

---

# Calculating pKa values in enzyme active sites

---

JENS ERIK NIELSEN AND J. ANDREW MCCAMMON

Howard Hughes Medical Institute, Departments of Pharmacology, Chemistry, and Biochemistry, University of California, San Diego, La Jolla, California 92093, USA

(RECEIVED March 31, 2003; FINAL REVISION May 29, 2003; ACCEPTED May 30, 2003)

## Abstract

The ionization properties of the active-site residues in enzymes are of considerable interest in the study of the catalytic mechanisms of enzymes. Knowledge of these ionization constants (pKa values) often allows the researcher to identify the proton donor and the catalytic nucleophile in the reaction mechanism of the enzyme. Estimates of protein residue pKa values can be obtained by applying pKa calculation algorithms to protein X-ray structures. We show that pKa values accurate enough for identifying the proton donor in an enzyme active site can be calculated by considering in detail only the active-site residues and their immediate electrostatic interaction partners, thus allowing for a large decrease in calculation time. More specifically we omit the calculation of site-site interaction energies, and the calculation of desolvation and background interaction energies for a large number of pairs of titratable groups. The method presented here is well suited to be applied on a genomic scale, and can be implemented in most pKa calculation algorithms to give significant reductions in calculation time with little or no impact on the accuracy of the results. The work presented here has implications for the understanding of enzymes in general and for the design of novel biocatalysts.

**Keywords:** Enzyme; pKa calculation; electrostatics; active site; catalysis

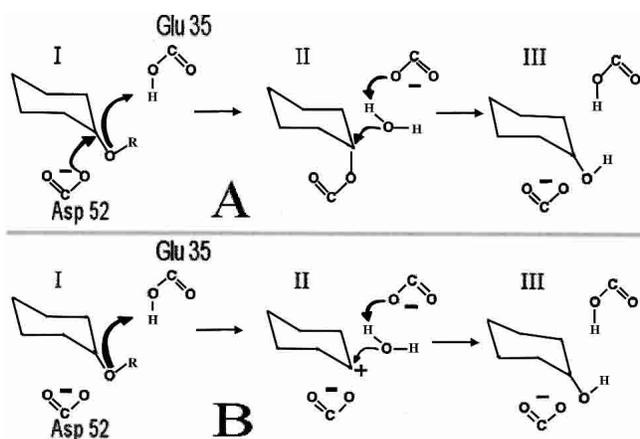
The pKa values (ionization constants) of the active-site residues in an enzyme are of importance to the functionality of the catalytic mechanism of the enzyme. Often a catalytic reaction is initiated by the transfer of a proton from a protein residue (the proton donor) to the substrate, and one of the steps in an enzymatic reaction mechanism is normally a nucleophilic attack on a substrate atom or the stabilization of a positively charged intermediate. For catalysis to take place, the proton donor must be protonated, and the nucleophilic attack and the stabilization of the intermediate must be performed by a protein residue with a free lone-pair or a negative charge. In the following we will refer to the latter residue as “the catalytic nucleophile.” A well known example of an enzyme which employs a proton donor and a catalytic nucleophile (Fig. 1) is hen egg white lysozyme

(HEWL; Philips 1967; Vocadlo et al. 2001). Although it is still unclear whether the catalytic reaction in HEWL proceeds via a covalent (Vocadlo et al. 2001) or carbenium ion intermediate (Philips 1967; see below), it is a well established fact that the chemistry of the catalytic mechanism requires that both the proton donor and the catalytic nucleophile be in a particular protonation state as outlined above. The pKa values of the proton donor and the catalytic nucleophilic residue are therefore limiting the pH range at which the enzyme can function. Many enzymes function optimally at pH values around 7, and use an aspartic acid or a glutamic acid as the proton donor. For these enzymes, it is necessary to shift the pKa values of the proton donor upward to a value that is much closer to neutral pH than the pKa values of 4.0 and 4.4 that Glu and Asp normally display. This is illustrated by the pKa values in the active site of *Bacillus circulans* xylanase, where Glu 172 is the proton donor and has a pKa value of 6.7 in the apo enzyme (McIntosh et al. 1996). Glu 78 is the catalytic nucleophile and has a pKa value of 4.6. Many other enzymes have Asp and Glu residues that act as proton donors and catalytic nucleophiles (Coutinho and Henriessat 1999), and these enzymes are

---

Reprint requests to: Jens Erik Nielsen, Howard Hughes Medical Institute, Departments of Pharmacology, Chemistry, and Biochemistry, MC 0365, University of California, San Diego, La Jolla, CA 92093, USA; e-mail: jnielsen@mccammon.ucsd.edu; fax: (858) 534-7042.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.03114903>.



**Figure 1.** The general catalytic mechanism for retaining glycosyl hydrolases (A), and the alternative mechanism for HEWL (B). (A) I: Protonation of the glycosidic oxygen by the proton donor (Glu 35) and attack on the glucose C1 by the nucleophile (Asp 52). Departure of the reducing end of the substrate. II: Activation of a water molecule, cleavage of C1-Asp 52 covalent bond. III: Regeneration of the initial protonation states. The residue numbering corresponds to that of HEWL. (B) The protonation of the glycosidic oxygen by Glu 35 results in the formation of an oxo-carbenium ion which is generated by an  $S_N1$  elimination of the R-OH group (I). The highly charged intermediate (II) is relaxed by reacting with a water molecule that neutralizes the positive charge on the substrate and regenerates the initial protonation state of Glu 35 (III).

