

SOFTWARE

Open Access

# Graphical analysis of pH-dependent properties of proteins predicted using PROPKA

Michał Rostkowski, Mats HM Olsson, Chresten R Søndergaard, Jan H Jensen\*

## Abstract

**Background:** Charge states of ionizable residues in proteins determine their pH-dependent properties through their  $pK_a$  values. Thus, various theoretical methods to determine ionization constants of residues in biological systems have been developed. One of the more widely used approaches for predicting  $pK_a$  values in proteins is the PROPKA program, which provides convenient structural rationalization of the predicted  $pK_a$  values without any additional calculations.

**Results:** The PROPKA Graphical User Interface (GUI) is a new tool for studying the pH-dependent properties of proteins such as charge and stabilization energy. It facilitates a quantitative analysis of  $pK_a$  values of ionizable residues together with their structural determinants by providing a direct link between the  $pK_a$  data, predicted by the PROPKA calculations, and the structure via the Visual Molecular Dynamics (VMD) program. The GUI also calculates contributions to the pH-dependent unfolding free energy at a given pH for each ionizable group in the protein. Moreover, the PROPKA-computed  $pK_a$  values or energy contributions of the ionizable residues in question can be displayed interactively. The PROPKA GUI can also be used for comparing pH-dependent properties of more than one structure at the same time.

**Conclusions:** The GUI considerably extends the analysis and validation possibilities of the PROPKA approach. The PROPKA GUI can conveniently be used to investigate ionizable groups, and their interactions, of residues with significantly perturbed  $pK_a$  values or residues that contribute to the stabilization energy the most. Charge-dependent properties can be studied either for a single protein or simultaneously with other homologous structures, which makes it a helpful tool, for instance, in protein design studies or structure-based function predictions. The GUI is implemented as a Tcl/Tk plug-in for VMD, and can be obtained online at [http://propka.ki.ku.dk/~luca/wiki/index.php/GUI\\_Web](http://propka.ki.ku.dk/~luca/wiki/index.php/GUI_Web).

## Background

The pH dependence of important protein properties such as binding affinity, catalytic activity, solubility, charge and stability is determined by ionizable residues [1-3]. Thus, it is of great importance for researchers to have access to a reliable description of these residues. Protonation states of ionizable groups can be described with titration curves and ionization constants ( $pK_a$  values). Because  $pK_a$  values are difficult to obtain experimentally, especially for large biological systems, several software packages have been developed to predict them based on the protein structure [4-6]. PROPKA [7-9] is one of the popular protein  $pK_a$  prediction software

packages mainly because of its speed and accuracy compared to other methods [4,6], but also because it offers a structural rationalization of the predicted  $pK_a$  values.

PROPKA computes the  $pK_a$  values of the ionizable residues in a protein by determining a perturbation to the model  $pK_a$  value,  $pK_{\text{model}}$ , due to the protein environment [7-9]:

$$pK_a = pK_{\text{model}} + \Delta pK_{\text{DS}} + \Delta pK_{\text{HB}} + \Delta pK_{\text{CC}} \quad (1)$$

This perturbation comes from the desolvation penalty (DS), back-bone and side-chain hydrogen bonds (HB), and interactions with other charged groups (CC). The functional form of these terms and the associated parameters are determined empirically, and the relationship between the perturbation and the structure is described

\* Correspondence: [jhjensen@kemi.ku.dk](mailto:jhjensen@kemi.ku.dk)  
Department of Chemistry, University of Copenhagen, Universitetsparken 5,  
2100 Copenhagen, Denmark

by simple distance and angle dependent functions in order to be evaluated with minimal computational effort, and to make analysis tractable also for large proteins or protein complexes. Results of the PROPKA calculations are saved in a formatted text file containing the  $pK_a$  and  $pK_{model}$  values for each ionizable residue as well as corresponding lists of all interactions contributing to the  $pK_a$  shifts (equation 1). The PROPKA output file also contains the total charge of the protein and the pH-dependent free energy of unfolding, both as functions of pH. The latter can be obtained from the difference in the total protein charge between the folded and unfolded state at a given pH [10,11]:

$$\Delta G_U(pH) = \Delta G_U(pH_{ref}) + 1.36 \int_{pH_{ref}}^{pH} (Q_U - Q_F) dpH \quad (2)$$

