Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 25 (2007) 691–699
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Analysing the pH-dependent properties of proteins using pK_a calculations

Jens Erik Nielsen *

School of Biomolecular and Biomedical Science, Centre for Synthesis and Chemical Biology, UCD Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland

Received 19 March 2006; received in revised form 13 May 2006; accepted 13 May 2006
Available online 19 May 2006

Abstract

The results of protein pK_a calculations are routinely being analysed to understand the pH-dependence of protein characteristics such as stability and catalysis. Systems of functionally important titratable groups are identified from protein from pK_a calculations, but the rationalisation of the behaviour of such systems is inherently problematic due to a lack of theoretical tools and methods.

I present a number of novel methods for analysing the results of protein pK_a calculations which have been embedded in a graphical user interface (pKaTool). In the present paper I present novel methods for assessing the reliability of protein pK_a calculations and for analysing the roles of individual residues in determining active site pK_a values and the pH-dependence of protein stability. The methods presented are freely available to academic researchers at http://enzyme.ucd.ie/Science/pKa/pKaTool.

Keywords: Protein pK_a calculations; Enzyme active site electrostatics; Protein stability; pK_a value decomposition

1. Introduction

Electrostatic interactions are of major importance for the stability, binding characteristics and function of proteins. The electrostatic characteristics of a protein are, almost exclusively, determined by the ionisation states of the side chains of its constituent amino acid residues, although also the dipoles of the main chain have been shown to contribute to the electrostatic field [1]. Proteins modulate the pK_a values of amino acid side chains to achieve a specific pH-dependence of characteristics such as stability, enzymatic activity and binding specificity, and it is often non-trivial to understand how a specific pH-dependent behaviour arises from the individual interactions in a system of titratable groups.

Software packages for calculating pK_a values from protein X-ray structures are abundant [2–10] and recently also Web-based services offer protein pK_a calculations at the click of a button [8,11,12]. Typically pK_a calculation packages calculate titration curves, which are interpreted computationally to give the user a single pK_a value for each titratable amino acid residue. The agreement between calculated and experimentally observed pK_a values is quite impressive for the best of pK_a calculation methods, and generally pK_a values can be computed within ±0.5 units of the experimental values in most cases.

However, obtaining a set of calculated pK_a values is rarely a goal in modern biology, and in most cases the calculated pK_a values must be analysed to give information on the pH-dependent characteristics of the protein in question. During the analysis the protein researcher will want to address questions such as ''How reliable are the calculated pK_a values?'', ''Why is the pK_a value my residue perturbed?'' and ''How can I change the pH-dependence of the characteristic that I am studying?''.

1.1. Interpreting calculated protein pK_a values

The task of deciphering the pH-dependent behaviour of a protein is far from complete once one has obtained calculated pK_a values for a given protein structure with a state-of-the-art pK_a calculation package. In the simplest application of pK_a calculations, enzymologists are interested in assigning the proton donor in an enzyme catalytic mechanism [13–15] by

Abbreviations: HEWL, hen egg white lysozyme; BLI, *Bacillus licheniformis* α-amylase

* Tel.: +353 1 716 6724; fax: +353 1 716 6898.
E-mail address: Jens.Nielsen@ucd.ie.

1093-3263/$ – see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2006.05.007

identifying an active site residue with an elevated pK_a value. It has already been demonstrated that proton donor identification is sensitive to the details of the PDB file used for the calculations [16], but in addition to these considerations it is important to inspect the calculated titration curves carefully to ascertain that the calculations are interpreted correctly.

It is of particular importance to verify that titration curves are indeed Henderson–Hasselbalch-like, since the reported pK_a values otherwise are meaningless. It is also important to keep in mind that the enzyme could be operating by a reverse protonation state mechanism [17,18], and that the substrate and/or conformational changes could perturb the active site electrostatic environment and thereby the calculated titration curves significantly. In addition, it is of interest to examine if the enzyme employs ''pK_a cycling'' [19], and it might also be of interest to establish which titratable groups influence the pK_a values of the catalytic residues so as to maintain the ''catalytically competent'' protonation state of the enzyme.

When interpreting calculated pK_a values for other purposes than proton donor identification, the number of analyses of interest increase, since one typically has to compare the pK_a values of an apo/unfolded form with those from a holo/folded form. This is the case for the pH-dependence of protein–ligand binding [20] and protein–protein binding, but also in the case of the pH-dependence of protein stability [21]. In these cases, and also for the purpose of proton donor identification, it is important to be able to perform a rigorous analysis of the behaviour of a system of titratable groups when perturbed in various ways.

Here I present novel methods for analysing the pH-dependent characteristics of a protein using pre-calculated pK_a values. These methods are embedded in a new interactive graphical tool (pKaTool) which makes it possible to study many aspects of the pH-dependence of protein characteristics at the click of a button. The graphical interface makes it possible for wet-lab scientists to study the biology of their protein without investing large amounts of time in running complicated calculations and installing complicated pK_a prediction software.

In the following I use pKaTool for analysing the pH-dependent characteristics of protein functional sites. I have chosen to present examples that illustrate how to one can analyse the pH-dependence of protein stability and catalytic activity.

I use two well-known enzymes (hen egg white lysozyme and *Bacillus licheniformis* α-amylase) to illustrate the use of these methods, but the methods are applicable to any protein system.

In addition to its scientific value, pKaTool is ideal for developing an intuitive understanding of the basis of pK_a shifts in proteins and provides functions for producing figures for illustrating the behaviour of systems of titratable groups.

