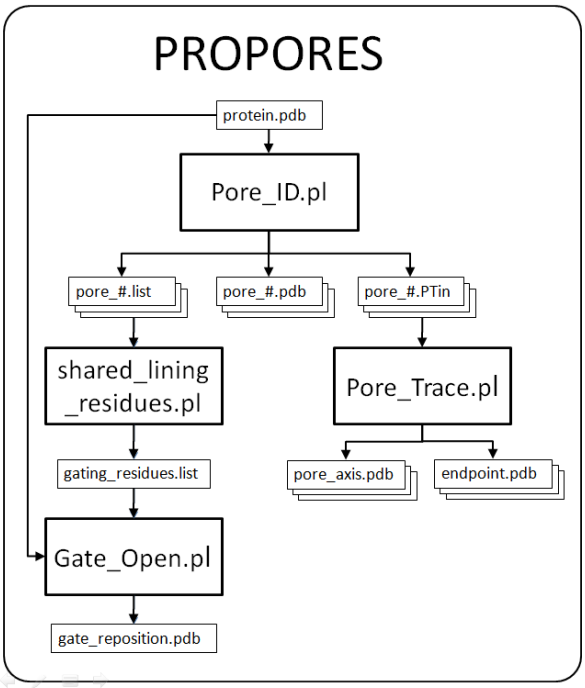
User Guide

Po-Hsien Lee
p.lee@mx.uni-saarland.de
Center for Bioinformatics, Saarland University, D-66041 Saarbruecken, Germany
August, 2011

1 Introduction

"PROPORES" is a novel toolkit for identifying pockets, cavities and channels of protein structures. The toolkit was developed in PERL programming language and includes "PoreID" for pore identification, "PoreTrace" for pore axes determination and "GateOpen" for opening the gate between neighboring pores. "PoreID" is a grid-based method that avoids orientation dependency of the results. It targets all kinds of pores (pockets, cavities and channels) and is automatic so that only the PDB file of the target protein has to be specified by the user. Several descriptors of pores such as volume, lining residues, axes, and radius profiles are provided by "PoreID" and "PoreTrace". In addition, for two neighboring pores, "GateOpen" can be used to test connectivity by rotating side chains of the gating residues. This was done to introduce a dynamic variable into the pore identification of the otherwise static protein structure.

2 Scheme of the toolkit



3 Content of the toolkit

The toolkit contains

One file for the GPL license

COPYING

One file for the instruction of installation

INSTALL

Two shell startup files

EnvVars.sh

EnvVars.csh

Four programs written in PERL.

Pore_ID.pl

Pore_Trace.pl

Gate_Open.pl

shared_lining_residues.pl

One document

UserGuide.pdf

Three folders

module_lib	6 perl modules : PDBreader.pm, Vector.pm, KD_tree_NN.pm, Parameter.pm, Dihedral.pm, Quat_rot.pm
ROTA_lib_30	Rotamer libraries of 17 amino acids
Examples	Examples of Pore_ID, Pore_Trace and Gate_Open

4 Installation

The toolkit has been tested on Linux systems. PERL executing environment is required for running this toolkit.

4.1 Download the toolkit “PROPORES.tar.gz” at the Prof. Volkhard Helms’ group website

http://gepard.bioinformatik.uni-saarland.de

4.2 Extract the compressed file

tar xzf PROPORES.tar.gz

4.3 Edit the file “EnvVars.sh” or “EnvVars.csh” according to your working shell for assigning the path of modules and libraries used in the PERL programs. Configure the environment variable “INSTALL_PATH” by the path where the toolkit is installed. For example:

INSTALL_PATH=/home/lee/ PROPORES

4.4 Source EnvVars.sh (or EnvVars.csh) before using the toolkit.

source EnvVars.sh

5 Usage, Input and Output

5.1 Pore_ID.pl

Pore_ID.pl is the program for pore identification. The usage is

```
usage : perl Pore_ID.pl -f pdb_file [-r 1.0 -s 1.2 -c 1.4 -n]
        -f      PDB file of the targeted protein
        -r      side length of the grid voxel (default: 1.0 angstrom)
        -s      probe radius (default: 1.2 angstrom)
        -c      trimming depth for shallow pore region on the protein
                surface (default: 1.4 angstrom)
        -n      prefix of output files (default: the prefix of the .pdb file)
```

The input file is the coordinates of the targeted protein in standard PDB format. Pore_ID.pl identifies pores surrounded by protein residues, so all heteroatoms (start by HETATM) in PDB file will be ignored.

The output contains three files for each identified pore. These are pore_#.pdb, pore_#.list and pore_#.PTin. # is the index of the identified pore.

- pore_#.pdb contains the identified pore whose pore grid voxels represented in PDB format.
- pore_#.list contains the lining residues of the identified pore. The format is

```
A       398     HIS
A       237     SER
A       306     GLU
…
```

This is also the format for the list of gating residues used in Gate_Open.pl. Columns are separated by tabs. The first column is the chain ID. If the protein has no chain ID, the single space must be shown at the first column. The second column is the index of the residue. The third column is the type of residues in three letter code.

- pore_#.PTin is the input file for Pore_Trace.pl. The format is

-17.264 -15.713 -5.626 2.123472 2 24.104

…

The first three columns are the coordinates of the pore grid voxel. The fourth column is the radius of the sphere that can be squeezed in without clashing with protein atoms. The fifth column is the type of the pore grid voxel. “0” is for the pore grid voxel surround only by other pore grid voxels. “1” is for the pore grid voxel contacting with the protein atoms. “2” is for the pore grid voxel belonged to surface patches. The last column is the square of distance between the voxel and the center of mass of the protein. The unit is $Å^2$.

5.2 Pore_Trace.pl

Pore_Trace.pl is the program for pore axis determination. The usage is

usage : perl Pore_Trace.pl -f pore_#.PTin

-f pore_#.PTin file (the output from Pore_ID calculation)

The input file for Pore_Trace.pl is pore_#.PTin that is one of the output files of pore identification by Pore_ID.

The output contains pore axes and endpoints in PDB format. There is an extra column in the PDB files of the pore axes. This column is for the pore radii and the unit is Å. For a pore with two endpoints, the output contains two file for endpoints (endpoint_0.pdb, endpoint_1.pdb) and one file for pore axis (pore_axis_0_to_1.pdb).

5.3 Gate_Open.pl

Gate_Open.pl is the program for opening the gate of two neighboring pores. The usage is

usage : perl Gate_Open.pl -f pdb_file -i gating_residue_list -t clash_tolerance

-f PDB file of the targeted protein

-i file contains the list of gating residues

-t the tolerance of steric clashes (default: 1.0 angstrom)

The input files of the Gate_Open.pl are the PDB file of the targeted protein

and the list of shared lining residues of two neighboring pores. The format of the list is the same as pore_#.list.

The output is a PDB file named “gate_reposition.pdb” whose gating residues have been repositioned.

5.4 Others

There is also a simple PERL script “shared_lining_residues.pl” to figure out the shared lining residues between pores. The output is the lists of gating residues which is also the input file for Gate_Open.pl. For example, “gating_residues_1_2.info” contains the gating residues between pore_1 and pore_2. The usage is

usage : perl shared_lining_residues.pl <prefix> <work_path>

The <prefix> is the common prefix of the input (prefix)_#.list files
The <work_path> is the absolute path of directory where (prefix)_#.list files exist