Very fast prediction and rationalization of pK_a values for protein–ligand complexes

Delphine C. Bas,[1] David M. Rogers,[2] and Jan H. Jensen[2]*

[1] Equipe de Chimie et Biochimie Théoriques, UMR 7565 - CNRS, Université Henri Poincaré, Nancy I,

Boulevard des Aiguillettes BP 239, 54506 Vandoeuvre-lès-Nancy Cedex, France

[2] Department of Chemistry, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark

ABSTRACT

The PROPKA method for the prediction of the pK_a values of ionizable residues in proteins is extended to include the effect of non-proteinaceous ligands on protein pK_a values as well as predict the change in pK_a values of ionizable groups on the ligand itself. This new version of PROPKA (PROPKA 2.0) is, as much as possible, developed by adapting the empirical rules underlying PROPKA 1.0 to ligand functional groups. Thus, the speed of PROPKA is retained, so that the pK_a values of all ionizable groups are computed in a matter of seconds for most proteins. This adaptation is validated by comparing PROPKA 2.0 predictions to experimental data for 26 protein–ligand complexes including trypsin, thrombin, three pepsins, HIV-1 protease, chymotrypsin, xylanase, hydroxynitrile lyase, and dihydrofolate reductase. For trypsin and thrombin, large protonation state changes ($|n| > 0.5$) have been observed experimentally for 4 out of 14 ligand complexes. PROPKA 2.0 and Klebe's PEOE approach (Czodrowski P, et al. J Mol Biol 2007;367:1347–1356) both identify three of the four large protonation state changes. The protonation state changes due to plasmepsin II, cathepsin D and endothiapepsin binding to pepstatin are predicted to within 0.4 proton units at pH 6.5 and 7.0, respectively. The PROPKA 2.0 results indicate that structural changes due to ligand binding contribute significantly to the proton uptake/release, as do residues far away from the binding site, primarily due to the change in the local environment of a particular residue and hence the change in the local hydrogen bonding network. Overall the results suggest that PROPKA 2.0 provides a good description of the protein–ligand interactions that have an important effect on the pK_a values of titratable groups, thereby permitting fast and accurate determination of the protonation states of key residues and ligand functional groups within the binding or active site of a protein.

INTRODUCTION

The prediction of pK_a values of protein–ligand complexes is of great practical importance to molecular modeling in rational drug design because the strength of ligand binding (i.e. the binding free energy) is dependent on the protonation states of the ionizable residues and functional groups in the active site. Moreover, upon ligand binding, the pK_a values of the ionizable groups may change, resulting in the uptake or release of protons. This leads to a pH-dependent correction to the binding free energy that can be computed using the pK_a values. Additionally, studies of enzyme reaction mechanisms will benefit from the correct assignment of the protonation states of the ionizable groups in the active site and those of the substrate.

There are several ways to compute pK_a values for proteins,[1–24] including utilizing free energies from molecular dynamics simulations,[1,2,19,21,23,24] and from numerical solutions of the linearized Poisson-Boltzmann equation (LPBE).[3–9,11,13] In the LPBE approaches the protein is described by the partial charges from a classical force field and embedded within a dielectric continuum with dielectric constants (ε) of 80 for the solvent and between 4 and 20 for the interior of the protein. pK_a shifts are obtained by computing the difference in electrostatic energy of a charged residue with that of its neutral form; the resulting shift is added to a model pK_a value. Typically these models[3–7] have root-mean-square errors with respect to experiment of less than 1 pH unit.

Difficulties arise when extending PB approaches to incorporate a ligand bound to a protein. In addition to pK_{model} values for the ionizable groups in the ligand, charges and other parameters that describe the ligand are required. Consequently, only a small number of computational studies on protein–ligand complexes have been performed.

Trylska *et al.*[25] have employed a linearised PB model to compute apparent pK_a values for HIV-1 protease complexed with four different inhibitors and also of HIV-1 protease complexed with a series of cyclic ureas.[26] The authors found that pK_a shifts calculated for the buried Asp dyad using their methodology were sensitive to the magnitude of the dielectric constant employed to describe the protein. Charges and atomic radii were taken from the polar-hydrogen CHARMM[27] and OPLS[28] parameter sets, respectively, for the single-site/$\varepsilon = 20$ calculations and from the PARSE[29] parameter set for the full-charge/$\varepsilon = 4$ calculations. The initial pK_a of the secondary amine in the ligand MVT-101 was estimated to be 10.0. Alexov[30] has studied the peptide pepstatin complexed with the proteases plasmepsin, cathepsin D and endothiapepsin using a modified PB technique, that takes into account conformational changes of side chains and polar protons following the Boltzmann distribution of the ionizable states. For the three complexes with pepstatin, in addition to proton uptake and release by residues, conformational changes due to binding were also examined along with the overall effect on the binding free energies. The semiempirical quantum chemistry program MOPAC[31] was employed to calculate the partial charges for pepstatin, which was assumed to have a net charge of -1 in its unbound form. Czodrowski *et al.*[32] have extended a LPBE method to predict pK_a shifts for two thrombin–ligand and one dihydrofolate reductase–ligand complex. Charges describing the proteins and the ligand functional groups were obtained by the PEOE (partial equalization of orbital electronegativity) PB procedure and model ligand pK_a values were taken from experiment. For the two thrombin complexes studied, their method was able to successfully differentiate (based on the overall protonation change) between the complex that captures protons from the buffer and the complex, which has no significant pK_a shifts and hence negligible protonation state changes. Their calculations on the dihydrofolate reductase complex with methotrexate yielded pK_a shifts for an Asp residue (Asp27) and a ligand aromatic nitrogen that are in accord with experiment, where a salt bridge is formed between these two neighboring groups, i.e. the Asp residue loses a proton and the nitrogen gains a proton. Dihydrofolate reductase complexed with methotrexate was also studied in the earlier work of Cannon *et al.*,[33] who employed different charge sets for the protein and for the ligand (OPLS and quantum chemically derived charges, respectively), but whose predictions did not agree with experiment. In more recent studies, Czodrowski *et al.* have applied their PEOE_PB approach to ligands bound to trypsin and thrombin[34] and to HIV-1 protease.[35]

Brooks and coworkers have studied the change of the pK_a value of an aromatic nitrogen atom in dihydrofolate (as the moiety undergoes reduction to tetrahydrofolate) upon binding to dihydrofolate reductase using free energy perturbation with molecular dynamics simulations.[36,37] Finally, Warshel and co-workers routinely include ligands (such as transition states or ions) in PSLD/S-LRA or FEP AC charging calculations.[38,39]

PROPKA[40] has been proven as a rapid and accurate empirically based method for predicting protein pK_a values in proteins[41] and protein-protein complexes.[42] We extend the PROPKA method to the calculation of pK_a values for ionizable residues and ligand functional groups in protein–ligand complexes, where the main challenge is the atom typing employed to describe the ligand functional groups.

The paper is organized as follows: First, we describe our extensions of the PROPKA method. Secondly, we validate the methodology by computing pK_a values for 26 protein–ligand complexes whose structures are deposited in the Protein Data Bank (PDB)[43] and where corresponding pK_a or related data is available in the literature. We focus on the prediction of pK_a values and changes in protonation state using static geometries, i.e. pK_a shifts and the uptake and release of protons. Finally, we summarize and discuss future directions.

COMPUTATIONAL METHODOLOGY

PROPKA 1.0

Before describing the new functionalities implemented in PROPKA in the next subsection, we first summarize the version of PROPKA (PROPKA 1.0) that is currently generally available.

PROPKA determines the pK_a of an ionizable group through the application of an environmental perturbation, ΔpK_a, to the unperturbed pK_a value of the group, pK_{Model}

$$pK_a = pK_{\text{Model}} + \Delta pK_a \tag{1}$$

where pK_{Model} and ΔpK_a are determined empirically. For each ionizable group, ΔpK_a comprises five terms that describe global (GlobalDes) and local desolvation (LocalDes), hydrogen bonds with side-chain groups (SDC-HB), hydrogen bonds with amide backbone (BKB-HB), and interactions with charged groups (ChgChg)

$$\Delta pK_a = \Delta pK_{\text{GlobalDes}} + \Delta pK_{\text{LocalDes}} + \Delta pK_{\text{SDC}-\text{HB}} \\ + \Delta pK_{\text{BKB}-\text{HB}} + \Delta pK_{\text{ChgChg}} \tag{2}$$

Simple distance functions and, for backbone hydrogen bonds, distance/angle functions with a constant empirical pK_a shift are employed to compute the above ΔpK_a terms. Each ionizable group interacts with other ionizable and nonionizable groups. pK_a values are computed for each ionizable group and in a final step are updated in an iterative procedure that takes into account changes in

relative pK_a values between interacting pairs of ionizable groups and hence any charge–charge interactions if applicable. This scheme neglects the effect of nonsigmoidal titration curves on predicted pK_a values. However, this effect is generally smaller than about the 1.0 pH unit error typical of most pK_a prediction methods.[42]

A description of the reasoning behind the construction of the form of the five terms mentioned above are described in detail for amino acid residues by Li *et al.*[40] and are summarized below. The choice of parameters and their physical meaning have been discussed in Ref. 40.

Hydrogen bonding

There are two terms that involve hydrogen bonding; hydrogen bonding with side chains and hydrogen bonding with amide backbone. For side-chain hydrogen bonds we employ a simple distance function to compute the pK_a shift

$$\Delta pK_{SDC-HB} = \begin{cases} C_{HB} & \text{if } D \leq d_1 \\ C_{HB} \cdot \dfrac{D - d_2}{d_1 - d_2} & \text{if } d_1 < D < d_2 \quad (3) \\ 0 & \text{if } d_2 \leq D \end{cases}$$

Shifts due to backbone hydrogen bonds are described by a distance/angle function

$$\Delta pK_{BKB-HB}$$
$$= \begin{cases} -\cos\theta \cdot C_{HB} & \text{if } D \leq d_1, \theta > 90° \\ -\cos\theta \cdot C_{HB} \cdot \dfrac{D - d_2}{d_1 - d_2} & \text{if } d_1 < D < d_2, \theta > 90° \\ 0 & \text{if } d_2 \leq D, \theta \leq 90° \end{cases}$$
$$(4)$$

C_{HB} is the constant pK_a shift, D is defined as the distance between atoms in the hydrogen bond, i.e. D is the distance between carboxyl oxygen atoms and the donor protons of Asn, Gln, Trp, His, Lys, and Arg side-chain groups for SDC-HB between these groups and, for BKB-HB, carboxyl oxygen atoms, and amide hydrogens, for example. For other SDC-HB D is the distance between the carboxyl oxygen atoms and the heavy atoms O, S, and N. The parameters d_1 and d_2 are, respectively, the optimum distance for hydrogen bonding (where ΔpK_{HB} is at a maximum) and the distance where the hydrogen bonding strength is effectively 0. For an amide BKB-HB with Asp and Glu, θ is defined as the larger of the angles between N—H—O atoms and as N—H—X (X = S or O) for amide hydrogen bonds with Cys and Tyr residues; for a carbonyl BKB-HB to His, θ is defined as the larger of the angles between C—O—H. When interacting groups are buried within the protein, their hydrogen bonding is considered stronger. These interactions are known as strong hydrogen bonds (STR-HB).