1.2. Calculating pK_a values of protein titratable groups

The pK_a value of a titratable group is defined as $-\log(K_A)$, where K_A is the equilibrium constant for a general acid–base reaction. The pK_a value thus describes the pH-dependence of the $r_{crg} = [A^-]/[HA]$ ratio for a given titratable group. In addition to being pH-dependent, r_{crg} depends on the free energy difference between the protonated (HA) and deprotonated (A^-) forms of the titrable group in question. In the equations above I have used a titratable group with a neutral protonated state, but the equations are also valid for titratable groups that have a positive charge in the protonated state. Proteins can alter the energy difference between HA and A^- by selectively interacting favourably or unfavourably with either of the two states. The interactions between the two states of titratable groups are traditionally divided into pH-independent interactions (hydrogen bonds, permanent charges and dipoles, desolvation effects), and pH-dependent interactions (electrostatic interactions between titratable groups). The pH-independent effects of the protein are typically added to pK_a value that the titratable group is assumed to have in solution (the model pK_a value) to give the so-called intrinsic pK_a value of the group in the protein. The intrinsic pK_a value of a titratable group in a protein is the pK_a value that the residue would have in the absence of all other titratable charges.

The intrinsic pK_a values of a system are used as a starting point for modelling the pH-dependent effects on pK_a values. This is typically done by calculating the titration curves of all titratable groups in a range of pH values, while taking into account the effect of the interactions between titratable groups. Titration curves can be calculated using a Monte Carlo-based sampling scheme [22], an explicit evaluation of the Boltzmann sum or a hybrid method [7] incorporating elements of the Tanford–Roxby algorithm [7,23]. The final pK_a value for each titratable group is found by identifying the pH where the group is half-protonated.

In the case of simplified pK_a calculation schemes, pH-dependent interactions are sometimes simply added to the intrinsic pK_a values, and in these cases titration curves are therefore not calculated explicitly.

The physical interactions that determine the intrinsic pK_a values, and the final pK_a values can be determined by a number of different approaches, but typically electrostatic energies are obtained from solutions to the Poisson–Boltzmann equation (PBE) [24,25], generalised Born (GB) models [26,27], or semi-empirical approaches [5,8]. All of these approaches give pK_a values of similar accuracy with the GB models being somewhat faster than PBE-based methods [27], and the semi-empirical methods being much faster than the GB and PBE approaches.

In summary, pK_a values can be obtained using a large number of software packages. Each software package has its own advantages and disadvantages, but common to all is that it is imperative to analyse the calculated results properly to gain an accurate picture of the biological implications of the results. In the following I present methods that are specifically designed for the pK_a calculation methods that calculate protein titration curves explicitly. I limit myself to studying this type of pK_a calculation methods because they accurately model the pH-dependent behaviour of a strongly coupled system of titratable groups.

2. Results

pK_a calculations can be used to identify elevated pK_a values in enzyme active sites, to analyse the pH-dependence of protein stability and ligand binding, and ultimately to explain the

pH-dependence of protein function. However, the output of standard pK_a calculation packages consists only of floating point pK_a values, and hence it is quite laborious to perform sophisticated analyses of these results without devoting a significant effort to running extra calculations and/or manual inspection of the results.

I present novel computational methods and procedures for maximizing the knowledge-gain from protein pK_a calculations. I focus on a number of problems of particular interest to protein scientists when interpreting protein function, and show how novel information on enzyme active sites and pH-stability profiles can be obtained with a set of simple analyses. The methods and procedures presented here have been incorporated in a graphical tool, which is freely available to academic researchers.

2.1. Interpreting calculated pK_a values—the basics

In the majority of cases the results of protein pK_a calculations are reported simply as a floating point values (pK_a values), although the primary calculational result is titration curves for all titratable groups. pK_a values are reported for benchmarking reasons (calculated pK_a values are compared to experimental pK_a values), for reasons of simplicity (a single value versus a full titration curve), and when using simplified computational models. In the vast majority of cases, calculated (and experimentally determined) titration curves are modelled quite adequately by a single pK_a value. However, in the complex titratable environment of enzyme active sites, calculated titration curves are often bi-phasic or show other irregular behaviour [28]. Additionally bi-phasic titration curves have been observed both in proteins and for simple chemical compounds [29]. Therefore, it is important to directly inspect the calculated titration curves of the residues of interest to ascertain that they can be modelled by a single pK_a value.

The first step of any analysis of calculated pK_a values should therefore be to calculate the fit of each titration curve to the Henderson-Hasselbalch (HH) equation ($\mathrm{pH} = b + a\log([\mathrm{A}^-]/[\mathrm{HA}])$). Such an analysis can often yield surprising results, and thus it is found that more than half of the calculated titration curves for the active site residues of *B. licheniformis* α-amylase (BLI) display values of a significantly different from 1 when fitted to the HH equation.

It has been found that active sites often contain non-HH shaped titration curves, and this phenomena has been exploited for active site identification [30]. Indeed a significant fraction of the most irregular titration curves for BLI are situated in and around the active site.

2.2. Are non-HH titration curves real?

Irregular titration curves are a product of strong electrostatic interaction energies, which can be made artificially strong by the limited modelling of protein dynamics employed in most protein pK_a calculation packages. Furthermore, the magnitude of electrostatic interactions is heavily influenced by the dielectric constant used for the calculation. A low dielectric constant (e.g. 2 or 4) will significantly increase electrostatic interaction energies and thus the irregularity of the calculated titration curves as compared to a calculation with the same model using a higher dielectric constant. Similarly, a less explicit modelling of protein dynamics will tend to increase electrostatic interaction energies by not allowing the protein to respond to protonation state changes (see [31] for a discussion of how protein dielectric constants relate to models of protein dynamics). In an example of this behaviour, a for His 15 in HEWL changes from 0.92 (almost perfect HH) to 0.72 when the protein dielectric constant used in the WHAT IF [2] pK_a calculation is lowered from 8 to 2.

