JCTC Journal of Chemical Theory and Computation

ARTICLE

pubs.acs.org/JCTC

Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pK_a Values

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

Department of Chemistry and Center for Computational Molecular Sciences, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark

Supporting Information

ABSTRACT: The new empirical rules for protein pK_a predictions implemented in the PROPKA3.0 software package (Olsson et al. *J. Chem. Theory Comput.* **2010**, *7*, 525−537) have been extended to the prediction of pK_a shifts of active site residues and ionizable ligand groups in protein−ligand complexes. We present new algorithms that allow pK_a shifts due to inductive (i.e., covalently coupled) intraligand interactions, as well as noncovalently coupled interligand interactions in multiligand complexes, to be included in the prediction. The number of different ligand chemical groups that are automatically recognized has been increased to 18, and the general implementation has been changed so that new functional groups can be added easily by the user, aided by a new and more general protonation scheme. Except for a few cases, the new algorithms in PROPKA3.1 are found to yield results similar to or better than those obtained with PROPKA2.0 (Bas et al. *Proteins: Struct., Funct., Bioinf.* **2008**, *73*, 765−783). Finally, we present a novel algorithm that identifies noncovalently coupled ionizable groups, where pK_a prediction may be especially difficult. This is a general improvement to PROPKA and is applied to proteins with and without ligands.

1. INTRODUCTION

The interactions between biological molecules are fundamental to all biological phenomena. The often high degree of specificity of the intermolecular interactions offers a high level of control in basic biological processes such as the humoral immune response (antibody−antigen interactions), gene regulation (protein−DNA interactions), and enzyme catalysis (initial enzyme−substrate binding). Similarly, most drugs are small molecules that bind to protein targets. In the early steps of the drug discovery process, thousands of candidate molecules are often screened for activity on the protein target in an expensive and time-consuming process. Accurate modeling of protein−ligand interactions is therefore of great importance and has practical applications in the pharmaceutical industry where virtual screening for drug candidates is of increasing importance.[1]

A great number of computer algorithms have been developed for modeling interactions between ligands and protein molecules.[2] However, much less effort has been devoted to the prediction of titrational events during protein−ligand binding.[3−5] When a ligand binds to a protein receptor, the chemical environment of the ligand and the active site residues is changed. This can result in titrational events on the ligand molecule[6] and in the active site, which influences the binding affinity of the ligand molecule to the protein receptor. Hence, the prediction of ligand pK_a values can be of great importance for the accurate modeling of these events.

In this Article, we announce the new version 3.1 of PROPKA,[5,7,8] which features a number of improvements with particular focus on prediction of ligand pK_a values. The previous version 2.0 of PROPKA models interactions between protein and ligand chemical groups. However, PROPKA2.0 has a number of shortcomings that we will seek remedy in the present PROPKA3.1.

While the previous version 2.0 of PROPKA supported modeling of interactions between protein and ligand molecules, no modeling of interactions between ligand groups was available. This shortcoming can lead to errors in predicted pK_a values for systems with multiple ligand molecules complexed with a protein, systems containing large ligand molecules including several titratable groups, or systems containing ligand molecules that are coordinated with monatomic ions.

Additionally, PROPKA3.1 takes inductive interactions between covalently coupled ligand groups into account. For two covalently coupled groups in an aromatic ring, it can be the case that a titrational event at one of the covalently coupled groups renders a titrational event at the other group much less likely. As the molecular environment of the two groups is often similar, it is often the chemical environment of the groups that determines which group will titrate. PROPKA3.1 models covalently coupled titrational events by identifying groups that are not titrating due to coupling with another titrational event and restricting these groups to the neutral state.

In PROPKA3.1, we have furthermore implemented algorithms for the modeling of noncovalently coupled titrational events. Two titratable groups, residing in a protein or ligand, that are spatially proximate can influence the titration of each other such that the two groups titrate in a coupled fashion and that the shape of the titration curves can deviate from the standard Henderson−Hasselbalch shape.[9,10] Such noncovalently coupled titration curves have, for example, been observed for *Bacillus circulans* xylanase[11] and thioredoxins.[12−14] These noncovalently coupled titration events can be considered a combination of two microscopic titrational events where either of the two groups can titrate first. PROPKA3.1 models noncovalently coupled titrational events by predicting the resulting pK_a values in both of the two possible titration orders.

Received: February 23, 2011
Published: May 27, 2011

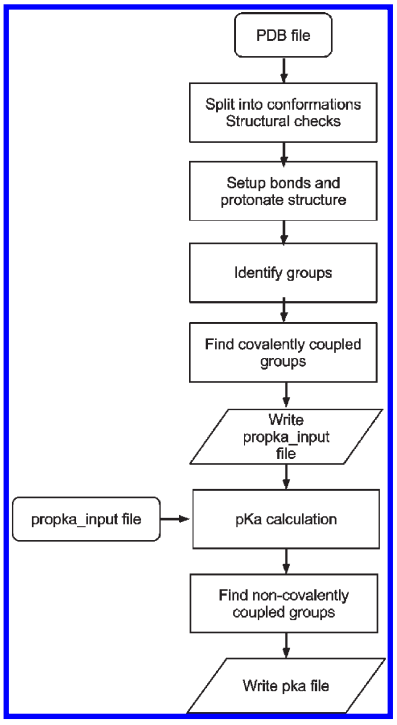


Figure 1. Flowchart of the calculation procedure in PROPKA3.1.

Finally, the implementation of the PROPKA versions 3.0 and 3.1 in Python[15] makes it a relatively easy task to modify and extend the code to include new chemical groups.

2. MATERIALS AND METHODS

2.1. Calculation Workflow. The workflow of PROPKA3.1 is sketched in Figure 1. When initialized with a .pdb file as argument, the workflow of PROPKA3.1 is as follows: (i) The structure PDB file is read into PROPKA3.1, taking into account multiple models and alternative locations, and a series of checks is done to ensure structural consistency, for example, that known protein residues contain the expected number of atoms. The assignment of Sybyl atom types to ligand atoms is done internally in PROPKA3.1 based solely on atom elements and coordinates. For a list of recognized Sybyl types, see Table 1. (ii) Chemical bonds are initialized on the basis of atom coordinates and element types. Atoms separated by less than 2.0 Å are generally assumed to be bonded. However, if one of the two atoms is a hydrogen atom, the threshold is 1.5 Å, and for disulfides the threshold is 2.5 Å. Information on π-electrons and conjugated bonds is set up using table values for protein atom names and ligand Sybyl types. Protons are added as described below. (iii) The chemical groups needed for the pK_a calculation are identified from the atoms that have been read in. For protein atoms, the identification of chemical groups is based on atom names, and for heterogeneous atoms, the identification is based on predicted Sybyl types. Besides the ligand groups recognized by PROPKA2.0,[5] two additional groups are recognized in PROPKA3.1: titratable oxygen atoms in phosphate groups (default model pK_a value 6.0 was set on the basis of a series of experimental pK_a values for molecules containing phosphate groups[16]) and thiol groups (default model pK_a value 10.0[17]); confer Table 2. (iv) Covalently coupled groups are identified as described below. (v) A .propka_input file is written out, allowing the user to modify the standard PROPKA3.1 setup. (vi) The pK_a value calculation is done using the algorithm described in ref 7 expanded to also include ligand atoms as described below. (vii) Noncovalently coupled groups are identified as described below. (viii) A resulting .pka file is written out.

Table 1. Recognized Sybyl Types

Sybyl name	atom
C.3	sp^3 carbon
C.2	sp^2 carbon
C.1	sp carbon
C.ar	aromatic carbon
N.3	sp^3 nitrogen
N.1	sp nitrogen
N.ar	aromatic nitrogen
N.am	nitrogen in amide
N.pl3	trigonal planar nitrogen
O.3	sp^3 oxygen
O.2	sp^2 oxygen
O.co2	oxygen in carboxylate
S.o2	sulfur in sulfone
S.3	sp^3 sulfur
P.3	sp^3 phosphorus

2.2. pK_a Shifts Due to Ligand Interactions. In the following, a brief summary of the underlying algorithms in PROPKA3.1 is given with focus on the extension of the parameter set to include ligand groups. For more details on the theoretical considerations that PROPKA3.1 is based on, please consult ref 7.

