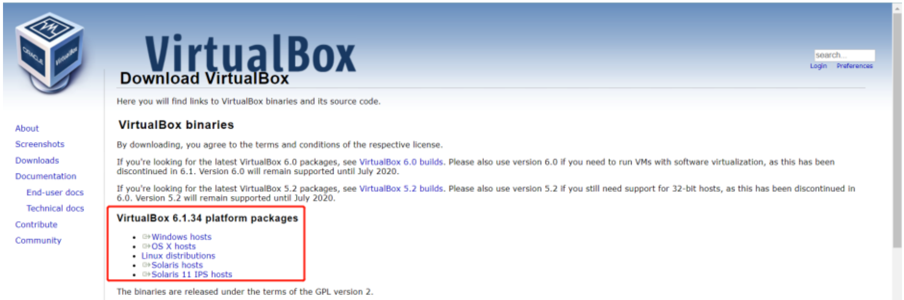
User manual of the *Mimilator*

1. **Install the *Ubuntu Linux system for VirtualBox***
   a. https://www.virtualbox.org/wiki/Downloads , select *Windows host* and the installation package will start to download. Follow the guidance to install *the Oracle VM VirtualBox* on your system



Download VirtualBox

Here you will find links to VirtualBox binaries and its source code.

VirtualBox binaries

By downloading, you agree to the terms and conditions of the respective license.

If you're looking for the latest VirtualBox 6.0 packages, see VirtualBox 6.0 builds. Please also use version 6.0 if you need to run VMs with software virtualization, as this has been discontinued in 6.1. Version 6.0 will remain supported until July 2020.

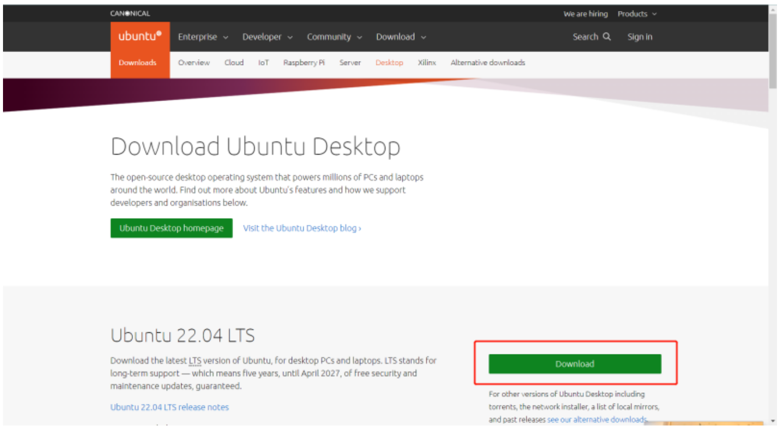
If you're looking for the latest VirtualBox 5.2 packages, see VirtualBox 5.2 builds. Please also use version 5.2 if you still need support for 32-bit hosts, as this has been discontinued in 6.0. Version 5.2 will remain supported until July 2020.

VirtualBox 6.1.34 platform packages

- Windows hosts
- OS X hosts
- Linux distributions
- Solaris hosts
- Solaris 11 IPS hosts

The binaries are released under the terms of the GPL version 2.

   b. https://ubuntu.com/download/desktop , download the Ubuntu iso file by clicking the download button



Download Ubuntu Desktop

The open-source desktop operating system that powers millions of PCs and laptops around the world. Find out more about Ubuntu's features and how we support developers and organisations below.

Ubuntu 22.04 LTS

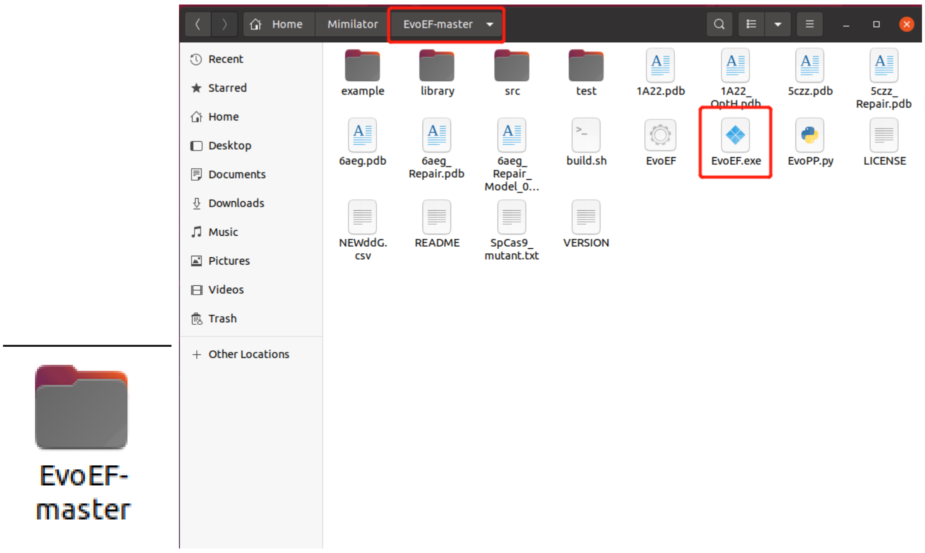
Download the latest LTS version of Ubuntu, for desktop PCs and laptops. LTS stands for long-term support — which means five years, until April 2027, of free security and maintenance updates, guaranteed.