Irregular titration curves are thus clearly dependent on the parameters and the model used for the calculations. However, if titration curves remain irregular after exploring the response of the system to changes in the dielectric constant and changes in the individual interaction energies (see below), then it likely that the irregular titration curves are not an artefact of the calculations.

2.3. Sensitivity to small structural perturbations

The accuracy of the positions of atoms in X-ray structures is subject to the resolution of the electron density map, the interpretation of the electron density map by the crystallographer, and the refinement procedure used when producing the final coordinates.

The pair-wise electrostatic interaction energies and intrinsic pK_a values are almost exclusively calculated from X-ray protein structures, and consequently these are also influenced by the accuracy of the atom positions in the X-ray structure used. However, there is no straight-forward relation between the accuracy of a protein pK_a calculation and the resolution of the X-ray structure used [16], and it is therefore of interest to examine the sensitivity of pK_a calculation results to small perturbations of the X-ray structure. Calculated pK_a values for slightly perturbed X-ray structures of Crambin, hen egg white lysozyme and *B. licheniformis* α-amylase (see Section 4) showed significant differences to those calculated from unperturbed structures. The variation of the pK_a values could be ascribed to changes both in the calculated intrinsic pK_a values, and changes in the pair-wise electrostatic interaction energies. 90% of the variation in the pair-wise electrostatic interaction energies can be described by a ±20% change in the interaction strength (data not shown), whereas the difference in the desolvation energies and background interaction energies (these determine the instrinsic pK_a values) can vary up to 50%. The variations are only rough guidelines, but describe the reliability of the coordinates of an average X-ray structure with a resolution around 2.0 Å quite well. For higher resolution X-ray structures smaller variations in energies can be expected.

Given that these energies can vary by this much, it is prudent to ask which consequences this will have on the calculated pK_a values for a given protein. In particular, it is of importance to examine if the biological conclusion reached from a pK_a calculation will change due the variation in the intrinsic pK_a values and pair-wise electrostatic interaction energies.

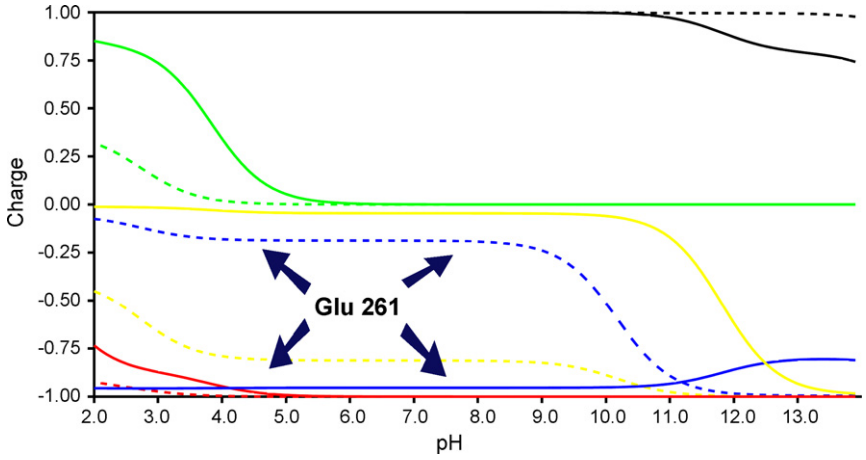

Fig. 1. The maximum change in the titration curve of E261 (blue) before (dashed) and after (full) perturbing the underlying energetic parameters of the system. The dramatic change in the protonation of E261 is a result of the strong coupling between D328 (yellow) and E261 combined with changes in the interactions between these two residues and the rest of the system (see text).

2.4. Sensitivity of proton donor identification

A common use for pK_a calculations in modern biology is proton donor identification in catalytic mechanisms. HEWL and BLI constitute two systems for which the proton donor can be predicted from calculated titration curves, and it is of interest to study if the proton donor identification is influenced by varying the intrinsic pK_a values and interaction energies as mentioned above. In the following I report the change in the titration curves of the two enzymes when considering a sub-set of the titratable residues in HEWL and BLI, and when performing a full combinatorial variation of all individual intrinsic pK_a values and pair-wise interaction energies. Such a sensitivity analysis can be performed within minutes with pKaTool when a sub-set of the titratable residues is used, while an analysis of the full system in such a way potentially takes days to complete using MD-based methods (see [16,32] for examples of an MD-based analyses of the effect of protein flexibility).

For HEWL I select a sub-set of titratable groups that form a tightly coupled cluster around E35. The cluster consists of residues D48, D52, R61 and D66, and the cluster reproduces the titrational behaviour of the catalytic residues quite closely. Furthermore, not a single combination of perturbed energies of the system influences the relative titrational behaviour of E35 and D52 (the catalytic nucleophile) significantly, and consequently from an analysis of this particular X-ray structure (PDBID: 2LZT), it can confidently be concluded that Glu 35 is the proton donor in the HEWL catalytic mechanism. However, this is not necessarily the case for other X-ray structures of HEWL since the energetic contributions to the pK_a value of E35 can be markedly different depending on the X-ray structure used [16]. Thus, a sensitivity analysis of a HEWL pK_a calculation from a different structure, might show that Glu 35 cannot be confidently identified as the proton donor.

In the case of BLI the system of titratable groups behaves radically different. The sub-set of residues consists of R229, D231, E261, H327 and D328 qualitatively reproduces the titration curves of the full system although several differences exist. A full combinatorial perturbation of the system parameters identifies a state where E261 is deprotonated at all pH values (Fig. 1) and consequently we cannot confidently identify E261 as the proton donor from these calculations. A single 20% change in the interaction energy between R229 and E261 is sufficient to yield identical titration curves for E261 and D328, thus invalidating the identification of the proton donor.