Desolvation

Desolvation contributions can have a marked effect on pK_a shifts for ionizable groups within the protein interior. For conjugate base groups, for example carboxylate, desolvation increases the energy of the negatively charged group and therefore raises the pK_a value; for positively charged (i.e. conjugate acid) groups, such as the side chain of His, desolvation increases their energy, therefore lowering their pK_a values. Desolvation is considered to have two contributions, a local and a global effect

$$\Delta pK_{Des} = \Delta pK_{GlobalDes} + \Delta pK_{LocalDes} \quad (5)$$

with

$$\Delta pK_{LocalDes} = N_{Local} \cdot C_{Local} \quad (6)$$

where N_{Local} is equal to the number of nonhydrogen atoms within a sphere of radius R_{Local} centered at the ionizable group and

$$\Delta pK_{GlobasDes} = (N_{15.5Å} - 400) \cdot C_{Global} \quad (7)$$

where $N_{15.5Å}$ is the number of nonhydrogen atoms within 15.5 Å of the centre of an ionizable group.

A similar principle is used to define a buried pair of ionizable groups, which is used to determine strong hydrogen bonding (STR-HB, described in the previous section) and charge–charge interactions described in the following section. For two neighboring groups, 1 and 2, a buried pair is defined such that

$$N_{15.5Å}(1) \geq 400 \text{ and } N_{15.5Å}(2) \geq 400 \quad (8)$$

or

$$[N_{15.5Å}(1) + N_{15.5Å}(2)] \geq 900 \quad (9)$$

Equation (9) incorporates the groups that are not buried ($N_{15.5Å}$ = 350, for example) but have a neighboring group that is deeply buried ($N_{15.5Å} \geq 550$, for example).
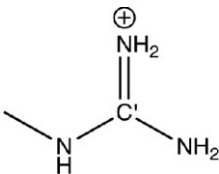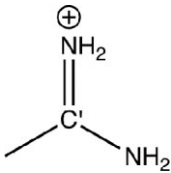
Charge–charge interactions
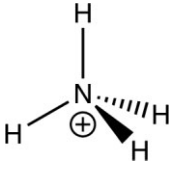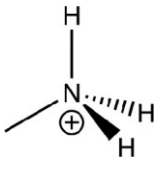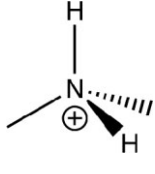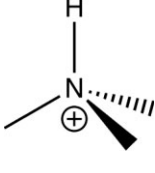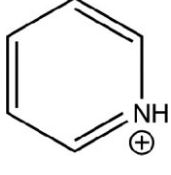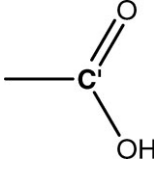
Charge–charge interactions are described by terms similar to the SDC-HB term

$$\Delta pK_{ChgChg} = \begin{cases} \pm C_{ChgChg} & \text{if } D \leq d_1 \\ \pm C_{ChgChg} \cdot \dfrac{D - d_2}{d_1 - d_2} & \text{if } d_1 < D < d_2 \\ 0 & \text{if } d_2 \leq D \end{cases}$$
$$(10)$$

We consider charge–charge interactions between buried pairs of ionizable residues, which are computed in the final, iterative part of PROPKA. The pK_a shift $\pm C_{ChgChg}$ has a value of 2.40 for all pair wise charge–charge interactions, where the residue with the lower relative pK_a will be shifted by -2.40, with the residue with the higher rel-

Table I
(A) Ionizable Ligand Groups and (B) Nonionizable Ligand Groups Recognized by PROPKA 2.0

Ligand group	SYBYL atom type	PROPKA		Structure	pK_{Model}
		Input	Internal		
(A)					
Guadininium group	N.pl3	Npl	Cg		11.5
Amidinium group	N.pl3	Npl	C2N		11.5
Ammonium	N.3/N.4	N3/N4	N30		10.0
sp^3 primary nitrogen	N.3	N3	N31		10.0
sp^3 secondary nitrogen	N.3	N3	N32		10.0
sp^3 tertiary nitrogen	N.3	N3	N33		10.0
Aromatic nitrogen	N.ar	Nar	Nar		5.0
Oxygen in carboxylate group	O.co2	Oco	Oco		4.5

(Continued)

Table I
(Continued)

Ligand group	SYBYL atom type	PROPKA		Structure
		Input	Internal	
(B) NH$_2$ group	N.pl3	Npl	Np1	
sp^3 oxygen	O.3	O3	Oh	
sp^3 oxygen	O.3	O3	O3 not identified as Oh	
Chlorine atom	Cl	Cl	Cl	
Fluorine atom	F	F	F	
Amide nitrogen	N.am	Nam	Nam	
sp nitrogen	N.1	N1	N1	
sp^2 oxygen	O.2	O2	O2	
Singly charged atom	n/a	1P/1N	1P/1N	$X^{1\pm}$
Doubly charged atom	n/a	2P/2N	2P/2N	$X^{2\pm}$

C′ identifies the guanidinium, amidinium, and carboxylate ligand groups.

ative pK_a shifted by 2.40. The distances d_1 and d_2 are 4.0 and 7.0 Å, respectively.

PROPKA 2.0

The two main obstacles that must be overcome in including the effect of ligands in PROPKA are (A) generating the parameters necessary to evaluate the pK_a shifts summarized in Eqs. (2) and (B) identifying the corresponding functional groups in the ligand. There are not enough experimentally determined pK_a values of protein–ligand complexes to parameterize the method. Instead we have chosen the ligand parameters to be as similar as possible to the corresponding parameters in PROPKA 1.0 whenever possible, as described next.

Table IA displays the ligand groups considered ionizable. In order to identify all the groups of a ligand (in the current implementation of the code), we employ OPEN BABEL[44,45] to generate SYBYL MOL2[46] atom-types for the ligand groups listed in Table I, part A. The ionizable groups are carboxylate, basic aromatic nitrogen (i.e. pyridine N), sp^3 hybridized nitrogen (i.e. primary, secondary and tertiary amines) and trigonal planar nitrogen, which includes the guanidinium group (R—NHCNH$_2$NH$_2^+$, c.f. Arg side chain) and the amidinium group (R—CNH$_2$NH$_2^+$). An sp^2 hybridized carbon is employed to identify the carboxylate, guanidinium, and amidino groups in both the workings of the code and in the PROPKA 2.0 output. Nonionizable ligand groups, which interact with the ionizable residues, are depicted in Table I, part B.

Addition of hydrogen atoms

Similar algorithms to those used to determine the positions of N—H protons of backbone amides, Asn, Gln, Trp, His, and Arg residues that may be involved in hydrogen bonding have been implemented to determine the positions of the N—H protons for amide, aromatic, trigonal planar and sp^3 nitrogen as well as the position of a ghost atom for the N1 atom-type in order to include the orientation effect in the hydrogen bonding term.

In these algorithms the H—N bond length is always 1.0 Å. Two atoms are considered linked via a bond if the distance between them is less than 2.0 Å.
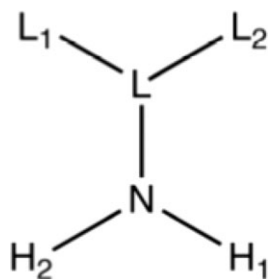
Scheme 1

Addition of hydrogen atoms for trigonal planar nitrogen.

- For an amide group, the proton position is assigned so that the vector H-N is parallel with the C—O bond vector.
- For aromatic nitrogen, the two closest aromatic carbon atoms are identified (C1 and C2), the coordinates of the middle M between C1 and C2 is calculated and the proton position is assigned so that the vector Nar-H is parallel to the vector M-Nar.
- For trigonal planar nitrogen the atom linked to the nitrogen and its two neighbors are identified as shown in Scheme 1 and represented by L, L_1, and L_2.

The position of the proton is assigned so that the N—H_1 vector is parallel to the vector L_1-L and the N-H_2 vector is parallel to the vector L_2-L.

According to this, the guadininium type group and the amidino group are identified using the corresponding carbon atom.

- For sp^3 nitrogen, the number of protons assigned will depend on the number of non-hydrogen atoms linked to the nitrogen:

If 0 or 1 atoms are linked to the nitrogen atom; we do not assign any proton due to the spherical symmetry of the group.

If 2 atoms are linked to the nitrogen atom; a new set of mutually orthogonal axes {x, y, z} is defined as shown in Scheme 2 and the coordinates of the proton are expressed in this new coordinate system using the spherical coordinates. Two protons are assigned, as represented by H_1 and H_2 (where the angle θ(H_1—N—H_2) is 125°). During the pK_a calculation only the closest one to the considered residue will be used.

If 3 atoms are linked to the nitrogen atom; the position of the centre of mass G of the three nonhydrogen atoms is calculated and the proton position is assigned so that the vector H-N is parallel to the vector G-N.

- For sp^1 nitrogen, the position of the atom L linked to the sp^1 nitrogen is used to calculate the L-N vector

and the position of the ghost atom G is assigned so that the N-G vector is parallel to the L-N vector.

Hydrogen bonding

So far PROPKA 2.0 can treat the 18 different functional groups listed in Table I. Sixteen of these groups can serve as either a hydrogen bond donor or acceptor or both (the two exceptions are "charged atom" in Table I, part B to be described later). Eight of the 16 groups capable of hydrogen bonding have obvious equivalents among the amino acids: guanidinium and amidino (Arg), ammonia and primary amine (Lys), carboxylate (Asp or Glu), planar NH_2 group (Asn or Gln), OH (Ser or Thr), and the amide group (protein backbone). For these cases the hydrogen bonding parameters used in Eqs. (3) and (4) are exactly the same as the corresponding amino acids.

The remaining functional groups fall into two categories: the remaining amines (secondary, tertiary, and aromatic amines) and hydrogen bond acceptors (ether, Cl, F, nitrile, and carbonyl). The amines differ from the primary amines in that an angle dependence is introduced for the hydrogen bond interactions since the hydrogen positions are conformationally immobile compared to primary amines. The hydrogen bond acceptors give rise to maximum pK_a shifts that are twice as large as the usual value of 0.8. For example, for an aromatic amine the maximum pK_a shift due to a hydrogen bond with a carbonyl group (C_{HB} = 1.6) is larger than the corresponding shift due to a OH group (C_{HB} = 0.8) because in the former case there is a repulsive interaction in the deprotonated case (right structure in Scheme 3), but not in the latter case since the OH can function as a hydro-
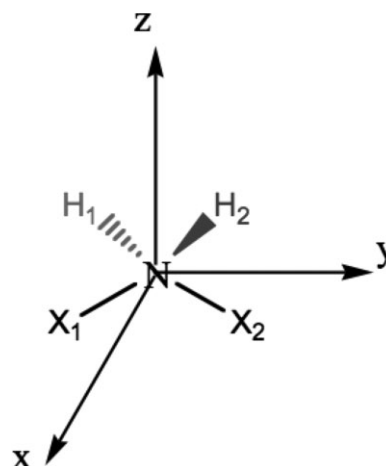
Scheme 2

Addition of hydrogen atoms for sp^3 nitrogen.
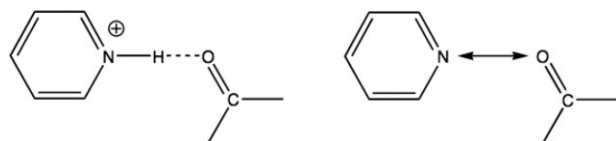
Scheme 3

Hydrogen bonding between an aromatic amine and a carbonyl group.

gen donor. Also, since the position of the amine proton is well defined, an angle-dependence is introduced in analogy with the treatment of the amide proton in PROPKA 1.0.

Table SI in the supplementary material defines the type of hydrogen bonding, the pK_a shift C_{HB}, d_1, d_2 and D for interactions between residues and ligand groups.

Desolvation

The desolvation effect of the ligand atoms on the protein residues are taken into account using Eqs. (5)–(7). For protein residues, in addition to protein atoms, atoms from the ligand are included in the summation over atoms within the spatial coordinates. For ligand groups, the summation is exclusively over protein atoms.