therefore expected to display an upward shift in the pKa value of the proton donor. Knowledge of the pKa values of the active-site residues can therefore aid in the identification of the reaction mechanism of the enzyme (Raquet et al. 1997; Lamotte-Brasseur et al. 2000; Ondrechen et al. 2001) and aid in the interpretation of experimental results (Raquet et al. 1997; Lamotte-Brasseur et al. 1999, 2000). Present pKa calculation packages obtain information on the pKa values of active-site residues by calculating the pKa values for all titratable groups in a protein. However, in many cases it is not necessary to obtain pKa values for more than two residues, because pKa calculations frequently are applied to answer questions such as “Which of these two acids is the proton donor?” or “Which of these two residues has the higher pKa?” (Nielsen and McCammon 2003). In addition, it is often necessary to obtain only a significant difference in pKa values between two residues in order to answer the above questions, and obtaining highly accurate pKa values for the two residues in question is therefore not so important as long as a clear distinction between the two can be made. Here we employed two simple criteria (Nielsen and McCammon 2003) to identify the proton donor from two candidate residues: We require the pKa value of the proton donor to be at least 5.0, and that the pKa value of the catalytic nucleophile is at least 1.5 units lower than the pKa value of the proton donor. These criteria are of course arbitrary and reflect our opinion on when we would be able to

confidently identify the proton donor from calculated pKa values.

We show that it is possible to calculate active-site pKa values just as accurately as in a full pKa calculation, by considering in detail only the environment of the active-site groups and their immediate electrostatic interaction partners, and by treating the rest of the titratable groups in an approximate way. Using our method and the criteria described above, it is thus possible to quickly and reliably identify the proton donor in an enzyme active site.

#### Calculating pKa values in proteins

Protein pKa calculations were performed as early as 1981 (Warshel 1981), and during the past two decades there has been much interest in developing and improving protein pKa calculation methodology and accuracy (Bashford and Karplus 1990; Gilson 1993; Yang et al. 1993; Antosiewicz et al. 1994; Karshikoff 1995; Demchuk and Wade 1996; Alexov and Gunner 1997; Raquet et al. 1997; Sham et al. 1997; Wlodek et al. 1997; van Vlijmen et al. 1998; Mehler and Guarnieri 1999; Sandberg and Edholm 1999; Nielsen and Vriend 2001). Present pKa calculation packages typically obtain electrostatic energies from solutions to the Poisson-Boltzmann equation (PBE), although other approaches for calculating protein pKa values exist (Sham et al. 1997; Mehler and Guarnieri 1999; Sandberg and Edholm 1999). We deal exclusively with the pKa calculation methodology based on PBE solvers, but the conclusions presented are directly applicable to most other pKa calculation methods. Numerous PBE solvers are available, and programs such as DelPhi II (Nicholls and Honig 1991), UHBD (Madura et al. 1995), and APBS (Baker et al. 2001) provide fast evaluation of various electrostatic energy terms. In pKa calculations, three types of energies are required: desolvation energies, background interaction energies, and site-site interaction energies (electrostatic interaction energies between pairs of titratable groups). The calculation methodology for calculating these energies has been described in detail elsewhere (Bashford and Karplus 1990; Yang et al. 1993) and will not be reviewed here. Briefly, desolvation energies and background interaction energies provide a description of how the immediate protein environment influences the pKa values of a particular titratable residue when disregarding the titration of all other titratable groups in the enzyme. The desolvation energy and background interaction energy can thus be used to calculate the so-called “intrinsic pKa” (Warshel 1981), which is the pKa value that a titratable group would have if all other titratable groups were fixed in their neutral state. Knowledge of the intrinsic pKa value for each residue and knowledge of all site-site interaction energies can be used to calculate the energy of every single possible protonation state of the protein at a particular pH. Since each titratable group has at least two protonation

states (histidine has four), a protein with  $N$  titratable groups will have at least  $2^N$  possible protonation states. For proteins with more than  $\sim 35$  titratable groups, the exact calculation of the charge on every titratable group at every pH value of interest using the Boltzmann sum becomes intractable, and consequently much work has been devoted to finding alternative procedures that produce accurate fractional charges at every pH value for any number of titratable groups. Some of the more popular methods include Monte Carlo sampling (Beroza et al. 1991) and the “cluster approach” (Gilson 1993), but other approaches have also been developed (Bashford and Karplus 1990; Yang et al. 1993). Once the fractional charges for a group are known at a range of pH values, it is trivial to determine the pKa values for the titratable groups, as the pH value at the midpoint of titration.

#### *Calculating pKa values for a subset of the titratable groups*

We describe a fast approach for calculating pKa values of active-site residues. It is well known that the immediate environment of a titratable group plays an important role in determining the pKa value of that residue, and that long-range interactions are less important (Lee et al. 1993; Sham et al. 1997). We therefore include a detailed description of the immediate environment of the titratable groups of interest, while we employ a less accurate description for groups further away from the interesting titratable groups. At the onset of the calculation, we define a set of titratable residues that contains the titratable groups for which accurate pKa values are needed. We call this set the “basis set.” We then employ an electrostatic cut-off value ( $E_{\text{cutoff}}$ ) to select a second set of titratable groups that interact strongly with the basis set. We define the “full subset” as the union of the basis set and this second set of residues. The full subset will be treated in detail in the calculations, whereas the remainder of the titratable groups in the protein are treated with less detail. More specifically, we calculate desolvation energies and background interaction energies only for the groups included in the full subset. Furthermore, we calculate the site-site interaction energy for a pair of titratable groups only when at least one of the groups in the pair is included in the full subset (for details on the selection procedure and on the calculation strategy see the Materials and Methods section). The rationale for this approach is that the individual contributions from titratable groups outside the full subset are very small (although the combined effect of several groups can be significant), and that it is therefore not necessary to obtain accurate pKa values for these residues. We assume that the groups that are not contained in the full subset have zero desolvation and background interaction energies, and we furthermore assume that these groups are interacting only with residues contained in the full subset. We therefore save the computation of the desolvation and

background interaction energies for the groups not contained in the full subset, and we also omit the very time-consuming calculation of the detailed site-site interaction energies involving neutral species for the groups not contained in the full subset.

