pKD: re-designing protein pK_a values

Barbara M. Tynan-Connolly and 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

ABSTRACT

The pK_a values in proteins govern the pH-dependence of protein stability and enzymatic activity. A large number of mutagenesis experiments have been carried out in the last three decades to re-engineer the pH-activity and pH-stability profile of enzymes and proteins. We have developed the pKD webserver (http://polymerase.ucd.ie/pKa_Design), which predicts sets of point mutations that will change the pK_a values of a set of target residues in a given direction, thus allowing for targeted re-design of the pH-dependent characteristics of proteins. The server provides the user with an interactive experience for re-designing pK_a values by pre-calculating ΔpK_a values from all feasible point mutations. Design solutions are found in less than 10 min for a typical design job for a medium-sized protein. Mutant ΔpK_a values calculated by the pKD web server are in close agreement with those produced by comparing results from full-fledged pK_a calculation methods.

INTRODUCTION

The pH-dependence of enzymatic activity and protein stability is of major importance for the biological function and industrial application of enzymes and proteins. pH-dependent protein characteristics have therefore been the subject of significant research efforts over the last two decades [e.g. (1–8)]. The pH-dependence of protein characteristics is determined by the pK_a values of amino acid residues in the unfolded and folded state of a protein. Consequently a growing number of theoretical methods have been constructed for calculating the pK_a values of protein titratable groups (9–20), and presently the best pK_a calculation methods are accurate to within 0.5 pK_a units when benchmarked against experimentally determined pK_a values. The application of pK_a calculation methods in biology is limited to a relatively small number of studies on catalytic mechanisms and ligand binding [e.g. (2–4,13,21–27)], and there is thus much scope for a wider application of pK_a calculation methodology to biological

problems. To our knowledge pK_a calculations have yet to be used for designing proteins with novel pH-dependent properties, although methods have been developed for dissecting the contribution of charged groups to protein pK_a values (25) (J. E. Nielsen, manuscript submitted). These methods can, in principle, be used for targeted re-design of protein pK_a values.

To motivate research into designing proteins with novel pH-dependent properties we have developed the pKD web server (http://polymerase.ucd.ie/pKa_Design), which interfaces with a novel algorithm (B. M. Tynan-Connolly and J. E. Nielsen, manuscript submitted) for re-designing protein pK_a values. The web server allows the user to specify a set of design criteria (target residues, desired Δ pK_a values, maximum number of mutations and minimum distance between target residue and any mutation) that specifies how the user want to change the pK_a values of the protein. The pKD web server subsequently calculates a number of design solutions consisting of sets of point mutations predicted to change the target pK_a values as specified by the design criteria. In addition to reporting the design solutions, the server produces 3D structural models of the proposed solutions and allows the user to inspect the pK_a values calculated from the wild-type structure. The results and inputs of the pKD server can be analysed with a novel graphical interface (pK_aTool) that facilitates the analysis of protein pK_a calculations and titration curves. p$K_a$tool is freely available to academic researchers at http://enzyme.ucd.ie/Science/pKa/pKaTool.

MATERIALS AND METHODS

The pKD server combines the functionality of a number of software packages using Python scripts. Figure 1 shows an overview of the functionality of the server and indicates where other software packages are employed. The parameters for the pKD pK_a calculations are described in detail below. The pKD server software is freely available to academic researchers by contacting pka@ucd.ie, or in some cases by download from http://enzyme.ucd.ie/Science/pKa.

Preparation of PDB files

PDB files are prepared for a pKD design by deleting all water molecules and non-protein atoms. Missing protein atoms are rebuilt using the position-specific rotamer libraries (28) in WHAT IF (29).

