PHEPS: web-based pH-dependent Protein Electrostatics Server

Alexander A. Kantardjiev and Boris P. Atanasov*

Biophysical Chemistry Group, Institute of Organic Chemistry, Bulgarian Academy of Sciences, Sofia-1113, Bulgaria

Received February 10, 2006; Revised March 1, 2006; Accepted March 20, 2006

ABSTRACT

PHEPS (pH-dependent Protein Electrostatics Server) is a web service for fast prediction and experiment planning support, as well as for correlation and analysis of experimentally obtained results, reflecting charge-dependent phenomena in globular proteins. Its implementation is based on long-term experience (PHEI package) and the need to explain measured physicochemical characteristics at the level of protein atomic structure. The approach is semi-empirical and based on a mean field scheme for description and evaluation of global and local pH-dependent electrostatic properties: protein proton binding; ionic sites proton population; free energy electrostatic term; ionic groups proton affinities (pK\textsubscript{a,i}) and their Coulomb interaction with whole charge multipole; electrostatic potential of whole molecule at fixed pH and pH-dependent local electrostatic potentials at user-defined set of points. The speed of calculation is based on fast determination of distance-dependent pair charge-charge interactions as empirical three exponential function that covers charge–charge, charge–dipole and dipole–dipole contributions. After atomic coordinates input, all standard parameters are used as defaults to facilitate non-experienced users. Special attention was given to interactive addition of non-polypeptide charges, extra ionizable groups with intrinsic pK\textsubscript{a}s or fixed ions. The output information is given as plain-text, readable by ‘RasMol’, ‘Origin’ and the like. The PHEPS server is accessible at http://pheps.orgchm.bas.bg/home.html.

INTRODUCTION

Electrostatic phenomena are widely manifested as a fundamental feature of protein structure–function relationships (1–5). Protein molecules are very complex dielectric systems but can be treated as ‘solid state’ nanometer particles, immersed in buffered water solutions. Many approaches [‘macroscopic’—continuum dielectrics (6,7) and ‘microscopic’—polarizability (8,10)] were developed with different degree of validity: from simple TK-‘dielectric cavity models’ and analytical solution of Poisson–Boltzmann equation to their non-linear numerical and sophisticated empirical generalized Born solutions (11–13). The application of detailed and complex model description leads to increased difficulties for experimentalists to understand and use such sophisticated models. At present there are number of popular program packages [(14,15) and others] and web servers (16,17) but they are of limited significance for everyday problems of experimentalists. To the best of our knowledge, there is not available web server for fast pH-dependent calculation and analysis of protein electrostatic properties. Such software is needed because proteins are polyelectrolytes and their system of ionizable groups is pH-dependent. Many programs and servers compute pK\textsubscript{a}s at ‘neutral pH’ yielding pH-independent pK\textsubscript{a}s, which leads to erroneous results and distorts our view of principal properties of protein molecules, important for their functions. It is known that pK\textsubscript{a} is directly related to free energy change of the corresponding protolytic reaction ($\Delta G_a = RT pK_a$) and that this $\Delta G_a$ is pH-dependent. It is well known that protein pK\textsubscript{a}s are also pH-dependent, because ionic groups are closely arranged in the molecule. There are excellent theoretical works describing pH-dependent protein electrostatics [(7,18,19) and so on] but they are not straightforward for application by experimentalists. For many years a method addressing aforementioned requirements was developed and applied successfully in Biophysical Chemistry Laboratory at IOCh. The theoretical results are unequivocally validated by comparison with experimental studies as shown in a number of peer-reviewed publications over the years. Typical examples for this are pK\textsubscript{a} prediction of lysozyme, BPTI and cytochrome c (21–23); spectrophotometric titration prediction and infrared carboxylic groups titration (24), enthalpy of protein ionization prediction (25), pH-dependent protein ultrasonic compressibility analysis (26); local electrostatic potentials (27); electrostatic contribution to a protein crystal

*To whom correspondence should be addressed. Tel: +359-2 960 6123; Fax: +359-2 870 0225; Email: boris@orgchm.bas.bg

© The Author 2006. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org
lattice energy (28) and so on. The method was also applied to clarify enzyme mechanisms (29) and proved to be an invaluable tool for fast evaluation of electrostatic interactions and their analysis in large biomolecular immunocomplexes (30). We hope that our server http://pheps.orecmh. bas.bg/home.html will be useful for experimentalist (protein scientist in need for fast evaluation of pH-dependent properties, enzymologists in need of pK values, spectroscopists and the like) as well as for in silico analysis by structural biologists and bioinformaticians. Being fast and easy to use this server is suitable for first acquaintance and training in the field.