In PROPKA3.1, pK_a values are modeled as:

$$\mathrm{p}K_a = \mathrm{p}K_a^{\mathrm{water}} + \Delta\mathrm{p}K_a^{\mathrm{water}\rightarrow\mathrm{protein}} \quad (1)$$

where $\mathrm{p}K_a^{\mathrm{water}}$ is the pK_a value of the titratable group in water (often referred to as the model pK_a value), and $\Delta\mathrm{p}K_a^{\mathrm{water}\rightarrow\mathrm{protein}}$ is the contribution to the pK_a value that stems from the protein environment of the group. While the former value is a table value for protein residues, it will, for ligand groups, depend on the chemical properties of the molecule that the group is a part of. Hence, the default model pK_a values in Table 2 for ligand groups are only proximate and should be replaced if a more precise value is known for the considered ligand molecule. The latter value is predicted by PROPKA3.1 on the basis of the supplied structure and consists of

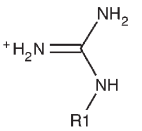
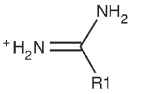
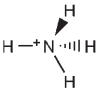
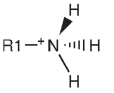
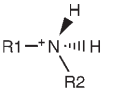
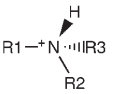
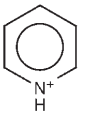
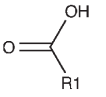
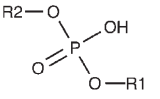
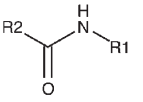
$$\Delta\mathrm{p}K_a^{\mathrm{water}\rightarrow\mathrm{protein}} = \Delta\mathrm{p}K_a^{\mathrm{desolv}} + \Delta\mathrm{p}K_a^{\mathrm{HB}} + \Delta\mathrm{p}K_a^{\mathrm{RE}} + \Delta\mathrm{p}K_a^{\mathrm{QQ}} \quad (2)$$

where $\Delta\mathrm{p}K_a^{\mathrm{desolv}}$ is the contribution due to desolvation effects, $\Delta\mathrm{p}K_a^{\mathrm{HB}}$ is the contribution due to hydrogen-bond interactions, $\Delta\mathrm{p}K_a^{\mathrm{RE}}$ is the contribution due to unfavorable electrostatic reorganization energies, and $\Delta\mathrm{p}K_a^{\mathrm{QQ}}$ is the contribution due to Coulombic interactions. Of these four terms only the hydrogen bond and Coulombic contributions are calculated using group-specific parameters. The remaining terms are calculated using generic parameters and do therefore not need to be extended with new parameters to allow modeling of ligand groups.

2.2.1. Intrinsic Electrostatic Interactions. Hydrogen-bond interaction energies are modeled as

$$\Delta\mathrm{p}K_a^{\mathrm{HB}} = \begin{cases} c^{\mathrm{HB}} w(r) \cos\theta & \text{if } \theta \geq 90^\circ \\ 0 & \text{else} \end{cases} \quad (3)$$

Table 2. Ligand Chemical Groups Recognized by PROPKA3.1 and Their Default Model pK_a Values

Chemical name	Group ID	Model pK_a value	Structure
Guanidinium	CG	11.50	
Amidinium	C2N	11.50	
Ammonium	N30	10.00	
sp^3 primary nitrogen	N31	10.00	
sp^3 secondary nitrogen	N32	10.00	
sp^3 tertiary nitrogen	N33	10.00	
Aromatic nitrogen	NAR	5.00	
Carboxyl	OCO	4.50	
Thiol	SH	10.00	
Phosphate	OP	6.00	
Amide	NAM	-	
Trigonal planar NH_2	NP1	-	
sp^3 oxygen	O3	-	
Chlorine	CL	-	
Flourine	F	-	
sp nitrogen	N3	-	
sp^2 oxygen	O2	-	
Hydroxyl	OH	-	

where c^{HB} is the maximal hydrogen-bond interaction energy (fitted to 0.85 pK_a units), r is the interatomic distance between the hydrogen-bond donor and acceptor, θ is the angle formed by the donor, the hydrogen atom, and the acceptor, and w is a

Table 3. Distance Parameters for Calculation on Intrinsic Electrostatic Interactions for Ligand Groups That Were Set On the Basis of Parameters of Equivalent Protein Groups[a]

chemical name	group ID	similar protein group
guanidinium	CG	arginine guanidinium
amidinium	C2N	arginine guanidinium
ammonium	N30	lysine/N-terminus nitrogen
sp^3 primary nitrogen	N31	lysine/N-terminus nitrogen
sp^3 secondary nitrogen	N32	lysine/N-terminus nitrogen
sp^3 tertiary nitrogen	N33	lysine/N-terminus nitrogen
aromatic nitrogen	NAR	histidine imidazole nitrogen
carboxyl	OCO	aspartic/glutamic acid carboxyl
phosphate	OP	
thiol	SH	cystine thiol
hydroxyl	OH	threonine/serine hydroxyl
trigonal planar NH_2	NP1	
sp^3 oxygen	03	
chlorine	CL	
flourine	F	
amide nitrogen	NAM	asparagine/glutamine amide nitrogen
sp nitrogen	Nl	
sp^2 oxygen	02	

[a] Table S.1 in the Supporting Information contains a list of all ligand intrinsic electrostatic interaction distance parameters. Interactions between ligand groups are allowed if the interacting ligand groups are separated by at least four chemical bonds.

distance-dependent weight function given by

$$w(r) = \begin{cases} 1 & \text{if } r \le r_{\min} \\ \dfrac{r_{\max} - r}{r_{\max} - r_{\min}} & \text{if } r_{\min} < r \le r_{\max} \\ 0 & \text{else} \end{cases} \quad (4)$$

where $r_{\min}$ and $r_{\max}$ are group-specific cutoff distances found by analysis of hydrogen-bond distances found in X-ray protein structures.

The generic maximal intrinsic interaction energy, c^{HB}, was set to 0.85 for all interactions involving ligand groups. However, to achieve consistency with PROPKA3.0, strong, buried interactions between histidine residues and ligand carboxylic groups are assigned an initial interaction energy, c^{HB}, of 1.6 pK_a units similarly to interactions between Asp/Glu and His residues.[5,7]

Because of the relatively limited amount of structures containing hydrogen-bond interactions between protein and specific ligand groups, cutoff distances for ligand groups have been set by extrapolating the existing parameters in PROPKA3.0 for protein groups. Interaction distance cutoffs are set to the values of similar protein residue groups, confer Table 3; for example, the distance parameters of guanidinium and amidinium groups are set to the values of the arginine residue. In those cases where no equivalent protein group exists, the default distance parameters of $c_{\min}$ = 3.0 Å and $c_{\max}$ = 4.0 Å were applied.

2.2.2. Coulombic Interactions. The contribution to the pK_a value of group i due to the Coulombic interaction between groups i and j is calculated as[7]

$$\Delta pK_{a,i}^{Q,Q} = \sigma_{ij} \frac{244}{\varepsilon r_{ij}} w(r_{ij}) \quad (5)$$

where 244 is the coefficient of Coulomb's law converted into pK_a units, $w(r_{ij})$ is a weight function, and r_{ij} is the distance between groups i and j. The dielectric constant, ε, is in this context a function of the degree of solvent exposure as explained in detail in ref 7. A unit step function, σ_{ij}, is determining the sign of the contribution to the pK_a value of group i. The unit step function for the Coulombic interaction between groups i and j is defined as

$$\sigma_{ij} = \begin{cases} -1 & \text{if } i \text{ is acid and } j \text{ is base or} \\ & \quad i \text{ and } j \text{ are bases and } pK_{a,i} < pK_{a,j} \\ +1 & \text{if } i \text{ is base and } j \text{ is acid or} \\ & \quad i \text{ and } j \text{ are acids and } pK_{a,i} > pK_{a,j} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

In some cases, the contribution of Coulombic interactions to the pK_a value depends on the total pK_a values of the interacting groups. An iterative approach is taken to resolve the total pK_a values in these cases. Table S.2 in the Supporting Information lists the interaction type (normal or iterative) for each possible interaction.

2.3. Protonation Scheme. A new generic protonation scheme has been implemented. The protonation scheme involves four steps. (i) Formal charges are set. For protein atoms, ARG-NH1, HIS-ND1, LYS-NZ, and the N-terminus are given a positive charge. For ASP-OD2, GLU-OE2, and the C-terminus, a negative charge is given. For heterogeneous atoms, charges are assigned on the basis of Sybyl atom types. Atoms assigned the Sybyl types N.pl3, N.3, N.4, and N.ar are given a positive charge, and one of the atoms in carboxyl groups (both oxygen atoms are assigned the Sybyl type O.co2) is given a negative charge. These charges are only used internally in the protonation algorithm and not in the pK_a calculation. (ii) The number of protons that should be added to each atom is set on the basis of the number of valence electrons, the number of bonds, the number of π-electrons in double and triple bonds, and the formal charge. (iii) The steric number and number of lone pairs are set for each atom. (iv) Protons are added on the basis of the assigned steric numbers using either trigonal or tetrahedral geometrical considerations. Only ligand molecules and protein groups included in the pK_a calculation are protonated by default.