Here,  $\Delta G_U(pH_{ref})$  is the unfolding free energy at a reference pH, and the latter term is the pH-dependent change in the unfolding free energy related to the change in protein charge  $Q$  between two folding states. Thus, the perturbed protein  $pK_a$  values are used to calculate the charge of the folded protein, whereas  $pK_{model}$  values are used for the unfolded state.

The results from the PROPKA calculations can be very helpful, and give detailed information about the influence of the protein environment on the ionizable groups. Nevertheless, the PROPKA output does not provide a direct link between obtained  $pK_a$  values and the three-dimensional structure of the studied system. In order to complete analysis of the ionizable residues one needs to make a separate search of these residues together with the interactions determining their  $pK_a$  values by hand, using software for visualizing biomolecules. Furthermore, studying raw text data for larger sets of structures can easily become a difficult, complex and time-consuming task.

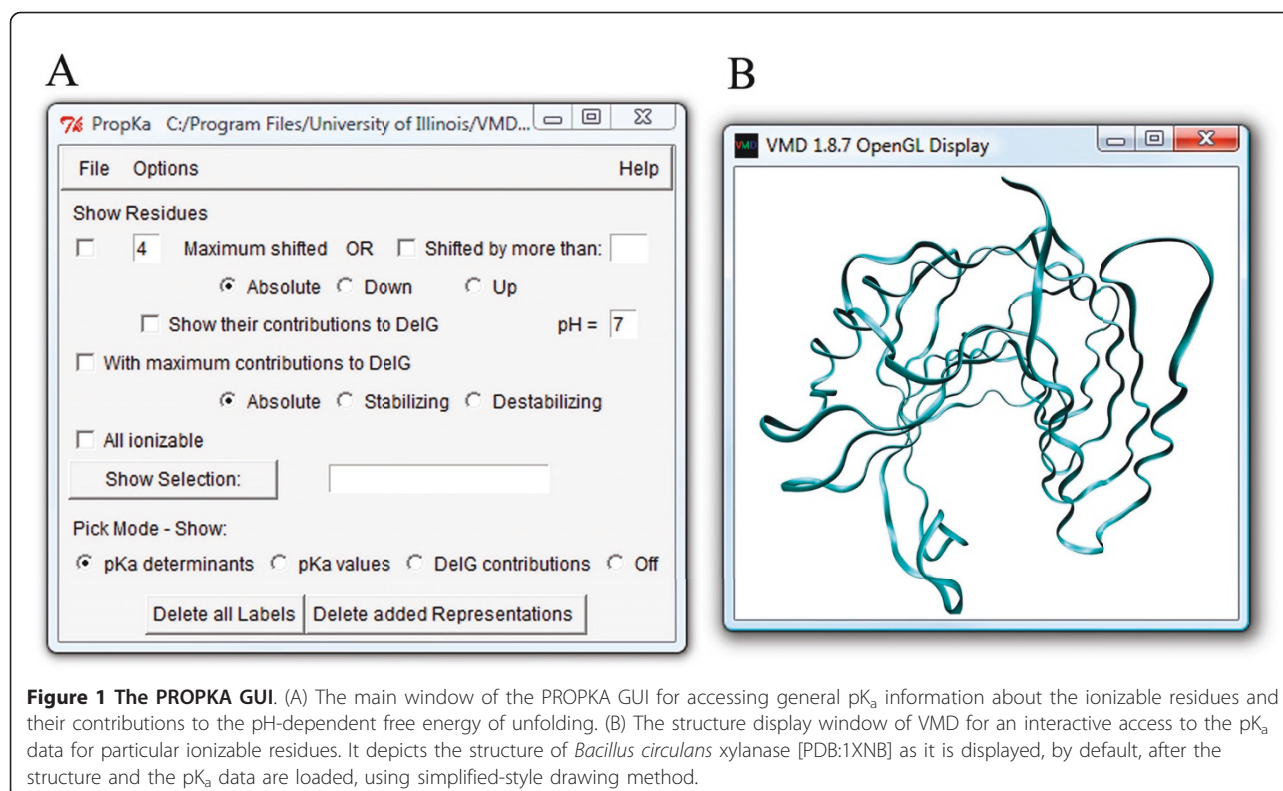
The PROPKA Graphical User Interface (GUI) presented in this paper is developed to facilitate exploration of the pH-dependent protein properties in a convenient manner by providing a direct link between the structure and the  $pK_a$  data, predicted by the PROPKA calculations, via the Visual Molecular Dynamics (VMD) program [12]. Our interface is an easy-to-use tool to identify and rationalize residues with unusual  $pK_a$  values or those significantly contributing to the free energy of unfolding. The PROPKA GUI is designed to facilitate the use of the PROPKA program and interpreting its results both for the user's convenience and to increase accessibility to the PROPKA approach for a wide range of researchers. Additionally, our GUI allows for comparative studies of the pH-dependent properties of many structures together, which can be used to rationalize the differences in these properties between homologous structures.