Charge–charge interactions

We consider charge-charge interactions between buried pairs of ionizable residues and groups and between buried ionizable residues and groups with buried charged atoms such as S^{1-}. For the latter case the interactions are computed in the noniterative part of the code; for the former the interactions are computed in the final iterative part. The interactions are computed using Eq. (10). There are four possible interactions between charged residues (Asp/Glu, His, Cys, Tyr, Lys and Arg) and ligand groups (carboxylate, aromatic N, sp^3 hybridized N, and trigonal planar N): negative with negative, positive with positive, negative with positive, and positive with negative. For the negative with negative and positive with positive interactions, the pK_a shift is applied to the group with the higher and the lower temporary pK_a, respectively. For the negative with positive and positive with negative interactions, the pK_a shift is applied to both groups (provided that the temporary pK_a of the negative group is lower than that of the positive group), where the group with the lower temporary pK_a will experience a decrease in pK_a and where the group with the higher temporary pK_a will experience an increase in pK_a. C_{ChgChg} takes a generic value of 2.4 and d_1 and d_2 are 4.0 and 7.0 Å, respectively, which are analogous to PROPKA 1.0. Supplementary material Table SII displays the charge centers for the ligand groups used to compute D.

Covalently bound ligands

For ligands covalently bound to Asp or Glu, and to Ser and Cys residues (via a carbon atom with one of the carboxylate oxygens, or O_γ oxygen or S_γ sulphur atoms, respectively), we ignore the side chain of the residue to which the ligand is bound during the calculation, while the residue's amide backbone contributes to the pK_a shift. A ligand is considered covalently bound via one of its carbon atoms to an Asp or Glu residue if the C—$O_{OD1/OD2/OE1/OE2}$ distance is less than 1.9 Å ($R_{COD1/OD2/OE1/OE2} \leq 1.9$ Å); to a Ser residue if the C—O_γ distance is less than 1.9 Å ($R_{CO\gamma} \leq 1.9$ Å); and to a Cys residue if the C—S_γ distance is less than 2.1 Å ($R_{CS\gamma} \leq 2.1$ Å). The C—O and the C—S distances correspond to the bond lengths obtained using Bondi radii[47] scaled by 1.2.

Model pK_a values

Model or input pK_a values for the ionizable ligand groups are, when available, taken from the literature; when they are unavailable we employ the pK_a plug-in featured in MARVIN[48] to predict initial model values. The default pK_{Model} values are listed for each of the ionizable groups in Table I, part A, and were derived from MARVIN pK_a calculations (for carboxylate, basic aromatic nitrogen and sp^3 hybridized nitrogen) and from empirical values (for trigonal planar nitrogen).

Main steps in a PROPKA 2.0 calculation

Figure 1 outlines the principal steps in generating a PROPKA 2.0 input file for a protein–ligand complex (the steps of modification of the PDB file are skipped and the PDB file is used as the input file). In the first step the PDB file (containing atomic coordinates for the protein, ligand(s), and other moieties such as cofactors) is processed, thereby yielding a PDB′ file with the relevant peptide chains, side-chain conformers and heteroatoms. OPEN BABEL is then employed to generate SYBYL MOL2 atom-types for the ligand(s) that are appended to (the end of) a temporary file containing the peptide chain(s) from the previous PDB′ file. Based on these atom-types, pK_{Model} values are added to the file to give a PROPKA 2.0 input file. PROPKA 2.0 is executed using this file as input, to generate PROPKA 2.0 output. The entire process is automated (via a Perl script). The steps enclosed by the box are the subject of this paper. The remaining steps are, with the exception of the PROPKA atom-types, independent of the pK_a prediction. Thus, other pK_{Model} assignment algorithms can be interfaced with PROPKA or modified manually in the input file by the user. If experimental pK_{Model} values are to be used, these are simply entered in the PROPKA 2.0 input file and PROPKA 2.0 is rerun. Alternatively, other pK_a prediction programs or databases can be used to generate
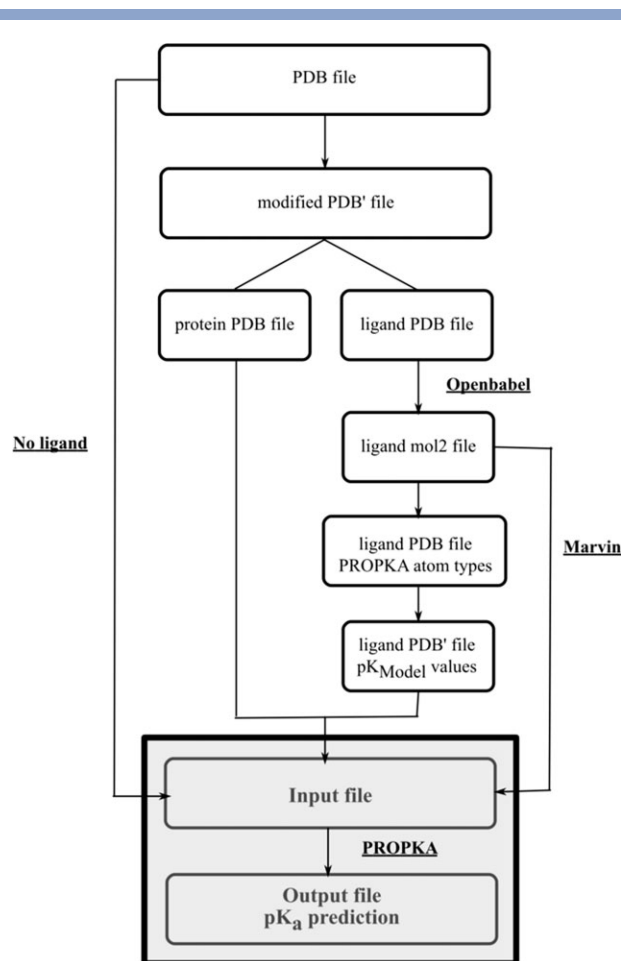
Figure 1
Chart outlining the main steps in a PROPKA 2.0 calculation. See text for more detail.

pK_{Model} values. Here we use the MARVIN web interface[48] to generate pK_{Model} values by uploading the ligand structure (in MOL2 format).

PROTEIN STRUCTURES

Proton uptake and release

Only a few pK_a values of ionizable groups in protein–ligand complexes have been measured by NMR. To further validate our predicted pK_a values we compute the protein uptake or release upon ligand binding, which has been measured for several protein–ligand complexes by isothermal titration calorimetry.

The predicted ligand-induced pK_a changes are compared to two types of experimental data. pK_a values for the free protein and ligand and for the protein–ligand complex allow the overall protonation change, n, upon

binding to be computed as a function of pH using the (micro) change for each ionizable residue i

$$n = \sum_{i}^{\text{ionizable residues}} n_i$$

$$= \sum_{i}^{\text{ionizable residues}} \left(\frac{10^{pK_{a,i}^c - pH}}{1 + 10^{pK_{a,i}^c - pH}} - \frac{10^{pK_{a,i}^f - pH}}{1 + 10^{pK_{a,i}^f - pH}} \right) \quad (11)$$

where $pK_{a,i}^c$ and $pK_{a,i}^f$ refer to pK_a values of the ionizable groups in the protein–ligand complex and in the free protein or ligand, respectively. The first and second terms enclosed by the brackets are the degree of protonation for a particular group i at a given pH in the complex and in the free protein or ligand, respectively.

Supplementary material Table SIII displays the proteins and ligands studied and their associated PDB codes and also the peptide chains and alternate side-chain conformers used for the calculations. The ligands are depicted in Figure S1 in the supplementary materials. All solvent and sulphate ions were excluded from the calculations on the protein–ligand complexes. Uncomplexed or free protein calculations employed the same X-ray coordinates as those with ligands present, i.e. all heteroatoms were excluded from the free protein calculations. For the HIV-1 protease with DMP-323 complex (PDB code 1QBS), S-hydroxycysteine (Cso) residues were mutated to cysteine for the pK_a calculations. In addition, α-aminobutyric acid (Aba) residues were mutated to cysteine residues in the apo HIV-1 protease X-ray structure 3HVP. Apo crystal structures were employed to describe the unliganded pepsin proteins and HIV-1 protease. For any amino acids absent in the apo crystal structures but present in the liganded crystal structures (and vice versa), the associated pK_a values were removed from the determination of the net proton uptake/release [cf. Eq. (11)].

RESULTS AND DISCUSSION

We have applied PROPKA 2.0 to the study of 26 protein–ligand complexes with available PDB X-ray crystal structures and experimental data for validation. The data is either a net change in protonation state upon ligand binding [cf. Eq. (11)] or pK_a values of ligand–protein complexes measured by NMR. Results are displayed in Tables II and III where calculated overall and micro changes in protonation state at the appropriate pH are shown along with complexed and uncomplexed pK_a values.

Trypsin and thrombin

The first protein we consider is the serine protease trypsin complexed with a set of closely related inhibitors derived from N^α-(2-naphthylsulphonyl)-L-3-amidino-

Table II
Experimental and Calculated Changes in Protonation State for the Trypsin and Thrombin Complexes Studied

Ligand (PDB)	$n_{i(pK_a)}$ complexed					pK_a uncomplexed		
	n_{exp}	n_{calc}	His57	ligCOO	ligAMINO	His57	ligCOO	ligAMINO
Trypsin								
1b (1K1I[49])	0.90	0.09	0.24 (7.58)	0.0 (1.95)	n/a	6.99	3.21	n/a
1bMe[a]	0	0.06	−0.12 (6.32)	n/a	n/a	7.06	n/a	n/a
1c	0	0.15	0.16 (7.38)	0.0 (2.81)	n/a	6.93	4.17	n/a
1cMe (1K1J[49])	0	−0.24	−0.10 (5.84)	n/a	n/a	6.91	n/a	n/a
1d (1K1L[49])	−0.53	−0.55	−0.08 (5.94)	n/a	−0.33 (4.92)	6.80	n/a	7.49
1dAc (1K1M[49])	0	−0.23	−0.09 (5.86)	n/a	n/a	6.85	n/a	n/a
2	0.93	0.69	0.88 (9.51)	0.0 (0.17)	n/a	6.87	3.40	n/a
3 (1K1N[49])	0	−0.21	−0.08 (5.72)	0.0 (3.51)	n/a	6.77	3.84	n/a
4 (1K1O[49])	0	0.03	−0.11 (6.09)	0.0 (3.26)	0.24 (7.91)	6.96	2.65	7.48
5 (1K1P[49])	0	0.07	−0.09 (5.98)	0.0 (2.79)	0.27 (8.58)	6.87	2.51	7.95
Thrombin								
2	0.88	0.34	0.33 (7.50)	0.0 (2.27)	n/a	4.97	3.40	n/a
3 (1YPK[50])	0	−0.01	0.0 (3.59)	0.0 (3.09)	n/a	4.62	3.84	n/a
4 (1K21[49])	0	−0.07	0.0 (4.07)	0.0 (2.74)	0.09 (7.65)	5.08	2.65	7.48
5 (1K22[49])	0	0.04	0.0 (3.96)	0.0 (2.83)	0.0 (7.95)	4.92	2.51	7.95

Calculated micro protonation changes and pK_a values are also shown. Intrinsic ligand pK_a values were taken from the literature. For predictions of proton uptake and release, pH is 7.8.
[a]ChgChg interactions between His57 with Cys25 and Cys41 removed.

phenylalanine for which Klebe and co-workers[34,49] have measured changes in protonation state and computed pK_a changes using their PEOE_PB scheme.