Calculating pKa values of subsets of titratable groups has been performed before. Gilson (1993) split protein titratable groups into clusters and calculated titration curves for each individual cluster by evaluating the Boltzmann sum over all possible ionization states for the cluster. Interactions between clusters were taken into account by a mean-field approach. Bashford and Karplus (1990) employed a “reduced site” approach based on the exclusion of groups that are very close to being fully deprotonated or fully protonated. A hybrid statistical mechanical/Tanford-Roxby method was employed by Yang et al. (1993). In the latter method, the titration curve for each individual titratable group is computed by calculating the effects of groups closer than a cut-off distance by the Boltzmann sum, whereas groups further away are taken into account using the Tanford/Roxby approximation (Tanford and Roxby 1972). All of the methods referenced above perform full pKa calculations to obtain accurate pKa values for all residues. The method presented here calculates pKa values only for the active-site groups in an enzyme, but is still able to produce pKa values that are as accurate as the pKa values obtained from a full pKa calculation. The method achieves this by ignoring desolvation energies and background interaction energies for groups interacting weakly with the active-site groups. Additionally we ignore site-site interaction energies between residues that are both weakly linked to the active-site groups. This procedure results in a significant reduction in calculation time, and the accurate results suggest that the details of the electrostatic environment of groups far away from the active site are less important for determining the pKa values of active-site residues.

In the following we benchmark our procedure on three enzymes in order to obtain appropriate values for the cut-off value  $E_{\text{cutoff}}$ . The results show that it is indeed possible to obtain accurate pKa values for active-site residues using quite high values for  $E_{\text{cutoff}}$ .

## **Materials and methods**

### *Preparing X-ray structures*

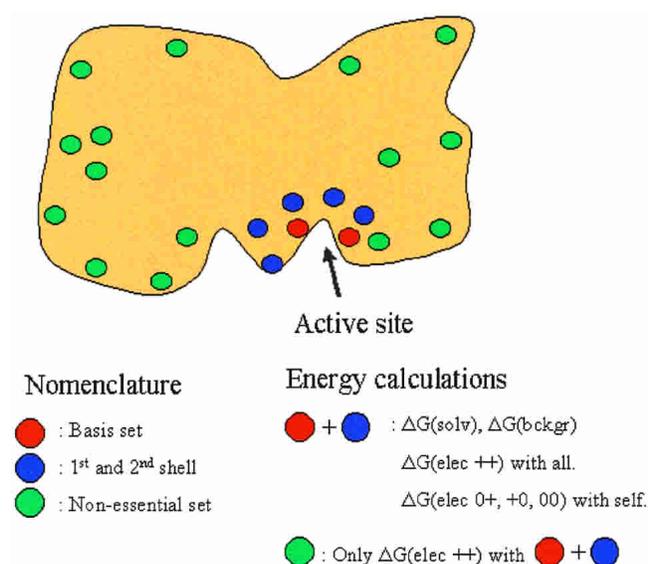
The structures for the proteins used in this study were downloaded from the Protein Data Bank (PDB; Berman et al. 2000). The specific files used were: HEWL (2LZT; Ramanadham et al. 1990), *Bacillus circulans* xylanase (BCX: 1XNB; Wakarchuk et al. 1994), and *Bacillus licheniformis*  $\alpha$ -amylase (BLA: 1BLI; Machiusi et al. 1998). All crystallographic water molecules were deleted prior to the pKa calculations, and all ions that were deemed to be a result of the crystallization conditions were removed.

### Selecting titratable groups for inclusion in the pKa calculations

We wish to obtain accurate pKa values for a subset of the titratable groups in a protein (the basis set). Whether a titratable group will be included fully in the pKa calculations (see Fig. 2) depends on the magnitude of the electrostatic interaction energy between the charged form of the residue and the charged form of the residues in the basis set. We calculate these energies by solving the PBE using DelPhi II (Nicholls and Honig 1991) for each of the residues in the basis set. We make two rounds of selections. In the first round of selection, a residue is selected for full treatment if the absolute interaction energy between it and any residue in the basis set is equal to or higher than  $E_{\text{cutoff}}$ . In the second round of selection, a residue is included if it has an absolute electrostatic interaction energy equal to or greater than  $2E_{\text{cutoff}}$  with any of the groups selected in the first round. The union of the basis set and of the titratable groups selected in the first and second rounds (the full subset) comprises the titratable groups that will be included fully in the pKa calculations. The remainder of the titratable groups in the protein constitutes the “nonessential set”. The selection procedure is very fast, and in the most unfavorable cases it constitutes no more than 5% of the total calculation time for HEWL and BCX, and only 0.2% of the total calculation time for BLA.

### Calculating pKa values

pKa values for the titratable groups in the protein were calculated as described previously (Nielsen and Vriend 2001), except that a single dielectric constant of 8 was used for the protein, instead of increasing the dielectric constant for residues with high B-factors and



**Figure 2.** The division of titratable groups used in the present pKa calculation scheme. Spheres indicate titratable groups. The figure shows a typical enzyme with most titratable groups at the surface and a higher concentration of titratable groups in and near the active site. The electrostatic energies calculated for each of the subsets are shown in the right text box.  $\Delta G(\text{elec ++})$  denotes charged-charged interaction energies,  $\Delta G(\text{elec } 0+, +0, 00)$  denotes neutral-charged, charged-neutral, and neutral-neutral interaction energies.

for residues with accessible alternative rotamers (Nielsen and Vriend 2001). The use of a single dielectric constant for the entire protein gives a slightly worse correlation with experimental pKa values, but illustrates the general applicability of the present method more convincingly. The OPLS force field (Jorgensen and Tirado-Rives 1988) was used as source of charges and radii. The linearized form of the Poisson-Boltzmann equation was solved using DelPhi II (Nicholls and Honig 1991) with the parameters set as described (Nielsen and Vriend 2001): solvent dielectric constant, 80; 65 cubed grid; final grid resolution, 0.25 Å/grd point for background interaction energies and 0.33 Å/grd point for desolvation energies; ion exclusion layer 2.0 Å; ionic strength, 0.144 M; surface probe radius, 1.4 Å.