---

*To whom correspondence should be addressed. Tel: +353 1 716 6724; Fax: +353 1 716 6898; Email: Jens.Nielsen@UCD.IE
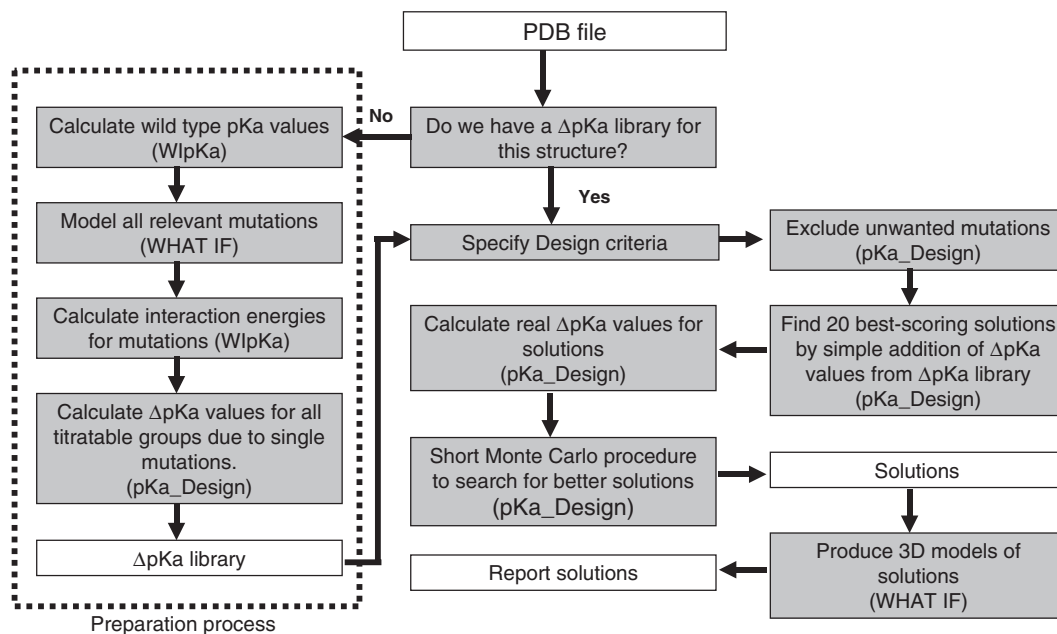
Figure 1. The workflow of the pKD server. The server incorporates the functionality of the WHAT IF pK_a calculation package (WIpK_a), construction of point mutations (WHAT IF) and the pK_a_Design algorithm.

Calculation of pK_a values of the wild-type structure

The pK_a values of the wild-type structure are calculated with the WHAT IF pK_a calculation package as described elsewhere (19), except that a uniform dielectric constant of eight is used for the protein, and that the neutral–charged, charged–neutral and neutral–neutral interaction energies are calculated only for residue pairs with an interaction energy greater than 10 kT. The latter approximation is employed to speed up the pK_a calculation step for large structures and can introduce errors in the reported pK_a values for tightly coupled titratable groups. Users that find this to be a problem for a particular structure should request a full pK_a calculation to be carried out by emailing pka@ucd.ie.

Modelling point mutations

Point mutations are modelled using position-specific rotamer libraries as implemented in WHAT IF (28). We mutate only residues that are at least 30% solvent exposed and fit well in the wild-type structure as deemed by automatic inspection of the rotamer library population. In addition we allow the mutation of buried charged residues to neutral residues of a similar size.

   Finally we allow the user to exclude mutations at sites less than a certain distance (user adjustable but recommended to be at least 5 Å) from the residue whose pK_a value is being redesigned.

Calculation of interaction energies for single point mutations

The interaction energies between a point mutation and all other residues are calculated using the WHAT IF pK_a calculation package as described above.

Calculation of Δ pK_a values resulting from point mutations

ΔpK_a values are calculated using a Monte Carlo sampling method (30) implemented as a C++ class, which is imported

into a python script. ΔpK_a values arising from single mutations are calculated by modifying the site–site interaction energy matrix using energies derived from an explicit model of the mutant protein structure. ΔpK_a values arising from multiple point mutations are calculated by modifying the site–site interaction energy matrix using energies calculated from single point mutation models. Thus no explicit 3D modelling of multiple point mutations takes place until the final solutions are found. Nevertheless we have shown (B. M. Tynan-Connolly and J. E. Nielsen, manuscript submitted) that ΔpK_a values calculated in this way are in excellent agreement with the ΔpK_a values found by comparing the results of full-fledged pK_a calculations on wild-type and mutant protein structures, provided that only a small fraction of protein residues are mutated.