Ubuntu 22.04 LTS release notes

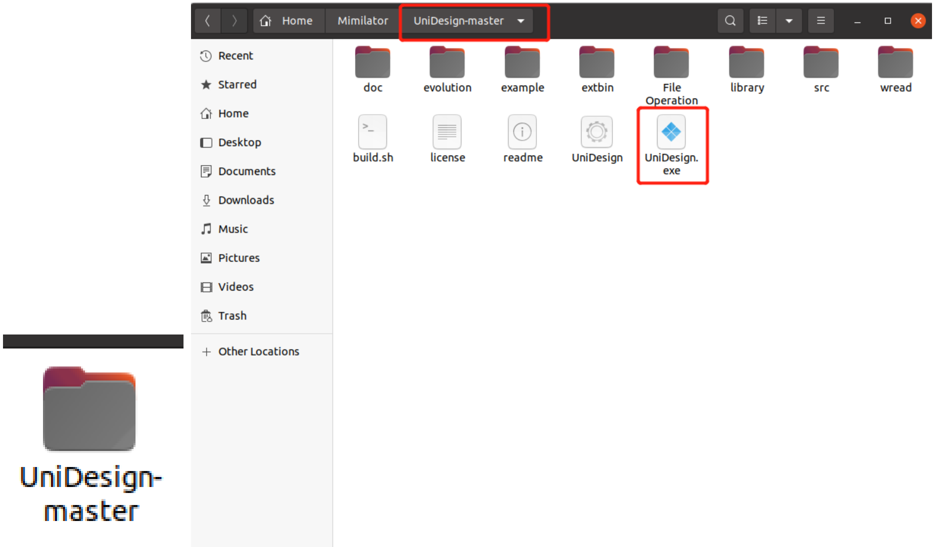
For other versions of Ubuntu Desktop including torrents, the network installer, a list of local mirrors, and past releases see our alternative downloads.

   c. Once the *Ubuntu Linux* is initialized, the Python and C++ modules may not be ready, the following commands should be ran firstly to install the Python and C++ modules:
      i. sudo apt-get update
      ii. sudo apt-get install python
      iii. sudo apt-get install g++
      iv. sudo apt-get install python3-pip
      v. pip install xlwt

2. **Install the *EvoEF and its extension* UniDesign**
   a. https://github.com/tommyhuangthu/EvoEF, download the *EvoEF* kernel from github, and follow **the installation guideline**. If everything is good, a folder named *EvoEF-master* which contains the file *EvoEF.exe* will be generated.

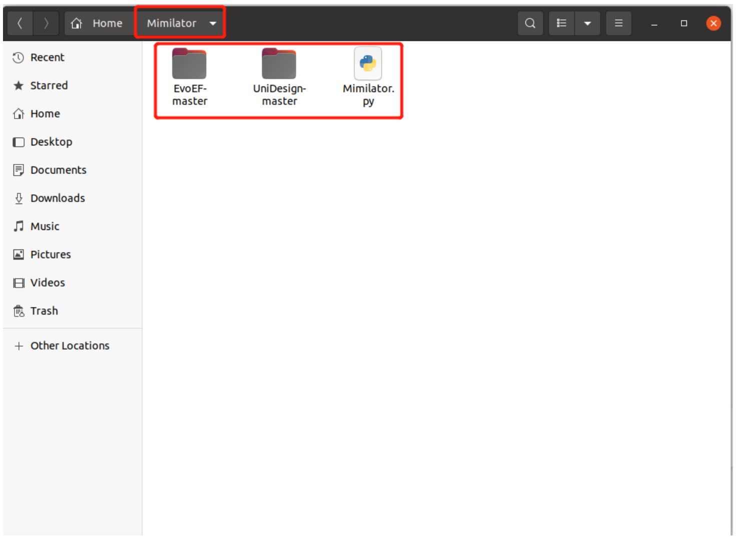


b. https://github.com/tommyhuangthu/UniDesign, download the *UniDesign* extension from github, and **follow the installation guideline**. If everything is good, a folder named *EvoEF-master* which contains the file *EvoEF.exe* will be generated.