METHODS

Protein self-consistent electrostatics

It is generally accepted that a model for protein electrostatics can be built on the assumption of continuum medium description, fixed atom approximation, protein–solvent boundary numerically described by atomic static accessibilities, SA, [variants of Lee-Richards algorithm (31)] and two type of charges: (i) permanent (pH-independent) partial charges (par) and (ii) proton-binding sites with pH-dependent titratable charges: (i) permanent (pH-independent) partial charges (par) and (ii) proton-binding sites with pH-dependent titratable charges (tit). The model accepts experimentally measured (par) and (ii) proton-binding sites with pH-dependent titratable charges: (i) permanent (pH-independent) partial charges and two type of properties, enzymologists in need of pK values, spectroscopists and the like) as well as for in silico analysis by structural biologists and bioinformaticians. Being fast and easy to use this server is suitable for first acquaintance and training in the field.

The pH-dependence of the electrostatic potential \( \Phi_{el,i}(pH) \) at the \( i \)-th proton binding site in PHEI was evaluated according to the following equation:

\[
\Phi_{el,i}(pH) = 2.3RT \sum_{j \neq i} Q_j(pH)W_{ij} \left[ 1 - \left( \frac{SA_i + SA_j}{2} \right) \right],
\]

where \( Q_j(pH) \) is defined by degree of dissociation or statistical mechanical proton population of given H\(^+\)-binding site; \( Q_j(pH) = (1 - \langle s_j \rangle) \) and \( Q_j(pH) = -\langle s_j \rangle \) for basic and acidic groups respectively, where

\[
\langle s_j \rangle = \frac{10^{(PH-pK_j)}}{1 + 10^{(PH-pK_j)}}.
\]

Thus using partial titration of each \( j \)-th group we can find the pH-dependent net-charge of the whole molecule, \( Z(pH) \), i.e. potentiometric titration curve:

\[
Z(pH) = \sum_j Q_j(pH).
\]

By definition if \( Z = 0 \) than pH = PI, i.e. the isoelectric point (the only pH at which the dipole moment of a protein molecule can be evaluated).

Thus starting with \( pK_{int,i} = pK_{mod,i} + \Delta pK_{Born,i} + \Delta pK_{par,i} \), where \( pK_{mod,i} \) is the pH of the \( i \)-th site according to model compounds—see set of \( pK_{mod,i} \) in (21,22,29); \( \Delta pK_{Born,i} \) is the Born self-energy of the \( i \)-th site buried within the ‘uncharged’ protein, and \( \Delta pK_{par,i} \) is the contribution of the \( i \)-th site interacting with the set of partial (permanent, fixed) atomic charges (see above).

\[
pK_{a,i}(pH) = pK_{int,i} + pK_{tit,i} = pK_{int,i} + \left[ \frac{1}{2.3RT} \right] \times \sum_{j \neq i} \left\{ Q_j(pH)(W_{ij} - C) \left[ 1 - \left( \frac{SA_i + SA_j}{2} \right) \right] \right\},
\]

where \( C \) is the Debye–Hückel term for ionic strength (I). The term \( pK_{int,i} \) is the \( pK_a \) shift of the \( i \)-th site caused by interactions with all other proton-binding groups and is evaluated according to efficient self-consistent iterative procedure (32). Coming to self-consisted pH-dependent ionization the free energy term \( G_{el}(pH) \) is calculated as follows.

\[
\Delta G_{el} = \sum_{j \neq i} Q_j Q_i(pH)W_{ij} \left[ 1 - \left( \frac{SA_i + SA_j}{2} \right) \right],
\]

as well as pH-dependent Coulomb energy of each \( i \)-th ionic group with whole charge multipole:

\[
E_{el,i}(pH) = \sum_{j \neq i} Q_j Q_i(pH)W_{ij} \left[ 1 - \left( \frac{SA_i + SA_j}{2} \right) \right].
\]

After applying this iterative algorithm the electrostatic system is converged and all basic pH-dependent properties are reported.
IMPLEMENTATION