Finding the optimal set of point mutations

The search for sets of point mutations that fulfil the design criteria is initiated by selecting 20 sets of combinations of single point mutations, whose cumulative ΔpK_a values are closest to the design criteria. Since ΔpK_a values only sometimes can be combined linearly (B. M. Tynan-Connolly and J. E. Nielsen, manuscript submitted), we calculate a more realistic set of ΔpK_a values for each solution as described above. Finally we perform a short Monte Carlo sampling to find solutions that do cannot be found from a linear combination of individual ΔpK_a values.

Scoring of design solutions

We apply a scoring function of the form

$$\text{score} = \sum \left(pKa_{\text{desired}} - pKa_{\text{achieved}}\right)^2 \qquad \mathbf{1}$$

to identify the design solutions in agreement with the design criteria. The sum in equation 1 is calculated over all pK_a criteria specified in the design setup phase. The function is optimized to identify a specific pK_a change rather than identifying the largest possible pK_a change. Therefore, if one is

Table 1. Example design solutions and predicted ΔpK_a values from pKD, the WHAT IF pK_a calculation package (19) and the H++ webserver (12) for a re-design of the pK_a of Glu 35 in HEWL (PDBID 2lzt)

Design criteria	Minimum distance to target	Mutations	pKD ΔpK_a	WI ΔpK_a	H++ ΔpK_a
Glu35 +2.0	5.0 Å	R112E+R114Q+N37E+W62D	1.2	0.9	1.0
Glu35 −2.0	10.0 Å	N103H+D48R+T43K	−0.6	−0.6	−0.7

The maximum number of mutations allowed was set to 5, WI pK_a calculations were carried out as described in Materials and Methods, and H++ calculations were carried out with the default settings on the website. The results illustrate that it often is not possible to achieve the desired pK_a change with the parameters and the structure in question.

interested in a maximum pK_a change then one should specify design criteria of ±20 to obtain solutions that yield the maximum pK_a shift possible for the residue in question.

RESULTS AND CONCLUSIONS

We have constructed a the pKD web server, which allows for the re-design of protein pK_a values by site-directed mutagenesis. Proteins with redesigned pK_a values will display a change in their pH-dependent characteristics, such as ligand binding, stability and, for enzymes, catalytic rate, and the pKD server is thus aimed at researchers who aim to understand or change the pH-dependent properties of proteins.

The pKD webserver asks the user to select one or more titratable groups for pK_a value re-design. It must specify how many mutations one is willing to construct and how close these mutations can be to the titratable groups of interest. Subsequently the pKD server predicts a set of point mutations that will change the pK_a values of the selected titratable groups in the given direction, by the desired amount. Furthermore the user is presented with a set of modelled 3D structures containing the proposed mutations, which can be analysed using pK_aTool (J. E. Nielsen, manuscript submitted).

It is thus possible for the user to achieve an interactive, in-depth understanding of the titrational behaviour of any given protein using these two freely available software tools. Future work will focus on integrating pK_aTool and the pKD server directly to allow for a desktop-based, convenient, interactive analysis facility of the titrational behaviour of wild-type and mutant proteins.

Table 1 and Figure 2 show a typical design solution from the pKD server, and illustrates that good agreement is obtained between the ΔpK_a values reported by the pKD server and those predicted by a full-fledged pK_a calculation packages. The pKD server uses a full physical model of the pH-dependent behaviour of titratable groups, which ensures that effects arising from differences in intrinsic pK_a values and complicated pair-wise electrostatic interaction networks are calculated as accurately as possible. Therefore, the solutions calculated with the pKD server are more physically realistic than solutions obtained using the classic relation $\Delta pKa = (\Delta\Phi/\ln 10)$ (8), which is known to break down when multiple strong electrostatic interactions are present and when dealing with titratable groups with perturbed intrinsic pK_a values.

