OXFORD

Structural bioinformatics

DelPhiPKa web server: predicting pK_a of proteins, RNAs and DNAs

Lin Wang[1,*], Min Zhang[2] and Emil Alexov[1,*]

[1]Computational Biophysics and Bioinformatics, Department of Physics and Astronomy, Clemson University, Clemson, SC 29634, USA and [2]Department of Computer Sciences, East Carolina University, Greenville, NC 27858, USA

*To whom correspondence should be addressed.

Associate Editor: Ivo Hofacker

Abstract

Summary: A new pK_a prediction web server is released, which implements DelPhi Gaussian dielectric function to calculate electrostatic potentials generated by charges of biomolecules. Topology parameters are extended to include atomic information of nucleotides of RNA and DNA, which extends the capability of pK_a calculations beyond proteins. The web server allows the end-user to protonate the biomolecule at particular pH based on calculated pK_a values and provides the downloadable file in PQR format. Several tests are performed to benchmark the accuracy and speed of the protocol.

Implementation: The web server follows a client-server architecture built on PHP and HTML and utilizes DelPhiPKa program. The computation is performed on the Palmetto supercomputer cluster and results/download links are given back to the end-user via http protocol. The web server takes advantage of MPI parallel implementation in DelPhiPKa and can run a single job on up to 24 CPUs.

Availability and implementation: The DelPhiPKa web server is available at http://compbio.clemson.edu/pka_webserver.

Contact: lwang3@clemson.edu or ealexov@clemson.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Biological processes are affected by the pH of the cellular environment. For example, enzyme catalysis, protein folding and binding, protein-protein interactions are all pH-dependent (Alexov *et al.*, 1997; Mitra *et al.*, 2011; Wang *et al.*, 2012). Different protonation states of titratable residues within molecules due to pH changes can result in significantly different complex conformations and binding free energies. Fast and accurate predictions for pK_a values of ionizable groups will facilitate investigations aiming at designing better drugs and more robust enzymes, which are active and stable over a wide range of pH.

Although there are many programs for predicting pK_a values of protein ionizable residues such as MCCE (Alexov *et al.*, 1997; Song *et al.*, 2009), UHBD (Madura *et al.*, 1995), PROPKA (Li *et al.*, 2005), H++ (Gordon *et al.*, 2005), which are based on various methods, to the best of our knowledge, there is no pK_a prediction web sever that provides pK_a calculations for RNAs and DNAs (Tang *et al.*, 2007)

bases as well. Here, we developed a new pK_a web server that implements the following features (i) calculates pK_a values for the protein, RNA and single stranded DNA, (ii) protonates the structure at user-specified pH according to calculated pK_a, (iii) provides the option of various force field parameters (AMBER, CHARMM, PARSE, GROMOS), (iv) calculates electrostatic free energy components with the Gaussian dielectric function without defining molecular surface and (v) provides different hydrogen conformations. The web server is implemented in DelPhiPKa program and built on the supercomputer cluster to take advantage of MPI parallelization, which improves the computation speed up to 20 times.

2 Methods

2.1 Back-end program DelPhiPKa

The web server is built on the pK_a calculation program—DelPhiPKa, which is written in the object-oriented C++ code. The methodology

of the back-end program is described in the supplementary material and more details can be found in Wang *et al.* (2015).

2.2 Overall architecture of the web server

The server contains two components, the first of which is user interface and second is the job submission/queue/running system. The user interface is implemented using PHP and HTML. It contains various parameters for user input that control the back-end program and results calculated by the algorithm. The job submission/queue system uses a job monitor module written in bash/python scripts to sends/retrieves the job information to/from the Palmetto supercomputer cluster.

2.3 User input parameters of the web server

The web server requires the following user input parameters and each parameter has the corresponding function:

- *Force field.* The server reads atomic charges and radii from the specific force field parameter file to calculate electrostatic potentials and to produce the output PQR file.
- *Remove HETATM.* The server can remove all HETATM information in the uploaded PDB file, making those HETATM not involved in the calculations.
- *HETATM in PQR format.* If selected, the server reads the HETATM information that is provided by users in PQR format within the uploaded PDB file. This is applied to the calculations involving ligands and solvent ions, for which the atomic information is not included in the standard topology parameter file.
- *Output protonated PQR file based on calculated pK_a results.* This gives the option for users to download the protonated structure in PQR format based on calculated pK_a results and specific pH.
- *Given pH value.* It's related to the previous parameter as the user-defined pH value to produce PQR output file.
- *Hydrogen of ASP/GLU attached atom.* These options provide users to choose the hydrogen positions of aspartic and glutamic acids (which atom should the hydrogen placed with) to be generated.
- *Variance of Gaussian distribution.* This is the sigma value in the Gaussian function and our benchmark results showed that with a low value (default is 0.70) the program performs better on surface residues while a high value (e.g. 0.93) is better for buried residues.
- *Reference dielectric.* This is the reference dielectric constant for the protein in the Gaussian function and the default is set 8.0 based on our benchmark results.
- *External dielectric.* This is the external dielectric constant for the solvent water in the Gaussian function and the default is set to be the widely adopted value of 80.