2.4. Coupling Effects. Two types of coupling effects, covalent and noncovalent, are modeled in PROPKA3.1. The coupling effects are described in the following subsections and demonstrated using specific examples on structures including HIV protease, ATP, and a pyrazine derivative in the Results and Discussion.

2.4.1. Noncovalent Coupling. Titratable groups that are proximate in the protein 3D structure can titrate in a coupled fashion. This phenomenon can manifest itself in macroscopic titration curves that do not conform to the standard Henderson—Hasselbalch shape. Microscopically, the titration of two noncovalently coupled acidic groups can occur with either of the two groups titrating first and the other group titrating at an elevated pH value. PROPKA3.1 identifies noncovalently coupled groups and models their titrational behavior. For a pair of noncovalently coupled titratable groups, PROPKA3.1 models both orders of titration by assigning the shift in pK_a value due to the interaction between the groups to either of the two groups.

The criteria listed in Table 4 must all be fulfilled before two groups are treated as being noncovalently coupled in PROPKA3.1. The difference of the intrinsic pK_a values of the two groups, $\Delta pK_{a,max}^{int}$, must not be higher than 2.0 pK_a units to ensure

Table 4. Criteria for Detection of Noncovalently Coupled Groups

name	symbol	value [pH units]
max difference in intrinsic pK_a	$\Delta pK_{a,max}^{int}$	2.0
min interaction energy	E_{min}	0.5
max free energy difference	ΔG_{max}	1.0
min pK_a shift on swap	$\Delta pK_{a,min}$	1.0
min pK_a	$pK_{a,min}$	0.0
max pK_a	$pK_{a,max}$	10.0

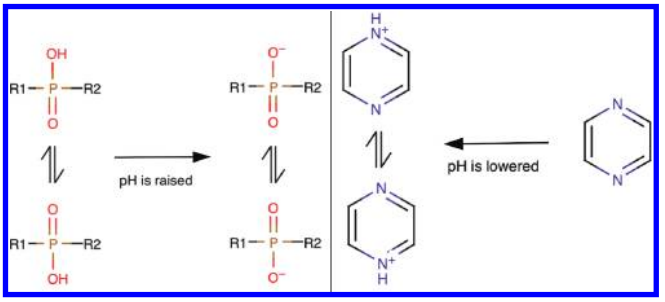


Figure 2. Examples of covalently coupled groups. Left panel: Phosphate group containing two oxygen atoms of which one can titrate and the other is double bonded to the phosphorus atom. Right panel: Pyrazine ring containing two nitrogen atoms for which titration of one nitrogen atom will significantly lower the pK_a value of the other nitrogen atom. Prepared using MarvinSketch.[19]

that the groups titrate at similar pH values. The intrinsic pK_a is defined as the theoretical pK_a value a titratable group would have in the protein environment if no electrostatic interactions with other titratable groups were present.[18] The interaction energy, E_{min}, between the two interacting groups must correspond to at least 0.5 pK_a units, and the difference in free energy between the two possible protonation states, ΔG_{max}, must not be greater than 1.0 pK_a units. The shift in pK_a value when swapping to the alternative protonation state, $\Delta pK_{a,min}$, must be greater than 1.0 pK_a unit. Only titratable groups with pK_a values between 0.0 and 10.0 are considered.

2.4.2. Covalent Coupling. We define the covalent coupling effect as the titration of one group that is significantly influencing the probability of titration of another group proximate in the chemical structure. Two examples of such systems of covalently coupled groups are given in Figure 2. It is the interactions with the environment that determines which of the groups in a covalently coupled system of groups that will titrate. We model covalent coupling effects with a simple scheme that hinders the titration of groups for which titration has been rendered unlikely by a covalently coupled group. The groups for which titration has been penalized are restricted to the neutral form.

Ligand titratable groups are as default assumed to be covalently coupled if they are separated by no more than three chemical bonds and assigned the same Sybyl atom type. However, the user is free to define systems of covalently coupled titratable groups.

An initial pK_a calculation is done including all groups in the molecular structure. Next, for each system of covalently coupled groups, the groups that are to be penalized are found. For bases, the group with the highest predicted pK_a (the most stable group in the charged form) will be the first one to adopt a proton as pH is lowered, and this group is allowed to titrate. The remaining groups in the system are penalized so that they are restricted to the neutral form and not allowed to titrate. For acids, the group with the highest pK_a value (the least stable group in the charged form) will be the last group to lose the proton as pH is raised and will be penalized so that it is restricted to the neutral form and not allowed to titrate. The remaining groups in the system are allowed to titrate. Finally, a new pK_a calculation is performed disregarding the penalized groups in each covalently coupled system.

In some cases where the charge is distributed between more than one titratable atom (see the below example of amino-chloro-pyrazine in complex with cell division protein kinase 2), a common charge center can be set. The common charge center, $\underline{r_{ccc}}$, is defined as the average of the positions of the covalently coupled groups:

$$\underline{r_{ccc}} = n^{-1} \sum_{i=1}^{n} \underline{r_i} \tag{7}$$

where n in the number of covalently coupled groups in the system, and $\underline{r_i}$ is the position of group i.

2.5. Prediction of Ligand Model pK_a Values. For the calculations of protonation state change on protein−ligand binding, experimental and previously calculated ligand model pK_a values were used.[5] The Schrödinger program Epik,[20] which is part of the Schrödinger Software Suite, was used for the prediction of ligand model pK_a values for the covalently coupling examples presented in the Results and Discussion.

2.6. Calculation of Protonation State Changes. The overall protonation change, n, is calculated as described in ref 5:

$$n = \sum_{i}^{N} n_i \; n = \sum_{i}^{N} \left(\frac{10^{pK_{a,i}^{c} - pH}}{1 + 10^{pK_{a,i}^{c} - pH}} - \frac{10^{pK_{a,i}^{f} - pH}}{1 + 10^{pK_{a,i}^{f} - pH}} \right) \tag{8}$$

where N is a set containing all titratable groups in the system, and $pK_{a,i}^{c}$ and $pK_{a,i}^{f}$ represent pK_a values of titratable group i in the complexed and free forms of the protein and the ligand molecule, respectively.

2.7. Protein Structures. A list of the structures from the Protein Data Bank[21] used in this study is presented in the Supporting Information in Table S.3.

In the plasmepsin II (PDB id: 1pfz and 1sme) and cathepsin D (1lyw and 1lyb) structures, the N-terminus undergoes a large displacement between the apo and holo structures. This movement can be attributed to the autocatalytic behavior of aspartic acids. To limit the effect of the N-terminus movement on the calculation results, we adopt the procedure of Alexov[22] and remove the N-terminus section from these structures. Furthermore, a flap region (residues 76−80) in the plasmepsin II holo structure was not resolved, and the corresponding region in the apo structure (1sme) was removed. All structures where repaired using the corall method of WHAT IF.[23]

Reduced cysteine bridges were manually reconstructed in the PROPKA3.1 input files for the complexes trypsin−1bMe (Cys173−Cys197 and Cys25−Cys41) and trypsin−1c (Cys173−Cys197). Ligand model pK_a values were set in accordance with experimentally determined values.[24]

3. RESULTS AND DISCUSSION

In the following sections, we will demonstrate the implemented covalent and noncovalent coupling algorithms with pK_a calculations on HIV protease, a pyrazine derivative, and ATP.