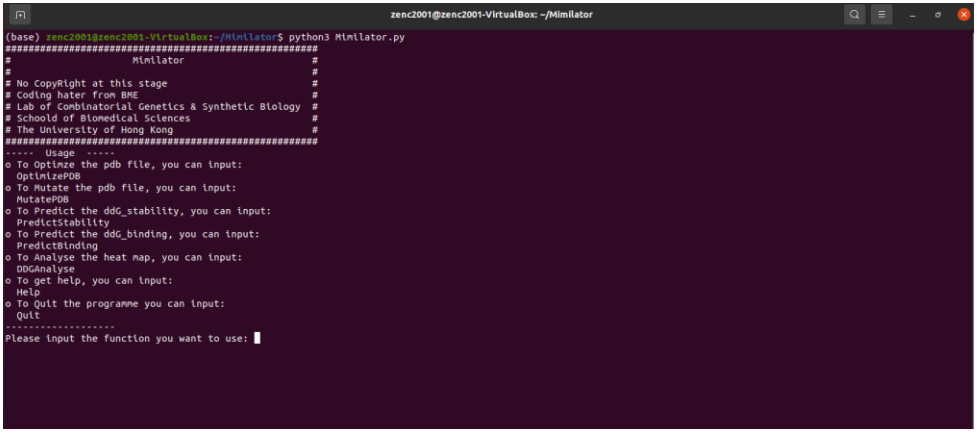


3. <u>Install the pipeline *Mimilator*</u>

a. Create a folder named *Mimilator*, move the *EvoEF-master* and *UniDesign-master* file into this folder, and download the python file *Mimilator.py* from https://github.com/KaiSan2001/BMED4010-FYP/tree/main/EvoEF to this folder



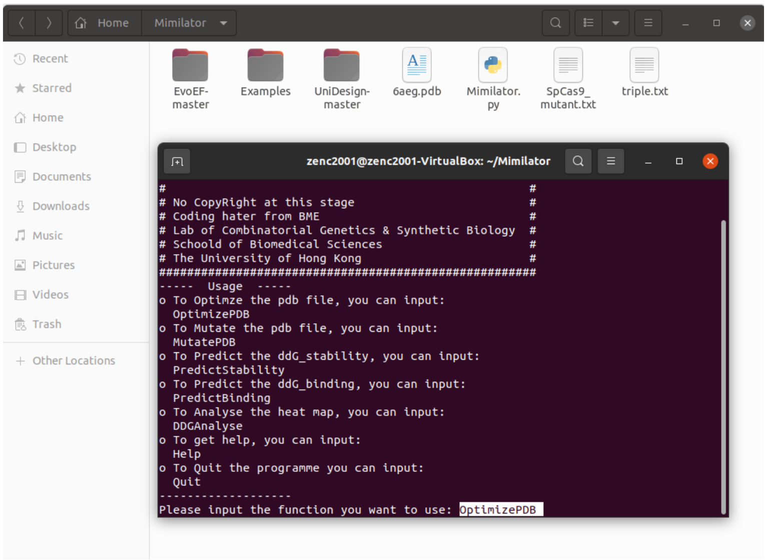
b. Open the terminal (Ctrl + Shit +T or use other approach), activate the ***Mimilator*** by inputting the command: *python3 Mimilator.py*, and the API as below will pop out:



```
(base) zenc2001@zenc2001-VirtualBox:~/Mimilator$ python3 Mimilator.py
######################################################
#                     Mimilator                      #
#                                                    #
# No CopyRight at this stage                         #
# Coding hater from BME                              #
# Lab of Combinatorial Genetics & Synthetic Biology  #
# Schoold of Biomedical Sciences                     #
# The University of Hong Kong                        #
######################################################
-----  Usage  -----
o To Optimze the pdb file, you can input:
  OptimizePDB
o To Mutate the pdb file, you can input:
  MutatePDB
o To Predict the ddG_stability, you can input:
  PredictStability
o To Predict the ddG_binding, you can input:
  PredictBinding
o To Analyse the heat map, you can input:
  DDGAnalyse
o To get help, you can input:
  Help
o To Quit the programme you can input:
  Quit
-------------------
Please input the function you want to use: 
```