## Implementation

The PROPKA GUI is written in the Tool Command programming Language with the Tk graphical user interface (Tcl/Tk) as a platform-independent plug-in for the VMD program. The VMD program was chosen as a host application for the PROPKA GUI as it offers a great versatility of options and tools for analyzing biological structures, and also because it provides the Tcl/Tk environment as an extension of the VMD core code functionality without the need of making any additional installations. Besides, Tcl/Tk gives a wide range of users an easy but powerful tool to develop their own programs or scripts, or to extend already existing ones. The PROPKA GUI requires the VMD package to be installed on the user's computer. VMD can be obtained online at <http://www.ks.uiuc.edu/Research/vmd/>. The current version of our GUI is available as a single file that has to be copied into the VMD plug-ins directory and adding only one line into the VMD starting script makes the PROPKA GUI available from the menu in the main VMD window. The PROPKA GUI source-code, which is freely distributed under the GNU General Public License (GPL), installation instructions, documentation, and a screencast tutorial are available on the web at [http://propka.ki.ku.dk/~luca/wiki/index.php/GUI\\_Web](http://propka.ki.ku.dk/~luca/wiki/index.php/GUI_Web).

The GUI extracts and visualizes data from the PROPKA output file. The  $pK_a$  calculations can be performed online at <http://propka.ki.ku.dk/>, or locally, via the GUI, if the PROPKA program is installed on the same computer. By default, the  $pK_a$  data from the PROPKA output file and corresponding structure, contained in a separate Protein Data Bank (PDB) [13] file, are loaded simultaneously.  $pK_a$  values and their determinants are assigned to the appropriate residues and can be accessed interactively either through the main PROPKA GUI window (Figure 1A) or through the structure display window of VMD (Figure 1B). It is also important to note that the data from the PROPKA output file is assigned to residues of the current top molecule in VMD, which allows for loading  $pK_a$  data for all proteins in VMD separately. This provides the user with an access to the  $pK_a$  data for many proteins within the same instance of VMD. Going further, such accessibility to the  $pK_a$  information together with the VMD MultiSeq tool [14], which allows for structural alignment of homologous proteins, makes the PROPKA GUI a convenient tool to rationalize the differences in the pH-dependent properties between structurally-related proteins.