We have calculated design solutions for a large number of protein structures and examined the dependence of the results on the algorithm parameters, and we have found the algorithm to reliably identify mutations that will change the pK_a values of target residues as judged by theoretical methods. We are
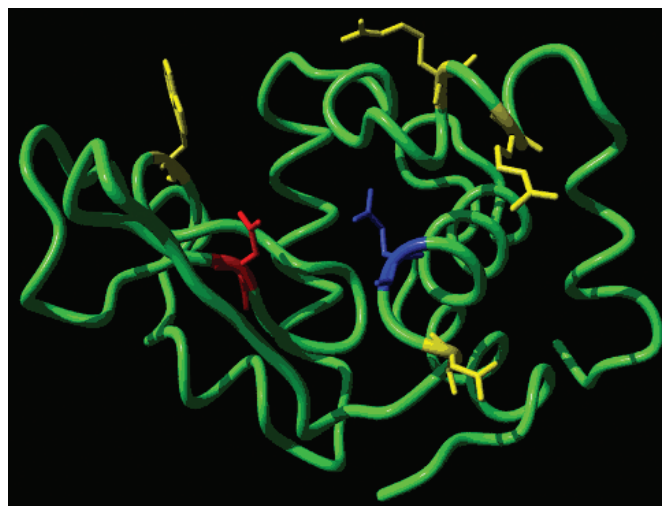
Figure 2. A 3D molecular representation of HEWL with the sites of mutations suggested by the pKD server in Table 1 highlighted in yellow. The proton donor (Glu 35) is shown in blue, whereas the catalytic nucleophile (Asp 52) is shown in red. Figure produced with the Yasara (http://www.yasara.org) and Pov-ray (http://www.povray.org) packages.

currently verifying these pK_a shifts in the lab, and simultaneously continuing our theoretical studies on factors that influence the pH-dependent properties of proteins.

We believe that the pKD server will be of great benefit to researchers that are interested in re-designing and understanding the pH-dependent characteristics of proteins, and we furthermore hope that the server will encourage more experiments aimed at understanding the complex links between protein structures and their pH-dependent characteristics.

ACKNOWLEDGEMENTS

REFERENCES

1. de Kreij,A., van den Burg,B., Venema,G., Vriend,G., Eijsink,V.G. and Nielsen,J.E. (2002) The effects of modifying the surface charge on the catalytic activity of a thermolysin-like protease. *J. Biol. Chem.*, **277**, 15432–15438.

2. Joshi,M.D., Sidhu,G., Nielsen,J.E., Brayer,G.D., Withers,S.G. and McIntosh,L.P. (2001) Dissecting the electrostatic interactions and pH-dependent activity of a family 11 glycosidase. *Biochemistry*, **40**, 10115–101139.

3. Lambeir,A.M., Backmann,J., Ruiz-Sanz,J., Filimonov,V., Nielsen,J.E., Kursula,I., Norledge,B.V. and Wierenga,R.K. (2000) The ionization of a buried glutamic acid is thermodynamically linked to the stability of *Leishmania mexicana* triose phosphate isomerase. *Eur. J. Biochem.*, **267**, 2516–2524.

4. Nielsen,J.E., Borchert,T.V. and Vriend,G. (2001) The determinants of alpha-amylase pH-activity profiles. *Protein Eng.*, **14**, 505–512.

5. Yang,A.S. and Honig,B. (1994) Structural origins of pH and ionic strength effects on protein stability. Acid denaturation of sperm whale apomyoglobin. *J. Mol. Biol.*, **237**, 602–614.

6. Perl,D., Mueller,U., Heinemann,U., Schmid,F.X. and Perl,D. (2000) Two exposed amino acid residues confer thermostability on a cold shock protein. *Nature Struct. Biol.*, **7**, 380–383.

7. Alexov,E. (2004) Numerical calculations of the pH of maximal protein stability. The effect of the sequence composition and three-dimensional structure. *Eur. J. Biochem.*, **271**, 173–185.

8. Jackson,S.E. and Fersht,A.R. (1993) Contribution of long-range electrostatic interactions to the stabilization of the catalytic transition state of the serine protease subtilisin BPN'. *Biochemistry*, **32**, 13909–13916.