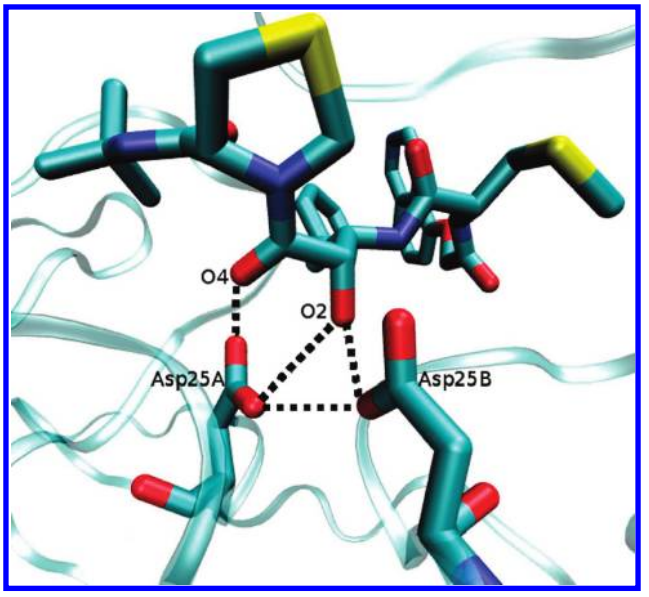


Figure 3. HIV protease complexed with the inhibitor KNI-272 (1hpx). The catalytic active aspartic acids at position 25 in the A and B chains and the ligand molecule are represented as sticks. Side-chain interactions for catalytic active aspartic acids are indicated with dotted lines. Prepared using VMD[25] with the PROPKA GUI plugin.[26]

3.1. Noncovalent Coupling: HIV Protease. PROPKA3.1 detects the catalytic active aspartic acids in HIV protease at position 25 in chains A and B to be noncovalently coupled. Figure 3 is a representation of HIV protease in complex with the KNI-272 inhibitor (1hpx) where the side-chain interactions predicted by PROPKA are represented as dotted lines. PROPKA predicts Asp25A to have side-chain interactions with Asp25B and the two oxygen atoms (O2 and O4) in the allophenylnorstatine fragment of the inhibitor. These side-chain interactions contribute with −0.85 (O2), −0.85 (O4), and −0.84 (Asp25B) pK_a units to the total pK_a value of Asp25A. For the pK_a value of Asp25B, O2 contributes −0.85 pK_a units, whereas Asp25A contributes with +0.84 pK_a units. This gives rise to the total pK_a values of 5.07 (Asp25A) and 9.28 (Asp25B). However, the two aspartic residues are identified as being coupled, indicating the existence of an alternative protonation state. If the contributions to the pK_a values due to the interactions between the aspartic acids are swapped, the alternative total pK_a values of 8.64 (Asp25A) and 5.70 (Asp25B) are found.

Experimental results show that the pK_a value of one of the catalytic active aspartic acids is 6.6 when in complex with KNI-272.[27] As the dyad is found to be monoprotonated, the experimental value should be compared to the highest of the predicted pK_a values. While these values (9.28 or 8.64) are somewhat higher than the experimental value, the PROPKA2.0 value of 10.28 has a greater discrepancy. The pK_a calculation on HIV protease complexed with KNI-272 is discussed in more detail later.

3.2. Covalent Coupling: Aromatic Nitrogen Atoms in a Pyrazine Derivative. For some chemical compounds, the titration of individual titratable sites is far from independent. Pyrazine rings, for example, contain two potentially titratable nitrogen atoms. However, once one of the nitrogen atoms has adopted a hydrogen atom, the probability of the other nitrogen atom adopting a hydrogen atom is significantly reduced. For the pyrazine molecule, pK_a values of the first and second protonation have been observed to be pK_a (1) 0.65 and pK_a (2) −6.6.[28] To model such compounds in PROPKA3.1, we must take these covalent couplings into account.

To illustrate this principle, we will investigate cyclin-dependent kinase 2 (CDK2) in complex with amino-chloro-pyrazine (1wcc[29]). A schematic overview of the protonation states of amino-chloropyrazine is given in Figure 4.

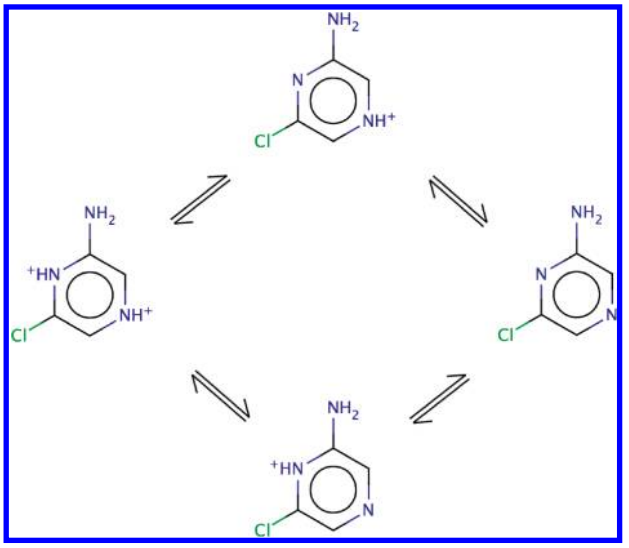


Figure 4. Modeled protonation states of 2-amino-6-chloro-pyrazine.

Table 5. Resulting pK_a Values for the Aromatic Nitrogen Atoms in 2-Amino-6-chloro-pyrazine[a]

	predicted pK_a values		model pK_a values		
calculation	N4	N8	N4	N8	CCC
A	1.91	0.44	5.00	5.00	no
B	−0.98	−4.99	2.09	−0.43	no
C	−0.78	−4.76	2.09	−0.43	yes

[a] Calculation A, the PROPKA default model pK_a value of 5.00; calculation B, custom ligand model pK_a values; calculation C, custom model pK_a values and common charge center (CCC) approach.

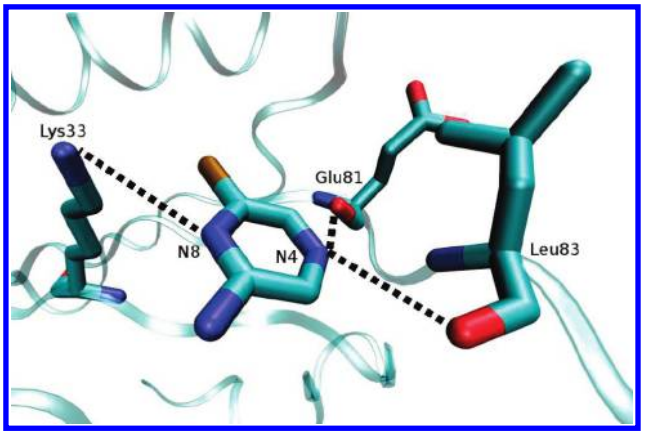


Figure 5. Modeled interactions between 2-amino-6-chloro-pyrazine and CDK2. Prepared using the PROPKA GUI plugin for VMD.

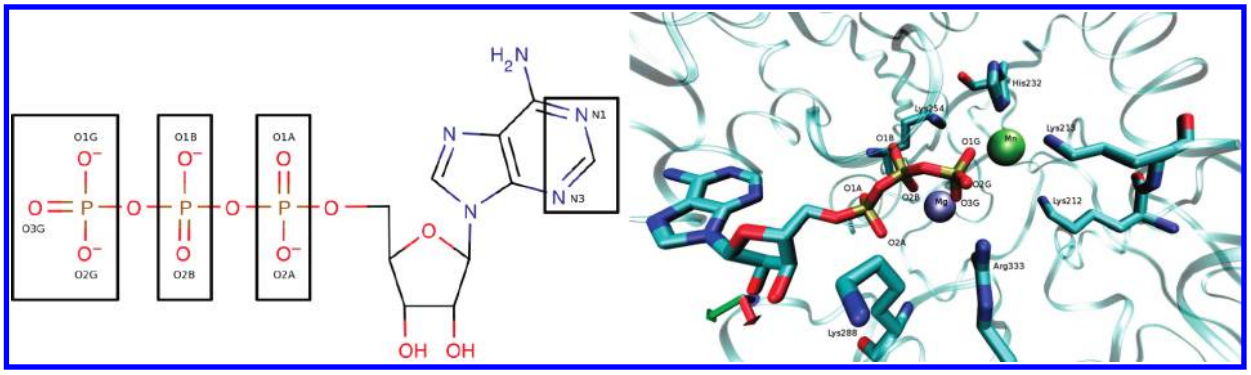


Figure 6. ATP. Left panel: Structure of ATP with covalently coupled groups marked by boxes. Prepared using MarvinSketch. Right panel: ATP in the binding site of phosphoenolpyruvate carboxykinase. Residues with strong interactions with the ATP phosphate groups are represented as sticks. The magnesium and manganese ions are represented as spheres. Prepared using the PROPKA GUI plugin for VMD.

Three calculations on the complex between CDK2 and amino-chloro-pyrazine have been done; see Table 5. In calculation A, the PROPKA default ligand model pK_a value of 5.00 for aromatic nitrogen atoms is applied. In calculation B, custom ligand model pK_a values of 2.09 (N4) and −0.43 (N8) calculated with Schrödinger's Epik program were utilized. Finally, in calculation C, the custom Epik ligand model pK_a values utilized and electrostatic interactions were calculated using the common charge center approach; see Materials and Methods. For both aromatic nitrogen atoms, the predicted pK_a value is significantly lowered due to large desolvation contributions (−2.51 pK_a units for N8 and −3.41 pK_a units for N4 in calculations A and B, and −2.84 pK_a units for both nitrogen atoms in calculation C). A protonation of the N8 nitrogen would result in unfavorable interactions with the proximate Lys33; see Figure 5. The lysine residue therefore contributes with a side-chain interaction contribution of −0.36 pK_a units in all calculations and a Coulombic interaction contribution of −1.56 pK_a units (calculations A and B) or −0.99 pK_a units (calculation C) to the predicted pK_a value of N8. The pK_a value of N4 is, on the other hand, increased due to interactions with the backbone carbonyl of Glu81 (0.24 pK_a units) and Leu83 (0.86 pK_a units). Hence, in all three calculations, N8 has the lowest pK_a value and is excluded from the final calculation.