However, the results of a sensitivity analysis is critically dependent on the sub-set chosen for the analysing, and one should include at least two layers of strongly interacting residues to obtain a reliable sensitivity analysis [33]. For BLI the behaviour of the cluster above changes significantly when D100 is included in the sub-set. In this case a change in a single energy no longer changes the titration curve of E261. A closer inspection of the energetics of the system reveal that D100 preferentially stabilizes the protonated state of E261 (as compared to the protonated state of D328) and therefore provides extra stability to the system.

It should be noted that many pK_a calculation packages attempt to model protein flexibility implicitly or explicitly [2,6,9,10,34–38] and thus remedy the sensitivity of the calculated pK_a values to small perturbations in the X-ray structures used. However, even when using a pK_a calculation algorithm, which models all protein dynamics explicitly, it is still of importance of assess the sensitivity of the subsystem in question to small structural perturbations. In these cases the method presented here can be used as a quick scan for potential problems since it presents a worst-case scenario. Subsequently any subsystem found to be unstable should be investigated further with a larger structural ensemble.

2.5. Identifying amino acid residues that influence active site pK_a values

Active site pK_a values determine the pH-activity profiles of enzymatic activity, and the pH-dependence of protein stability is also determined by the perturbation of protein pK_a values. Since it often is of interest to re-engineer pH-stability profiles

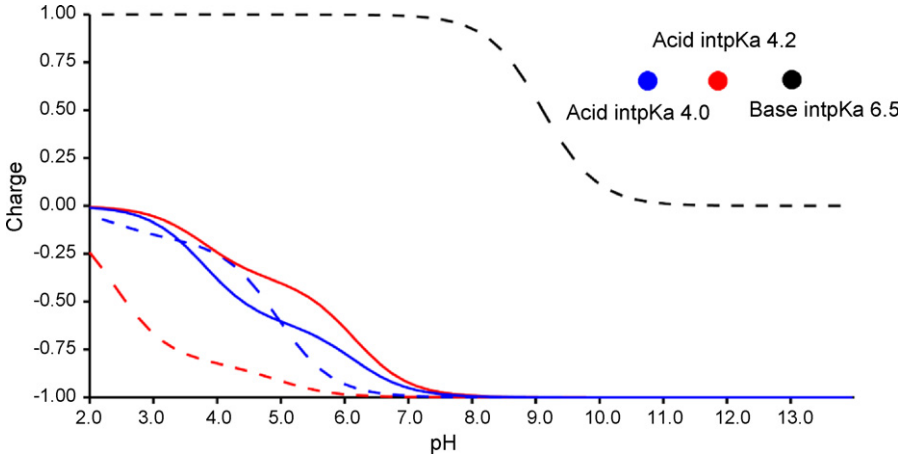

Fig. 2. A hypothetical system of three titratable groups on a straight line. The centre group is an acid with intrinsic pK_a 4.2, and the two outer groups are an acid and a base with intrinsic pK_a values of 4.0 and 6.5, respectively. Assuming a uniform dielectric constant the magnitude of the electrostatic interaction energy between the outer groups will be exactly half of that between any of the outer groups and the centre group. The present system the interaction energies between neighbouring groups are 4.0 kT, whereas the interaction energy between the two outer groups is 2.0 kT. The dashed lines (before) illustrate the behaviour of the system when all three groups are present, and the full lines (after) show the behaviour when the base has been removed from the system. It is seen that removing the base has the counter-intuitive effect of making the blue acid more negative at certain pH values.

and pH-activity profiles of enzymes, it is essential to rationalise the determinants of protein pK_a values in order to construct the correct point mutations. It is furthermore of high academic interest to decipher the sometimes complicated pH-dependent effects in natural and engineered enzymes in an evolutionary context [14,18].

For small systems of titratable groups it is straight-forward to estimate the impact of a titratable group on the pK_a value of another, but even for three-group systems this analysis becomes non-intuitive when strong electrostatic interactions are involved (see Fig. 2). Since titratable groups in active sites generally interact strongly, it becomes essential to have a quantitative methods for analysing the behaviour of a system consisting of strongly coupled titratable groups.

2.6. Decomposing pK_a values

Essentially scientists want to measure the effect of each titratable group on the pK_a values of all the remaining titrable groups, and in this way identify the key amino acids that determine the pH-dependent behaviour of the protein. However, the effect of a titratable group (X) on a specific target pK_a value (e.g. the pK_a value of a proton donor) depends on the context of X and the target. If a protein has only two titratable groups (X and the target) then the pH-dependent behaviour of the protein depends exclusively on interaction energy between X and the target, and their individual intrinsic pK_a values. However, in a real protein there will be many other titratable groups interacting with X and the target. These other titratable groups might enhance or diminish the influence of X on the target, and in such cases the effect of X becomes as much a function of its environment as a function of its direct electrostatic interaction with the target.

In Section 4.6, I describe two methods for measuring the influence of a titratable group on a target pK_a value. Briefly I measure the influence of X on the target when the two groups are isolated from the protein, and when the two groups are in the protein. In both cases I measure the integrated change in the titration curve of the target when I remove X. I call the corresponding energy change $\Delta pK_{a_{ex\text{-}system}}$, $\Delta pK_{a_{in\text{-}system}}$, respectively. The difference between $\Delta pK_{a_{ex\text{-}system}}$ and $\Delta pK_{a_{in\text{-}system}}$ ($\Delta\Delta pK_{a_{system}}$) describes the difference in the effect of X on the pK_a value of the target caused by other parts of the system.

In the following I examine the values of $\Delta pK_{a_{ex\text{-}system}}$, $\Delta pK_{a_{in\text{-}system}}$ and $\Delta\Delta pK_{a_{system}}$ for the active site residues of BLI.

