The AlloType is written in C++ and relies on *Eigen* and *Boost* library, which can be downloaded and included easily. The source code is provided in the '/src' directory. Users can modify or compile it as you wish. One compiled program is also provided in the 'src/' directory named 'AlloType.exe' based on g++ compiler.

To help users better manipulate this program, one demo is provided in the 'examples/' directory. In this demo, the allosteric protein, met repressor (metJ), was selected. metJ is a DNA-binding protein and the allosteric modulator, SAM, can enhance the binding affinity of DNA. Therefore, this system is an allosteric activation system. The crystal structures of *apo* state (P), allosteric modulator binding state (PA), and ternary state (PAS) were collected in the 'www.rcsb.pdb', named '1cmb.pdb', '1cmc.pdb', and '1cma.pdb', respectively. These files were then checked to ascertain that the residue number was identical and no lacking segment exists. The 'Mg^{2+}' was deleted in the PA state.

After that, one can launch the program to perform prediction. The interactive I/O was designed for guidance step by step. First, one can see:

Enter dataset path:

Enter protein family name:

These two sentences locate the path of the PDB files, that is '*dataset_path*/*protein_family_name*'. One can also leave alone the second sentence, though. Next, the PDB name of these files, the spring constant (set to 1), and the cutoff (set to 9) are entered:

Enter PDB name for apo state:1cmb

Enter residues to be excluded:

Enter spring constant: 1

Enter cutoff: 9

[Info] Spring constant = 1.0 Kcal/(mol A^2).

[Info] Cutoff = 9.0 A.

[Info]Successfully loaded protein at path 'dataset_path/protein_family_name/1cmb.pdb'

[Info] Coordinate matrix, distance matrix, contact map have been generated for this protein.

Enter PDB name for binding state:

Enter PDB name for allostery state:1cmc

Enter residues to be excluded:

Enter spring constant: 1

Enter cutoff: 9

[Info] Spring constant = 1.0 Kcal/(mol A^2).

[Info] Cutoff = 9.0 A.

[Info]Successfully loaded protein at path 'dataset_path/protein_family_name/1cmc.pdb'

[Info] Coordinate matrix, distance matrix, contact map have been generated for this protein.

[Info] Residue distance to ligand has been calculated.

Enter PDB name for complex state:1cma

Enter residues to be excluded:

Enter spring constant: 1

Enter cutoff: 9

[Info] Spring constant = 1.0 Kcal/(mol A^2).

[Info] Cutoff = 9.0 A.

[Info] Successfully loaded protein at path 'dataset_path/protein_family_name/1cma.pdb'

[Info] Coordinate matrix, distance matrix, contact map have been generated for this protein.

[Info] Residue distance to ligand has been calculated.

This process generates coordinate matrix, distance matrix, and contact map for each state. The water molecules, named 'HOH', are deleted automatically. If there exist other irrelevant ions, or molecules that need not consider, one can input their names into "*Enter residues to be excluded:*", and they will be deleted. Note that if residues are not the basic 20 amino acids, one need to change their residue name to the basic form, for

example, ‘CYX’ to ‘CYS’. One can also modify the source code to include their name into the automatically recognition list.

Then, the hessian matrix and covariance matrix were constructed and written to the same path. The binary file, ‘1cmb.pdb.hessian’ and “1cmb.pdb.covariance”, was saved, as the diagonalization process is rate-limiting, one need not repeat this process and can directly provide these generated binary files and move to the next step. However, the related functions in the source code did not integrate into this compiled program, and one may need to modify the source code to activate it. The RMSD relative to *apo* state was also calculated.

[Info] Finish constructing Hessian matrix.

[Info] Finish constructing Covariance matrix.

[Info] Fitting process succeed.

[Info] Calculating displacement succeed.

[Result] RMSD from allostery state PDB file: 1.6921 A.

[Info] Fitting process succeed.

[Info] Calculating displacement succeed.

[Result] RMSD from complex state PDB file: 1.9768 A.

[Info] Matrix (binary data) has been written to 1cmb.pdb.hessian.

[Info] Matrix (binary data) has been written to 1cmb.pdb.covariance.