PROPKA2.0 applied to the same structure wrongly predicts that both aromatic nitrogen atoms titrate at similar pH values. The pK_a values are predicted to be 0.99 (N4) and 0.76 (N8).

3.2.1. ATP. A more complicated example is adenosine-5′-triphosphate (ATP), which with its three phosphate groups urges for careful considerations during the setup of a PROPKA calculation. Each of the phosphate groups contain two or three oxygen atoms that could potentially adopt a proton. However, only four of the total of seven phosphate oxygen atoms can titrate as the remaining three phosphate oxygen atoms have double bonds to a phosphorus atom. The challenge is therefore to determine which of the phosphate oxygen atoms should be allowed to titrate.

The new algorithm for treatment of covalently coupled titratable groups allows the users to let PROPKA3.1 make this decision. The oxygen atoms in each phosphate group are set up to be covalently coupled as shown in the left panel in Figure 6. The ability of titration is hence removed for one oxygen atom in each of the phosphate groups.

Applying this approach to ATP complexed with phosphoenolpyruvate carboxykinase (1aq2,[30] Figure 6, right panel) identifies the double-bonded, nontitratable phosphate oxygen atoms as O3G with the potential pK_a value −11.66 (as compared to O1G, −12.92 and O2G, −12.82), O2B with pK_a value −4.30 (as compared to O1B, −5.31), and O1A with pK_a value 1.69 (as compared to O2A, 0.57); confer the left panel of Figure 6.

The potential pK_a values for the oxygen atoms that are restricted to the neutral state are consistently around one pK_a unit higher than the pK_a values of the oxygen atoms that are found to titrate. This difference in pK_a value is relatively small, and the choice of nontitratable oxygen atoms can seem somewhat ambiguous. We do, however, not consider this ambiguity important. Instead, it is important to note that PROPKA automatically has assigned the double bonds to the oxygen atoms for which it is predicted to be most energetically favorable for the entire system. Additionally, the ambiguity stemming from a user-defined setup of the ATP molecule is removed.

3.3. Protonation State Changes on Complexation. The following sections describe the results using PROPKA3.1 on calculations originally done with PROPKA2.0. For more in-depth information on the experimental results, etc., please consult ref 5.

3.3.1. Trypsin and Thrombin. Table 6 contains the results for calculations of protonation state changes done on serine protease trypsin complexed with a set of related inhibitors derived from N^{α}-(2-naphthylsulphonyl)-L-3-amidino-phenylalanine and thrombin complexed with four inhibitors. Klebe and co-workers[24,31] have measured protonation state changes and calculated pK_a changes using the PEOE_PB scheme for these complexes. Also included for comparison are the calculations originally done with PROPKA2.0.

For the binding of the ligand 1b to trypsin, an overall protonation state change of 0.66 was calculated in accordance with the experimental value of 0.90. This is an improvement as compared to the calculated value by PROPKA2.0 of 0.09. The majority of the overall calculated protonation state change has its origin in a pK_a shift of His57 from 7.47 in the apo form to 9.35 in the holo form. The shift in pK_a value stems from interactions between the ligand carboxyl group and His57. While the increased desolvation contribution gives rise to a negative contribution to the pK_a shift of His57 (−2.69 in the holo form as compared to −1.82 in the apo form), the desolvation contribution also activates the COO-HIS exception (confer Materials and Methods) for both Asp102 and the ligand carboxyl giving a contribution of 1.60 for each carboxyl group. For trypsin complexed with the ligands 1bMe, 1c, and 1cMe, negligible overall protonation state changes were predicted in accordance

Table 6. Experimental and Calculated Total Changes in Protonation State on Complexation for the Trypsin and Thrombin Complexes Studied[a]

		PROPKA		n (pK_a) complexed			pK_a uncomplexed		
name	n_{exp}	3.1	2.0	His57	ligCOO	ligAMINO	His57	ligCOO	ligAMINO
Trypsin									
1b (1kli[24])	0.90	0.66	0.09	0.65(9.35)	0.00 (0.99)	n/a	7.47	3.21	n/a
1bMe	0.00	0.12	0.06	0.06(7.57)	n/a	n/a	7.45	n/a	n/a
1c	0.00	−0.05	0.15	−0.01 (7.49)	0.00 (4.19)	n/a	7.50	4.17	n/a
1cMe (1klj[24])	0.00	−0.30	−0.24	−0.21(6.85)	n/a	n/a	7.46	n/a	n/a
1d (1kll[24])	−0.53	−0.54	−0.55	−0.24(6.32)	n/a	−0.26(6.65)	7.37	n/a	7.49
1dAc (1klm[24])	0.00	−0.22	−0.23	−0.18(6.89)	n/a	n/a	7.42	n/a	n/a
2	0.93	0.71	0.69	0.68(9.39)	0.00 (1.30)	n/a	7.43	3.40	n/a
3 (1kln[24])	0.00	0.00	−0.05	−0.09 (7.18)	0.00(3.14)	n/a	7.39	3.84	n/a
4 (1klo[24])	0.00	0.21	0.03	−0.16(6.93)	0.00(0.94)	0.28(7.99)	7.38	2.65	7.48
5 (1klp[24])	0.00	0.22	0.07	−0.12(7.15)	0.00(1.09)	0.27(8.56)	7.43	2.51	7.95
Thrombin									
2	0.88	0.61	0.34	0.80(9.22)	0.00 (−0.90)	n/a	7.08	3.40	n/a
3 (1ypk[32])	0.00	−0.15	−0.01	−0.13(5.92)	0.00 (1.65)	n/a	7.01	3.84	n/a
4 (1k21[24])	0.00	0.12	−0.07	−0.15(6.46)	0.00(0.75)	0.27(7.96)	7.18	2.65	7.48
5 (1k22[24])	0.00	−0.08	0.04	−0.15(6.37)	0.00 (0.61)	0.20(8.36)	7.15	2.51	7.95
rmsd		0.18	0.29						

[a] For His57 and ligand carboxyl (ligCOO) and amino (ligAMINO) groups (where applicable), pK_a values in the complexed and uncomplexed structures are listed along with contributions to the change in protonation state on complexation.

with experimental results. For the ligand 1d, the calculated value of −0.54 corresponds very well to the experimental value of −0.53. Similarly, the calculated values for trypsin−2 (0.71) and thrombin−2 (0.61) correspond well with the experimental values of 0.93 and 0.88, respectively. For thrombin−2, a clear improvement to the value calculated using PROPKA2.0 of 0.34 is observed.

In summary, an improvement in the predictions of PROPKA3.1 as compared to the predictions of PROPKA2.0 was observed for trypsin−1b and thrombin−2. The remaining results were similar to those obtained with PROPKA2.0. The overall rms deviations between experimental protonation state changes on complexation and calculated values were 0.18 for PROPKA3.1 and 0.29 for PROPKA2.0.

3.3.2. Proteins Complexed with Pepstatin. The protonation state change on ligand binding for three proteins (plasmepsin II, cathepsin, and endiothiapepsin) complexed with pepstatin is calculated and compared to experimental data and PROPKA2.0 results, cf., Table 7.

For plasmepsin II (1pfz) complexed with pepstatin (1sme), PROPKA3.1 predicts the overall protonation state change on ligand binding to be 2.67. This value is higher than the PROPKA2.0 value of 1.50 but still in accord with the experimental value of 1.7. The main predicted contributions to the overall protonation state change stem from Asp303 (1.00), Asp214 (0.99), and His318 (0.54). While the pK_a shifts of Asp303 and Asp12 can be attributed to increased desolvation on ligand binding and to conformational changes, the pK_a shift of His318, which is distant from the active site, stems from an interaction with Asp190 found in the holo structure but not in the apo structure. His164 was found to give a small contribution (−0.20) in the opposite direction of what would be expected from the experimental pK_a values.