2.7. Decomposing the active site of BLI

The active site of BLI consists of a set of strongly coupled titratable groups. Experimental evidence has shown D231 to be the nucleophile in the catalytic mechanism, while E261 is likely to be the proton donor [39]. A full protein pK_a calculation for this system identifies E261 as the proton donor since it has a high pK_a value. Initial inspection of the system shows that E261 has an intrinsic pK_a of 5.0, and that strong interactions with other titratable groups elevate its pK_a value by an additional 13 units. The question is now which groups are responsible for increasing the pK_a value of E261.

Since it is infeasible to perform an interactive perturbation analysis on the full system of 167 titratable groups in BLI, I select a sub-set of titratable residues that interact strongly with D231 or E261. Using a cut-off value of 4 kT/e for the electrostatic interaction produces a final set of titratable groups that contains Y56, D100, R229, D231, E261, H327 and D328. Calculation of the titration curves for this subsystem reveals that the interactions present in this subsystem manage to reproduce the titration curves for E261 and D231 to a good approximation.

Table 1 shows the $\Delta pK_{a_{ex\text{-}system}}$, $\Delta pK_{a_{in\text{-}system}}$ and the $\Delta\Delta pK_{a_{system}}$ values for E261 in the active site of BLI. Broadly speaking the pK_a value is elevated by Asp 100, Asp 231 and Asp 328 by a total of 10.8 pK_a units in-system (9.4 ex-system). The pK_a value is depressed by a total of 3.9 units by Arg 229

Table 1
Decomposition of the pK_a value of Glu 261 in the active site of *Bacillus licheniformis* α-amylase

Residue removed	$\Delta pK_{a_{in\text{-}system}}$, Glu 261	$\Delta pK_{a_{ex\text{-}system}}$, Glu 261	$\Delta\Delta pK_{a_{system}}$, Glu 261
Asp 100	−1.2	−1.2	0.0
Arg 229	3.9	3.2	0.7
Asp 231	−1.5	−1.4	0.1
Glu 261	–	–	–
His 327	0.0	0.0	0.0
Asp 328	−7.1	−6.8	0.3

in-system (3.2 ex-system). This should result in a pK_a value elevation of 6.9 units in-system and 6.2 units ex-system. The total ΔpK_a due to electrostatic effects is 6.4 in the reduced system, which reflects that the effect of some groups on E261 is context-dependent (i.e. their $\Delta\Delta pK_{a_{system}}$ is non-zero). Arg 229 is a prime example, with the effect of Arg 299 being 0.7 larger ''in-system'' than ''ex-system''. A detailed analysis of such system-dependent enchancements of pK_a effects is beyond the scope of the present paper, but essentially Arg 229 modulates the titrational behaviour of E261 and D328 by selectively interacting most favourably with the charged state of E261.

The titrational behaviour and the roles of individual protein residues can be analysed in this way, and active site pK_a values can be re-engineering by removing titratable residues and by changing acids to bases and vice versa using pKaTool. The insertion of novel titratable residues close to the active site raises a set of different problems, which are beyond the scope of the present paper but are addressed with the recent release of the pKD Webserver at http://enzyme.ucd.ie/pKD [12,40].

2.8. Identifying residues responsible for changes in protein stability at high- and low-pH

The pH-dependence of protein stability is determined by the difference in the titrational behaviour of protein residues between the denatured and folded form [21,41], and protein pK_a calculations can be used to get an accurate prediction of the electrostatic contribution to the pH-dependent stability of proteins. Proper analysis of the results of protein pK_a calculations, can furthermore pin-point specific residues whose pK_a values should be changed in order to re-engineer the pH-stability profile. Fig. 3 shows the electrostatic contribution to the pH-dependence of protein stability as analysed with pKaTool. The figure depicts the total electrostatic contribution to protein stability. This kind of analysis allows for the straightforward identification of residues that stabilises the protein at various pH values, and those that destabilise the protein.

2.9. pKaTool

The methods presented here have been embedded in a freely available graphical tool (pKaTool) along with several other functions to allow easy interpretation of calculated pK_a values at the click of a button.

pKaTool consists of two main modules: a module for visualising and manipulating the titration curves from protein pK_a calculations, and a module for analysing the behaviour of small systems of titratable groups (pKaSystem). pKaTool allows the user to inspect calculated titration curves, provides detailed information on the energies contributing to perturbed pK_a values, contains functions for finding titratable groups with specific properties, and allows the user to dissect the

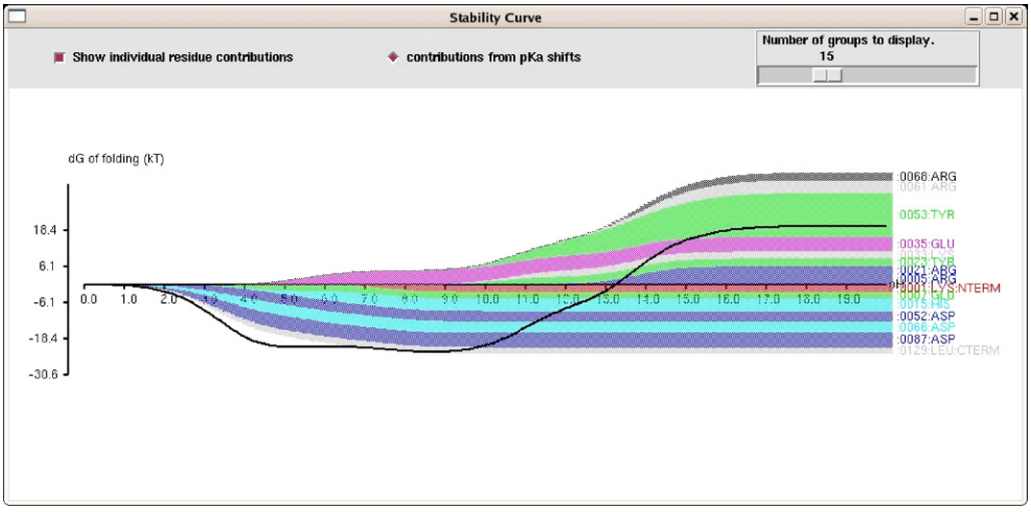

Fig. 3. The electrostatic contribution to the pH-dependence of protein stability calculated for HEWL (PDBID: 2LZT). Only the 15 titratable groups that contribute most to the stability profile are shown. The black line represents the contribution to stability due to differences in pK_a values between the unfolded and folded form of the protein. The shaded areas show individual residue contributions to the pH-stability profile originating from perturbed pK_a values.

