DelPhiPKa User Manual

Computational Biophysics & Bioinformatics

Prof. Emil Alexov Lab at Clemson University

Table of Contents

1. Introduction

1.1 What is DelPhiPKa

DelPhiPKa is a DelPhi based open source C++ program, allowing to predict pKa's of ionizable groups of proteins, RNAs and DNAs. Some the unique approaches and features include:

- Use Gaussian based dielectric function to mimic conformational changes associated with ionization changes.
- Calculate the electrostatic energy without defining the molecular surface.
- Provide an option of various force field parameters.
- Provide different hydrogen conformations.
- Protonate the structure at particular pH using calculated pKa values for ionizable residues

1.2 What is DelPhiPKa Web Server

The DelPhiPKa web server is built on DelPhiPKa program and distributed on the Palmetto supercomputer cluster held in Clemson University. The web server allows researchers to use the pKa calculation program without installing the standalone code.

Since DelPhiPKa implements MPI library, the web server allows users to submit the job for parallel computing on 8 and up to 24 CPUs.

The web server provides the download of the pKa calculated results, the titration curves and the protonated structure in PQR format based on pKa predictions and user specified pH.

2. Installation

2.1 The compilation environment

The DelPhiPKa program is designed to be compiled and run on Linux/Unix and Mac OSX operation system. To compile the code, a C++ compiler and several libraries are required:

1. C++ Compiler (https://gcc.gnu.org)

We used GNU GCC to compile the code. The compilation is tested with Clang and Intel compilers on OSX. Use version 4.4 or above to compile, which includes C++11 features.

2. Boost Library (http://www.boost.org)

Boost library is used in DelPhi C++ code. Since DelPhi C++ is a part of DelPhiPKa program, boost library is required for compilation. Use version 1.55.0 or above.

3. OpenMPI (http://www.open-mpi.org)

The DelPhiPKa program implements MPI library to parallelize the energy calculation module and the titration module. To obtain the best efficiency, use Open-MPI library to compile the code. Use version 1.8.1 or above.

If you do want to compile the sequential code, do the following:

    a. Edit *prime_environment.h* file in src/delphiPKa directory.
    b. Delete or comment out these two lines:

```
#define MPI_PARALLEL
#include <mpi.h>
```

c. Edit *Makefile* and change `CC=mpic++` to `CC=g++` or `CC=c++`, which depends on your compiler

4. GSL Library (http://www.gnu.org/software/gsl/)

The GSL library is used for fitting the titratation curves and it is required for compilation. Use version 1.15 or above.

5. Command Line Tool and Xcode package (For OS X user only)

For users of OSX 10.8 and above, you need to download and install Command Line Tool and Xcode (optional) to compile the program. Clang compiler is the default C++ compiler comes with Xcode package and has been fully tested.

2.2 How to compile the program

With the required libraries and C++ compiler above, run

`make`

in the directory contains *Makefile* to compile the program.

3. Basic Tutorial

3.1 What are the files in *param* folder.

Files in *param* folder are force-field parameter file and topology file. Currently it contains AMBER, CHARMM, PARSE, GROMOS force fields. The format is designed to be identical as DelPhi utilized *.crg atomic charges and *.siz atomic radii files.

Topology file contains heavy atom bond connectivity, hydrogen positions, residue types. It also contains reference pKa value for each ionizable residue group.

The force field parameter files and topology file are required to run the DelPhiPKa program. Individual entry can be edited for specific purpose.

3.2 Edit the runtime control file *run.prm*

The entries in control file *run.prm* set the runtime parameters used in the program. Four entries must be edited before running the program. You can leave the rest of parameters as default, or edit them as you desire.

Required editing:

PDB file name
Specify the PDB name. Currently, DelPhiPKa only supports standard PDB format. If other format is used, for example PQR format is used, the program will only read xyz coordinates and charge/radius values will be skipped.

Charge parameter
Specify the atomic charges parameter file. If the *param* folder is located in other directory, you need to specify the corresponding directory.

Radius parameter

Specify the atomic radii parameter file. Modify the directory if needed as above.

Topology parameter

Specify the topology parameter file. Modify the directory if needed as above.

Other control entries (can be left as default):

Remove HETATM

Remove all HETATM information in PDB file, making those HETATM not involved in the calculations. Default is T (true).

Remove water molecule

Remove all water molecule in PDB file. Default is T (true).