2.4 Results page and download links

The server provides pK_a results with three download items. *pKa.csv,* which contains the calculated pK_a value for each individual ionizable residue with electrostatic energy components for its protonated and de-protonated states; *titration.txt,* which reports each individual ionizable residue probability of protonation from pH of 0 to 14; *(input_pdbname).pqr,* which is the protonated structure in PQR format with each individual ionizable residue in its protonated or de-protonated state based on the calculated pK_a values and user selected pH.

3 Performance

3.1 Test of accuracy

We performed the benchmark with calculated pK_a values against experimental results with three force fields (AMBER/CHARMM/ PARSE) to test the accuracy of pK_a predictions. The dataset we used is from Protein pK_a Database (Toseland *et al.,* 2006), which contains 302 experimental measured pK_a values for titratable residues from 32 proteins. The parameters we used for this dataset are $\sigma = 0.70$, $\varepsilon_{ref} = 8$, $\varepsilon_{extern} = 80$ and others with default values. With three force fields, we achieved an average RMSD of 0.77 (Supplementary Fig. 1S) and about 90% of the overall predictions less than 1.0 pK unit error. For surface residues (72% of the dataset), we achieved an average RMSD$\approx$0.55 (Supplementary Table 1S).

We also benchmarked calculations on two RNAs with experimental measured results by NMR spectroscopy: branch-point helix (BPH) and lead-dependent ribozyme (LDZ) with 5 and 7 measured pK_a's for adenosine residues, respectively. The topology information and corresponding parameters used for nucleic acids are described in the supplementary material. Results in Table 2S showed the pK_a values were successfully identified in the correct order and 90% of predictions are less than 0.6 pK units error.

3.2 Test of speed

The webserver is built based on DelPhiPKa program, which implemented with MPI parallelization for the distributed computing. The server could run the job on the Palmetto supercomputer cluster with 8 or 24 CPUs. Here, we performed a speedup test of the parallelization. According to the Supplementary Fig 2S, the speedup could reach 7.5× and 19.4× with 8 and 24 CPUs (by increasing the number of CPUs, the huge memory usage and more communications between CPUs degrade the speedup).

Funding

Conflict of Interest: none declared.

References

Alexov,E. *et al.* (1997) Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. *Biophys. J.,* **72,** 2075.

Gordon,J.C. *et al.* (2005) H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res.,* **33,** W368–W371.

Li,H. *et al.* (2005) Very fast empirical prediction and rationalization of protein pKa values. *Proteins Struct. Funct. Bioinf.,* **61,** 704–721.

Madura,J.D. *et al.* (1995) Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. *Comput. Phys. Commun.,* **91,** 57–95.

Mitra,R.C. *et al.* (2011) In silico modeling of pH-optimum of protein–protein binding. *Proteins Struct. Funct. Bioinf.,* **79,** 925–936.

Song,Y. *et al.* (2009) MCCE2: improving protein pKa calculations with extensive side chain rotamer sampling. *J. Comput. Chem.,* **30,** 2231.

Tang,C.L. *et al.* (2007) Calculation of pKas in RNA: on the structural origins and functional roles of protonated nucleotides. *J. Mol. Biol.,* **366,** 1475–1496.

Toseland,C.P. *et al.* (2006) PPD v1. 0—an integrated, web-accessible database of experimentally determined protein pKa values. *Nucleic Acids Res.,* **34,** D199–D203.

Wang,L. *et al.* (2013) In silico investigation of pH-dependence of prolactin and human growth hormone binding to human prolactin receptor. *Commun. Comput. Phys.,* **13,** 207.

Wang,L. *et al.* (2015) pKa Predictions for proteins, RNAs and DNAs with the Gaussian Dielectric Function Using DelPhiPKa. *Proteins,* **83,** 2186–2197.