pH-dependent stability of the protein. The input required for pKaTool consists of a completed pK_a calculation for a specific PDB file, or alternatively pK_a values can be calculated with a ''quick-and-dirty'' pK_a calculation facility in the program.

pKaSystem is a sub-module of pKaTool which can be used as stand-alone or as an in-detail analysis facility within pKaTool. pKaSystem allows the user to manipulate the individual energetic terms of a system of titratable groups. Thus the intrinsic pK_a values and pair-wise electrostatic interaction energies of a system of titratable groups can be changed individually, and the resulting changes in the titration curves can be observed immediately. pKaSystem furthermore provides functions for assessing the sensitivity of a system of titratable groups to small changes in the underlying energies, functions for calculating pH-dependent stability, and contains functions for decomposing the pK_a values of all groups. pKaTool and pKaSystem are integrated with the pKD Webserver at http://enzyme.ucd.ie/pKD [12] for redesigning protein pK_a values. Future developments will focus on enabling pKaTool to communicate directly with the pKD server so that potential site-directed mutations for changing pH-dependent characteristics of a protein can be constructed *in silico* and subsequently the resulting perturbed pK_a values can be inspected with pKaTool. pKaTool is freely available to academic researchers at http://enzyme.ucd.ie/Science/pKa/pKaTool.

3. Conclusion

Protein pK_a calculations are becoming a standard part of the protein scientist's toolkit, and with the publication of several Web-based resources for performing pK_a calculations it is now possible for everyone to calculate pK_a calculations for an X-ray structure. Advances in computer power and algorithms have furthermore made it possible to perform large-scale studies of calculated pK_a values [42,43], and have provided significant insight into the pH-dependence of protein characteristics.

With the large amounts of theoretical data now readily available, it is prudent to establish guidelines for assessing the reliability and stability of this data. pKaTool, which is presented in this paper, provides simple functions for assessing and analysing the reliability of standard FDPB-based pK_a calculations, and indeed any pK_a calculation method that produces intrinsic pK_a values, pair-wise electrostatic interaction energies and titration curves.

From the results presented here it is clear that systems of titratable groups display complicated and non-additive behaviour when perturbed. When analysing protein pK_a calculations and ascribing roles to certain titratable groups, it is therefore necessary to test one's hypotheses using a method similar to the decomposition method used on BLI.

There is still much to be learned about the behaviour of systems of titratable groups in proteins, and their role in the evolution of proteins and enzymes. Hopefully pKaTool will encourage more research and interest in the properties of systems of titratable groups, which ultimately will yield novel methods for redesigning the pH-dependent properties of proteins and enzymes.

4. Materials and methods

4.1. Software

The pKanalyse tool was written using Python [44] and its Tk/Tcl binding Tkinter. The program can be obtained free of charge for academic researchers from http://enzyme.ucd.ie/Science/pKa/pKaTool.

4.2. Calculation of pK_a values

pK_a values, intrinsic pK_a values and pair-wise interaction energies for hen egg Crambin, hen egg white lysozyme (HEWL) and *B. licheniformis* α-amylase (BLI) were calculated with the WHAT IF pK_a calculation package as described previously [2] with the modification that a single uniform dielectric constant of 8 was used for the protein.

The PDB entry 1CRN was used for Crambin, 2LZT was used for HEWL, while 1BLI was used for BLI. All X-ray structures were stripped of water molecules and ions deemed to be crysta-llisation artefacts before being submitted to pK_a calculations.

4.3. pK_a value calculations in pKaTool

pK_a values in pKaTool are calculated using either the Tanford–Roxby method, Monte Carlo simulations or by evaluating the Bolztmann sum explicitly as chosen by the drop-down menu. Fractional charges are calculated from pH 2.0 to 14.0 at regular intervals as specified by the pH-step parameter. pK_a values are determined as the lowest pH value where the fractional degree of protonation drops below 0.5.

4.4. Selection of strongly interacting sites

Strongly interacting sites are selected simply by selecting titratable groups that have charged–charged electrostatic interaction energies greater than a certain cut-off value with one of more of the groups that are being studied.

4.5. Generation of slightly perturbed protein structures

Slightly perturbed protein structures were generated using the ''random'' function of WHAT IF [45], which displaces the coordinates of all atoms by in a protein by a random distance between 0 and 0.25 Å in a random direction.

4.6. Decomposing pK_a values

The effect of a titratable group (X) on the titration curve of another group (Y) embedded in a larger system (N, consisting of X, Y and other titratable groups Z) of titratable groups can be determined in two ways:

1. By calculating the titration curve of Y in N, and subsequently subtracting the titration curve of Y calculated in N-X.

2. By calculating the titration curve of Y in a system consisting only of X and Y, and subsequently subtracting the titration curve of Y on its own.

Method 1 gives the "in-system" decomposition energy ($\Delta G_{\text{in-system}}$), and method 2 yields the "ex-system" decomposition energy ($\Delta G_{\text{ex-system}}$). Both of these energies can be converted into ΔpK_a values ($\Delta pK_{a_{\text{in-system}}}$ and $\Delta pK_{a_{\text{ex-system}}}$) to give a more intuitively understandable quantity.