Desolvation energies and background interaction energies were calculated only for the groups contained in the full subset. All other desolvation energies and background interaction energies were set to zero. Site-site interaction energies between the charged forms of titratable groups (charged-charged interaction energies) were calculated only if at least one of the titratable groups was a member of the full subset. Site-site interaction energies between the residues in the full subset and all other titratable groups are thus calculated and included in the pKa calculation. Charged-neutral and neutral-neutral interaction energies (Yang et al. 1993) were calculated only if the charged-charged interaction energy between the residues was stronger than 1 kT/e, and only if both groups of the pair were included in the full subset. T was at all times set to 298.15 K. Titration curves and pKa values were calculated for all titratable groups as described (Nielsen and Vriend 2001), and the titration of the groups in the nonessential set was thus allowed to influence the pKa values of the residues in the full subset. However, the omission of the charged-charged interaction energy, the desolvation energies, and the background interaction energies for a pair of titratable groups that both reside in the nonessential set introduces significant errors in the calculation of pKa values for the residues in the nonessential set.

The accuracy of the pKa values of residues in the basis set depends on the error of the pKa values for residues not contained in the basis set, and on the electrostatic interaction between the residues in the basis set and the rest of the residues. A smaller value of  $E_{\text{cutoff}}$  will include more titratable groups in the full pKa calculation procedure and thus produce results that are more similar to a “standard pKa calculation” (i.e., a pKa calculation where all groups are included fully). A higher value of  $E_{\text{cutoff}}$  will result in fewer groups being included fully and will thus produce results that are likely to be more different from the results of a full pKa calculation.

### Results

The fast calculation of active-site pKa values are of importance for the understanding of enzyme catalytic mechanisms. In the following, we establish an appropriate value for the cutoff parameter  $E_{\text{cutoff}}$  by examining the calculated pKa values for hen egg white lysozyme (HEWL), *Bacillus circulans* xylanase (BCX), and *Bacillus licheniformis*  $\alpha$ -amylase (BLA). All three enzymes belong to the family of retaining glycosyl hydrolases (see the CaZy database <http://afmb.cnrs-mrs.fr/CAZY/>; Coutinho and Henrissat 1999). HEWL and the two latter enzymes have all been proposed to employ the catalytic mechanism of retaining glycosyl hydrolases (Fig. 1A), and additionally an alternative mechanism involving a carbenium intermediate (Philips

1967) has been suggested for HEWL (Fig. 1A). The catalytic mechanism of the retaining glycosyl hydrolases (Davies and Henrissat 1995) involves two protein residues: a catalytic nucleophile and a proton donor. The reaction is initiated by the protonation of the glycosidic oxygen by the proton donor. The next step is a nucleophilic attack on the C1 on the nonreducing side of the glycosidic bond, followed by departure of the aglycon part of the substrate. Finally, a water molecule activated by the proton donor performs a nucleophilic attack on the C1, thereby breaking the covalent protein–substrate bond and regenerating the initial protonation states. The alternative mechanism for HEWL involves a highly charged carbenium intermediate, but is otherwise equivalent to the mechanism for the retaining glycosyl hydrolases, and the differences between the two mechanisms do not in any way affect the conclusions or results of this paper.

#### Establishing an appropriate value for $E_{\text{cutoff}}$

##### *Hen egg white lysozyme*

HEWL is a monomeric single-domain enzyme, which consists of an all- $\alpha$  region and a  $\beta$ -rich region. The active site is situated in a cleft between the two regions, and the two key active-site residues are Glu 35 and Asp 52 (Fig. 1). HEWL is a retaining glycosyl hydrolase (Family 22 in the CaZy database), and the pH activity profile is determined by the pKa values of Glu 35 (the proton donor) and Asp 52 (the catalytic nucleophile). Two different catalytic mechanisms have been proposed for HEWL (Philips 1967; Vocadlo et al. 2001), and the conclusions and interpretations presented in the present report are equally valid for both mechanisms, since these differ only in the nature of the intermediate (Fig. 1A,B).

Table 1 shows the results for the calculations of HEWL. It is seen that Glu 35 is correctly identified as the proton

donor even in the case where pKa values are calculated only for Glu 35 and Asp 52. The inclusion of more groups in the calculations changes the pKa values of the active-site residues insignificantly. This shows that the active site in HEWL constitutes a relatively uncoupled electrostatic system, where only the structural details of the two active-site acids and the electrostatic interaction with these two residues are important for maintaining the catalytically competent protonation state, whereas the structural details of the remainder of the titratable groups (and thereby their intrinsic pKa values) do not have to be calculated explicitly. This is true even though the electrostatic field originating from the remainder of the titratable groups has a profound impact on the pKa values in the active site (Dao-Pin et al. 1989; J.E. Nielsen, unpubl.).

