Thank you for downloading **P**rotein-protein **A**ssociation **R**ate **C**onstant prediction Package.

--------------------

Version Dec. 2015

--------------------

Dec. 12 2015 the first version + a small modification

--------------------

Contents Summary

--------------------

This package contains 2 Perl Scripts, 2 Fortran and 4 C++ source codes, 4 executable file, a Matlab Workspace file, and several documents and .pdb files with corresponding result/output files:

* install\_PARC.pl: before execute the main script, please install the whole package to make sure all the source codes would be compiled successfully and correctly.

$ perl install\_PARC.pl

* PARC2015.pl: the main script (including the code for preprocessing procedure)

$ perl PARC2015.pl -pdbid xxxx Y Z -trj n0 -atom1 n1 -atom2 n2 -res1 n3 -res2 n4

-pdbid xxxx A B (necessary; xxxx: 4 digit pdb ID, e.g. 1e4k; Y: chain name(s) of subunit 1, e.g. AB; Z: chain name(s) of subunit 2, e.g. C)

-trj (optional; No. of trajectories; default value 10,000)

-w1 d.dd (optional; the weight of electrostatic interaction; default value 1)

-w2 d.dd (optional; the weight of hydrophobic interaction; default value 0.04)

-T d.dd (optional; Temperature factor; default value 5)

-xi d.dd (optional; the Debye length; default value 9.5)

-atom1 n (necessary; No. of atoms on subunit 1)

-atom2 n (necessary; No. of atoms on subunit 2)

-res1 n (necessary; No. of residues on subunit 1)

-res2 n (necessary; No. of residues on subunit 2)

-dx1 d.dd (optional; will be automatically calculated based on molecular weight)

-dx2 d.dd (optional; will be automatically calculated)

-dr1 d.dd (optional; will be automatically calculated)

-dr2 d.dd (optional; will be automatically calculated)

* parc2015: the main executable program for simulating the association between 2 subunits of interacting protein complex
* parc2015trj: the alternative main executable program for simulating the association between 2 subunits of interacting protein complex, but the protein translation & rotation trajectories will be recorded
* PARC\_surf: An auxiliary executable program to compute the Electrostatic potential on the binding surface
* PARC\_nat: An auxiliary executable program to compute the total Electrostatic potential
* PARC2015.mat: a Matlab Workspace file, which contains the trained decision tree model using 49 complexes training set
* other documents or files: 3 C++ and 2 Fortran source codes for compiling the executable files as well as a couple of .pdb files and their corresponding output files

--------------------

Execution Procedures

--------------------

1. Run install\_PARC.pl to compile the executable files

$ perl install\_PARC.pl

1. Run PARC2015.pl to simulate the protein-protein association

$ perl PARC2015.pl -pdbid 1brs A D -trj 1000 -atom1 864 -atom2 693 -res1 108 -res2 87

1. Obtain the preliminary predicted rate constant and the 3 input indicators for the Machine-learning tool in an output file

$ vi results/results\_xxxxxx.txt

1. Open PARC2015.mat, the Matlab Workspace file containing the trained decision tree model, and input the three variables (please find them in the output result file) into the trained model, e.g.

fx >> x = table(1.5517, 61.5385, 45.8333)

fx >> y = predict(all, x{1,1:3})

1. Divide the preliminary predicted kon by a factor, f. Depends on the output of decision tree, i.e. y, the prediction of whether the preliminary predicted kon is likely to over-estimated or not, the f would be either 19.89, if y were "y" or 0.34 if y were "n". In this example above, y = "n"; therefore, the final predicted kon would be 3.28e7/0.34 = 9.65e7/Ms

--------------------

Results/Outputs

--------------------

A standard preliminary results file should look like the text attached below and contains the information including the Electrostatic potential of the binding surface and the whole protein, the percentages of interface residues on the flexible loops of both 2 interacting subunits, the maximum time duration, the number of trajectories and the preliminary predicted association rate constant. Please find it in the results folder.

1brs

electrostatic potential of the binding surface: -66.906

electrostatic potential of the whole protein: -43.1165

the ratio r(elec): 1.5517

%interface residues on the flexible loops of chain A: 61.5385 %

%interface residues on the flexible loops of chain D: 45.8333 %

500 trajectories have been recorded & collected.

PDB id: 1brs; Max. duration: 1000 ns; no. trj: 10000/10000; No. of successfully associated: 26

Predicted kon of 1brs: 3.284570e+07 /Ms

--------------------

Note

--------------------

PLEASE Make Sure the **NUMBER** of **atoms** and **residues** match the input .pdb file correctly and exactly, OR the execution and computation process may produce errors and the predicted results could be incorrect!

--------------------

Contact Information

--------------------

We are more than happy to take any questions, comments, or suggestions.

If it does not work for you or you have identified any bugs, please do not hesitate to contact us:

Paul Zhong-Ru Xie Ph. D.

amianxie@gmail.com

Thank you for using PARC package and providing us with your precious feedback!