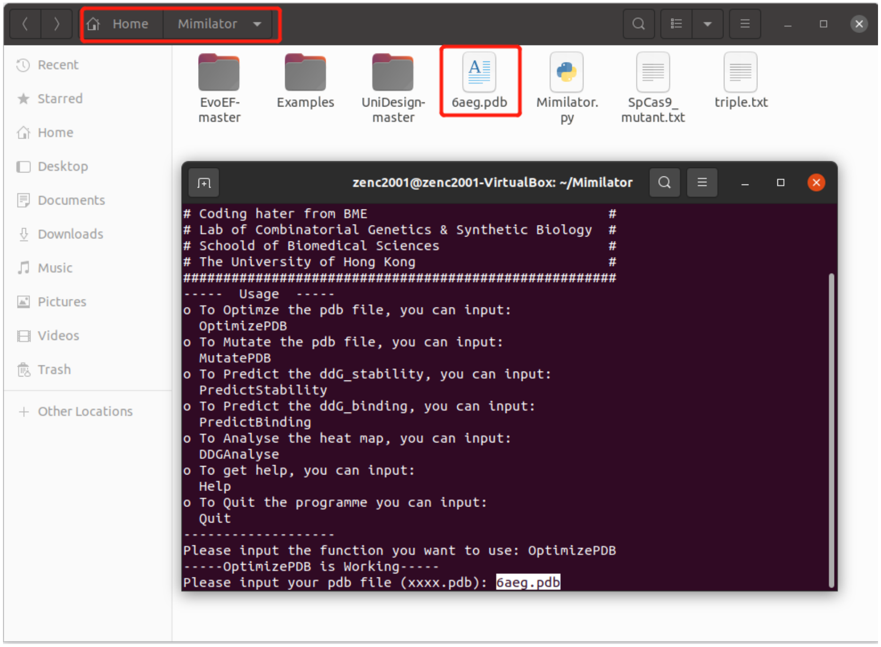
c. There are four major functions in the *Mimilator*. The function OptimizePDB, MutatePDB, and PredictStability, are served for ΔΔG_Stability prediction, they should be combined and used in below order:

   i. **Repair the protein model (PDB file).** The model may contains a lot of steric clashes which do not exist in the natural situation, to predict a more accurate ΔΔG value, optimization is required by input ***OptimizePDB*** to the terminal:



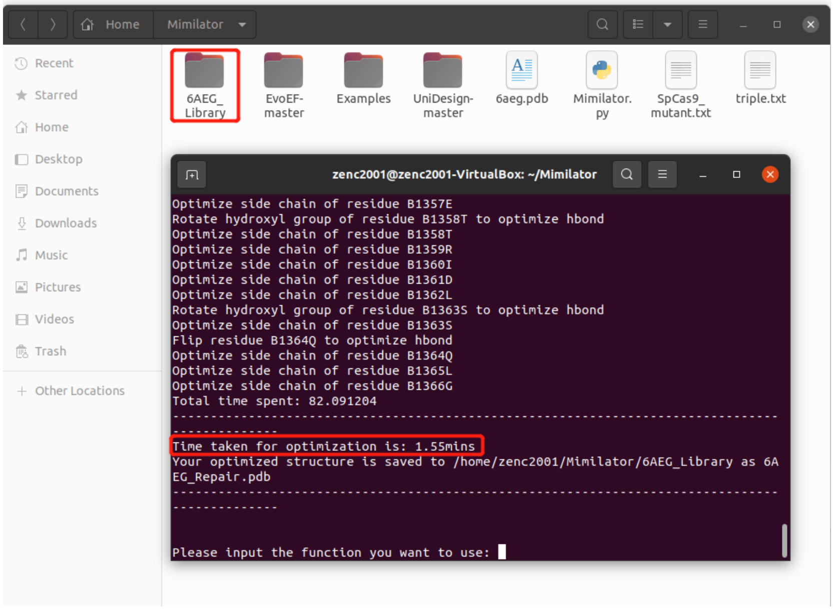
```
#                                                    #
# No CopyRight at this stage                         #
# Coding hater from BME                              #
# Lab of Combinatorial Genetics & Synthetic Biology  #
# Schoold of Biomedical Sciences                     #
# The University of Hong Kong                        #
######################################################
-----  Usage  -----
o To Optimze the pdb file, you can input:
  OptimizePDB
o To Mutate the pdb file, you can input:
  MutatePDB
o To Predict the ddG_stability, you can input:
  PredictStability
o To Predict the ddG_binding, you can input:
  PredictBinding
o To Analyse the heat map, you can input:
  DDGAnalyse
o To get help, you can input:
  Help
o To Quit the programme you can input:
  Quit
------------------
Please input the function you want to use: OptimizePDB
```

After that, a model input is required. Take the SpCas9 model (PDBID: 6aeg) as an example, move the pdb file of SpCas9 to the current working directory (~/Mimilator) and text ***6aeg.pdb*** to the terminal:



```
# Coding hater from BME                              #
# Lab of Combinatorial Genetics & Synthetic Biology  #
# Schoold of Biomedical Sciences                     #
# The University of Hong Kong                        #
######################################################
-----  Usage  -----
o To Optimze the pdb file, you can input:
  OptimizePDB
o To Mutate the pdb file, you can input:
  MutatePDB
o To Predict the ddG_stability, you can input:
  PredictStability
o To Predict the ddG_binding, you can input:
  PredictBinding
o To Analyse the heat map, you can input:
  DDGAnalyse
o To get help, you can input:
  Help
o To Quit the programme you can input:
  Quit
------------------
Please input the function you want to use: OptimizePDB
-----OptimizePDB is Working-----
Please input your pdb file (xxxx.pdb): 6aeg.pdb
```