To illustrate the use of PROPKA 2.0, we consider trypsin complexed with ligand 1b (Supplementary Fig. S1) and detail the determinants of the pK_a value predicted for His57. The key protein residues along with the ligand are depicted in Figure 2. His57 has 591 and 37 non-hydrogen atoms within 15.5 Å and 4.5 Å, respectively, of its centre, and is therefore defined as "buried," with $\Delta pK_{GlobalDes} = -1.91$ and $\Delta pK_{LocalDes} = -2.59$ [cf. Eqs. (7) and (6)]. His57 forms a hydrogen bond with the side chain of Asp102, which is also "buried," therefore giving rise to a strong hydrogen bond, $\Delta pK_{STR-HB} = 1.60$. His57 has charge-charge interactions with Asp102 and the carboxylate group of the ligand, both of which are predicted to be deprotonated when His57 titrates, yielding $\Delta pK_{ChgChg} = 2.40 + 1.58$, respectively. Summation of these ΔpK_a terms shifts the pK_a for His57 from the intrinsic value of 6.50 to 7.58 in the environment of the protein–ligand complex. When the ligand is removed the number of non-hydrogen atoms within 15.5 and 4.5 Å drops by 36 and 9 atoms, respectively, which decreases $\Delta pK_{GlobalDes}$ and $\Delta pK_{LocalDes}$ to −1.55 and 1.96, respectively. The charge-charge interaction with the carboxyl group of the ligand is removed, while the interaction with Asp102 remains unchanged. Summation of these changes in ΔpK_a shifts the pK_a for His57 from the complexed value of 7.58 to 6.99 when the ligand is removed. Equation (11) is then used to compute the net uptake/release of protons upon ligand binding, which, for this example, is a net gain of 0.09 protons at pH 7.8. The His57 residue is predicted to uptake 0.24 protons, which

is partially cancelled by a release of 0.14 protons by the N-terminus. The experimentally determined proton uptake is 0.9 and considerably larger than the predicted value. As discussed by Czodrowski et al.,[34,35] net protonation changes are extremely sensitive to errors in pK_a values that are close to the pH of interest, as in the case here. For example a relatively small underestimation of the pK_a value (i.e. 7.58 rather than 8.08) leads to a considerable error in the prediction of the net proton uptake/release (0.24 rather than 0.52) at pH 7.8.

Table II summarizes the experimental and computed data (using PROPKA 2.0) including pK_a values for key protein residues and ionizable functional groups on the ligand. His57 plays a key role in the active site and its protonation state may change upon binding to a ligand. Upon binding of the ligands 1b, 1d and 2 (napsagatran) the overall protonation state changes. PROPKA predicts the correct protonation and deprotonation upon binding with 1d and 2 at a pH of 7.8. The experimentally observed proton uptake for complex 1b was not reproduced for the reasons discussed above. Following Czodrowski et al., net proton uptake/release values of less than ~0.3 are taken to be negligible.

For the trypsin:1d complex, His57 and the ligand's amino group are predicted to release 0.1 and 0.3 protons, respectively, at pH 7.8. For this complex PROPKA predicts pK_a values of 4.92 and 6.25 for the ligand's piperazine amino nitrogen group when employing experimental[49] (7.49) and MARVIN (8.82) pK_a values for the uncomplexed ligand, respectively. These values are lower than the model input values and the experimental upper limit, a pK_a of 7.49, implying that this group loses protons upon binding and is predicted by PROPKA to be

Table III

Experimental (Where Available) and Calculated Changes in Protonation State for the Remaining Protein-Ligand Complexes Studied

Protein (PDB)	n_{exp} (n_{PROPKA})	pK_a Exp (PROPKA)			
		Asp34/Asp214	Asp34/Asp214	His164	His318
Plasmepsin II (1PFZ[51])		4.7 (4.92)	4.7 (3.98)	6.0 (6.50)	(6.36)
+ pepstatin (1SME[52])	[a]1.7 (1.50)	3.0 (9.51)	6.5 (3.60)	7.5 (4.34)	(7.09)
		Asp33/Asp231	His77	Glu260	Asp323
Cathepsin D (1LYW[53])		<6.5 (0.63/6.17)	<6.5 (7.46)	<6.5 (4.15)	(4.97)
+ pepstatin (1LYB[54])	[a]2.9 (2.49)	>6.5 (3.71/9.54)	>6.5 (6.50)	>6.5 (5.31)	(7.35)
		Asp32/Asp215	Asp12	Asp30	
Endiothiapepsin (4APE[55])		(10.11/3.57)	(5.55)	(6.69)	
+ pepstatin (4ER2[55])	1.06 (0.77)	>7.0 (9.91/3.46)	(9.26)	(7.21)	
		Asp25/Asp225	Asp30/Asp230	His69/His269	
HIV-1 protease (5HVP[56])		(4.14/9.98)	(5.64/5.18)	(8.24/6.43)	
+ pepstatin (5HVP)	NMR	<2.5/>6.5 (4.04/10.20)	(5.92/6.44)	(7.34/6.43)	
HIV-1 protease (1HHP[57])		(8.68/3.56)	(4.40/4.40)	(8.32/8.32)	
+ pepstatin (5HVP)	NMR	<2.5/>6.5 (4.04/10.20)	(5.92/6.44)	(7.34/6.43)	
HIV-1 protease (3HVP[58])		(8.15/2.87)	(4.38/4.38)	(7.89/7.89)	
+ pepstatin (5HVP)	NMR	<2.5/>6.5 (4.04/10.20)	(5.92/6.44)	(7.34/6.43)	
		Asp25/Asp25′	Asp30/Asp30′	KNI-272 N1	
HIV-1 protease (1HPX[59])		6.0 (3.77/9.26)	(5.18/5.32)	4.8 (4.8[c])	
+ KNI-272 (1HPX)	[b]-0.23 (0.18)	6.6 (10.28/3.97)	3.88/3.78 (5.45/5.73)	2.9 (4.52)	
HIV-1 protease (1HHP)		6.0 (8.68/3.56)	(4.40/4.40)	4.8 (4.8[c])	
+ KNI-272 (1HPX)	[b]-0.23 (1.27)	6.6 (10.28/3.97)	3.88/3.78 (5.45/5.73)	2.9 (4.52)	
HIV-1 protease (3HVP)		6.0 (8.15/2.87)	(4.38/4.38)	4.8 (4.8[c])	
+ KNI-272 (1HPX)	[b]-0.23 (0.97)	6.6 (10.28/3.97)	3.88/3.78 (5.45/5.73)	2.9 (4.52)	
		Asp25/125	Asp30/Asp130		
HIV-1 protease (1QBS[60])		6.0 (3.79/9.07)	(4.85/5.20)		
+ DMP-323 (1QBS)	NMR	8.19 (3.26/8.56)	3.99 (5.17/5.49)		
HIV-1 protease (1HHP)		6.0 (8.68/3.56)	(4.40/4.40)		
+ DMP-323 (1QBS)	NMR	8.19 (3.26/8.56)	3.99 (5.17/5.49)		
HIV-1 protease (3HVP)		6.0 (8.15/2.87)	(4.38/4.38)		
+ DMP-323 (1QBS)	NMR	8.19 (3.26/8.56)	3.99 (5.17/5.49)		
		His57			
Chymotrypsin (7GCH[61])		7.5 (6.94)			
+ N-Acetyl-L-Leu-DL-Phe-CF$_3$ (7GCH)	NMR	12.0 (8.90)			
		His57			
Chymotrypsin (6GCH[61])		7.5 (6.56)			
+ N-Acetyl-DL-Phe-CF$_3$ (6GCH)	NMR	10.8 (10.33)			
		Glu172			
Xylanase (1BVV[62])		6.7 (7.45)			
+ 2FXb (1BVV)	NMR	4.2 (6.64)			
		Asp35/Glu172			
Xylanase N35D (1C5I[63])		3.7/8.4 (4.15/12.16)			
+ 2FXb (1C5I)	NMR	1.9–3.4 or > 9.0 (4.40/12.35)			
		His235			
Hydroxynitrile lyase(2YAS[64])		2.5 (2.92)			
+ thiocyanate (2YAS)	NMR	8.0 (3.76)			
		Asp27	methotrexate N1		
DHFR (4DFR[65])		6.6 (4.98)	5.7 (5.7[c])		
+ methotrexate (4DFR)	NMR	(0.86)	10.7 (6.89)		
DHFR (4DFR)		6.6 (4.98)	5.7 (4.40)		
+ methotrexate (4DFR)	NMR	(3.26)	10.7 (3.19)		

Calculated pK_a values for selected residues and groups are also shown. Intrinsic ligand pK_a values were taken from MARVIN or (where available) experiment. For predictions of proton uptake and release. pH is 7.0 unless stated.
[a]pH = 6.5.
[b]pH = 5.0.
[c]Experimental ligand pK_a value.

the major contributor to the release of protons. The large negative pK_a shift for 1d's piperazine amino nitrogen is due to a charge-charge interaction with His57 (−2.25 pH units), which serves to deprotonate the nitrogen and hence remove an unfavorable charge-charge interaction with a protonated His57. The ligand rather than His57 is predicted to lose the proton, because it has a lower pK_a value in the absence of this interaction. The negative pK_a shift experienced by His57 in the trypsin:1d complex is largely due to the absence of a charge–charge interaction with the ligand's carboxyl group, as observed for trypsin:1b, consequently the negative desolvation contribu-
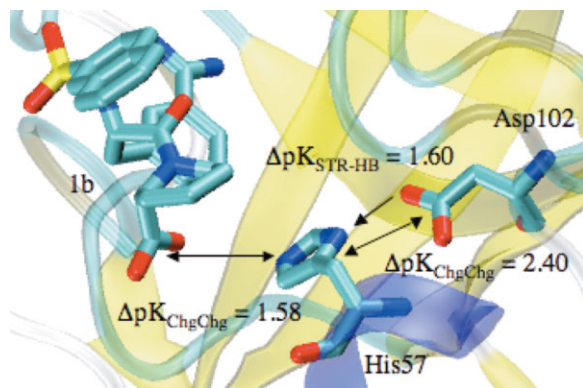
Figure 2
The determinants of the pK_a value of His57 in the trypsin:1b complex, in addition to $\Delta pK_{GlobalDes} = -1.91$ and $\Delta pK_{LocalDes} = -2.59$ (coordinates from PDB, code 1K1I).

tion (-4.29 pH units) dominates the shift in addition to a small negative contribution from a side-chain hydrogen bond to the ligand's piperazine nitrogen (-0.11 pH units).

0.93 protons are gained from the buffer when trypsin binds to ligand 2 at pH 7.8; PROPKA predicts a net protonation gain of 0.69 for this event and, according to our calculations, His57 is the titratable site primarily responsible for the uptake of protons. The large pK_a shift predicted for trypsin's His57 upon binding to 2 is due to a side-chain hydrogen bond and a charge–charge interaction with the ligand's carboxylate group (which contribute 1.60 and 2.20 pH units, respectively). The results obtained using MARVIN values for the ligand's model pK_a values (2.43) are similar, where the net protonation change at pH 7.8 is predicted to be 0.58. For the remaining ligands, which give negligible or zero net protonation changes upon binding, PROPKA predicts net losses and small net gains of protons to and from the buffer.

In a recent computational study, Czodrowski et al.[34,35] have applied their PEOE_PB procedure to the same trypsin complexes discussed above. Their results for two of the three complexes that undergo protonation changes are in good agreement with experiment where they predict net proton uptakes of 0.51 and 0.57 for trypsin complexed with 1b and 2 at pH 7.8, respectively. Their prediction for the trypsin:1d complex is less accurate (where they predict that -0.14 protons are released). Our PROPKA results are consistent with their predictions and their analyses of the dominating groups involved in (de)protonation, where our predictions for the trypsin complexes involving 1d and 2 are in excellent agreement with experiment.