Figure 7 shows predicted protonation state changes for the complexation of plasmepsin II with pepstatin at various pH values. While the prediction has a clear discrepancy with the experimental values, it is comparable to previously calculated results.[5,22]

For cathepsin (1lyw) complexed with pepstatin (1lyb), the overall protonation state change is predicted to be 2.39, which is in good agreement with the experimental value of 2.9 and very close to the PROPKA2.0 value of 2.49. The main contributions stem from Asp323 (0.92) and His56 (0.68). The pK_a shift of Asp323 is caused by an increased desolvation due to the binding of pepstatin, whereas the pK_a shift of His56 is due to conformational changes more distant from the active site.

For endiothiapepsin (4ape) complexed with pepstatin (4er2), only a small protonation state change is predicted (0.08), whereas the PROPKA2.0 value is 0.77 and the experimental value is 1.06. The original PROPKA2.0 calculation found the pK_a value of Asp30 to be shifted from 6.69 to 7.21 (corresponding to a protonation state change of 0.29) and the pK_a value of Asp12 to be shifted from 5.55 to 9.26 (corresponding to a protonation state change of 0.96) during complexation at pH 7.0. These shifts were mainly due to desolvation effects and not reproduced in PROPKA3.1, which has an improved desolvation model.

To summarize, the main result of this subsection is that the overall protonation state values predicted by PROPKA3.1 are similar to the PROPKA2.0 values for cathepsin, but worse for plasmepsin II and endiothiapepsin.

3.3.3. HIV Protease. For HIV protease complexed with pepstatin (5hvp), the pK_a values of the catalytic dyad consisting of Asp25A and Asp25B are predicted to be 10.16 and 5.73, respectively. This is partly in accord with experimental findings where one of the aspartic acids has a pK_a value less than 2.5 and the other has a pK_a value greater than 6.5 as measured with

Table 7. Experimental (Where Available) and Calculated Protonation State Changes for Other Protein−Ligand Complexes Studied[a]

name	n_{exp} (PROPKA3.1; 2.0)	experimental pK_a (PROPKA3.1) for selected residues			
		Asp34A/Asp214A	Asp34A/Asp214A	His164A	Asp303A
plasmepsin II (1pfz[33])		4.7 (4.90/3.67)	4.7 (4.90/3.67)	6.0 (6.45)	(3.69)
+ pepstatin (1sme[34])	1.7[b] (2.67; 1.50)	3.0 (4.73/8.77)	6.5 (4.73/8.77)	7.5 (6.06)	(9.14)
		Asp33A/Asp231B	His77A	Glu260B	Asp323B
cathepsin D (1lyw[35])		<6.5 (4.66/7.51)	<6.5 (6.80)	<6.5 (4.58)	(5.34)
+ pepstatin (1lyb[36])	2.9[b] (2.39; 2.49)	>6.5 (4.42/8.98)	>6.5 (6.76)	>6.5 (5.87)	(8.42)
		Asp32A/Asp215A	Aspl2A	Asp30A	
endiothiapepsin (4ape[37])		(9.23/4.42)	(6.70)	(8.07)	
+ pepstatin (4er2[37])	1.06 (0.08; 0.77)	>7.0 (9.33/4.54)	(7.23)	(8.35)	
		Asp25A/Asp25B	Asp30A/Asp30B	His69A/His69B	
HIV protease (5hvp[38])		(4.78/9.56)	(5.10/3.92)	(6.96/6.25)	
+ pepstatin (5hvp)	NMR	<2.5/>6.5 (10.16/5.73)	(5.40/3.52)	(6.96/6.25)	
HIV protease (1hhp[39])		(5.32/7.19)	(4.38/4.38)	(7.23/7.23)	
+ pepstatin (5hvp)	NMR	<2.5/>6.5 (10.16/5.73)	(5.40/3.52)	(6.96/6.25)	
HIV protease (3hvp[40])		(7.60/4.79)	(4.78/4.78)	(6.83/6.83)	
+ pepstatin (5hvp)	NMR	<2.5/>6.5 (10.16/5.73)	(5.40/3.52)	(6.96/6.25)	
HIV protease (1hpx[41])		6.0 (8.89/5.17)	4.8 (4.42/4.63)	4.8	
+ KNI-272 (1hpx)	−0.23[c] (0.01; 0.18)	6.6 (5.07/9.28)	3.88/3.78 (4.62/4.91)	2.9 (4.40)	
HIV protease (1hhp)		6.0 (5.32/7.19)	4.8 (4.38/4.38)	4.8	
+ KNI-272 (1hpx)	−0.23[c] (0.04; 1.27)	6.6 (5.07/9.28)	3.88/3.78 (4.62/4.91)	2.9 (4.40)	
HIV protease (3hvp)		6.0 (7.60/4.79)	4.8 (4.78/4.78)	4.8	
+ KNI-272 (1hpx)	−0.23[c] (−0.24; 0.97)	6.6 (5.07/9.28)	3.88/3.78 (4.62/4.91)	2.9 (4.40)	
HIV protease (1qbs[42])		6.0 (8.21/6.00)	(4.82/4.82)		
+ DMP-323 (1qbs)	NMR	8.19 (5.88/8.52)	3.99 (5.07/5.09)		
HIV protease (1hhp)		6.0 (5.32/7.19)	(4.38/4.38)		
+ DMP-323 (1qbs)	NMR	8.19 (5.88/8.52)	3.99 (5.07/5.09)		
HIV protease (3hvp)		6.0 (7.60/4.79)	(4.78/4.78)		
+ DMP-323 (1qbs)	NMR	8.19 (5.88/8.52)	3.99 (5.07/5.09)		
		His57F/His57I			
chymotrypsin (7gch[43])		7.5 (7.33/7.33)			
+ *N*-acetyl-L-Leu-DL-Phe-CF_3 (7gch)	NMR	12.0 (8.07/8.07)			
chymotrypsin (6gch[43])		7.5 (7.54/7.48)			
+ *N*-acetyl-DL-Phe-CF_3 (6gch)	NMR	10.8 (9.46/9.41)			
		Glu172A			
xylanase (1bvv[44])		6.7 (7.52)			
+ 2FXb (1bvv)	NMR	4.2 (6.80)			
		Asp35A/Glu172A			
xylanase N35D (1c5i[45])		3.7/8.4(5.58/11.45)			
+ 2FXb (1c5i)	NMR	1.9−3.4 or >9.0 (10.83/5.70)			
		His235A			
hydroxynitrile lyase (2yas[46])		2.5 (6.72)			
+ thiocyanate (2yas)	NMR	8.0 (8.30)			
		Asp27B	methotrexate N1		
DHFR (4dfr[47])		6.6 (5.82)	5.7		
+ methotrexate (4dfr)	NMR	(−0.76)	10.7 (5.97)		
DHFR (4dfr)		6.6 (5.82)	4.4		
+ methotrexate (4dfr)	NMR	(−0.76)	10.7 (4.67)		

[a] Unless otherwise stated, pH is 7.0. [b] pH is 6.5. [c] pH is 5.0. Experimental and calculated pK_a values are listed for selected residues.

NMR.[48] PROPKA3.1 correctly predicts that Asp25A and Asp25B are noncovalently coupled and alternative pK_a values (by swapping the interaction between the two aspartic residues) are 5.81 and 10.08 for Asp25A and Asp25B, respectively. In the apo form, the pK_a values of Asp25A and Asp25B are predicted to be 4.78 and 9.56, respectively. Utilization of two alternative structures (1hhp and 3hvp) to describe the apo structure lowers the higher pK_a value in the catalytic dyad, making the pK_a shift on pepstatin binding larger.

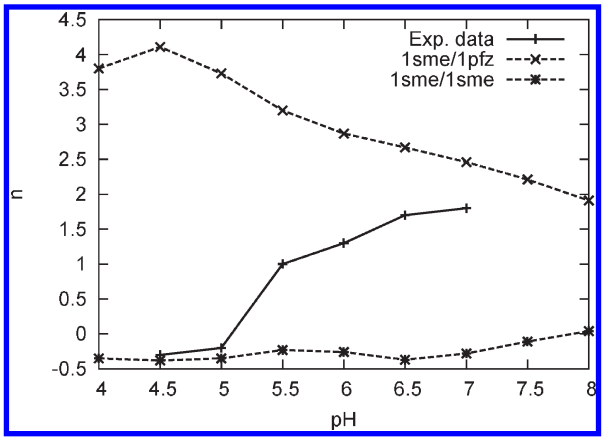


Figure 7. Protonation state change as a function of pH for complexation of plasmepsin II with pepstatin. Predictions were done solely on the basis of the plasmepsin II−pepstatin complex structure (1sme) as well as in combination with the plasmepsin II apo structure (1pfz).