The difference between $\Delta G_{\text{in-system}}$ and $\Delta G_{\text{ex-system}}$ is called $\Delta\Delta G_{\text{system}}$, and the corresponding sign-corrected pK_a value is called $\Delta\Delta pK_{a_{\text{system}}}$. $\Delta\Delta pK_{a_{\text{system}}}$ is defined as $-\gamma(\Delta pK_{a_{\text{in-system}}} - \Delta pK_{a_{\text{ex-system}}})$, where γ is -1 for acids and $+1$ for bases. A non-zero $\Delta\Delta pK_{a_{\text{system}}}$ for a particular interaction of X with Y, but also of the interactions between the remainder of the groups in the protein. Positive of $\Delta\Delta pK_{a_{\text{system}}}$ bigger than 0 indicate that the system enchances the effect of X, while negative values indicate that the system diminishes, or reverses (values below -1) the effect of X on Y.

4.7. Estimating the sensitivity of pK_a calculations to small structural errors

A system of N titratable groups have N intrinsic pK_a values and $N \times (N-1)$ unique electrostatic interaction energies, resulting in $N + N \times (N-1)$ defining energies. All of these energies are derived from X-ray structures and are consequently influenced by small structural perturbations. Assume that each intrinsic pK_a value is likely to vary by $\pm Q$ pK_a units, while each interaction energy can vary by $\pm$W kT/e. In this case we can examine the maximum possible variation of the system by constructing $M = 2^N + 2^{(N(N-1))}$ states which fully explore the configuration space for the system. Since the number of states to be explored increases rapidly with the number of titratable groups ($M = 65\,536$ for $N = 4$; $M = 33 \times 10^6$ for $N = 5$) it is only possible to perform a full exploration for a very small number of titratable groups). For a higher number of groups pKaTool contains the option of performing a random sampling of configuration space to obtain an estimate of the sensitivity of the system.

Similarly it is possible to examine the sensitivity of the system to changes in each single parameter to assess the importance of each intrinsic pK_a or interaction energy in turn. Since the number of perturbations in this analysis increases only with the number of defining energies, it is often more feasible to perform this simpler analysis on larger systems of titratable groups.

4.8. Calculating the pH-dependence of protein stability

The pH-dependence of protein stability can be calculated by integrating the proton uptake when the protein folds over all pH values [21,46], e.g.

$$\Delta G_{\text{pH}} = \int_{\text{pH\_start}}^{\text{pH}} (q_{\text{D}} - q_{\text{F}})\text{dpH} \quad (1)$$

q_{D} refers to the charge on the denatured form of the protein, wheras q_{F} refers to the charge of the folded (native) form of the protein. Charges for the folded form of the protein are given by the titration curves calculated from a protein X-ray structure, whereas charges for the unfolded form are assumed to be determined by the model pK_a values of the titratable groups in the protein. It is known that this is not strictly true [35,47], but for the purposes of fast identification and decomposition of stability effects this assumption becomes reasonable.

Acknowledgements

This work was supported by an SFI PYIRA award (04/YI1/M537) and a UCD President's award to JN.

References

[1] M.R. Gunner, et al., Backbone dipoles generate positive potentials in all proteins: origins and implications of the effect, Biophys. J. 78 (3) (2000) 1126–1144.
[2] J.E. Nielsen, G. Vriend, Optimizing the hydrogen-bond network in Poisson–Boltzmann equation-based pK(a) calculations, Proteins 43 (4) (2001) 403–412.
[3] J. Antosiewicz, J.A. McCammon, M.K. Gilson, Prediction of pH-dependent properties of proteins, J. Mol. Biol. 238 (3) (1994) 415–436.
[4] D. Bashford, K. Gerwert, Electrostatic calculations of the pK_a values of ionizable groups in bacteriorhodopsin, J. Mol. Biol. 224 (2) (1992) 473–486.
[5] E.L. Mehler, F. Guarnieri, A self-consistent, microenvironment modulated screened coulomb potential approximation to calculate pH-dependent electrostatic effects in proteins, Biophys. J. 77 (1) (1999) 3–22.
[6] H.W. van Vlijmen, M. Schaefer, M. Karplus, Improving the accuracy of protein pK_a calculations: conformational averaging versus the average structure, Proteins 33 (2) (1998) 145–158.
[7] A.S. Yang, et al., On the calculation of pK_as in proteins, Proteins 15 (3) (1993) 252–265.
[8] H. Li, A.D. Robertson, J.H. Jensen, Very fast empirical prediction and rationalization of protein pK(a) values, Proteins 61 (4) (2005) 704–721.
[9] J. Mongan, D.A. Case, J.A. McCammon, Constant pH molecular dynamics in generalized Born implicit solvent, J. Comput. Chem. 25 (16) (2004) 2038–2048.
[10] E.G. Alexov, M.R. Gunner, Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties, Biophys. J. 72 (5) (1997) 2075–2093.
[11] J.C. Gordon, et al., H++: a server for estimating pK_as and adding missing hydrogens to macromolecules, Nucl. Acids Res. 33 (2005) pW368–pW371 (Webserver issue).
[12] B. Tynan-Connolly, J.E. Nielsen, pKD: re-designing protein pK_a values, Nucl. Acids Res. (in press).
[13] J. Lamotte-Brasseur, A. Dubus, R.C. Wade, pK(a) calculations for class C beta-lactamases: the role of Tyr-150, Proteins 40 (1) (2000) 23–28.
[14] B.S. Tolbert, et al., The active site cysteine of ubiquitin-conjugating enzymes has a significantly elevated pK(a): functional implications, Biochemistry 44 (50) (2005) 16385–16391.
[15] B. Synstad, S. Gaseines, D.M.F. Aalten, G. Vriend, J.E. Nielsen, V.G.H. Eijsink, Mutational and computational analysis of the role of conserved residues in the active site of a family 18 chitinase, Eur. J. Biochem. 271 (2) (2004) 253–262.
[16] J.E. Nielsen, J.A. McCammon, On the evaluation and optimisation of protein X-ray structures for pK_a calculations, Protein Sci. 12 (2) (2003) 313–326.
[17] W.L. Mock, D.J. Stanford, Arazoformyl dipeptide substrates for thermolysin. Confirmation of a reverse protonation catalytic mechanism, Biochemistry 35 (23) (1996) 7369–7377.
[18] M.D. Joshi, et al., 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. 299 (1) (2000) 255–279.