9. Li,H., Robertson,A.D. and Jensen,J.H. (2005) Very fast empirical prediction and rationalization of protein pK_a values. *Proteins*, **61**, 704–721.

10. Antosiewicz,J., McCammon,J.A. and Gilson,M.K. (1994) Prediction of pH-dependent properties of proteins. *J. Mol. Biol.*, **238**, 415–436.

11. Bashford,D. and Gerwert,K. (1992) Electrostatic calculations of the pK_a values of ionizable groups in bacteriorhodopsin. *J. Mol. Biol.*, **224**, 473–486.

12. Gordon,J.C., Myers,J.B., Folta,T., Shoja,V., Heath,L.S., Onufriev,A. and Gordon,J.C. (2005) H++: a server for estimating pK_a's and adding missing hydrogens to macromolecules. *Nucleic Acids Res.*, **33**, W368–W371.

13. Lamotte-Brasseur,J., Lounnas,V., Raquet,X. and Wade,R.C. (1999) pK_a calculations for class A beta-lactamases: influence of substrate binding. *Protein Sci.*, **8**, 404–409.

14. Sham,Y.Y., Chu,Z.T. and Warshel,A. (1997) Consistent calculations of pK_a's of ionizable residues in proteins: semi-microscopic and microscopic approaches. *J. Phys. Chem.*, **101**, 4458–4472.

15. van Vlijmen,H.W., Schaefer,M. and Karplus,M. (1998) Improving the accuracy of protein pK_a calculations: conformational averaging versus the average structure. *Proteins*, **33**, 145–158.

16. Warwicker,J. (2004) Improved pK_a calculations through flexibility based sampling of a water-dominated interaction scheme. *Protein Sci.*, **13**, 2793–2805.

17. Yang,A.S., Gunner,M.R., Sampogna,R., Sharp,K. and Honig,B. (1993) On the calculation of pK_a's in proteins. *Proteins*, **15**, 252–265.

18. 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.

19. Nielsen,J.E. and Vriend,G. (2001) Optimizing the hydrogen-bond network in Poisson–Boltzmann equation-based pK_a calculations. *Proteins*, **43**, 403–412.

20. 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.

21. Alexov,E., Miksovska,J., Baciou,L., Schiffer,M., Hanson,D.K., Sebban,P. and Gunner,M.R. (2000) Modeling the effects of mutations on the free energy of the first electron transfer from QA- to QB in photosynthetic reaction centers. *Biochemistry*, **39**, 5940–5952.

22. Alexov,E.G. and Gunner,M.R. (1999) Calculated protein and proton motions coupled to electron transfer: electron transfer from QA- to QB in bacterial photosynthetic reaction centers. *Biochemistry*, **38**, 8253–8270.

23. 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.

24. Trylska,J., Antosiewicz,J., Geller,M., Hodge,C.N., Klabe,R.M., Head,M.S. and Gilson,M.K. (1999) Thermodynamic linkage between the binding of protons and inhibitors to HIV-1 protease. *Protein Sci.*, **8**, 180–195.

25. Tolbert,B.S., Tajc,S.G., Webb,H., Snyder,J., Nielsen,J.E., Miller,B.L. and Basavappa,R. (2005) The active site cysteine of ubiquitin-conjugating enzymes has a significantly elevated pK_a: functional implications. *Biochemistry*, **44**, 16385–16391.

26. Nielsen,J.E. and McCammon,J.A. (2003) Calculating pK_a values in enzyme active sites. *Protein Sci.*, **12**, 1894–1901.

27. Synstad,B., Gaseidnes,S., Van Aalten,D.M., Vriend,G., Nielsen,J.E. and Eijsink,V.G. (2004) Mutational and computational analysis of the role of conserved residues in the active site of a family 18 chitinase. *Eur. J. Biochem.*, **271**, 253–262.

28. Chinea,G., Padron,G., Hooft,R.W., Sander,C. and Vriend,G. (1995) The use of position-specific rotamers in model building by homology. *Proteins*, **23**, 415–421.

29. Vriend,G. (1990) WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph*, **8**, 52–6, 29.

30. 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. USA*, **88**, 5804–5808.