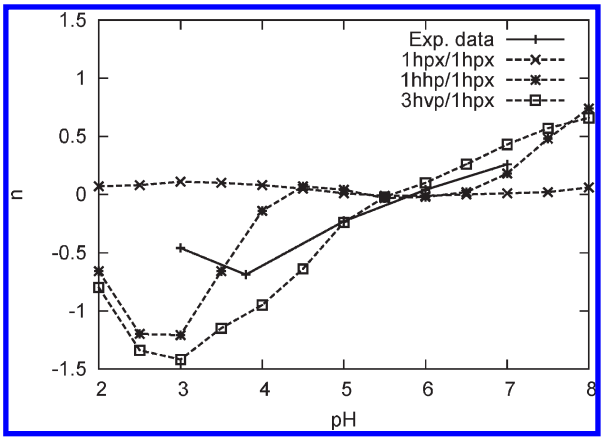


Figure 8. Protonation state change for HIV protease on complexation with the inhibitor KNI-272 as a function of pH. The predictions were done solely on the basis of the structure of the HIV protease−KNI-272 complex (1hpx) as well as in combination with HIV protease apo structures (1hhp and 3hvp).

For HIV-1 protease complexed with the inhibitor KNI-272 (1hpx), experimental results show that the pK_a value of one of the catalytic active aspartic acids changes from 6.0 in the apo form to 6.6 when in complex with KNI-272.[27] As the dyad is found to be monoprotonated both before and after complexation, we must compare these experimental values to the highest of the predicted pK_a values for the dyad, which are 8.89 (apo) and 9.28 (holo). While these predictions are higher than the experimental value, they are an improvement as compared to the PROPKA2.0 values of 9.26 (apo) and 10.28 (holo). The catalytic dyad is predicted to be noncovalently coupled; see Noncovalently Coupling: HIV Protease. The pK_a value of the isoquinoline nitrogen on KNI-272 has been observed to shift from 4.8 to 2.9 during complexation. Using the experimental model pK_a value of 4.8, PROPKA3.1 predicts the pK_a value after complexation to be 4.40. This is comparable to the PROPKA2.0 value of 4.52.

The overall protonation state change was predicted to be 0.01, 0.04, and −0.24 when the apo structure was represented by 1hpx (KNI-272 removed), 1hhp, and 3hvp, respectively. These values agree well with the experimental value of −0.23. For comparison, the values predicted with PROPKA2.0 were 0.18, 1.27, and 0.97, respectively.

Experimental protonation state changes as a function of pH for HIV protease complexed with the inhibitor KNI-272 are presented in Figure 8 along with calculated values. It is seen that calculations done with 1hpx representing the apo form underestimate the protonation state changes due to pepstatin binding, whereas better accordance with experimental results is achieved for calculations where the apo form is represented by the structure with the PDB id 3hvp.

For HIV protease complexed with DMP-323 (1qbs), the pK_a values of the catalytic dyad are predicted to be 5.88 and 8.52. This is in accord with the experimental result that the dyad has a pK_a value of 8.19.[49] While the experimental finding that both aspartic acids have a pK_a value above 7.2 is not reproduced, there is a better agreement than for the PROPKA2.0 values of 3.26 and 8.56.

In summary, a better agreement with the experimental overall protonation state change for the complexation of HIV protease with KNI-272 was achieved with PROPKA3.1 as compared to the previously reported PROPKA2.0 values. Furthermore, an improvement in the predicted pK_a values of the catalytic dyad was observed for HIV protease complexed with KNI-272 and DMP-323.

3.3.4. Ligands Covalently Bound to the Protein Receptor. We have applied PROPKA3.1 to chymotrypsin covalently bound, via its Ser195 residue, to two peptidyl trifluoromethyl ketone derived inhibitors. The oxygen in the ligand hemiketal adduct, formed during complexation from the ligand ketone and the serine hydroxyl, is negatively charged. For chymotrypsin, His57 undergoes a shift from 7.54 (chain F)/7.48 (chain I) to 9.46 (chain F)/9.41 (chain I) during complexation with the NAc−Phe−CF_3 inhibitor (6gch). This is in accordance with the experimental shift from 7.5 to 10.8. For the NAc−Leu−Phe−CF_3 inhibitor, the shift is predicted to be smaller (7.33/7.33 to 8.07/8.07 for chains F/I) in contrast to the experimental shift (7.5 to 12.0) but in agreement with the PROPKA2.0 shift (6.94 to 8.90).

For the covalent complexation of xylanase (1bvv) via its Glu78 with 2-deoxy-2-fluoro-β-xylobioside (2FXb), a pK_a shift of Glu172 from 7.52 to 6.80 was predicted. While smaller than the experimental shift (6.7 to 4.2), it is in agreement with the PROPKA2.0 shift (7.45 to 6.64). Using the xylanase Asn35Asp mutant structure (1c5i), the pK_a shift of the Asp35/Glu172 dyad is predicted to be from 5.58/11.45 to 10.83/5.70. The Asp35/Glu172 dyad is correctly detected as being noncovalently coupled.

3.3.5. Ligands Containing Thiol Groups. The pK_a value of His235 of hydroxynitrile lyase has been found to shift from 2.5 to ∼8 during binding of thiocyanate.[50] The PROPKA3.1 predictions of the pK_a value of His235 of 8.30 (holo form) and 6.72 (apo form) capture the direction of the shift and the experimental pK_a value in the holo form; however, the pK_a of the apo form is significantly higher than the experimental value. The nitrile group elevates the pK_a value of His235 by 0.41 pK_a units due to an intrinsic electrostatic interaction, and the thiol group elevates the pK_a value of His235 by 0.26 pK_a unit (intrinsic electrostatic interaction) and 1.05 pK_a units (Coulombic interaction).

3.3.6. Dihydrofolate Reductase. We apply PROPKA3.1 to dihydrofolate reductase complexed with methotrexate. NMR experiments have shown that that the aromatic N1 nitrogen of

methotrexate is protonated in the complex up to pH values in excess of 10.[51−54] In the complex, the N1 nitrogen forms a salt bridge with the deprotonated aspartic acid Asp27. PROPKA3.1 predicts the pK_a value of Asp27 to be 5.82 (experimental value 6.6) in the apo form and −0.76 in holo form. The pK_a value of the N1 nitrogen is predicted to shift from 5.7 to 5.97 during complexation. PROPKA3.1 does therefore not capture a large shift observed experimentally. PROPKA2.0 predicted the pK_a value of the N1 nitrogen in the holo form to be 6.89. Setting the model pK_a value of the N1 nitrogen to a calculated value of 4.4 does not significantly change the predicted shift in pK_a value.

4. CONCLUSION

In the preparation of PROPKA3.1, we have focused on improving the treatment of ligand molecules and coupling effects. It is our hope that the improvements will assist users in PROPKA calculations on complicated systems. The implemented algorithm for noncovalent coupling will automatically detect coupled groups and notify the user of the existence of alternative pK_a values. The implemented algorithm for covalent coupling improves the PROPKA treatment of complicated ligand molecules where titrational events are strongly linked. Furthermore, the algorithm will remove some of the ambiguity that arises when users manually set up complicated molecules for PROPKA calculations.

We have tested new implementations on a set of ligand-containing structures also used for the preparation of PROPKA2.0. Except for a few cases, the new algorithms in PROPKA3.1 were found to yield results similar to or better than those obtained with PROPKA2.0. With the improved treatment of complicated ligand molecules, pK_a calculation on structures containing DNA is now within reach, and this will be the goal of our continued efforts.

The PROPKA3.1 code is available from http://propka.ki.ku.dk.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information. Three tables: Table S.1, a list of ligand interaction parameters; Table S.2, types of interactions (normal or iterative) used for all combinations of interacting groups; and Table S.3, Protein Data Bank structures used in this study. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: chresten@chem.ku.dk (C.R.S.), jhjensen@chem.ku.dk (J.H.J.).

■ ACKNOWLEDGMENT

Atomic coordinates for trypsin in complex with the ligands 1bMe, 1c, and 2 were kindly supplied by Paul Czodrowski and Gerhard Klebe. Atomic coordinates for thrombin in complex with the ligand 2 were kindly supplied by David Banner. This work was supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 202167 and the Danish Council for Strategic Research through a research grant from the Program Commission on Strategic Growth Technologies (2106-07-0030).