In the next section, users will interactively input some commands to perform the force calculations. Related commands and notes are listed below:

pocketA, pocketS, pocketAS: these commands tell the program to enter into the class of A, S or AS to calculate the corresponding properties.

gen-pocket radius: this command generates pocket residues according to the radius that users define, and 4.5 Å is appropriate. The program will judge which residues are in contact (within 4.5 Å) with ligand heavy atoms.

show: list pocket residue ID. The ID of residues are rearranged from 1 according to the order of residues in PDB file.

add residue: artificially add residue ID to pockets as you wish

del residue: artificially delete residue ID from pockets as you wish

gen-force: generate pocket forces in each pocket residues

show-force: show pocket forces in each pocket residues. The force of each residue is stored as a 3D vector ([x y z]).

test: calculate RMSD and Pearson correlations between the fitted and real structure.

back: back to parent directory

energy: calculate free energy of each state (PS, PA, PAS) and predict the allosteric regulation direction. The '*force.txt*' and '*ddG_per_residue.txt*' were generated recording the force and allosteric free energy change of each residue.

In this example, the interactive commands are provided below. First, enter into *pocketA* and *pocketAS* to generate pocketA (and pockeS) residues as well as pocketA (and pocketS) forces using "*gen-pocket 4.5*" and "*gen-force*". After obtaining forces, the *energy* is provided to calculate free energy of each state ("*Free energy for binding state structure*"), which can be separated into the protein part ("*Pro*") and pocket part ("*Pocket*"), and finally the allosteric energy ("*Change of free energy*") will be shown to predict the allosteric regulation type. The "*S_ave*" refers to average work done by *pocketS* forces.

>>> pocketA

[1] >>> gen-pocket 4.5

[1] >>> gen-force

[1] >>> test

[Result] Pocket residues: 39 42 43 56 59 60 61 63 64 65 66 67 70 71 143 146 147 160 163 164 165 167 168 169 171 174 175

[Result] RMSD between real structure and structure calculated according to current pocket: 1.4332 A.

[Result] Pearson between real structure and structure calculated according to current pocket: 0.621188 A.

[1] >>> back

>>> pocketAS

[2] >>> gen-pocket 4.5

[2] >>> gen-force

[2] >>> test

[Result] Pocket residues: 17 21 22 23 24 25 27 40 52 53 54 118 119 120 121 122 126 127 128 129 131 144 156 157 158

[Result] RMSD between real structure and structure calculated according to current pocket: 1.59786 A.

[Result] Pearson between real structure and structure calculated according to current pocket: 0.708366 A.

[2] >>> back

>>> energy

[Info] Matrix has been written to force.txt.

[Info] Add residue 17 to pocket.

[Info] Add residue 21 to pocket.

[Info] Add residue 22 to pocket.

[Info] Add residue 23 to pocket.

[Info] Add residue 24 to pocket.

[Info] Add residue 25 to pocket.

[Info] Add residue 27 to pocket.

[Info] Add residue 40 to pocket.

[Info] Add residue 52 to pocket.

[Info] Add residue 53 to pocket.

[Info] Add residue 54 to pocket.

[Info] Add residue 118 to pocket.

[Info] Add residue 119 to pocket.

[Info] Add residue 120 to pocket.

[Info] Add residue 121 to pocket.

[Info] Add residue 122 to pocket.

[Info] Add residue 126 to pocket.

[Info] Add residue 127 to pocket.

[Info] Add residue 128 to pocket.

[Info] Add residue 129 to pocket.

[Info] Add residue 131 to pocket.

[Info] Add residue 144 to pocket.

[Info] Add residue 156 to pocket.

[Info] Add residue 157 to pocket.

[Info] Add residue 158 to pocket.

[Info] Matrix has been written to ddG_per_residue.txt.

[Result] All predict free energy results:

Free energy for binding state structure S: -203.6520 Kcal/mol.

Pro: 101.3760 Kcal/mol.

Pocket: -305.0280 Kcal/mol.

Free energy for allostery state structure A: -48.3490 Kcal/mol.

Pro: 50.4457 Kcal/mol.

Pocket: -98.7947 Kcal/mol.

Free energy for complex state structure AS: -259.0868 Kcal/mol.

Pro: 131.7048 Kcal/mol.

Pocket: -390.7916 Kcal/mol.

Change of free energy: : -7.0858 Kcal/mol.

S_ave: : -20.6030 Kcal/mol.

A_ave: : -6.1787 Kcal/mol.

AS_ave: : -26.3958 Kcal/mol.

>>>

Note that for systems containing P, PA and PAS, the coordinates of allosteric ligand (A) in PAS were removed as the pocketA information was already obtained in PA. Only the information of pocketS would be obtained from PAS.

The outputs above are stored in "*results.txt*" for better comparison, and a template of input is provided in "*cmd.txt*". Users can directly copy commands in this file and paste into the terminal window.