The web server is a front end of our program package PHEI, developed over many years in our Biophysical Chemistry Lab. Its current version is written in PERL and C/C++ by one of us (A.K.). Our package is capable of much more functionality and only basic electrostatic properties are presented online, the rest being under consideration for the next release (Conclusions and Future). The web implementation is driven by CGI/PERL routines. The only input file is a coordinate file in Protein Data Bank (PDB) format (33)—either user supplied or just as a PDB ID, following retrieval from our local PDB database. Following submission, the user is given some basic Protein Data Bank (PDB) format (33)—either user supplied or just as a PDB ID, following retrieval from our local PDB database. Following submission, the user is given some basic information about the protein molecule (chains; number of residues; ratio of ionogenic to all groups, , and warned about certain inconsistencies in structure, related to subsequent calculation (interruption in residue numbering which might influence appearance of terminal charges). The user is given the possibility to edit initial setup of ionogenic groups (attention to CYS in SSBONDs and excluding covalently modified groups). This is accomplished by convenient interactive selection of used set of groups. This gives opportunity for simulation of ‘electrostatic mutagenesis’. Full ‘charge mutant analysis’ is supposed for next versions. The same screen visualizes the PDB file in a text field which allows for direct editing: adding missing terminal charges, fixed (non-titratable) whole or partial charges and titratable groups with user defined . All other parameters used as input are predefined or automatically calculated. After initial setup completion the calculation proceeds through aforementioned steps—evaluation of accessibilities and Born term , perturbation of partial charges and finally the iterative procedure for self-consistent evaluation of titratable .

To calculate , the following energy conversion units were used: 1 kcal = 4.186 kJ = 1.68RT units (at 298 K) = 0.735 kcal/mol. The units of (pH) are kcal/mol·e = 43.176 mV or 30.24 mC/m². All calculations are provided at ionic strength (is) 0.1.

The obtained results are organized in two groups: (i) GLOBAL (Z(pH), , (pH) and at fixed pH) and (ii) LOCAL ( , , ). For each of them there is a link to own page. The contents of each page is comprised of the result itself, related derivatives (e.g. , , and so on) as well as a short description and examples for visualization of this type of data. All output data files are in standard plain text format. Visualization is straightforward with any 2D plotting software and molecular graphics programs (RasMol, JMol, PyMol and so on).

RESULTS

Global pH-characteristics

(1) pH-dependent protein net charge (pH) and its derivatives: Isoelectric point (pI) Z = 0 and protein buffer capacity .

It is equivalent to experimental potentiometric titration curve (34) and reflects basic global electrostatic characteristic of protein proton binding (35). The definition of pI is pH at which Z = 0. Protein buffer capacity is an important parameter for design of precise ion-exchange and electrophoresis (37) experiments. The difference between Z(pH) of analogous but perturbed states (e.g. apo-holo (30), oxidized-reduced, free-ligated and the like) can be useful in analysis the nature of such perturbation and identify pH region where it has maximal effect on proton binding. Other relevant issues are: the net charge of protein under condition of electro-spray mass-spectrometry (38); the critical Z-values at extreme pH in water (39,40) and in vacuum (41) at which protein ‘denature’ and many others (Supplementary Figure S1).

(2) pH-dependent electrostatic free energy term and its derivatives: , pH and pH at .

Quantitative estimate for charge dependent stability is basic electrostatic characteristic of protein molecules (2). By evaluating it is possible to determine pH-dependent specific ion and/or cofactor binding (30). Similarly ‘electron affinity’ can be evaluated from difference .

Other relevant issues are: the net charge of protein under pH region where it has maximal effect on proton binding. Use of stability of pH-induced conformational states and evaluation of energetic barrier between them (44). Presence or absence of stricture ruled charge asymmetry is reflected in difference (also from their pH min — pI) (Supplementary Figure S2).

(3) Electrostatic potential, (pH) at selected pH for all j-th protein non-hydrogen atoms in a PDB-formatted file and can be visualized in color scale by RasMol.

The electrostatic potential at each point within (45), on the molecular surface (46) and at near vicinity in solvent (47) for a protein molecule is its fundamental electrostatic characteristic (8,48). In fact all above quantities are derivative of \( \Phi_{el} = f(pH, ligands) \). Using present PHEPS version output file, it is straightforward to visualize (or EP) at each protein non-hydrogen atoms by switching on ‘color by temperature’ using color scale (dark blue: positive EP; green: zeroed EP; and red: negative EP) applicable to entire variety of RasMol model representations (Supplementary Figure S3).

Local pH-characteristics

(4) pH-dependent proton population or degree of ionization of each i-th ion group (Si).

The results for ion groups in order of increasing sequence numbers are presented in the form of column formatted file (all in one table).

For pH (pH) can be related to NMR pH-dependent chemical shifts, (pH) (49,50) or other individual titration characteristics—FTIR carboxylate titration (51); differential Tyr UV-titration (52); calorimetric/enthalpy titration (53) and so on (Supplementary Figure S4).

(5) pH-dependent proton affinity (pKa,i(pH)) at each individual i-th ion site: The results for each i-th ionizable group (their ) is available in another table. The set of for each i-th ionizable group (their ) is available in another table.

Predicted can be compared directly to experimentally obtained. Plotting is a fast way to differentiate
The results and protonic/ionic charges—individual sites and their sum. Calculated package PHEI were developed and have been tested for stability (60,62) and many others (Supplementary Figure S7). The intermediate species of enzyme catalytic cycles (59), protein for evaluation of the effect of whole protein electrostatic field (59,60). This characteristic is of indispensable use dipoles of different kind of ligands in static and dynamic inter-molecular interactions with charged groups (ions and like (55,56) (Supplementary Figure S5).