[19] L.P. McIntosh, et al., The pK_a of the general acid/base carboxyl group of a glycosidase cycles during catalysis: a 13C NMR study of bacillus circulans xylanase, Biochemistry 35 (31) (1996) 9958–9966.
[20] E. Alexov, 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 56 (3) (2004) 572–584.
[21] C. Tanford, Protein denaturation. Part C. Theoretical models for the mechanism of denaturation, Adv. Protein Chem. 25 (1970) 1–95.
[22] P. Beroza, et al., 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. USA 88 (13) (1991) 5804–5808.
[23] C. Tanford, R. Roxby, Interpretation of protein titration curves. Application to lysozyme, Biochemistry 11 (1972) 2193–2198.
[24] A. Warshel, Calculations of enzymatic reactions: calculations of pK_a, proton transfer reactions, and general acid catalysis reactions in enzymes, Biochemistry 20 (11) (1981) 3167–3177.
[25] I. Klapper, et al., Focusing of electric fields in the active site of Cu–Zn superoxide dismutase: effects of ionic strength and amino-acid modification, Proteins 1 (1) (1986) 47–59.
[26] A. Onufriev, D. Bashford, D.A. Case, Modification of the generalized Born model suitable for macromolecules, J. Phys. Chem. B 104 (2000) 3712–3720.
[27] M. Feig, et al., Performance comparison of generalized Born and Poisson methods in the calculation of electrostatic solvation energies for protein structures, J. Comput. Chem. 25 (2) (2004) 265–284.
[28] M.J. Ondrechen, J.G. Clifton, D. Ringe, THEMATICS: a simple computational predictor of enzyme function from structure, Proc. Natl. Acad. Sci. USA 98 (22) (2001) 12473–12478.
[29] A. Onufriev, D.A. Case, G.M. Ullmann, A novel view of pH titration in biomolecules, Biochemistry 40 (12) (2001) 3413–3419.
[30] J. Ko, et al., Statistical criteria for the identification of protein active sites using theoretical microscopic titration curves, Proteins 59 (2) (2005) 183–195.
[31] C.N. Schutz, A. Warshel, What are the dielectric ''constants'' of proteins and how to validate electrostatic models? Proteins 44 (4) (2001) 400–417.
[32] E. Alexov, Role of the protein side-chain fluctuations on the strength of pair-wise electrostatic interactions: comparing experimental with computed pK(a)s, Proteins 50 (1) (2003) 94–103.
[33] J.E. Nielsen, J.A. McCammon, Calculating pK_a values in enzyme active sites, Protein Sci. 12 (2003) 1894–1901.
[34] J. Warwicker, Improved pK_a calculations through flexibility based sampling of a water-dominated interaction scheme, Protein Sci. 13 (10) (2004) 2793–2805.
[35] A.H. Elcock, Realistic modeling of the denatured states of proteins allows accurate calculations of the pH dependence of protein stability, J. Mol. Biol. 294 (4) (1999) 1051–1062.
[36] H.X. Zhou, M. Vijayakumar, Modeling of protein conformational fluctuations in pK_a predictions, J. Mol. Biol. 267 (4) (1997) 1002–1011.
[37] S.T. Wlodek, J. Antosiewicz, J.A. McCammon, Prediction of titration properties of structures of a protein derived from molecular dynamics trajectories, Protein Sci. 6 (2) (1997) 373–382.
[38] Y.Y. Sham, Z.T. Chu, A. Warshel, Consistent calculations of pK_as of ionizable residues in proteins: semi-microscopic and microscopic approaches, J. Phys. Chem. 101 (1997) 4458–4472.
[39] J.E. Nielsen, T.V. Borchert, Protein engineering of bacterial alpha-amylases, Biochim. Biophys. Acta 1543 (2) (2000) 253–274.
[40] B.M. Tynan-Connolly, J.E. Nielsen, Re-designing protein pK_a values (submitted for publication).
[41] K. Linderstrom-Lang, Om proteinstoffernes ionisation, Compt. Rend. Trav. Lab. Carlsberg 15 (7) (1924) 1–29.
[42] J. Kim, J. Mao, M.R. Gunner, Are acidic and basic groups in buried proteins predicted to be ionized? J. Mol. Biol. 348 (5) (2005) 1283–1298.
[43] E. Alexov, Numerical calculations of the pH of maximal protein stability. The effect of the sequence composition and three-dimensional structure, Eur. J. Biochem. 271 (1) (2004) 173–185.
[44] G.V. Rossum, Python, 2003.
[45] G. Vriend, WHAT IF: a molecular modeling and drug design program, J. Mol. Graph. 8 (1) (1990), 52–56, 29.
[46] A.S. Yang, B. Honig, On the pH dependence of protein stability, J. Mol. Biol. 231 (2) (1993) 459–474.
[47] M. Oliveberg, V.L. Arcus, A.R. Fersht, pK_A values of carboxyl groups in the native and denatured states of barnase: the pK_A values of the denatured state are on average 0.4 units lower than those of model compounds, Biochemistry 34 (29) (1995) 9424–9433.