochim. Biophys. Acta, Bioenerg.* **2000**, *1458*, 63−87.
(24) Bombarda, E.; Ullmann, G. M. *J. Phys. Chem. B* **2010**, *114*, 1994−2003.
(25) Bombarda, E.; Ullmann, G. M. *Faraday Discuss.* **2011**, *148*, 173−193.
(26) Ullmann, G. M.; Bombarda, E. *Biol. Chem.* **2013**, *394*, 611−619.
(27) Honig, B.; Sharp, K. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301−332.




DOI: 10.1021/acs.jcim.5b00262
*J. Chem. Inf. Model.* 2015, 55, 2288−2296

(28) Warshel, A.; Papazyan, A. *Curr. Opin. Struct. Biol.* **1998**, *8*, 211−217.
(29) Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. *Biochim. Biophys. Acta, Proteins Proteomics* **2006**, *1764*, 1647−1676.
(30) Ullmann, G. M.; Bombarda, E. In *Protein Modelling*; Náray-Szabó, G., Ed.; Springer: Heidelberg, Germany, 2014.
(31) Alexov, E. G.; Gunner, M. R. *Biophys. J.* **1997**, *72*, 2075−2093.
(32) Georgescu, R. E.; Alexov, E. G.; Gunner, M. R. *Biophys. J.* **2002**, *83*, 1731−1748.
(33) Bashford, D.; Gerwert, K. *J. Mol. Biol.* **1992**, *224*, 473−486.
(34) Beroza, P.; Case, D. *Methods Enzymol.* **1998**, *295*, 170−189.
(35) You, T.; Bashford, D. *Biophys. J.* **1995**, *69*, 1721−1733.
(36) Neese, F. *Wiley Interdiscip. Rev.-Comput. Mol. Sci.* **2012**, *2*, 73−78.
(37) Beroza, P.; Fredkin, D. R.; Okamura, M. Y.; Feher, G. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 5804−5808.
(38) Ullmann, R. T.; Ullmann, G. M. *J. Comput. Chem.* **2012**, *33*, 887−900.
(39) Feliks, M.; Ullmann, G. M. *J. Phys. Chem. B* **2012**, *116*, 7076−7087.
(40) Feliks, M.; Martins, B. M.; Ullmann, G. M. *J. Am. Chem. Soc.* **2013**, *135*, 14574−14585.
(41) Matsumoto, M.; Nishimura, T. *ACM Trans. Model. Comput. Simul.* **1998**, *8*, 3−30.
(42) Ho, J.; Coote, M. L. *Theor. Chem. Acc.* **2010**, *125*, 3−21.
(43) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669−681.
(44) Klamt, A.; Schuurmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799−805.
(45) Cramer, C. J.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 760−768.
(46) van der Walt, S.; Colbert, S. C.; Varoquaux, G. *Comput. Sci. Eng.* **2011**, *13*, 22−30.
(47) Herzberg, O.; Sussman, J. *J. Appl. Crystallogr.* **1983**, *16*, 144−150.
(48) Ramanadham, M.; Sieker, L.; Jensen, L. *Acta Crystallogr., Sect. B: Struct. Sci.* **1990**, *46*, 63−69.
(49) Shu, X.; Royant, A.; Lin, M. Z.; Aguilera, T. A.; Lev-Ram, V.; Steinbach, P. A.; Tsien, R. Y. *Science* **2009**, *324*, 804−807.
(50) Yu, D.; Gustafson, W.; Han, C.; Lafaye, C.; Noirclerc-Savoye, M.; Ge, W.; Thayer, D.; Huang, H.; Kornberg, T.; Royant, A.; Jan, L.; Jan, Y.; Weiss, W.; Shu, X. *Nat. Commun.* **2014**, *5*, 3626−3634.
(51) Borucki, B.; von Stetten, D.; Seibeck, S.; Lamparter, T.; Michael, N.; Mroginski, M. A.; Otto, H.; Murgida, D. H.; Heyn, M. P.; Hildebrandt, P. *J. Biol. Chem.* **2005**, *280*, 34358−34364.
(52) Li, F.; Burgie, E. S.; Yu, T.; Heroux, A.; Schatz, G. C.; Vierstra, R. D.; Orville, A. M. *J. Am. Chem. Soc.* **2015**, *137*, 2792−2795.
(53) Pérez, F.; Granger, B. E. *Comput. Sci. Eng.* **2007**, *9*, 21−29.
(54) Geng, W.; Jacob, F. *Comput. Phys. Commun.* **2013**, *184*, 1490−1496.
(55) Hunter, J. D. *Comput. Sci. Eng.* **2007**, *9*, 90−95.
(56) Bachega, J. F.; Timmers, L. F.; Assirati, L.; Bachega, L. R.; Field, M. J.; Wymore, T. *J. Comput. Chem.* **2013**, *34*, 2190−2196.
(57) Donnini, S.; Tegeler, F.; Groenhof, G.; Grubmuller, H. *J. Chem. Theory Comput.* **2011**, *7*, 1962−1978.
(58) Kuramitsu, S.; Hamaguchi, K. *J. Biochem.* **1979**, *87*, 1215−1219.
(59) Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 10219−10225.
(60) Bartik, K.; Redfield, C.; Dobson, C. *Biophys. J.* **1994**, *66*, 1180−1184.
(61) Miteva, M.; Tuffery, P.; Villoutreix, B. *Nucleic Acids Res.* **2005**, *33*, W372−W375.


DOI: 10.1021/acs.jcim.5b00262
*J. Chem. Inf. Model.* 2015, 55, 2288−2296