The optimization will start working. Wait for some minutes, a new fold named PDBID_Library (here is the 6aeg_Library) will form in the current directory, and the corresponding repaired pdb file is stored in it:



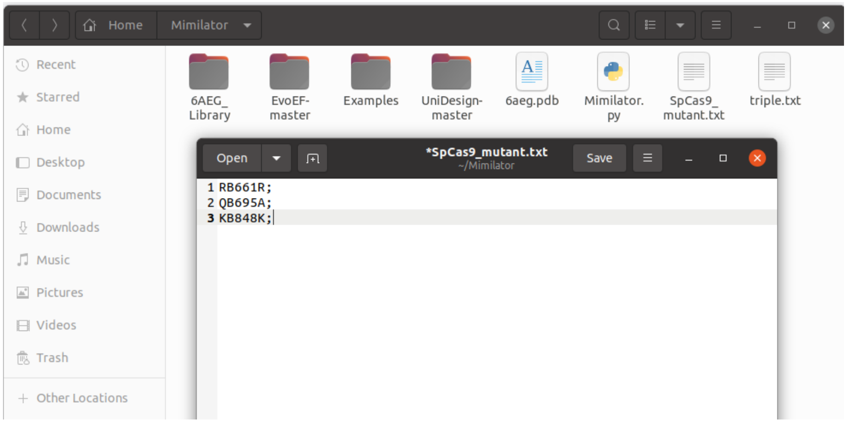
```
Optimize side chain of residue B1357E
Rotate hydroxyl group of residue B1358T to optimize hbond
Optimize side chain of residue B1358T
Optimize side chain of residue B1359R
Optimize side chain of residue B1360I
Optimize side chain of residue B1361D
Optimize side chain of residue B1362L
Rotate hydroxyl group of residue B1363S to optimize hbond
Optimize side chain of residue B1363S
Flip residue B1364Q to optimize hbond
Optimize side chain of residue B1364Q
Optimize side chain of residue B1365L
Optimize side chain of residue B1366G
Total time spent: 82.091204
--------------------------------------------------------------------------------
--------------
Time taken for optimization is: 1.55mins
Your optimized structure is saved to /home/zenc2001/Mimilator/6AEG_Library as 6A
EG_Repair.pdb
--------------------------------------------------------------------------------
--------------


Please input the function you want to use: 
```

The PDBID_Library folder will be used to store the downstream mutant files and the ΔΔG_Stability values

ii. **Mutate the protein model**. Firstly, a mutant file (xxxx_mutant.txt) is required for processing the mutation, the mutant file should be strictly formatted as following:

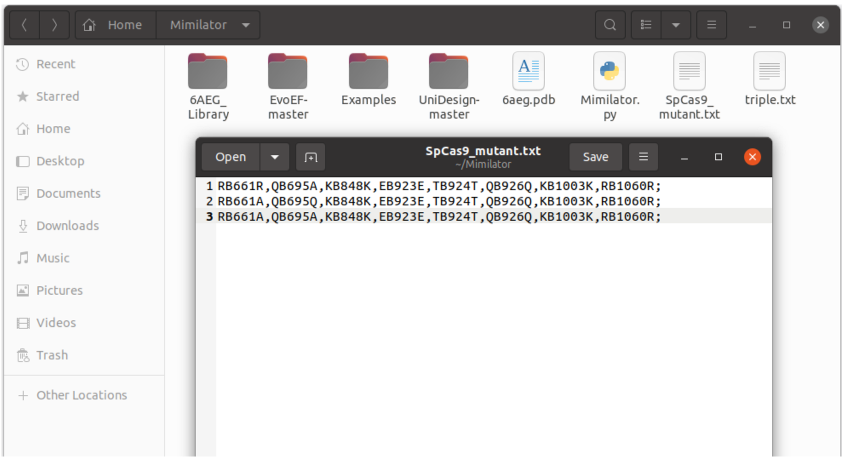
1. Each mutation is written in one line ending with ";", and multiple mutants are divided by ",". Note that there's no gap/space between single mutations. For each single mutation, the first alphabet is the wild-type amino acid, the second is the identifier of the chain that the amino acid is attached to, the number is the position of the amino

   acid in the chain, and the last alphabet is the amino acid after mutation.
2. If there is only one mutation in each round, the mutant file can be written as:



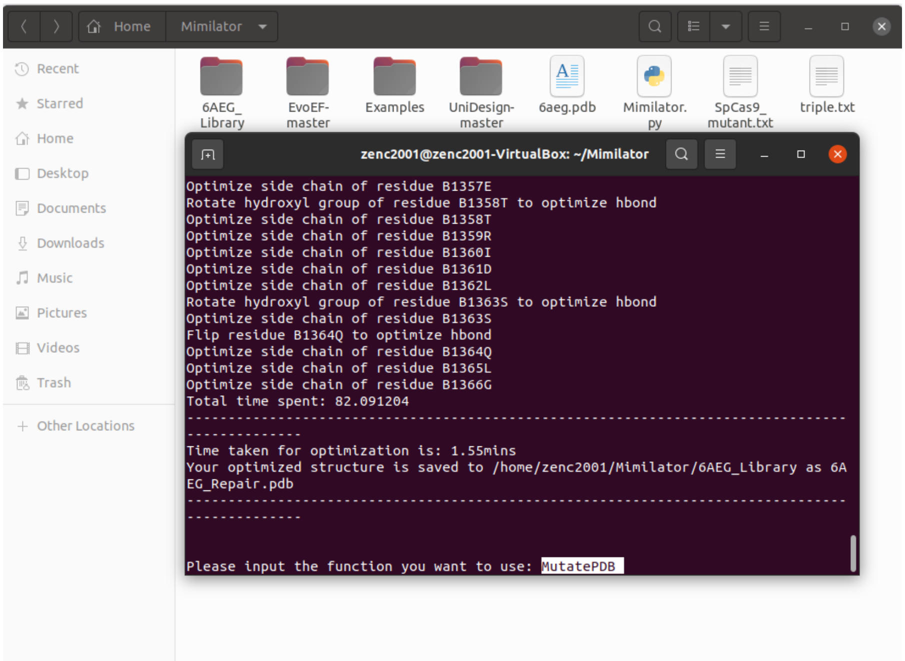
```
RB661R;
QB695A;
KB848K;
```

3. If there are multiple mutation in each round, the mutant file can be written as:



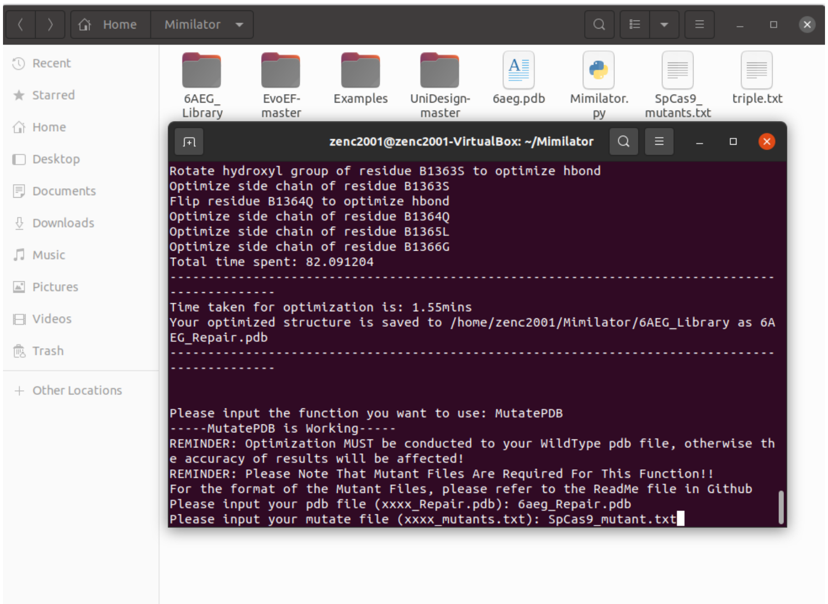
```
RB661R,QB695A,KB848K,EB923E,TB924T,QB926Q,KB1003K,RB1060R;
RB661A,QB695Q,KB848K,EB923E,TB924T,QB926Q,KB1003K,RB1060R;
RB661A,QB695A,KB848K,EB923E,TB924T,QB926Q,KB1003K,RB1060R;
```

After preparing the mutant file, move it to the current working directory (~/Mimilator), in the *Mimilator* API input ***MutatePDB***:



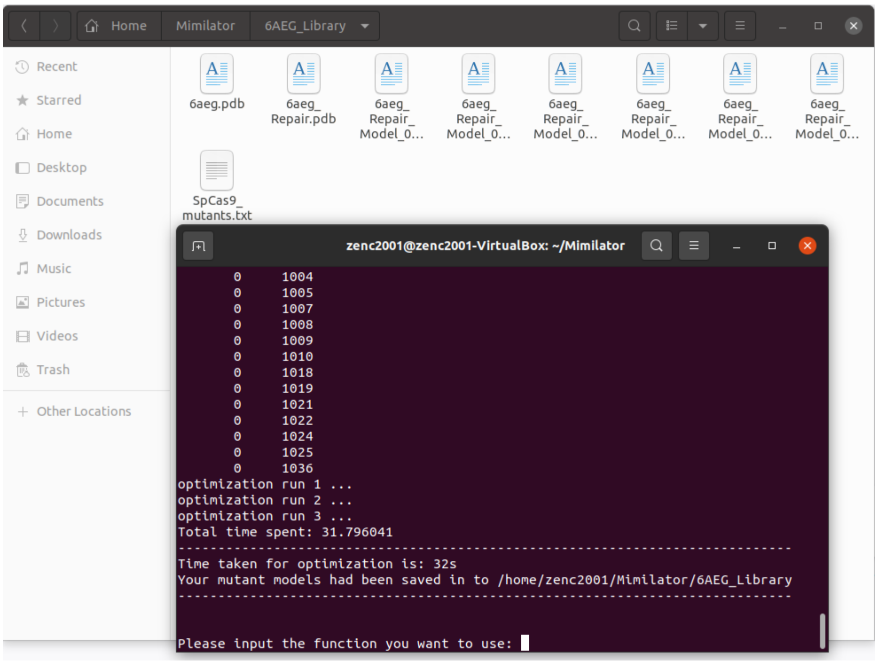
```
Optimize side chain of residue B1357E
Rotate hydroxyl group of residue B1358T to optimize hbond
Optimize side chain of residue B1358T
Optimize side chain of residue B1359R
Optimize side chain of residue B1360I
Optimize side chain of residue B1361D
Optimize side chain of residue B1362L
Rotate hydroxyl group of residue B1363S to optimize hbond
Optimize side chain of residue B1363S
Flip residue B1364Q to optimize hbond
Optimize side chain of residue B1364Q
Optimize side chain of residue B1365L
Optimize side chain of residue B1366G
Total time spent: 82.091204
--------------------------------------------------------------------------------
-------------
Time taken for optimization is: 1.55mins
Your optimized structure is saved to /home/zenc2001/Mimilator/6AEG_Library as 6A
EG_Repair.pdb
--------------------------------------------------------------------------------
-------------


Please input the function you want to use: MutatePDB
```

Input the optimized/repaired protein model (PDBID_Repaired.pdb) and the mutant file (xxxx_mutants.txt) to do the mutation:



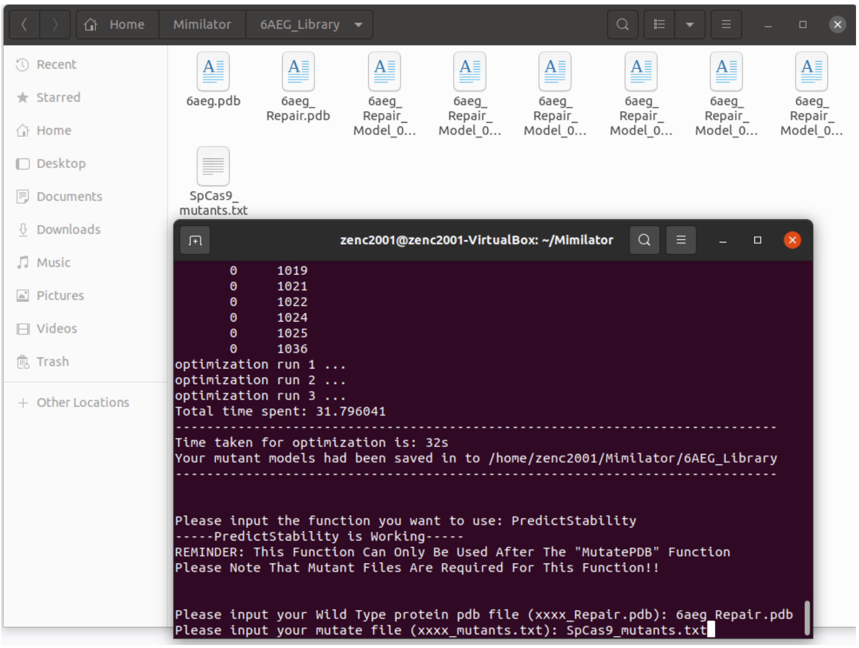
```
Rotate hydroxyl group of residue B1363S to optimize hbond
Optimize side chain of residue B1363S
Flip residue B1364Q to optimize hbond
Optimize side chain of residue B1364Q
Optimize side chain of residue B1365L
Optimize side chain of residue B1366G
Total time spent: 82.091204
--------------------------------------------------------------------------------
--------------
Time taken for optimization is: 1.55mins
Your optimized structure is saved to /home/zenc2001/Mimilator/6AEG_Library as 6A
EG_Repair.pdb
--------------------------------------------------------------------------------
--------------


Please input the function you want to use: MutatePDB
-----MutatePDB is Working-----
REMINDER: Optimization MUST be conducted to your WildType pdb file, otherwise th
e accuracy of results will be affected!
REMINDER: Please Note That Mutant Files Are Required For This Function!!
For the format of the Mutant Files, please refer to the ReadMe file in Github
Please input your pdb file (xxxx_Repair.pdb): 6aeg_Repair.pdb
Please input your mutate file (xxxx_mutants.txt): SpCas9_mutant.txt
```