In addition to trypsin, four of the inhibitors, 2, 3 (CRC220), 4 (inogatran), and 5 (melagatran), bound to the serine protease thrombin were studied.[49] For these four complexes, PROPKA predicts a negligible change in the overall protonation state for three of the complexes and an uptake of protons for the thrombin:2 complex (Table II), which is in accord with experiment. Czodrowski et al.[32] have applied the PEOE approach to the study of thrombin complexed with the two inhibitors 2 and 3. Their predictions agree with the experimental data in that the thrombin:2 complex undergoes a change in protonation (0.6 protons are predicted to be taken up at pH 7.8), which is attributed to an uptake of protons by the carboxylate group in the ligand. No change was observed (and predicted) for the thrombin:3 complex. The difference observed for the thrombin complexes with 2 and 3 may be due to the location of the titratable ligand groups in the active site, i.e. if the ligand carboxylate is solvent exposed it will tend not to lose a proton. In their more recent study, Czodrowski et al.,[34,35] based on their computed results for trypsin, concluded that the thrombin:2 complex uptakes protons to His57 and that the data suggest that the carboxylate will remain deprotonated, which is consistent with the PROPKA predictions. Clearly, predicting pK_a shifts and protonation changes for individual ionizable groups has its usefulness.

In summary, large ($|n| > 0.5$) protonation state changes have been observed experimentally for 4 of 14 ligand complexes with trypsin or thrombin. PROPKA 2.0 and Klebe's PEOE approach both identify three of the four large protonation state changes.

Proteins complexed with pepstatin

Proton uptake/release have been measured for binding of the peptide pepstatin to the following proteins: plasmepsin II, cathepsin D[66] and endothiapepsin (Table III).[67] For the plasmepsin II:pepstatin complex the overall uptake of protons at pH 6.5 is 1.7[66]; at pH 7 the complex uptakes 1.8 protons. Formation of the cathepsin D:pepstatin complex results in the uptake of 2.9 protons from the buffer at pH 6.5. 1.06 protons are taken up by the protein in the endothiapepsin:pepstain complex per molecule of the inhibitor at neutral pH; at a pH of 3.1 no protons are exchanged with the buffer.

PROPKA 2.0 underestimates the proton uptake predicted for each of the three pepsin complexes with pepstatin when the liganded X-ray crystal structures are used for the pK_a calculations on the unliganded proteins (data not shown). Therefore to include the effect of structural changes on the predicted pK_a shifts we employed apo PDB crystal structures to represent the unliganded proteins: 1PFZ (chain A and conformer set A), 1LYW (chains A and B) and 4APE for plasmepsin II, cathepsin D and endothiapepsin, respectively. The resulting changes in protonation are in good agreement with the experimental values given in parentheses: 1.50 (1.7) and 2.49 (2.9) protons are taken up at pH 6.5 for pepstatin complexed with plasmepsin II and cathepsin D, respectively;

and 0.77 (1.06) protons are gained at pH 7.0 for the endothiapepsin:pepstatin complex.

The protonation state changes can be rationalized as follows. The Asp34/Asp214 dyad in plasmepsin II undergoes pK_a shifts from 4.92 and 3.98 to 9.51 and 3.60 upon ligand binding, giving a micro protonation state change of 0.97 at pH 6.5; the His164 residue experiences a negative shift from 6.50 to 4.34, resulting in the loss of −0.49 protons. In addition, for each of the pepsins, residues unassociated with the binding site may undergo large pK_a shifts. For plasmepsin II, His318 undergoes a pK_a shift from 6.36 to 7.09 upon pepstatin binding (hence gaining 0.38 protons), which is consistent with the experimentally determined pK_a shift of 6.0 to 7.5 and assigned to His164. The change in the pK_a of His318 is due to a side-chain hydrogen bond with Asp190 (of 0.80 pH units) in the liganded structure. Proton uptake/release has also been determined for lower pH values, and the comparison to PROPKA predictions are shown in Figure 3. Clearly, the agreement is far from satisfactory: the predicted curve has a negative slope whereas the experimental curve has a positive slope. One possible explanation is that the protein structure undergoes significant structural changes at lower pH. The structure of the apo form was obtained at pH 7.5 while the corresponding information is not available for the complex. We note that Alexov[30] has obtained similar results as described below.

For cathepsin D, the Asp33/Asp231 dyad undergoes shifts from 0.63 to 3.71 and 6.17 to 9.54, implying a micro protonation change of 0.68 at pH 6.5; His77, Glu260, and Asp323 pK_a values shift from 7.46 to 6.50, 4.15 to 5.31, and 4.97 to 7.35, respectively, giving micro protonation state changes of −0.40, 0.06, and 0.85. Two residues, Glu5 and His56, located far from the binding site in cathepsin D undergo shifts from 3.65 and 4.11 to 6.62 and 6.28, respectively, upon pepstatin binding and consequently gain 0.57 and 0.37 protons. The change in pK_a for Glu5 is mainly due to increased desolvation contributions (2.18 pH units in the liganded structure) and the removal of a backbone hydrogen bond with itself (−1.12 pH units in the apo structure). The pK_a shift experienced by His56 is primarily due to a stronger backbone hydrogen bond with Ala129 in the liganded structure (1.13 pH units), and also a backbone hydrogen bond with Ala52 (0.67 pH units) along with less negative desolvation contributions.

The Asp32/Asp215 dyad in endothiapepsin undergoes negligible pK_a shifts upon pepstatin binding, whereas Asp30 shifts from 6.69 to 7.21 accounting for 0.29 of the protons taken up at pH 7.0. Asp12 in endothiapepsin undergoes a pK_a shift from 5.55 to 9.26 upon binding to pepstatin, resulting in a micro protonation change of 0.96. At pH 3.1 PROPKA predicts a net loss of −0.35 protons to the buffer.

Alexov[30] has studied pepstatin complexed with the three proteases using a modified LPBE technique. Alex-
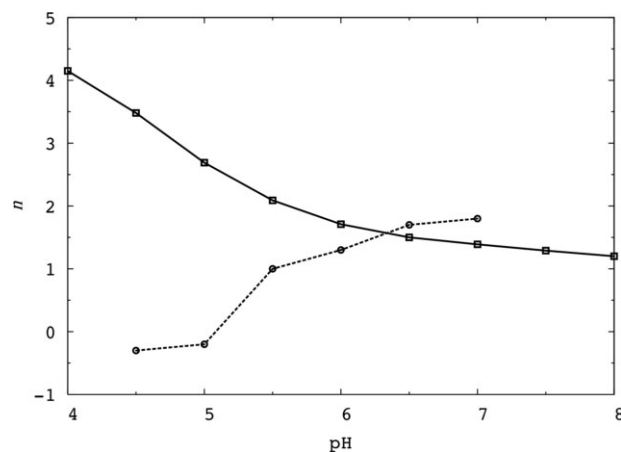
Figure 3
Calculated (solid line) and experimental (dashed line) protonation change (n) as a function of pH for plasmepsin II complexed with pepstatin.

ov's best predictions of net proton uptake for the three complexes employed apo X-ray structures for the free proteins, which is consistent with our results. The proton uptake predictions are approximately 3.2, 3.5, and 1.1 at pH 7.0, respectively, for cathepsin D, plasmepsin II, and endothiapepsin. Using the same X-ray structure for the complex and for the free protein, Alexov underestimated the net uptake of protons for each of the examples at pH 7.0, which is in accord with the PROPKA predictions (data not shown). For plasmepsin II:pepstatin our calculated change in protonation as a function of pH (see Fig. 3) is comparable to Alexov's prediction [Fig. 3(b)[30]], although our maximum is shifted to a lower pH by around 2.5 pH units.

Our results along with those of Alexov indicate that structural differences between the unliganded and liganded protein can have a significant influence on the calculated pK_a shifts.

In summary PROPKA 2.0 reproduces the protonation state changes due to plasmepsin II, cathepsin D and endothiapepsin binding to pepstatin to within 0.4 proton units at pH 6.5 and 7.0, respectively (Table III). The PROPKA results indicate that structural changes due to ligand binding contribute significantly to the proton uptake/release, as do residues far away from the binding site, primarily due to the change in the local environment of a particular residue and hence the change in the local hydrogen bonding network.
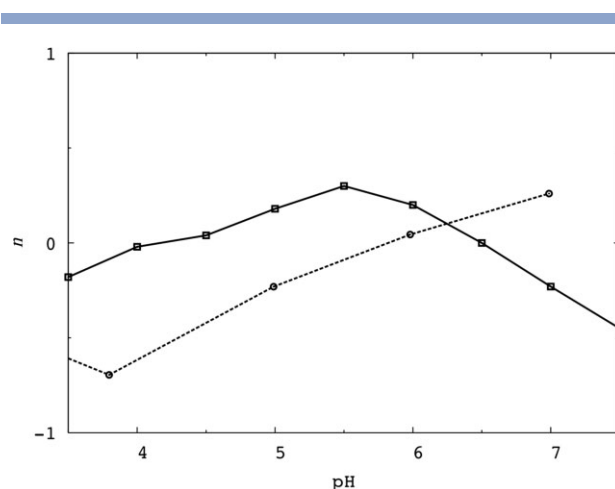
HIV-1 protease

Next we consider HIV-1 protease complexed with three different inhibitors, pepstatin, KNI-272 and DMP-323. HIV-1 protease is a homodimeric enzyme where

each monomer chain is comprised of 99 residues and contributes one Asp residue (Asp25) to a catalytic dyad located in a buried region of the protein, and which may be singly or doubly protonated.

For HIV-1 protease complexed with pepstatin the observed pK_a values of the Asp25/Asp25′ dyad have been measured by NMR and are, for one of the residues, less than 2.5 and, for the other, greater than 6.5.[68] PROPKA 2.0 predicts that the complexed pK_a values for the dyad are 4.04 and 10.20 for conformer 1 (Table III) and 10.08 and 4.49 for conformer 2, which are in accord with experiment. PROPKA predicts non-negligible pK_a shifts for other residues upon binding at pH 6.5. The most notable shifts are a positive shift for Asp230, which is partly due to a side-chain hydrogen bond with pepstatin's carboxylate group, and a negative shift for His69, which is due to weaker backbone hydrogen bonds with Ile66 and Cys67 in the complex compared to the free protein. Utilization of two apo X-ray structures (1HHP and 3HVP) to describe the free protein result in larger pK_a shifts for the Asp25/Asp225 dyad, and Asp30, Asp230, and His269 residues (Table III). The predicted net proton uptake is reduced to 0.16 protons at pH 6.5.

Experimental results for HIV-1 protease complexed with KNI-272 show that one of the pair of Asp25/Asp25′ residue's pK_a changes from 6.0 in the uncomplexed protein to 6.6 when liganded.[69] Experimental data also indicates that the pK_a of the isoquinoline nitrogen in the inhibitor shifts from 4.8 to 2.9 upon binding to the enzyme. The dyad remains mono-protonated upon formation of the complex. Using literature pK_a values for the isoquinoline group in the ligand (4.8), PROPKA predicts combined shifts from 9.26 to 10.28 and 4.80 to 4.52 for the Asp25/Asp25′ dyad and isoquinoline nitrogen, respectively, upon complex formation (Table III). Using MARVIN pK_a values for the uncomplexed ligand (4.52), analogous shifts from 9.26 to 9.48 and 4.55 to 4.27, respectively, are obtained. The predictions for the dyad are combinations of shifts from 3.77 and 9.26 to 10.28 and 3.97 for Asp25 and Asp25′, respectively. At a pH of 5.0, the total protonation change is predicted to be 0.18, which is at odds with experiment where approximately −0.23 protons are released to the buffer. Analysis of the PROPKA terms that contribute to the pK_a shifts show that the negligible negative pK_a shift experienced by the isoquinoline nitrogen is facilitated by a small and unfavorable local desolvation effect of −0.28 pH units. The positive pK_a shift for the Asp25/Asp25′ dyad may be due to favorable desolvation effects. Figure 4 displays the computed and the experimental change in the number of protons associated with the complex upon ligand binding as a function of pH. As noted for plasmepsin II with pepstatin and upon comparison with the available experimental data (over a small pH range), the PROPKA protonation change predictions appear to be shifted to lower pH values. Employing the apo X-ray structures to repre-

Figure 4

Calculated (solid line) and experimental (dashed line) protonation change (n) as a function of pH for HIV-1 protease complexed with KNI-272.

sent the free protein, the net uptake of protons is predicted to increase (Table III), with the majority of the protons gained by Asp30 and Asp230.