All residues and graphical objects displayed using the PROPKA GUI such as ionizable residues,  $pK_a$  determinants, ligands, etc., depending on their type, are shown using pre-defined sets of VMD settings and representations. These representations can easily be accessed and



**Figure 1** The PROPKA GUI. (A) The main window of the PROPKA GUI for accessing general p<sub>K<sub>a</sub></sub> information about the ionizable residues and their contributions to the pH-dependent free energy of unfolding. (B) The structure display window of VMD for an interactive access to the p<sub>K<sub>a</sub></sub> data for particular ionizable residues. It depicts the structure of *Bacillus circulans* xylanase [PDB:1XNB] as it is displayed, by default, after the structure and the p<sub>K<sub>a</sub></sub> data are loaded, using simplified-style drawing method.

modified in the “Graphical Representations” window of VMD. In order to make the PROPKA GUI more convenient to use, the user can also easily display the desired VMD selections, or remove previously shown, directly from the GUI. By default, all labels displaying the desired p<sub>K<sub>a</sub></sub> information in the structure display window of VMD are drawn using different sets of colors for each molecule. Moreover, corresponding labels for different loaded structures, depending on their molecule ID in VMD, are shifted relative to each other to increase their readability in the case of overlapping residues, which considerably facilitates using the GUI for comparative protein studies. Additionally, the information shown in the structure display window is also printed in the VMD text console.