Tested proteins. All these features of our program package PHEI were developed and have been tested for many years. The method was applied to numerous proteins [Supplementary Table 1, (20–30)]. Calculated pK\textsubscript{i;f} was compared with experimental estimates of pK\textsubscript{s} (21,22) and correlation was made of calculated Z(pH) to published experimental curves.

CONCLUSION AND FUTURE DEVELOPMENT

We hope this server will be useful to anyone who needs fast and detailed analysis of pH-dependent properties of a protein with known atomic structure and a tool for protein electrostatic design (61,62). We are ready to share our experience in the field with other protein scientists and are open for discussion.

Features in preparation for next PHEPS version are as follows:

- For each AA in sequence order n (backbone, side chain and residue) with respective (B-factor), static accessibility (SA), and G, now implemented.
- 3D-contour EP map generation (in static and dynamic regime)—search of saddle and other critical points on multidimensional maps.
- Correct determination of dipole (at pI) and electric (at any other pH) moments (μ\textsubscript{d} and μ\textsubscript{e}, respectively) using 3D-EP—grid—their scalar and vector values.

- Thorough ‘electrostatic mutation’ analysis with ‘Δ = mut – wild’ as function of pH—data for all mentioned above characteristics: ΔZ, ΔS, ΔpK\textsubscript{i}, ΔG\textsubscript{el}, ΔG\textsubscript{mut}, ΔΦ\textsubscript{el}.
- EP gradients (electrostatic forces, EF) at pH control, located at defined atoms and sites (user selected fragments, domains, subunits).

Many of these features are implemented in our program package PHEI, but their online access will be realized after extensive testing.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Profs B. Honig and E. Alexov for kind donation of computers, one of which hosts our server. We thank L. Roumenina for her kind help in text edition and correction. This work is partially supported by grant X-1310 of National Fund ‘Scientific Research’, Sofia, Bulgaria. The Open Access publication charges for this article were waived by Oxford University Press.

Conflict of interest statement. None declared.

REFERENCES

1. Perutz, M.F. (1978) Electrostatic effects in proteins. Science, 201, 1178–1191.
2. Honig, B. and Nicholls, A. (1995) Classical electrostatics in biology and chemistry. Science, 268, 1144–1149.
3. Warshel, A. (1981) Electrostatic basis of structure–function correlation in proteins. Acc. Chem. Res., 14, 284–290.
4. Bashford, D. and Karplus, M. (1990) pK\textsubscript{s} of ionization groups in proteins: atomic detail from a continuum electrostatic model. Biochemistry, 29, 10219–10225.
5. Antosiewicz, J., McCammon, J.A. and and Gilson, M.K. (1994) Prediction of pH-dependent properties of proteins. J. Mol. Biol., 238, 415–436.
6. Simonson, T. (2001) Macromolecular electrostatics: continuum models and their grooving pains. Curr. Opin. Struct. Biol., 11, 243–253.
7. Bashford, D. (2004) Macroscopic electrostatic models for protonation state in proteins. Front. Biosci., 9, 1082–1099.
8. Warshel, A. and Papazyan, A. (1998) Electrostatic effects in macromolecules: fundamental concepts and practical modeling. Curr. Opin. Struct. Biol., 8, 211–217.
9. Braun-Sand, S. and Warshel, A. (2005) ‘Electrostatics of proteins: models and principles’. In Buchner, J. and Kiefhaber, T. (eds), Protein Folding Handbook, Part I. New York: Wiley-VCH GmbH & Co.
10. Aguist, J. (1996) Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions. J. Comput. Chem., 17, 1587–1597.
11. Matthews, J.B. and Gurd, F.R.N. (1986) Calculation of electrostatic interactions in proteins. Methods Enzymol., 130, 413–453.
12. Petersen, M.T.N. and Petersen, S.B. (2001) How to lipases and esterases work: the electrostatic contribution. J. Biotecnol., 85, 115–147.
13. Feig, M., Onufriev, A., Lee, M.S., Im, W., Case, E.A. and Brooks, C.L.III (2004) Performance comparison of generalized Born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. J. Comput. Chem., 25, 265–284.
14. Naylor, M., Hitz, B.C. and Honig, B. (1999) GRASS: a server for the graphical representation and analysis of structures. Protein Sci., 8, 676–679.
15. Dolinsky, D.J., Nielson, J.E., McCammon, A.J. and Backer, N.A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatic calculations. Nucleic Acid Res., 32, W665–W667.