##### *Bacillus circulans xylanase*

*Bacillus circulans* xylanase (BCX) belongs to Family 11 in the CAZY database, and has a  $\beta$ -jelly roll fold. The active site is located in a deep cleft between two layers of  $\beta$ -sheets, and employs the catalytic mechanism for retaining glycosyl hydrolases (Fig. 1A). In BCX, Glu 172 is the proton donor and Glu 78 is the catalytic nucleophile. McIntosh et al. (1996) used NMR titration data to show that the pKa value for the proton donor cycles during catalysis. Table 2 shows that we correctly identify Glu 172 as the proton donor in the catalytic mechanism. Furthermore, the pKa values of Glu 78 and Glu 172 are quite resistant to increasing the value of  $E_{\text{cutoff}}$ . At a value of 4.0 kT/e for  $E_{\text{cutoff}}$ , the sum of the difference in pKa values between the limited pKa calculation and the full pKa calculation is only 1.07 pKa units. The pKa values for Glu 78 and Glu 172 can be calculated correctly when the full subset consists of only these two residues. However, in this case Glu 172 is predicted to titrate back to its protonated state with a pKa value of 12.5, and then titrate to its deprotonated state again with a pKa value of

**Table 1.** Calculated pKa values for HEWL

Cutoff	pKa Glu 35	pKa Asp 52	Difference pKa Glu 35	Difference pKa Asp 52	# of groups in the full subset	Running time (as % of a full run)/total # of PBE runs <sup>a</sup>
Experimental values	6.20	3.68	0.63	1.78	—	—
Full pKa calc.	5.57	1.90	—	—	32/32 (100%)	100%/784
0.100 kT/e	5.57	1.90	0.0	0.0	32/32 (100%)	100%/784
0.200 kT/e	5.51	1.90	0.06	0.0	17/32 (53%)	69%/476
0.250 kT/e	5.51	1.87	0.06	0.03	15/32 (47%)	66%/420
0.500 kT/e	5.51	1.86	0.06	0.04	9/32 (28%)	53%/252
1.000 kT/e	5.51	3.00	0.06	1.10	7/32 (22%)	47%/196
2.000 kT/e	5.48	1.76	0.09	0.14	2/32 (6%)	22%/56
$\infty$ (only E35 & D52)	5.48	1.76	0.09	0.14	2/32 (6%)	22%/56

The pKa values of the two active site acids (Glu 35 and Asp 52) in hen egg white lysozyme (HEWL) calculated with different values for the cutoff parameter  $E_{\text{cutoff}}$ .

<sup>a</sup> Note that the number of PBE runs decreases more than the total running time. This is due to an initial overhead in the hydrogen-bond optimization procedure which is employed in the WHAT IF pKa calculations.

**Table 2.** Calculated pKa values for *Bacillus circulans* xylanase

Cutoff	pKa Glu 78	pKa Glu 172	Difference pKa Glu 78	Difference pKa Glu 172	# of groups in the full subset	Calculation time (as % of a full run)/total # of PBE runs <sup>a</sup>
Experimental values	4.60	6.70	1.97	0.72	—	
Full pKa calc.	2.63	5.98	—	—	40/40 (100%)	100%/1120
0.100 kT/e	2.63	5.98	0.0	0.0	40/40 (100%)	100%/1120
0.250 kT/e	2.62	5.96	0.01	0.02	36/40 (90%)	98%/1008
0.500 kT/e	2.56	5.94	0.07	0.04	24/40 (60%)	86%/672
1.000 kT/e	2.36	5.95	0.27	0.03	12/40 (30%)	51%/336
2.000 kT/e	2.09	5.69	0.54	0.29	7/40 (18%)	33%/196
4.000 kT/e	1.84	5.70	0.79	0.28	4/40 (10%)	14%/112
∞ (only E78 & E172)	1.85	5.64/17.2 <sup>b</sup>	0.78	0.34	2/40 (5%)	7%/56

The pKa values of the two active site acids Glu 78 and Glu 172 in *Bacillus circulans* xylanase calculated with different values for the cut-off parameter  $E_{\text{cutoff}}$ .  
<sup>a</sup> Note that the number of PBE runs decreases more than the total running time. This is due to an initial overhead in the hydrogen-bond optimization procedure which is employed in the WHAT IF pKa calculations.

<sup>b</sup> The titration curve for Glu 172 in this calculation displays a back-titrational behavior which is due to the incomplete description of the interactions between the titratable groups in the immediate surroundings of the residue. This results from the use of a too-high value for  $E_{\text{cutoff}}$  (∞ in this case).

17.2. This unphysical behavior is a result of the omission of the interaction energy between neighboring tyrosine residues close to the active site. BCX thus presents an example of a more tightly coupled system of titratable groups where it is essential to obtain correct pKa values also for the titratable groups in the immediate vicinity of the active-site acids.

#### *Bacillus licheniformis* α-amylase (BLA)

BLA has the common fold observed for all known α-amylases: a central TIM-barrel (domain A) is positioned roughly in between a β-rich domain (B) and a C-terminal domain (C). The active site is located in the interface between domains A and B, and forms an elongated cleft where glucose polymers (amylose and amylopectin) can bind (see Nielsen and Borchert 2000 for more information on α-amylases). The α-amylases are retaining glycosyl hydrolases, but the active site contains an additional acid that could play a direct role in the catalytic mechanism. This third residue was therefore included in the basis set to acquire information on the role of this residue. The basis set for BLA is

therefore Asp 231, Glu 261, and Asp 328. The experimental values of the active-site acids are not known for any member of the α-amylase family (or for the closely related cyclodextrin glucanotransferase family), and it is thus not possible to compare the calculated pKa values with the experimentally measured pKa values. However, the catalytic mechanism for the α-amylases is well established, and it was recently shown that Glu 261 is the proton donor, whereas Asp 231 is the catalytic nucleophile (Uitdehaag et al. 1999; Rydberg et al. 2002). The role of Asp 328 has not been identified, but roles in elevating the pKa value of Glu 261 (Klein et al. 1992; Knegt et al. 1995; Strokopytov et al. 1995), in substrate binding (Strokopytov et al. 1995; Uitdehaag et al. 1999), and a possible role in regulating the trans-glycosylating activity of α-amylases (Rydberg et al. 2002) have been suggested.