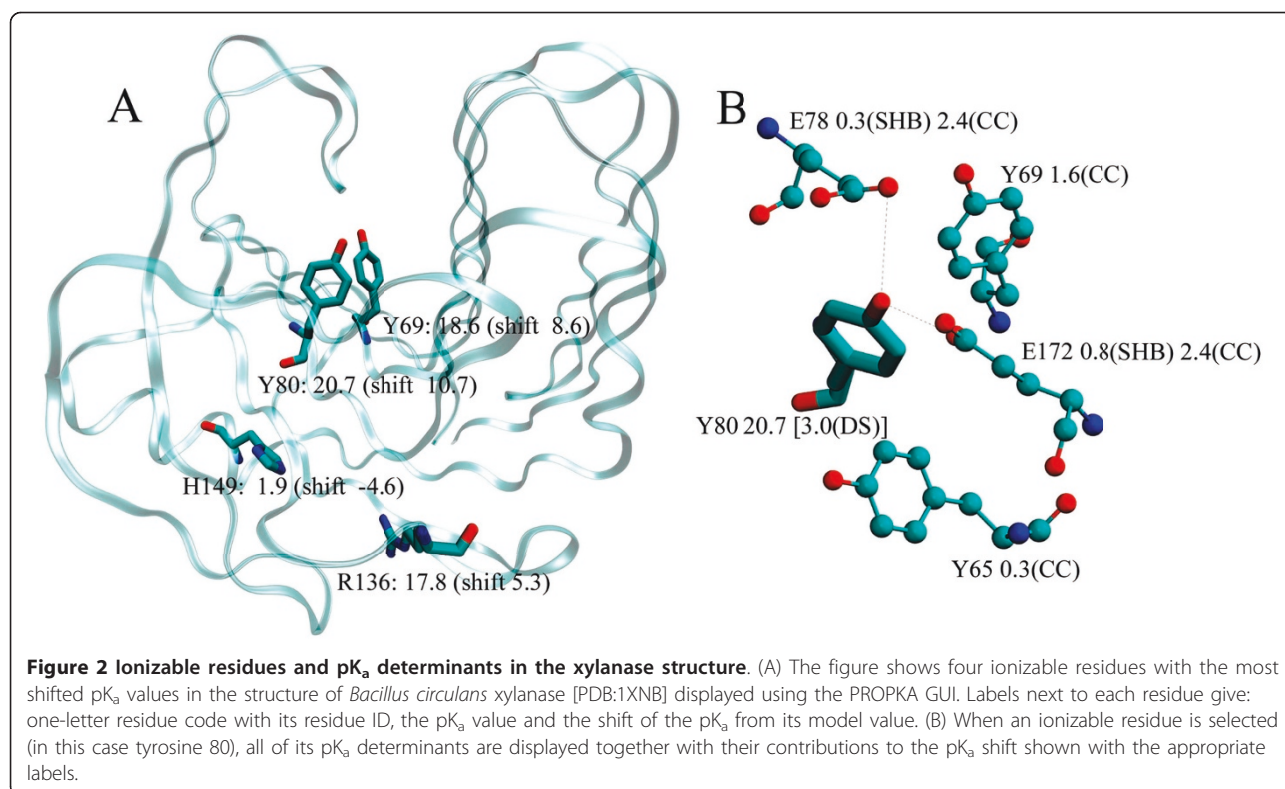
## Results and Discussion

The PROPKA GUI compares the computed p<sub>K<sub>a</sub></sub> values to p<sub>K<sub>model</sub></sub> values and can display residues with the largest p<sub>K<sub>a</sub></sub> shifts. Based on equation 2, the GUI also computes and displays the contribution of each ionizable residue to the pH-dependent part of the free energy of unfolding. The GUI can therefore be used to display residues contributing the most to the unfolding energy. Moreover, it provides an interactive access to the p<sub>K<sub>a</sub></sub> determinants listed in the PROPKA output file through the structure display window of VMD.

## Basic use

After installation of the PROPKA GUI plug-in, its main window (Figure 1A) can be accessed from the “Extensions” → “Analysis” menu in the main VMD window. By default, when the data from the PROPKA calculations and the appropriate PDB file are loaded, the structure is displayed automatically in the structure display window with a simplified-style drawing method (Figure 1B).

A user-defined number of residues with the most shifted p<sub>K<sub>a</sub></sub> values, or with p<sub>K<sub>a</sub></sub> shifts larger than a given threshold, can be displayed for the current top molecule in VMD simply by selecting the appropriate check box in the main GUI window. Figure 2A depicts the four residues with the largest p<sub>K<sub>a</sub></sub> shifts in *Bacillus circulans* xylanase (BCX), [PDB:1XNB] [15], computed by PROPKA2. These residues are: tyrosine 80 and 69, arginine 136 and histidine 149 with p<sub>K<sub>a</sub></sub> shifts of 10.7, 8.6, 5.3 and -4.6 pH units, respectively. This way, the user can easily visualize residues with the most perturbed p<sub>K<sub>a</sub></sub> values, which can often facilitate identification of the key residues as, for example, in case of the active site residues [16,17]. In the same way, the residues contributing the most to the pH-dependent free energy of unfolding, at a given pH, or just the most stabilizing or destabilizing residues can be shown. It is also possible to display all ionizable residues in the protein at once or



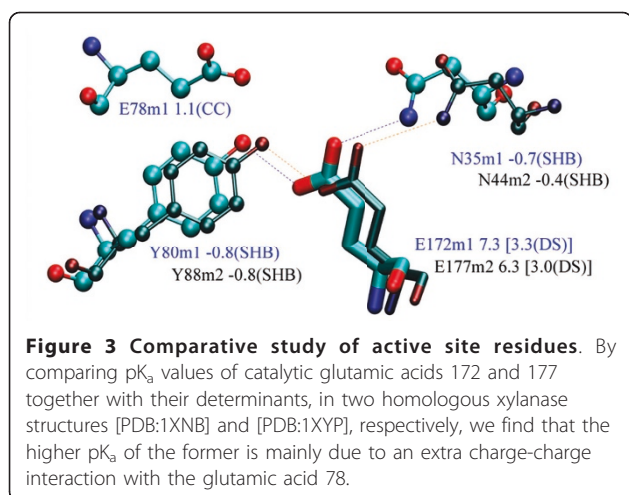
only the ones specified by the user. Moreover, the protein charge and the free energy of unfolding can be plotted as a function of pH through the “Options” menu, using the MultiPlot plug-in pre-installed in VMD.