The Asp25/Asp25′ dyad is found to have a high pK_a value when complexed with DMP-323; a value of 8.19.[70] PROPKA predicts a value of 8.56 for one of the pair and 3.26 for the other, indicating that the dyad is singly protonated, which is at odds with experiment where both pK_a values are considered to be above 7.2, therefore protonated and thereby form hydrogen bonds with the symmetrically bound ligand's two diol groups. For the other carboxylic residues our results are in reasonable agreement (data not shown). As noted for the complex with KNI-272, use of apo X-ray structures gives rise to larger overall protonation changes for the complex with DMP-323.

Trylska et al.[25] have applied a LPBE approach to HIV-1 protease complexed with KNI-272, DMP-323 (and an analogue XK-263), and MVT-101. For HIV-1 protease: DMP-323 complex their best results (using full-charge/ε = 4 calculations) agree well with experiment where the Asp dyad is predicted to be doubly protonated; for the complex with KNI-272 their method's predictions are consistent with the experimental data where Asp25 is considered protonated and Asp125 is deprotonated. In a more recent and similar study to their work on trypsin and thrombin complexes, Czodrowski et al.[35] applied their PEOE_PB approach to two of the HIV-1 protease complexes discussed above. Using a dielectric constant of 10, they successfully predicted that the Asp dyad in HIV-1 protease bound to KNI-272 and DMP-323 is singly and doubly protonated, respectively. The net proton gains are predicted to be 0.01 and 0.63, respectively. At a pH of 7.0, PROPKA predicts the loss of 0.23 protons for the complex with KNI-272 and no gain of protons for the

complex with DMP-323. At the pH employed by Czodrowski et al.[35] (5.0), PROPKA predicts an uptake of 0.18 and 0.27 protons for these complexes, where, for both cases, the two Asp30 residues uptake the protons. These comparisons indicate that PROPKA does poorly at predicting the protonation changes for the Asp25 catalytic dyad in HIV-1 protease upon ligand binding.

Overall, PROPKA 2.0 predicts that the catalytic Asp dyad in HIV-1 protease remains singly protonated upon binding to the ligands described above, which is in accord with the experiment for binding to pepstatin and KNI-272. Experimentally, binding to DMP-323 results in the Asp dyad becoming doubly protonated, which is not predicted by PROPKA.

Ligands covalently bonded to host proteins

We have applied PROPKA 2.0 to predict the pK_a values for protein–ligand complexes where the ligand is covalently bound to the protein and for which pK_a values have been measured by NMR. Two examples in the literature are chymotrypsin bound, via its Ser195 residue, to two peptidyl trifluoromethyl ketone (TFK) derived inhibitors,[71] and xylanase bonded, via its Glu78 residue, to 2-deoxy-2-fluoro-β-xylobioside (2FXb).[72]

For the former example and for each of the two X-ray structures employed, which describe a tetrahedral hemiketal adduct, the ligand carbonyl oxygen atom neighboring the bond to Ser195 O_γ is negatively charged and is incorporated in the pK_a calculations using the 1N atom-type since it satisfies the buried atom/group criterion. In both cases PROPKA 2.0 predicts a significant increase in the His57 pK_a, in agreement with experiment. However, the His57/Asp102 dyad in chymotrypsin complexed with the NAc-Leu-Phe-CF$_3$ inhibitor is more basic (i.e. has a higher pK_a) than the complex with the NAc-Phe-CF$_3$ inhibitor. PROPKA predicts the reverse, where the complex with the ligand without a Leu has the higher relative pK_a, a value of 10.33 as opposed to 8.90 (Table III), which may be due to stronger side-chain hydrogen bonds with the ligand's CF$_3$ group that are predicted for this complex (data not shown). When the biological unit of chymotrypsin is employed, i.e. a homodimer where each monomer is comprised of three chains, PROPKA predicts the correct pK_a ordering for His57 upon reaction with the two ligands: His57 has pK_a values of 8.90 and 8.46 for the adducts formed with the TFK with and without Leu, respectively. For the former case the negative global desolvation contribution is smaller than the latter by 1.89 pH units. For comparison, Schutz and Warshel have predicted pK_a values for the NAc-Phe-CF$_3$-chymotrypsin complex in the range 10.3–12.9.[73]

The acid/base catalyst in xylanase, the Glu172 residue, is observed to have a pK_a of 6.7 in the free protein, which, upon binding to 2FXb, shifts to 4.2.[72] PROPKA yields a pK_a shift from 7.45 to 6.64 (due to the removal

of a charge-charge interaction with Glu78 of 1.06 pH units), which is consistent with experiment. Joshi et al.[72] employed a Poisson-Boltzmann technique to calculate pK_a shifts for important residues to complement their experimental work. The authors calculated a pK_a shift of 5.9 to 3.7 for Glu172, which was ascribed to the removal of an unfavorable charge interaction and reduced desolvation and background interactions upon substrate binding. The Asn35Asp mutant of xylanase was also studied by Joshi et al. Upon substrate binding the Glu172 residue undergoes a pK_a shift from 8.4 to between 1.9 to 3.4 or greater than 9.0 (Asp35/Glu172 are a tightly coupled pair where Asp35 has a pK_a of 3.7 in the free protein). The reported PB calculations yielded shifts from 1.0 to 0.2 and 7.5 to 9.0 for Asp35 and Glu172, respectively. Using PDB coordinates 1C5I, PROPKA predicts shifts for these two residues from 4.15 to 4.40 and from 12.16 to 12.35, respectively. The large pK_a values for Glu172 are due to a side-chain hydrogen bond (1.60 pH units) and a coulombic interaction (2.40 pH units) with Asp35.

Charged ligands

Hydroxynitrile lyase:thiocyanate complex: Stranzl et al.[74] have shown by NMR spectroscopy that the pK_a value of His235 in hydroxynitrile lyase increases from 2.5 to ~8 upon binding of thiocyanate. The current version of PROPKA 2.0 does not treat S atoms as an ionizable group. However, the pK_a of thiocyanate is sufficiently low (>1) that it can be treated as a (nonionizable) charge using the 1N atom type (Table I, part B). The predicted pK_a value of His235 in the un-complexed form of hydroxynitrile lyase is 2.9 and in good agreement with experiment. However, the addition of thiocyanate increases the pK_a value by only 0.9 pH units to 3.8, due to charge–charge interactions. Inspection of the complex structure (2YAS) reveals a possible explanation. In the uncomplexed form, Ser80 serves as a hydrogen bond donor to His235, but upon complex formation this hydrogen bond donation is shifted to a hydrogen bond with thiocyanate (Scheme 4). This forces the N_{E2} atom of
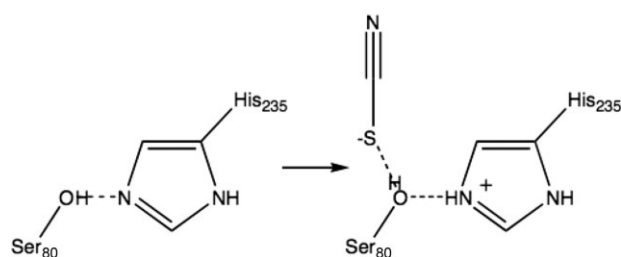
Scheme 4

Hydroxynitrile lyase : thiocyanate complexation.

His235 to serve as a hydrogen donor, thereby significantly increasing its pK_a. This effect is neglected in PROPKA since hydrogen positions are not determined for Ser residues. When the OH (O$_\gamma$) atom of Ser80 is considered to be an O.3 atom-type, a side-chain hydrogen bond (SDC-HB) results between His235 and Ser80 (of 0.85 pH units) that raises the predicted pK_a to 4.62 for His235 in the complex (from 3.77 in the unliganded protein), which is in somewhat better agreement with the experimental value of ~8.

Dihydrofolate reductase

Dihydrofolate reductase complexed with methotrexate has been studied by both experiment (NMR[75–78]) and by theory.[32] NMR studies have shown that the aromatic N1 nitrogen atom of methotrexate is protonated in the complex formed with dihydrofolate reductase, up to pH values in excess of 10; in solution the pK_a value of the N1 nitrogen atom is lower (pK_a of 5.7). In the dihydrofolate reductase:methotrexate complex, it is believed that Asp27 and the ligand's N1 nitrogen atom form a salt bridge, in which Asp27 would be deprotonated.

Using the more sequence complete chain (B) and using a reference pK_a value of 5.7 for the N1 aromatic nitrogen in the free ligand's pteridine ring, PROPKA 2.0 predicts that this group undergoes a shift to 6.89 and that Asp27 (which it forms a salt bridge with) shifts from 4.98 to 0.86 upon binding. These results are consistent with experiment and with the computational results of Czodrowski et al.[32] Interestingly, when the free pK_a value of N1 nitrogen takes a value predicted by MARVIN (a pK_a of 4.40) a negative shift to 3.19 is predicted for this group upon binding, which is at odds with the above. Moreover, the shift to a lower pK_a for Asp27 (to a value of 3.26) is less pronounced. Analysis of the PROPKA output suggests that this discrepancy is due to a charge-charge interaction between methotrexate's N1 nitrogen atom and Asp27 (of ±2.40 pH units and which favors protonation of N1 and deprotonation of Asp27), which is present in the calculation employing an initial reference pK_a of 5.7 for N1 and absent in the calculation using all MARVIN ligand pK_a values (i.e. pK_a of 4.40 for N1). This interaction is absent in the latter case because the temporary pK_a value (as determined by PROPKA) for Asp27 is greater than the temporary pK_a value of the ligand's N1 atom, i.e. a pK_a of 3.26 as opposed to 3.19, which is the opposite to the former case. That is, the prediction of pK_a may be sensitive to the initial intrinsic pK_a values assigned to each group.

Ligand pK_{Model} values

To validate the model pK_a values listed for each ligand functional group in Table I, part A we re-calculated the pK_a shifts for each of the protein–ligand complexes employing the model values for the appropriate group in each of the ligands. The results are discussed in light of the results previously discussed earlier.

For the trypsin complexes, where significant changes are observed in protonation state (Table II), the model parameters give protonation uptake/release values of 0.09, −0.92, and 0.67 for complexes with 1b, 1d, and 2, respectively, at a pH of 7.8. These are in line with the results obtained with the experimental or reference ligand pK_a values (Table II) with the exception of the trypsin:1d complex where there is a two-fold increase in the protons lost to the buffer. This difference may be due to the pK_a shift of the ligand's piperazine amino nitrogen from 10.0 in the free ligand to 7.43 in the complex, where the associated micro change in protonation is −0.69 at pH 7.8. The predictions for the remaining complexes with trypsin have results similar to those in Table II with the exceptions of trypsin:4 and trypsin:5 where the protonation changes are predicted to be −0.20 and −0.19, which may be due to the negligible pK_a shift of each ligand's amino group, which have little contribution to the net protonation changes at pH 7.8. Similar conclusions can be drawn about the thrombin:4 complex. The model values yield an uptake of 0.34 protons for the thrombin:2 complex, as above.