For BLA, the pKa values of the active-site acids can be calculated correctly (compared to the values calculated from the full pKa calculation) if a value for  $E_{\text{cutoff}}$  of less or equal to 2.0 kT/e is used (Table 3). A value of 2.0 kT/e includes only 7% (12 of 167) of the titratable groups, and it is thus

**Table 3.** Calculated pKa values for *Bacillus licheniformis* α-amylase

Cutoff	pKa Asp 231	pKa Glu 261	pKa Asp 328	% of groups in the full subset	Calculation time (as % of a full run)/total # of PBE runs <sup>a</sup>
Experimental values	low	high	unknown	—	
Full pKa calc.	1.77	>10.0	<0.0	167/167 (100%)	10%/4676
0.500 kT/e	1.36	>10.0	<0.0	82/167 (49%)	75%/2296
1.000 kT/e	0.40	>10.0	<0.0	44/167 (26%)	45%/1232
2.000 kT/e	<0.0	>10.0	<0.0	12/167 (7%)	14%/336
4.000 kT/e	<0.0	>10.0	>10.0	7/167 (4%)	8%/196
∞ (only D231, E261 & D328)	>10.0	>10.0	<0.0	3/167 (2%)	4%/84

The pKa values of the three active site acids Asp 231, Glu 261, and Asp 328 in *Bacillus licheniformis* α-amylase calculated with different values for the cutoff parameter  $E_{\text{cutoff}}$ .

<sup>a</sup> Note that the number of PBE runs decreases more than the total running time. This is due to an initial overhead in the hydrogen-bond optimization procedure which is employed in the WHAT IF pKa calculations.

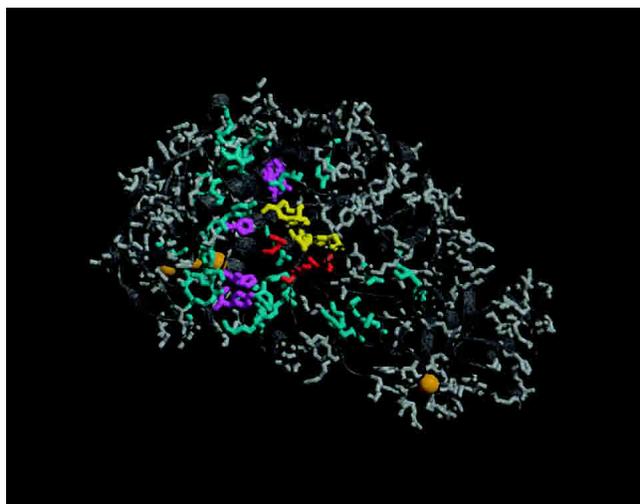
possible to obtain an adequate description of the system by including only a very small number of titratable groups (Fig. 3).

When using a value of 4.0 kT/e for  $E_{\text{cutoff}}$ , the pKa value for Asp 328 is predicted to be higher than 10, and thus not in agreement with the results of the full pKa calculation. The pKa values of both Asp 231 and Asp 328 are in disagreement with the full pKa calculation results when only the three active-site acids are included in the full subset. These results point to a much higher importance of the structural details for more titratable groups in the active site of BLA compared to HEWL and BCX.

To investigate the possible role of Asp 328 in elevating the pKa value of Glu 261, we set all interactions with Asp 328 to zero and recalculated the titration curves for Asp 231 and Glu 261. Removing the influence of Asp 328 lowers the predicted pKa value of Glu 261 by almost 7 units (from 18.1 to 11.2). However, Glu 261 is still predicted to be protonated at all pH values where BLA is enzymatically active, and according to our calculations Asp 328 is therefore not essential in maintaining the catalytically competent protonation state in BLA. A more comprehensive study on the electrostatics of the active-site residues in the  $\alpha$ -amylases is currently being undertaken by the authors.

## Discussion

We have demonstrated that it is possible to calculate accurate pKa values for active-site residues by treating in detail



**Figure 3.** Division of the titratable groups in *Bacillus licheniformis*  $\alpha$ -amylase. Side chains are shown only for titratable groups. The three active-site residues Asp 231, Glu 261, and Asp 328 (BLA numbering) are colored red. The orange spheres represent inorganic ions present in the X-ray structure. Groups colored yellow are the additional residues that constitute the full subset when  $E_{\text{cutoff}}$  is 2.0 kT/e. Additional titratable groups that are included in the full subset when  $E_{\text{cutoff}}$  is lowered to 1.0 kT/e are colored magenta. The cyan groups are included when  $E_{\text{cutoff}}$  is lowered to 0.5 kT/e. Groups that always are in the nonessential set are colored gray.

only a very limited number of titratable groups. The results presented for hen egg white lysozyme, *Bacillus circulans* xylanase, and *Bacillus licheniformis*  $\alpha$ -amylase show that a reliable identification of the proton donor in the active site can be achieved by using a value for  $E_{\text{cutoff}}$  equal to or less than 2.0 kT/e. Furthermore, we have been able to confirm that Asp 328 in BLA helps to elevate the pKa value of Glu 261, although we find Asp 328 to be nonessential for maintaining the catalytically competent protonation state.