More detailed pK<sub>a</sub> data can be accessed via the structure display window when the mouse picking mode is set to one of its “Label” actions. By default, when an ionizable residue or ionizable ligand atom is selected, all of its pK<sub>a</sub> determinants are displayed. In addition to the pK<sub>a</sub> value and the desolvation contribution for the selected residue, contributions to the pK<sub>a</sub> shift for all determinants are shown with the appropriate labels. Instead of displaying determinants, one can also choose to show only the pK<sub>a</sub> value or the contribution to the free energy of unfolding at a given pH. When the GUI interactive mode is disabled, VMD can be used in the standard way for analyzing the structure, making measurements of interatomic distances, angles, etc. As an example, we try to rationalize why the pK<sub>a</sub> value of tyrosine 80 in BCX is so extremely up-shifted compared to its model pK<sub>a</sub> value (20.7 compared to 10). By “clicking” on the residue with the mouse, we find that tyrosine 80 interacts strongly with three neighboring ionizable residues: glutamic acid 78, tyrosine 69, and second glutamic acid 172 (see Figure 2B). These contribute to raising the pK<sub>a</sub> value through unfavorable charge-charge interactions (CC) by 2.4, 1.6 and 2.4 pH units, respectively.

Increase of the pK<sub>a</sub> value, but to a smaller extent, is also achieved by charge-charge interaction with tyrosine 65 and by hydrogen bonds to the side chains (SHB) of the mentioned glutamic acids. In addition, tyrosine 80 is buried in the protein, and therefore shielded from the solvent, which raises its pK<sub>a</sub> value by additional 3 pH units due to the desolvation energy (DS).

#### Comparing structures

Having all abovementioned options for accessing the pK<sub>a</sub> data in hand, the PROPKA GUI is also a useful tool for more complex and demanding analysis such as carrying out comparative studies of the pH-dependent properties for homologous proteins. After loading structures to compare together with the pK<sub>a</sub> information, and aligning their coordinates, using for example the MultiSeq tool from VMD, the differences in pK<sub>a</sub> values of particular residues can be rationalized simply by displaying these residues and their pK<sub>a</sub> determinants. An example of such comparison is shown in Figure 3 for the catalytic glutamic acids 172 (PROPKA2-computed pK<sub>a</sub> value of 7.3) and 177 (pK<sub>a</sub> = 6.3) for two xylanase structures, [PDB:1XNB] and [PDB:1XYP] [18], respectively. A cursory look on the pK<sub>a</sub> determinants of these two residues clearly shows that the difference results mainly from the additional, repulsive interaction with the charged group of the other catalytic nucleophile,



glutamic acid 78, in the [PDB:1XNB] structure [11]. Such studies of homologous systems help to understand the key features underlying the differences in the protein properties. For example, they can help us to understand which residues and interactions are responsible for the extraordinary stability of extremophiles, or, for instance, which residues are crucial for certain reaction mechanism in enzyme-catalyzed reactions.

#### Future development

Currently, the second version of the PROPKA GUI is under development. The main improvements will extend the basic GUI functionality by automated and user-friendly procedures for protein structure comparisons in order to better understand their pH-dependent properties. It will provide the user with a more advanced, but still convenient tool for a quick and robust analysis of structural differences determining different ionization constants of corresponding residues for large sets of homologous structures. Then, if needed, the tool can be used to suggest and verify desired modifications to the studied structures within seconds.

#### Conclusions

Our newly developed PROPKA GUI is a powerful and convenient plug-in for VMD providing a direct link between the PROPKA-computed  $pK_a$  values, their determinants and the three-dimensional structures. The GUI significantly improves ease of use of the PROPKA approach, and facilitates quick and easy investigation of the pH-dependent properties of proteins such as charge and stabilization energy as well as the separate  $pK_a$  values and interactions determining them. It can easily be used to identify and rationalize ionizable residues with perturbed  $pK_a$  values or contributing to the pH-dependent stabilization energy the most, either for a single protein or in comparison with other structures. This

makes our GUI a helpful tool, for example, in the structure-based function prediction or protein design studies. Moreover, the PROPKA GUI is an open source code written in Tcl/Tk that can easily be customized whenever needed.

#### Availability and requirements

- **Project name:** PROPKA GUI
- **Project home page:** [http://propka.ki.ku.dk/~luca/wiki/index.php/GUI\\_Web](http://propka.ki.ku.dk/~luca/wiki/index.php/GUI_Web)
- **Operating system(s):** Platform independent
- **Programming language:** Tcl/Tk
- **Other requirements:** VMD program installed
- **License:** GNU General Public License
- **Restrictions to use by non-academics:** None

#### Acknowledgements

This work was supported by the Danish Council for Strategic Research through a research grant from the Program Commission on Strategic Growth Technologies (2106-07-0030). CRS was supported by the ERUDESP EU collaborative project.