HETATM in PQR format

If you want to take into account ions or ligands (HETATM) involved in the calculation, make this entry as T (true) and make Remove HETATM option as F (false). Thus, ions and ligands will be treated as permanent charges. The program will not output pKa values for those, however their existence as permanent charges will affect the pKa's of ionizable residues on the macromolecule. If you want to use the feature, the corresponding HETATM lines in the PDB file have to be modified in PQR format. As the charges and radii information for these atoms are not included in topology file, users are responsible for editing them in PQR format. For more details, refer to the Advanced Tutorial section.

Do Protonation

If you want to generate the protonated structure in PQR format, make this entry as T (true).

Do Energy Calculation

Run the energy calculation module and generate *energy.txt* and *pairwise.txt* files, which calculates the electrostatic polar energy (in *energy.txt*), the desolvation energy (in *energy.txt*), the charge-charge pairwise interaction energy (in *pairwise.txt*). The default is T (true). If you have previously calculated *energy.txt* and *pairwise.txt* output files, and make this entry as F (false), then the program will skip running this module and read energy terms from those two files and continue to do pKa calculations.

Do pKa's Calculation

Generate the titration curves and calculate pKa values. Default is T (true).

Output PQR file (with Topology)

Add hydrogens to the PDB and add the corresponding atomic charge and radius to each atom (PQR file). This step does not need to do energy and pKa calculation. It is the fast way to obtain the protonated structure. Default is F (false).

Output PQR file (with pKa result)

Similar as the previous entry, but protonate each ionizable residue based on its calculated pKa value at the user defined pH. At particular pH, each ionizable residue can be either in its protonated or deprotonated state depends on the pKa value. Default is F (false).

At given pH value

Associated with the previous entry, a user defined pH value is considered.

Gaussian surface

Set "1" to use the smooth Gaussian dielectric model to calculate electrostatic potentials; "0" to use homogenous dielectric model.

Variance of Gaussian distribution

This is *sigma* in the Gaussian distribution formula, which determines how the Gaussian function assigns the dielectric constant for the protein and protein-water interface. The protocol is based on how atoms are packed. If atoms are tightly packed, a low value is assigned for the epsilon; if atoms

are loosely packed, a high value is assigned for the epsilon. The assigned value is between *Internal Dielectric* and *External Dielectric* in the next entry. According to our benchmark results against experimental data, 0.70 is set as default, because with this value we obtained the best RMSD for surface residues on native proteins. If your target is buried residue or mutant protein (mutation site is buried), set the value to 0.90-0.95. Currently, there is no unique value for this entry.

Internal Dielectric

The reference dielectric constant in the Gaussian distribution formula for the protein interior. Default is 8.0 according to our benchmark results against experimental data.

External Dielectric

The dielectric constant in the Gaussian distribution formula for the water.

Cluster Delimitation Threshold (A)

This is the threshold for the distance within each network. Its recommended value is greater than 10, but less than 15 for efficiency. Default is 12 (angstrom).

Hydrogen of GLU Attached to Atom

The hydrogen position to be placed with, can be either OE1 or OE2 atom of the glutamic acid (GLU). Default is OE1 atom.

Hydrogen of ASP Attached to Atom

The hydrogen position to be placed with, can be either OD1 or OD2 atom of the aspartic acids (ASP). Default is OD1 atom.

pH Initial Value

The initial pH value to start titration. Default is 0.

pH End Value

The final pH value to end titration. Default is 14.

pH Interval

The pH interval during titration. Default is 1.0

3.3 How to run the program

With required force field parameter and topology files, and proper *run.prm* file, you are able to run the program.

If compiled with Open-MPI implementation, run: (x is the number of CPUs you want to use)

```
mpirun -np x delphiPKa run.prm
```

If compiled the sequential version, run:

```
delphiPKa run.prm
```

3.4 Results and output files

If the job runs successfully, it generates several output files.

pKa.csv

This csv file gives the pKa value for each ionizable residue with associated energy terms (the unit here is kcal/mol). The energy terms include electrostatic polar energy for individual residue in its protonated (+/-) state and neutral state, the desolvation energy for individual residue in its protonated (+/-) state and neutral state.

titra.csv

This csv file is the titration curve. It contains the probability of each residue in its ionization state at particular pH from 0 to 14.

energies.txt

This file contains the polar energy terms and the desolvation energy terms (the unit is kcal/mol), they are the same as in *pKa.csv* file.

pairwise.txt

This file contains the charge-charge pairwise interaction energy terms (the unit is kt).