Although the three enzymes studied employ the same catalytic machinery and all hydrolyze sugar polymers, the architectures of the three enzymes are quite different, and the conclusions presented here are therefore expected to be applicable to a wide range of enzymes where the catalytic mechanism involves a proton donor and a negatively charged catalytic nucleophile. The algorithmic improvements can furthermore be easily implemented in a wide range of pKa calculation algorithms and thus significantly reduce the calculation time needed for obtaining accurate pKa values for active-site residues.

The results show that the protonation states of a functional catalytic mechanism can be calculated when treating only a very limited number of groups in detail. This implies that enzyme active sites electrostatically are composed of a very small number of groups. In BLA, for example, it is possible to calculate the correct protonation states by treating only 7% (12 of 167) of the titratable groups in the enzyme in detail. Considering only the issue of having catalytically competent protonation states, our findings suggest that active sites are relatively autonomous entities, which function independently of the remainder of the enzyme. The electrostatic field of the remainder of the enzyme naturally plays a role in stabilizing the protonation states of the active-site residues, and we do indeed include site-site interaction energies between the residues in the full subset and all other titratable groups. However, the results presented here suggest that these interactions provide more of a “background electrostatic field” for the active site, and the position and structural details of the titratable groups outside the active site do not appear to be of importance. This is in accordance with the findings of the Warshel group (Lee et al. 1993; Sham et al. 1997) which suggest that long-range electrostatic interactions are less important for determining pKa values.

The findings of the present study are of importance for designing novel biocatalysts (Bolon and Mayo 2001) and may help in interpreting the results of artificial evolution studies (Powell et al. 2001). Studies to further elucidate the electrostatic autonomy of active sites are ongoing.

## Acknowledgments

J.E.N. is supported in part by a fellowship from the Danish Natural Research Science Council. Additional support was provided by

NIH, NSF, San Diego Supercomputing Center, National Biomedical Computation Resource, Center for Theoretical Biological Physics, HHMI, and the W.M. Keck Foundation.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