For the proteins complexed with pepstatin (Table III), the model pK_a ligand parameters yield results that are practically identical to those obtained using the MARVIN value (for pepstatin's carboxylate group). This observation is also true for the ligands complexed with HIV-1 protease (Table III).

For the remaining protein–ligand complexes (Table III) only the complex of methotrexate with dihydrofolate reductase has significantly different pK_a shift predictions with respect to the previous calculations when the model ligand pK_a values are employed. For the first conformer set (A) and for chain B, the model parameters yield a positive pK_a shift from 5.0 to 6.19 for the ligand's aromatic nitrogen, labeled N1, upon complexation and a large negative shift from 4.98 to −3.31 for Asp27, which are in accord with the results using the experimental pK_a value for N1, and also with experiment and with a computational prediction from another group.[32] For this example our model pK_a values yield better results than our predictions when MARVIN pK_a ligand values were employed. The shift to −3.31 for Asp27 is due to coulombic interactions with the four aromatic nitrogen atoms present in the ligand, each of which have an equal model pK_a value of 5.0.

The above results show that when employing the pK_a model ligand parameters, the computed pK_a shifts are, with two exceptions, similar to the results using experimental and/or MARVIN ligand pK_a values. However, we note that the prediction of proton uptake and release will not be sensitive to, and therefore not a good test of, the

pK_{Model} values unless the resulting pK_a value changes involve pK_a values close to the pH of interest.

SUMMARY AND FUTURE DIRECTIONS

The PROPKA method for the prediction of the pK_a values of ionizable residues in proteins is extended to include the effect of non-proteinaceous ligands on protein pK_a values as well as predict the change in pK_a values of ionizable groups on the ligand itself. This new version is called PROPKA 2.0, and can be used to compute the pK_a values of all ionizable groups in a matter of seconds for most proteins. Several challenges had to be met in creating PROPKA 2.0: (1) Identifying functional groups in ligands, such as ionizable groups or hydrogen bond donors and acceptors. (2) Assigning intrinsic pK_a values to the ligand functional groups, and (3) adapting and extending the original empirical rules underlying PROPKA 1.0 to ligand functional groups. The first two issues are dealt with primarily by using auxiliary programs so that the process can easily be modified by the user (e.g. interfaced to databases of ligand pK_a values) and interfaced to PROPKA 2.0 via an input file. The current programs are seamlessly interfaced to maintain the user-friendliness of PROPKA 1.0 as much as possible to the casual user. At the same time, the expert user can handle awkward cases through manual editing of the input file.

The new version of PROPKA (PROPKA 2.0) is, as much as possible, developed by adapting the empirical rules underlying PROPKA 1.0 to ligand functional groups. In the PROPKA approach there are three different sources of pK_a perturbation: desolvation, hydrogen bonding, and charge–charge interactions. The desolvation and charge–charge terms are calculated exactly as before, except that it is now also possible to treat doubly charged functional groups (where the maximum pK_a shift due the charged group is simply twice that of singly charged functional groups) as well as charged groups that are non-ionizable (e.g. Zn^{2+}). The hydrogen bonding terms are kept as similar as possible to the original terms as described in the text.

The PROPKA 2.0 predictions are compared to experimental data for 26 protein–ligand complexes. We have shown that PROPKA 2.0 is able to predict pK_a values and changes in protonation states that are in accordance with experiment for the majority of the 26 protein–ligand complexes studied. For trypsin and thrombin, large protonation state changes ($|n| > 0.5$) have been observed experimentally for 4 of 14 ligand complexes. PROPKA 2.0 and Klebe's PEOE approach both identify three of the four large protonation state changes. The protonation state changes due to plasmepsin II, cathepsin D and endothiapepsin binding to pepstatin are predicted to within 0.4 proton units at pH 6.5 and 7.0, respectively. The PROPKA 2.0 results indicate that structural changes due to ligand binding contribute significantly to the proton uptake/release, as do residues far away from the binding site, primarily due to the change in the local environment of a particular residue and hence the change in the local hydrogen bonding network. PROPKA 2.0 predicts that the catalytic Asp dyad in HIV-1 protease remains singly protonated upon binding to each of the ligands studied, which is in accord with experiment for binding to pepstatin and KNI-272, but not for the HIV-1 protease:DMP-323 complex, where the Asp dyad is doubly protonated. PROPKA 2.0 can account for a ligand that is covalently bound to a protein's Ser, Asp, or Glu residue, and describe nonionizable charged ligands and charged atoms, such as salts. For chymotrypsin bound, via its Ser195 residue, to two differing peptidyl trifluoromethyl ketone derived inhibitors (where a carbonyl oxygen atom in each ligand becomes negatively charged during complex formation), PROPKA 2.0 predicts an increase in the pK_a value of His57 upon binding, as observed by experiment. Xylanse is bound, via its Glu78 residue, to 2FXb. For this complex, PROPKA 2.0 predicts a negative shift for the pK_a value of Glu172 upon binding, which is in line with experiment. When PROPKA 2.0 was applied to predict the pK_a shifts for the dihydrofolate reductase:methotrexate complex, differing results were obtained depending on the initial choice of the pK_{Model} values used for the ligand's titratable sites. The two PROPKA predictions that agree with experiment, where a salt bridge is formed between Asp27 and a ligand aromatic nitrogen atom, employed pK_{Model} values for the aromatic nitrogen atom from experiment and from the model values proposed in Table I, part A.

PROPKA 2.0 has certain limitations, which we plan to address in future work. Among the limitations are: (A) Nonproteinaceous atoms (excluding water) are treated as a single ligand. Thus, any pK_a shifts due to inter-ligand interactions are neglected. (B) It is assumed that pK_a shifts due to intra-ligand interactions are included in the pK_{Model} values, and are not considered explicitly. Two important implications of this are: (B1) Ligands that contain two or more titratable sites whose titration is coupled. For example, pyrazine (SMILES format: c1cnccn1) has two ionizable aromatic nitrogen atoms with MARVIN pK_{Model} values of 0.6. However, upon protonation of one site the pK_a of the other site is lowered considerably, and becomes practically non-ionizable. (B2) An experimental pK_a value of the free ligand reflects the conformation most probable in solution. If the conformation is very different in the complex with the protein this can affect the intrinsic pK_a value of the ligand. (C) As with PROPKA 1.0, PROPKA 2.0 does not explicitly include the effects of side chain motion or automatically considers other conformations.

PROPKA 2.0 is an open source code that is freely distributed under the BSD style license (http://www.opensource.org/licenses/bsd-license.php). A web interface is also available and can be found at http://propka.ki.ku.dk/. We hope that PROPKA 2.0 will be a useful tool for the interpretation of pH-dependent effects of ligand binding, thereby generating more experimental data - badly needed for a thorough validation of new methods for pK_a prediction of protein–ligand complexes.

ACKNOWLEDGMENTS

REFERENCES

1. Sham YY, Chu ZT, Warshel A. Consistent calculations of p$K(a)$'s of ionizable residues in proteins: semi-microscopic and microscopic approaches. J Phys Chem B 1997;101:4458–4472.
2. Warshel A. Electrostatic basis of structure-function correlation in proteins. Accounts Chem Res 1981;14:284–290.
3. Demchuk E, Wade RC. Improving the continuum dielectric approach to calculating p$K(a)$s of ionizable groups in proteins. J Phys Chem 1996;100:17373–17387.
4. Nielsen JE, Vriend G. Optimizing the hydrogen-bond network in Poisson-Boltzmann equation-based p$K(a)$ calculations. Proteins 2001; 43:403–412.
5. Karshikoff A. A simple algorithm for the calculation of multiple-site titration curves. Protein Eng 1995;8:243–248.
6. Antosiewicz J, McCammon JA, Gilson MK. The determinants of p$K(a)$s in proteins. Biochemistry 1996;35:7819–7833.
7. Antosiewicz J, Briggs JM, Elcock AH, Gilson MK, McCammon JA. Computing ionization states of proteins with a detailed charge model. J Comput Chem 1996;17:1633–1644.
8. Antosiewicz J, McCammon JA, Gilson MK. Prediction of Ph-dependent properties of proteins. J Mol Biol 1994;238:415–436.
9. Bashford D, Karplus M. Pkas of ionizable groups in proteins—atomic detail from a continuum electrostatic model. Biochemistry 1990;29:10219–10225.
10. Mehler EL, Guarnieri F. A self-consistent, microenvironment modulated screened Coulomb potential approximation to calculate pH-dependent electrostatic effects in proteins. Biophys J 1999;77:3–22.
11. Ullmann GM, Knapp EW. Electrostatic models for computing protonation and redox equilibria in proteins. Eur Biophys J Biophys Lett 1999;28:533–551.
12. Wisz MS, Hellinga HW. An empirical model for electrostatic interactions in proteins incorporating multiple geometry-dependent dielectric constants. Proteins 2003;51:360–377.
13. Yang AS, Gunner MR, Sampogna R, Sharp K, Honig B. On the Calculation of p$K(a)$S in proteins. Proteins 1993;15:252–265.
14. Havranek JJ, Harbury PB. Tanford-Kirkwood electrostatics for protein modeling. Proc Natl Acad Sci USA 1999;96:11145–11150.
15. Nielsen J. Calculating pK_a values in enzyme active sites. Protein Sci 2003;12:1894–1901.
16. Barth P, Alber T, Harbury PB. Accurate, conformation-dependent predictions of solvent effects on protein ionization constants. Proc Natl Acad Sci USA 2007;104:4898–4903.
17. Godoy-Ruiz R, Perez-Jimenez R, Garcia-Mira MM, del Pino IMP, Sanchez-Ruiz JM. Empirical parametrization of pK values for carboxylic acids in proteins using a genetic algorithm. Biophys Chem 2005;115:263–266.
18. He Y, Xu J, Pan XM. A statistical approach to the prediction of p$K(a)$ values in proteins. Proteins Struct Funct Bioinf 2007;69:75–82.
19. Khandogin J, Brooks CL. Toward the accurate first-principles prediction of ionization equilibria in proteins. Biochemistry 2006;45:9363–9373.
20. Nielsen JE. Analysing the pH-dependent properties of proteins using p$K(a)$ calculations. J Mol Graphics Model 2007;25:691–699.
21. Simonson T, Carlsson J, Case DA. Proton binding to proteins: p$K(a)$ calculations with explicit and implicit solvent models. J Am Chem Soc 2004;126:4167–4180.
22. Tynan-Connolly BM, Nielsen JE. Redesigning protein p$K(a)$ values. Protein Sci 2007;16:239–249.
23. Del Buono GS, Figueirido FE, Levy RM. Intrinsic p$K(a)$s of ionizable residues in proteins: an explicit solvent calculation for lysozyme. Proteins Struct Funct Genet 1994;20:85–97.
24. Warshel A, Sussman F, King G. Free energy of charges in solvated proteins: microscopic calculations using a reversible charging process. Biochemistry 1986;25:8368–8372.
25. Trylska J, Antosiewicz J, Geller M, Hodge CN, Klabe RM, Head MS, Gilson MK. Thermodynamic linkage between the binding of protons and inhibitors to HIV-1 protease. Protein Sci 1999;8:180–195.
26. Mardis KL, Luo R, Gilson MK. Interpreting trends in the binding of cyclic ureas to HIV-1 protease. J Mol Biol 2001;309:507–517.
27. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM—a program for macromolecular energy. minimization, and dynamics calculations. J Comput Chem 1983;4:187–217.
28. Jorgensen WL, Tiradorives J. The OPLS potential functions for proteins—energy minimizations for crystals of cyclic-peptides and crambin. J Am Chem Soc 1988;110:1657–1666.
29. Sitkoff D, Sharp KA, Honig B. Accurate calculation of hydration free-energies using macroscopic solvent models. J Phys Chem 1994;98:1978–1988.
30. Alexov E. Calculating proton uptake/release and binding free energy taking into account ionization and conformation changes induced by protein-inhibitor association application to plasmepsin, cathepsin D and endothiapepsin-pepstatin complexes. Proteins Struct Funct Bioinf 2004;56:572–584.
31. Accelrys Insight II, 1st ed. San Diego: Accelrys; 1990.
32. Czodrowski P, Dramburg I, Sotriffer CA, Klebe G. Development, validation, and application of adapted PEOE charges to estimate p$K(a)$ values of functional groups in protein–ligand complexes. Proteins Struct Funct Bioinf 2006;65:424–437.
33. Cannon WR, Garrison BJ, Benkovic SJ. Consideration of the pH-dependent inhibition of dihydrofolate reductase by methotrexate. J Mol Biol 1997;271:656–668.
34. Czodrowski P, Sotriffer CA, Klebe G. Protonation changes upon ligand binding to trypsin and thrombin: structural interpretation based on pK_a calculations and ITC experiments. J Mol Biol 2007; 367:1347–1356.