(pdb_name)_1.pqr

This PQR file is the protonated structure based on the topology parameters.

(pdb_name)_2.pqr

This PQR file is the protonated structure based on the calculated pKa value for each ionizable residue.

4. Advanced Tutorial

4.1 Edit the topology file

Topology file is a parameter file contains information for each residue such as heavy atom bond connectivity, hydrogen positions, residue types. It also contains reference pKa value for each ionizable group. Users can access and edit this file for their specific purpose.

The line that starts with "#" is skipped.

The line starts with "$" is read as atom information. The format is

```
#   res atom obtal conf batm batm batm batm
#   |---|----|-----|----|----|----|----|----|·
$   ASP  CB   sp3   SD   CA   CG   HB1  HB2
```

res: residue type (e.g. ASP)

atom: atom type (e.g. CB)

obtal: atomic orbital type (e.g. sp3 hybrid orbitals)

conf: structure type (e.g. SD, side-chain)

batm: bond atom type (e.g. CA-CG-HB1-HB2)

Hydrogen naming rule like HB1 and HB2 is used in the program. If you want to use the naming rule like 1HB, 2HB instead of the default one, you can modify all HB1 to 1HB in the topology file.

If you want a specific atom (for example atom XX) to be bonded to CB atom of ASP instead of CG atom in the example, you can replace CG atom with the one that you desire (the XX atom) and also modify the CG entry with the name of your desired atom (the XX atom) and the corresponding bonded atoms in that entry.

The line starts with "*" is read as the reference pKa value for individual ionizable side-chain. The default value is set according to our benchmark results against experimental data. For specific purpose, users can access and edit these values.

4.2 Edit HETATM in PQR format

DelPhiPKa is able to treat ions and ligands (which are HETATM in PDB) as permanent charges and calculate their effects on protein ionizable residues. This can be applied to model cases involving structures with HETATMS. To achieve so, you need to modify your PDB file and make HETATM (ions/ligands) into PQR format. Here is an example, the original PDB file contains zinc and calcium ions:

```
ATOM    2694  N   GLY A 263     -2.736   7.378 -10.285  1.00 13.73           N
ATOM    2695  CA  GLY A 263     -3.940   6.565 -10.246  1.00 15.73           C
ATOM    2696  C   GLY A 263     -4.985   6.958 -11.273  1.00 19.15           C
ATOM    2697  O   GLY A 263     -4.828   7.910 -12.047  1.00 20.70           O
TER
HETATM  2703 ZN    ZN A 301      6.369  24.132  -0.778  1.00  8.95          ZN
HETATM  2704 ZN    ZN A 302     19.334  21.221  -1.079  1.00  7.89          ZN
HETATM  2705 CA    CA A 303     16.638  10.931  -6.150  1.00  9.52          CA
HETATM  2706 CA    CA A 304     18.964  21.655  14.129  1.00  9.28          CA
HETATM  2707 CA    CA A 305     16.174  28.276   7.457  1.00  8.25          CA
```

And they are modified into PQR format as:

```
ATOM    2694  N   GLY A 263     -2.736   7.378 -10.285  1.00 13.73           N
ATOM    2695  CA  GLY A 263     -3.940   6.565 -10.246  1.00 15.73           C
ATOM    2696  C   GLY A 263     -4.985   6.958 -11.273  1.00 19.15           C
ATOM    2697  O   GLY A 263     -4.828   7.910 -12.047  1.00 20.70           O
TER
HETATM  2703  ZN   ZN A 301      6.369  24.132  -0.778  2.0000 1.7300
HETATM  2704  ZN   ZN A 302     19.334  21.221  -1.079  2.0000 1.7300
HETATM  2705  CA   CA A 303     16.638  10.931  -6.150  2.0000 2.3000
HETATM  2706  CA   CA A 304     18.964  21.655  14.129  2.0000 2.3000
HETATM  2707  CA   CA A 305     16.174  28.276   7.457  2.0000 2.3000
```

As the atomic charge for ZN is 2.0000 and for CA is 2.0000. The atomic radius for ZN is 1.7300 and for CA is 2.3000.

Be cautious with this feature, because making the entries into PQR format is crucial for the calculation. If you do these entries incorrectly or leave as the original (for example), the program will read 1.00 as the atomic charge and 8.95 as the atomic radius for Zinc ion, which would cause serious errors in the calculations.

Last Updated: August, 2015. Dr. Lin Wang, Computational Biophysics and Bioinformatics, Department of Physics, Clemson University.