## References

- Alexov, E.G. and Gunner, M.R. 1997. Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. *Biophys. J.* **72**: 2075–2093.
- Antosiewicz, J., McCammon, J.A., and Gilson, M.K. 1994. Prediction of pH-dependent properties of proteins. *J. Mol. Biol.* **238**: 415–436.
- Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. 2001. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci.* **98**: 10037–10041.
- Bashford, D. and Karplus, M. 1990. pKa's of ionizable groups in proteins: Atomic detail from a continuum electrostatic model. *Biochemistry* **29**: 10219–10225.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Res.* **28**: 235–242.
- Beroza, P., Fredkin, D.R., Okamura, M.Y., and Feher, G. 1991. Protonation of interacting residues in a protein by a Monte Carlo method: Application to lysozyme and the photosynthetic reaction center of *Rhodobacter sphaeroides*. *Proc. Natl. Acad. Sci.* **88**: 5804–5808.
- Bolon, D.N. and Mayo, S.L. 2001. Enzyme-like proteins by computational design. *Proc. Natl. Acad. Sci.* **98**: 14274–14279.
- Coutinho, P.M. and Henrissat, B. 1999. Carbohydrate-active enzymes: An integrated database approach. In *Recent advances in carbohydrate bioengineering* (eds. H.J. Gilbert et al.), pp. 3–12. The Royal Society of Chemistry, Cambridge, UK.
- Dao-Pin, S., Liao, D.-I., and Remington, S.J. 1989. Electrostatic fields in the active sites of lysozymes. *Proc. Natl. Acad. Sci.* **86**: 5361–5365.
- Davies, G. and Henrissat, B. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure* **3**: 853–859.
- Demchuk, E. and Wade, R.C. 1996. Improving the continuum dielectric approach to calculating pKas of ionizable groups in proteins. *J. Phys. Chem.* **100**: 17373–17387.
- Gilson, M.K. 1993. Multiple-site titration and molecular modeling: Two rapid methods for computing energies and forces for ionizable groups in proteins. *Proteins* **15**: 266–282.
- Jorgensen, W.L. and Tirado-Rives, J. 1988. The OPLS potential functions for proteins: Energy minimizations for crystals for cyclic peptides and crambin. *J. Am. Chem. Soc.* **110**: 1657–1666.
- Karshikoff, A. 1995. A simple algorithm for the calculation of multiple site titration curves. *Protein Eng.* **8**: 243–248.
- Klein, C., Hollender, J., Bender, H., and Schulz, G.E. 1992. Catalytic center of cyclodextrin glycosyltransferase derived from X-ray structure analysis combined with site-directed mutagenesis. *Biochemistry* **31**: 8740–8746.
- Knegtel, R.M., Strokopytov, B., Penninga, D., Faber, O.G., Rozeboom, H.J., Kalk, K.H., Dijkhuizen, L., and Dijkstra, B.W. 1995. Crystallographic studies of the interaction of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251 with natural substrates and products. *J. Biol. Chem.* **270**: 29256–29264.
- Lamotte-Brasseur, J., Lounnas, V., Raquet, X., and Wade, R.C. 1999. pKa calculations for class A  $\beta$ -lactamases: Influence of substrate binding. *Protein Sci.* **8**: 404–409.
- Lamotte-Brasseur, J., Dubus, A., and Wade, R.C. 2000. pK(a) calculations for class C  $\beta$ -lactamases: The role of Tyr-150. *Proteins* **40**: 23–28.
- Lee, F.S., Chu, Z.T., and Warshel, A. 1993. Microscopic and semimicroscopic calculations of electrostatic energies by the POLARIS and Enzymix programs. *J. Comp. Chem.* **14**: 161–168.
- Machius, M., Declerck, N., Huber, R., and Wiegand, G. 1998. Activation of *Bacillus licheniformis*  $\alpha$ -amylase through a disorder→order transition of the substrate-binding site mediated by a calcium-sodium-calcium metal triad. *Structure* **6**: 281–292.
- Madura, J.D., Briggs, J.M., Wade, R.C., Davis, M.E., Luty, B.A., and McCammon, J.A. 1995. Electrostatics and diffusion of molecules in solution—Simulations with the University of Houston Brownian Dynamics Program. *Comput. Phys. Commun.* **91**: 57–95.
- McIntosh, L.P., Hand, G., Johnson, P.E., Joshi, M.D., Korner, M., Plesniak, L.A., Ziser, L., Wakarchuk, W.W., and Withers, S.G. 1996. The pKa of the general acid/base carboxyl group of a glycosidase cycles during catalysis: A <sup>13</sup>C-NMR study of *Bacillus circulans* xylanase. *Biochemistry* **35**: 9958–9966.
- Mehler, E.L. and Guarnieri, F. 1999. A self-consistent, microenvironment modulated screened coulomb potential approximation to calculate pH-dependent electrostatic effects in proteins. *Biophys. J.* **77**: 3–22.
- Nicholls, A. and Honig, B. 1991. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J. Comp. Chem.* **12**: 435–445.
- Nielsen, J.E. and Borchert, T.V. 2000. Protein engineering of bacterial  $\alpha$ -amylases. *Biochim. Biophys. Acta* **1543**: 253–274.
- Nielsen, J.E. and McCammon, J.A. 2003. On the evaluation and optimisation of protein X-ray structures for pKa calculations. *Protein Sci.* **12**: 313–326.
- Nielsen, J.E. and Vriend, G. 2001. Optimizing the hydrogen-bond network in Poisson-Boltzmann equation-based pK(a) calculations. *Proteins* **43**: 403–412.
- Ondrechen, M.J., Clifton, J.G., and Ringe, D. 2001. THEMATICs: A simple computational predictor of enzyme function from structure. *Proc. Natl. Acad. Sci.* **98**: 12473–12478.
- Philips, D.C. 1967. The hen egg white lysozyme molecule. *Proc. Natl. Acad. Sci.* **57**: 484–495.
- Powell, K.A., Ramer, S.W., Del Cardayre, S.B., Stemmer, W.P., Tobin, M.B., Longchamp, P.F., and Huisman, G.W. 2001. Directed evolution and biocatalysis. *Angew. Chem. Int. Ed. Engl.* **40**: 3948–3959.
- Ramanadham, M., Sieker, L.C., and Jensen, L.H. 1990. Refinement of triclinic lysozyme: II. The method of stereochemically restrained least squares. *Acta Crystallogr. B* **46**: 63–69.
- Raquet, X., Lounnas, V., Lamotte-Brasseur, J., Frere, J.M., and Wade, R.C. 1997. pKa calculations for class A  $\beta$ -lactamases: Methodological and mechanistic implications. *Biophys. J.* **73**: 2416–2426.
- Rydberg, E.H., Li, C., Maurus, R., Overall, C.M., Brayer, G.D., and Withers, S.G. 2002. Mechanistic analyses of catalysis in human pancreatic  $\alpha$ -amylase: Detailed kinetic and structural studies of mutants of three conserved carboxylic acids. *Biochemistry* **41**: 4492–4502.
- Sandberg, L. and Edholm, O. 1999. A fast and simple method to calculate protonation states in proteins. *Proteins* **36**: 474–483.
- Sham, Y.Y., Chu, Z.T., and Warshel, A. 1997. Consistent calculations of pKas of ionizable residues in proteins: Semimicroscopic and microscopic approaches. *J. Phys. Chem.* **101**: 4458–4472.
- Strokopytov, B., Penninga, D., Rozeboom, H.J., Kalk, K.H., Dijkhuizen, L., and Dijkstra, B.W. 1995. X-ray structure of cyclodextrin glycosyltransferase complexed with acarbose. Implications for the catalytic mechanism of glycosidases. *Biochemistry* **34**: 2234–2240.
- Tanford, C. and Roxby, R. 1972. Interpretation of protein titration curves. Application to lysozyme. *Biochemistry* **11**: 2193–2198.
- Uitdehaag, J.C., Mosi, R., Kalk, K.H., van der Veen, B.A., Dijkhuizen, L., Withers, S.G., and Dijkstra, B.W. 1999. X-ray structures along the reaction pathway of cyclodextrin glycosyltransferase elucidate catalysis in the  $\alpha$ -amylase family. *Nat. Struct. Biol.* **6**: 432–436.
- van Vlijmen, H.W., Schaefer, M., and Karplus, M. 1998. Improving the accuracy of protein pKa calculations: Conformational averaging versus the average structure. *Proteins* **33**: 145–158.
- Vocadlo, D.J., Davies, G.J., Laine, R., and Withers, S.G. 2001. Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature* **412**: 835–838.
- Wakarchuk, W.W., Campbell, R.L., Sung, W.L., Davoodi, J., and Yaguchi, M. 1994. Mutational and crystallographic analyses of the active site residues of the *Bacillus circulans* xylanase. *Protein Sci.* **3**: 467–475.
- Warshel, A. 1981. Calculations of enzymatic reactions: Calculations of pKa, proton transfer reactions, and general acid catalysis reactions in enzymes. *Biochemistry* **20**: 3167–3177.
- Wlodek, S.T., Antosiewicz, J., and McCammon, J.A. 1997. Prediction of titration properties of structures of a protein derived from molecular dynamics trajectories. *Protein Sci.* **6**: 373–382.
- Yang, A.S., Gunner, M.R., Sampogna, R., Sharp, K., and Honig, B. 1993. On the calculation of pKas in proteins. *Proteins* **15**: 252–265.