35. Czodrowski P, Sotriffer CA, Klebe G. Atypical protonation states in the active site of HIV-1 protease: a computational study. J Chem Inf Model 2007;47:1590–1598.

36. Rod TH, Brooks CL. How dihydrofolate reductase facilitates protonation of dihydrofolate. J Am Chem Soc 2003;125:8718–8719.

37. Khavrutskii IV, Price DJ, Lee J, Brooks CL. Conformational change of the methionine 20 loop of Escherichia coli dihydrofolate reductase modulates pK(a) of the bound dihydrofolate. Protein Sci 2007;16:1087–1100.

38. Varnai P, Warshel A. Computer simulation studies of the catalytic mechanism of human aldose reductase. J Am Chem Soc 2000;122:3849–3860.

39. Burykin A, Schutz CN, Villa J, Warshel A. Simulations of ion current in realistic models of ion channels: the KcsA potassium channel. Proteins Struct Funct Genet 2002;47:265–280.

40. Li H, Robertson AD, Jensen JH. Very fast empirical prediction and rationalization of protein pK(a) values. Proteins Struct Funct Bioinf 2005;61:704–721.

41. Davies MN, Toseland CP, Moss DS, Flower DR. Benchmarking pK_a Prediction. BMC Biochem 2006;7:18.

42. Mason AC, Jensen, J. H. Protein–protein binding is often associated with changes in protonation state. Proteins Struct Funct Bioinf 2008;71:81–91.

43. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucl Acids Res 2000;28:235–242.

44. Guha R, Howard MT, Hutchison GR, Murray-Rust P, Rzepa H, Steinbeck C, Wegner J, Willighagen EL. The blue obelisk-interoperability in chemical informatics. J Chem Inf Model 2006;46:991–998.

45. The Open Babel Package, version 2.1.1 http://openbabel.sourceforge. net/ (accessed Oct 2007).

46. Tripos International, 1699 South Hanley Road, St. Louis, Missouri, 63144, USA.

47. Bondi A. Van der Waals volumes + radii. J Phys Chem 1964;68:441.

48. ChemAxon Kft., Máramaros köz 3/a, Budapest, 1037 Hungary. Marvin and Calculator Plugin Demo, http://www.chemaxon.com/demosite/ marvin/index.html.

49. Dullweber F, Stubbs MT, Musil D, Sturzebecher J, Klebe G. Factorising ligand affinity: a combined thermodynamic and crystallographic study of trypsin and thrombin inhibition. J Mol Biol 2001; 313:593–614.

50. Jasmine Fokkens GK. A Simple protocol to estimate differences in protein binding affinity for enantiomers without prior resolution of racemates. Angewandte Chemie Int Ed 2006;45:985–989.

51. Bernstein NK, Cherney MM, Loetscher H, Ridley RG, James MNG. Crystal structure of the novel aspartic proteinase zymogen proplasmepsin. Nat Struct Mol Biol 1999;6:32–37.

52. Silva AM, Lee AY, Gulnik SV, Majer P, Collins J, Bhat TN, Collins PJ, Cachau RE, Luker KE, Gluzman IY, Francis SE, Oksman A, Goldberg DE, Erickson JW. Structure and inhibition of plasmepsin II, a hemoglobin-degrading enzyme from Plasmodium† falciparum. Proc Natl Acad Sci 1996;93:10034–10039.

53. Lee AY, Gulnik SV, Erickson JW. Conformational switching in an aspartic proteinase. Nat Struct Mol Biol 1998;5:866–871.

54. Baldwin ET, Bhat TN, Gulnik S, Hosur MV, Sowder Rc, II, Cachau RE, Collins J, Silva AM, Erickson JW. Crystal structures of native and inhibited forms of human cathepsin D: implications for lysosomal targeting and drug design. Proc Natl Acad Sci 1993;90:6796–6800.

55. Pearl L, Blundell T. The active-site of aspartic proteinases. FEBS Lett 1984;174:96–101.

56. Fitzgerald PM, McKeever BM, VanMiddlesworth JF, Springer JP, Heimbach JC, Leu CT, Herber WK, Dixon RA, Darke PL. Crystallographic analysis of a complex between human immunodeficiency virus type 1 protease and acetyl-pepstatin at 2.0-A resolution. J Biol Chem 1990;265:14209–14219.

57. Spinelli S, Liu QZ, Alzari PM, Hirel PH, Poljak RJ. The 3-dimensional structure of the aspartyl protease from the HIV-1 ISOLATE BRU. Biochimie 1991;73:1391–1396.

58. Wlodawer A, Miller M, Jaskolski M, Sathyanarayana BK, Baldwin E, Weber IT, Selk LM, Clawson L, Schneider J, Kent SB. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. Science 1989;245:616–621.

59. Baldwin ET, Bhat TN, Gulnik S, Liu BS, Topol IA, Kiso Y, Mimoto T, Mitsuya H, Erickson JW. Structure of HIV-1 protease with KNI-272, a tight-binding transition-state analog containing allophenylnorstatine. Structure 1995;3:581–590.

60. Lam PYS, Ru Y, Jadhav PK, Aldrich PE, DeLucca GV, Eyermann CJ, Chang CH, Emmett G, Holler ER, Daneker WF, Li L, Confalone PN, McHugh RJ, Han Q, Li R, Markwalder JA, Seitz SP, Sharpe TR, Bacheler LT, Rayner MM, Klabe RM, Shum L, Winslow DL, Kornhauser DM, Jackson DA, Erickson-Viitanen S, Hodge CN. Cyclic HIV protease inhibitors: synthesis, conformational analysis, P2/P2′ structure-activity relationship, and molecular recognition of cyclic ureas. J Med Chem 1996;39:3514–3525.

61. Brady K, Wei A, Ringe D, Abeles RH. Structure of chymotrypsin-trifluoromethyl ketone inhibitor complexes: comparison of slowly and rapidly equilibrating inhibitors. Biochemistry 1990;29:7600–7607.

62. Sidhu G, Withers SG, Nguyen NT, McIntosh LP, Ziser L, Brayer GD. Sugar ring distortion in the glycosyl-enzyme intermediate of a family G/11 xylanase. Biochemistry 1999;38:5346–5354.

63. Joshi MD, Sidhu G, Pot I, Brayer GD, Withers SG, McIntosh LP. Hydrogen bonding and catalysis: a novel explanation for how a single amino acid substitution can change the pH optimum of a glycosidase. J Mol Biol 2000;299:255–279.

64. Zuegg J, Gruber K, Gugganig M, Wagner UG, Kratky C. Three-dimensional structures of enzyme-substrate complexes of the hydroxynitrile lyase from Hevea brasiliensis. Protein Sci 1999;8:1990–2000.

65. Bolin JT, Filman DJ, Matthews DA, Hamlin RC, Kraut J. Crystal structures of Escherichia coli and Lactobacillus casei dihydrofolate reductase refined at 1.7 A resolution. I. General features and binding of methotrexate. J Biol Chem 1982;257:13650–13662.

66. Xie D, Gulnik S, Collins L, Gustchina E, Suvorov L, Erickson JW. Dissection of the pH dependence of inhibitor binding energetics for an aspartic protease: direct measurement of the protonation states of the catalytic aspartic acid residues. Biochemistry 1997;36:16166–16172.

67. Gomez J, Freire E. Thermodynamic mapping of the inhibitor site of the aspartic protease endothiapepsin. J Mol Biol 1995;252:337–350.

68. Smith R, Brereton IM, Chai RY, Kent SBH. Ionization states of the catalytic residues in HIV-1 protease. Nat Struct Biol 1996;3:946–950.

69. Velazquez-Campoy A, Luque I, Todd MJ, Milutinovich M, Kiso Y, Freire E. Thermodynamic dissection of the binding energetics of KNI-272, a potent HIV-1 protease inhibitor. Protein Sci 2000;9:1801–1809.

70. Yamazaki T, Nicholson LK, Torchia DA, Wingfield P, Stahl SJ, Kaufman JD, Eyermann CJ, Hodge CN, Lam PYS, Ru Y, Jadhav PK, Chang CH, Weber PC. NMR and X-ray evidence that the hiv protease catalytic aspartyl groups are protonated in the complex formed by the protease and a nonpeptide cyclic urea-based inhibitor. J Am Chem Soc 1994;116:10791–10792.

71. Cassidy CS, Lin J, Frey PA. A new concept for the mechanism of action of chymotrypsin: the role of the low-barrier hydrogen bond. Biochemistry 1997;36:4576–4584.

72. Joshi MD, Sidhu G, Nielsen JE, Brayer GD, Withers SG, McIntosh LP. Dissecting the electrostatic interactions and pH-dependent activity of a family 11 glycosidase. Biochemistry 2001;40:10115–10139.

73. Schutz CN, Warshel A. The low barrier hydrogen bond (LBHB) proposal revisited: the case of the asp. his pair in serine proteases. Proteins Struct Funct Bioinf 2004;55:711–723.
74. Stranzl GR, Gruber K, Steinkellner G, Zangger K, Schwab H, Kratky C. Observation of a short, strong hydrogen bond in the active site of hydroxynitrile lyase from *Hevea brasiliensis* explains a large pK(a) shift of the catalytic base induced by the reaction intermediate. J Biol Chem 2004;279:3699–3707.
75. Cocco L, Groff JP, Temple C, Montgomery JA, London RE, Matwiyoff NA, Blakley RL. C-13 nuclear magnetic-resonance study of protonation of methotrexate and aminopterin bound to dihydrofolate-reductase. Biochemistry 1981;20:3972–3978.
76. Cocco L, Roth B, Temple C, Montgomery JA, London RE, Blakley RL. Protonated state of methotrexate, trimethoprim, and pyrimethamine bound to dihydrofolate-reductase. Arch Biochem Biophys 1983;226:567–577.
77. Stone SR, Morrison JF. The ph-dependence of the binding of dihydrofolate and substrate-analogs to dihydrofolate-reductase from Escherichia-coli. Biochimica Et Biophysica Acta 1983;745:247–258.
78. London RE, Howell EE, Warren MS, Kraut J, Blakley RL. Nuclear-magnetic-resonance study of the state of protonation of inhibitors bound to mutant dihydrofolate-reductase lacking the active-site carboxyl. Biochemistry 1986;25:7229–7235.