■ REFERENCES

(1) Reddy, A. S.; Pati, S. P.; Kumar, P. P.; Pradeep, H.; Sastry, G. N. *Curr. Protein Pept. Sci.* **2007**, *8*, 329–351.
(2) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat. Rev. Drug Discovery* **2004**, *3*, 935–949.
(3) Czodrowski, P.; Dramburg, I.; Sotriffer, C. A.; Klebe, G. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 424–437.
(4) Lamotte-Brasseur, J.; Lounnas, V.; Raquet, X.; Wade, R. C. *Protein Sci.* **1999**, *8*, 404–409.
(5) Bas, D. C.; Rogers, D. M.; Jensen, J. H. *Proteins: Struct., Funct., Bioinf.* **2008**, *73*, 765–783.
(6) Aguilar, B.; Anandakrishnan, R.; Ruscio, J. Z.; Onufriev, A. V. *Biophys. J.* **2010**, *98*, 872–880.
(7) Olsson, M. H.; Søndergard, C. R.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.
(8) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins: Struct., Funct., Bioinf.* **2005**, *61*, 704–721.
(9) Klingen, A. R.; Bombarda, E.; Ullmann, G. M. *Photochem. Photobiol. Sci.* **2006**, *5*, 588–596.
(10) Søndergaard, C. R.; McIntosh, L. P.; Pollastri, G.; Nielsen, J. E. *J. Mol. Biol.* **2008**, *376*, 269–287.
(11) McIntosh, L. P.; Hand, G.; Johnson, P. E.; Joshi, M. D.; Korner, M.; Plesniak, L. A.; Ziser, L.; Wakarchuk, W. W.; Withers, S. G. *Biochemistry* **1996**, *35*, 9958–9966.
(12) Qin, J.; Clore, G. M.; Gronenborn, A. M. *Biochemistry* **1996**, *35*, 7–13.
(13) Chivers, P. T.; Prehoda, K. E.; Volkman, B. F.; Kim, B.-M.; Markley, J. L.; Raines, R. T. *Biochemistry* **1997**, *36*, 14985–14991.
(14) Jeng, M.; Holmgren, A.; Dyson, H. *Biochemistry* **1995**, *34*, 10101–10105.
(15) van Rossum, G. The Python Language Reference (accessed Feb 14, 2009); http://docs.python.org/py3k/reference/index.html.
(16) Williams, R. pK_a Data Compiled by R. Williams; research.chem.psu.edu/brpgroup/pKa_compilation.pdf.
(17) Smith, M. Functional Groups. *Organic Chemistry: An Acid-Base Approach*; CRC Press: Boca Raton, FL, 2011; p 142.
(18) Warshel, A. *Biochemistry* **1981**, *20*, 3167–3177.
(19) ChemAxon, MarvinSketch 5.2.6 (accessed Jan 15, 2010), 2009; http://www.chemaxon.com/products/marvin/marvinsketch/.
(20) Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 681–691.
(21) Berman, H.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.; Weissig, H.; Shindyalov, I.; Bourne, P. *Nucleic Acids Res.* **2000**, *28*, 235–242.
(22) Alexov, E. *Proteins: Struct., Funct., Bioinf.* **2004**, *56*, 572–584.
(23) Vriend, G. *J. Mol. Graphics* **1990**, *8*, 52–56.
(24) Dullweber, F.; Stubbs, M. T.; Musil, D.; Stürzebecher, J.; Klebe, G. *J. Mol. Biol.* **2001**, *313*, 593–614.
(25) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
(26) Rostkowski, M.; Olsson, M.; Søndergaard, C.; Jensen, J. H. *BMC Struct. Biol.* **2011**, *11*, 6.
(27) Velazquez-Campoy, A.; Luque, I.; Todd, M. J.; Milutinovich, M.; Kiso, Y.; Freire, E. *Protein Sci.* **2000**, *9*, 1801–1809.
(28) Joule, J. A.; Mills, K. The Diazines: Pyridazine, Pyrimidine, and Pyrazine: Reactions and Synthesis. *Heterocyclic Chemistry*, 5th ed.; Wiley-Blackwell: Chichester, UK, 2010; p 254.
(29) Hartshorn, M. J.; Murray, C. W.; Cleasby, A.; Frederickson, M.; Tickle, I. J.; Jhoti, H. *J. Med. Chem.* **2005**, *48*, 403–413.
(30) Tari, L.; Matte, A.; Goldie, H.; Delbaere, L. *Nat. Struct. Biol.* **1997**, *4*, 990–994.
(31) Czodrowski, P.; Sotriffer, C. A.; Klebe, G. *J. Mol. Biol.* **2007**, *367*, 1347–1356.
(32) Fokkens, J.; Klebe, G. *Angew. Chem., Int. Ed.* **2006**, *45*, 985–989.
(33) Bernstein, N. K.; Cherney, M. M.; Loetscher, H.; Ridley, R. G.; James, M. N. *Nat. Struct. Mol. Biol.* **1999**, *6*, 32–37.

(34) Silva, A. M.; Lee, A. Y.; Gulnik, S. V.; Maier, P.; Collins, J.; Bhat, T. N.; Collins, P. J.; Cachau, R. E.; Luker, K. E.; Gluzman, I. Y.; Francis, S. E.; Oksman, A.; Goldberg, D. E.; Erickson, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10034–10039.
(35) Lee, A. Y.; Gulnik, S. V.; Erickson, J. W. *Nat. Struct. Mol. Biol.* **1998**, *5*, 866–871.
(36) Baldwin, E. T.; Bhat, T. N.; Gulnik, S.; Hosur, M. V.; Sowder, R. C.; Cachau, R. E.; Collins, J.; Silva, A. M.; Erickson, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6796–6800.
(37) Pearl, L.; Blundell, T. *FEBS Lett.* **1984**, *174*, 96–101.
(38) Fitzgerald, P. M.; McKeever, B. M.; VanMiddlesworth, J. F.; Springer, J. P.; Heimbach, J. C.; Leu, C. T.; Herber, W. K.; Dixon, R. A.; Darke, P. L. *J. Biol. Chem.* **1990**, *265*, 14209–14219.
(39) Spinelli, S.; Liu, Q.; Alzari, P.; Hirel, P.; Poljak, R. *Biochimie* **1991**, *73*, 1391–1396.
(40) Wlodawer, A.; Miller, M.; Jaskólski, M.; Sathyanarayana, B.; Baldwin, E.; Weber, I.; Selk, L.; Clawson, L.; Schneider, J.; Kent, S. *Science* **1989**, *245*, 616–21.
(41) Baldwin, E.; Bhat, T.; Gulnik, S.; Liu, B.; Topol, I.; Kiso, Y.; Mimoto, T.; Mitsuya, H.; Erickson, J. *Structure* **1995**, *3*, 581–590.
(42) Lam, P. Y. S.; et al. *J. Med. Chem.* **1996**, *39*, 3514–3525.
(43) Brady, K.; Wei, A.; Ringe, D.; Abeles, R. *Biochemistry* **1990**, *29*, 7600–7607.
(44) Sidhu, G.; Withers, S.; Nguyen, N.; McIntosh, L.; Ziser, L.; Brayer, G. *Biochemistry* **1999**, *38*, 5346–5354.
(45) Sun, Z.; Dotsch, V.; Kim, M.; Li, J.; Reinherz, E.; Wagner, G. *EMBO J.* **1999**, *18*, 2941–2949.
(46) Zuegg, J.; Gruber, K.; Gugganig, M.; Wagner, U.; Kratky, C. *Protein Sci.* **1999**, *8*, 1990–2000.
(47) Bolin, J.; Filman, D.; Matthews, D.; Hamlin, R.; Kraut, J. *J. Biol. Chem.* **1982**, *257*, 13650–13662.
(48) Smith, R.; Brereton, I. M.; Chai, R. Y.; Kent, S. B. *Nat. Struct. Mol. Biol.* **1996**, *3*, 946–950.
(49) Yamazaki, T.; Nicholson, L. K.; Wingfield, P.; Stahl, S. J.; Kaufman, J. D.; Eyermann, C. J.; Hodge, C. N.; Lam, P. Y. S.; Torchia, D. A. *J. Am. Chem. Soc.* **1994**, *116*, 10791–10792.
(50) Stranzl, G. R.; Gruber, K.; Steinkellner, G.; Zangger, K.; Schwab, H.; Kratky, C. *J. Biol. Chem.* **2004**, *279*, 3699–3707.
(51) Cocco, L.; Groff, J. P.; Temple, C.; Montgomery, J. A.; London, R. E.; Matwiyoff, N. A.; Blakley, R. L. *Biochemistry* **1981**, *20*, 3972–3978.
(52) Cocco, L.; Roth, B.; Temple, C.; Montgomery, J. A.; London, R. E.; Blakley, R. L. *Arch. Biochem. Biophys.* **1983**, *226*, 567–577.
(53) Stone, S.; Morrison, J. *Biochim. Biophys. Acta* **1983**, *745*, 247–258.
(54) London, R.; Howell, E.; Warren, M.; Kraut, J.; Blakley, R. *Biochemistry* **1986**, *25*, 7229–7235.