#### Authors' contributions

MR contributed to design, developing and testing software, and drafted the manuscript. MHMO and CRS contributed to design and software testing, provided support with the PROPKA program, and were involved in revising the manuscript. JHJ conceived the PROPKA GUI, contributed to design, and was involved in revising the manuscript. All authors read and approved the final manuscript.

Received: 29 September 2010 Accepted: 26 January 2011

Published: 26 January 2011

#### References

1. Sharp KA, Honig B: **Electrostatic Interactions in Macromolecules - Theory and Applications.** *Annu Rev Biophys Biophys Chem* 1990, **19**:301-332.
2. Pace CN, Grimsley GR, Scholtz JM: **Protein Ionizable Groups: pK Values and Their Contribution to Protein Stability and Solubility.** *J Biol Chem* 2009, **284**:13285-13289.
3. Nakamura H: **Roles of electrostatic interaction in proteins.** *Q Rev Biophys* 1996, **29**:1-90.
4. Davies MN, Toseland CP, Moss DS, Flower DR: **Benchmarking pK(a) prediction.** *BMC Biochem* 2006, **7**:18.
5. Lee AC, Crippen GM: **Predicting pKa.** *J Chem Inf Model* 2009, **49**:2013-2033.
6. Stanton CL, Houk KN: **Benchmarking pK(a) prediction methods for residues in proteins.** *J Chem Theory Comput* 2008, **4**:951-966.
7. Li H, Robertson AD, Jensen JH: **Very fast empirical prediction and rationalization of protein pKa values.** *Proteins* 2005, **61**:704-721.
8. Bas DC, Rogers DM, Jensen JH: **Very fast prediction and rationalization of pKa values for protein-ligand complexes.** *Proteins* 2008, **73**:765-783.
9. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH: **PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions.** *J Chem Theory Comput* .
10. Yang AS, Honig B: **On the Ph-Dependence of Protein Stability.** *J Mol Biol* 1993, **231**:459-474.
11. Kongsted J, Ryde U, Wydra J, Jensen JH: **Prediction and rationalization of the pH dependence of the activity and stability of family 11 xylanases.** *Biochem* 2007, **46**:13581-13592.
12. Humphrey W, Dalke A, Schulten K: **VMD: Visual molecular dynamics.** *J Mol Graph* 1996, **14**:33-38.
13. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al: **The Protein Data Bank.** *Nucleic Acids Res* 2000, **28**:235-242.
14. Roberts E, Eargle J, Wright D, Luthey-Schulten Z: **MultiSeq: unifying sequence and structure data for evolutionary analysis.** *BMC Bioinf* 2006, **7**:382.

15. Wakarchuk WW, Campbell RL, Sung WL, Davoodi J, Yaguchi M: **Mutational and Crystallographic Analyses of the Active-Site Residues of the Bacillus-Circulans Xylanase.** *Protein Sci* 1994, **3**:467-475.
16. Elcock AH: **Prediction of functionally important residues based solely on the computed energetics of protein structure.** *J Mol Biol* 2001, **312**:885-896.
17. Ondrechen MJ, Clifton JG, Ringe D: **THEMATICS: A simple computational predictor of enzyme function from structure.** *Proc Natl Acad Sci USA* 2001, **98**:12473-12478.
18. Torronen A, Rouvinen J: **Structural Comparison of two Major Endo-1,4-Xylanases from Trichoderma-Reesei.** *Biochem* 1995, **34**:847-856.

doi:10.1186/1472-6807-11-6

**Cite this article as:** Rostkowski *et al.*: Graphical analysis of pH-dependent properties of proteins predicted using PROPKA. *BMC Structural Biology* 2011 **11**:6.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