The pipeline will perform the *in-silico* engineering simulation to mutate the PDB file, for each round of mutation, a corresponding mutant file will be generated and stored in the PDBID_Library folder:



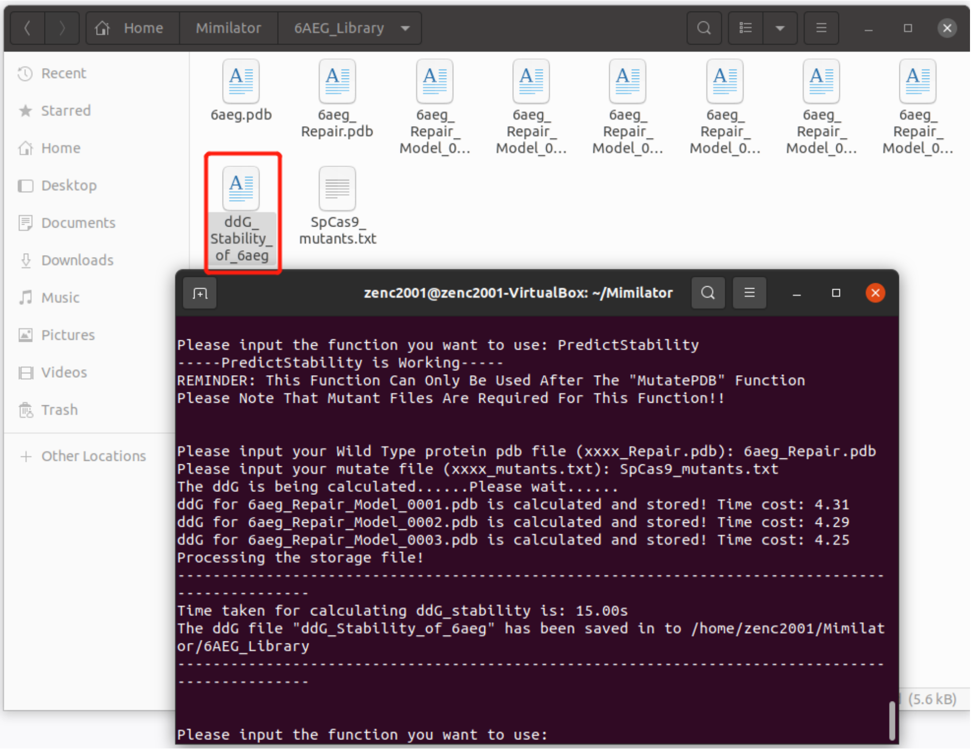
```
       0     1004
       0     1005
       0     1007
       0     1008
       0     1009
       0     1010
       0     1018
       0     1019
       0     1021
       0     1022
       0     1024
       0     1025
       0     1036
optimization run 1 ...
optimization run 2 ...
optimization run 3 ...
Total time spent: 31.796041
-----------------------------------------------------------------------------
Time taken for optimization is: 32s
Your mutant models had been saved in to /home/zenc2001/Mimilator/6AEG_Library
-----------------------------------------------------------------------------


Please input the function you want to use:
```

iii. **Calculate the ΔΔG_Stability.** In the *Mimilator* API, input the function ***PredictStability***, the pipeline will require a repaired protein model and mutant file. Text the name of repaired protein model and mutant file as previous section, and the pipeline will start to calculate the ΔΔG_Stability for each mutant model:



```
        0      1019
        0      1021
        0      1022
        0      1024
        0      1025
        0      1036
optimization run 1 ...
optimization run 2 ...
optimization run 3 ...
Total time spent: 31.796041
-----------------------------------------------------------------------------
Time taken for optimization is: 32s
Your mutant models had been saved in to /home/zenc2001/Mimilator/6AEG_Library
-----------------------------------------------------------------------------


Please input the function you want to use: PredictStability
-----PredictStability is Working-----
REMINDER: This Function Can Only Be Used After The "MutatePDB" Function
Please Note That Mutant Files Are Required For This Function!!


Please input your Wild Type protein pdb file (xxxx_Repair.pdb): 6aeg_Repair.pdb
Please input your mutate file (xxxx_mutants.txt): SpCas9_mutants.txt
```

After a few minutes of operation, a file contains the ΔΔG_Stability for each mutant model will be generated and stored in the folder PDBID_Library in a name of *ddG_Stability_of_PDBID:*

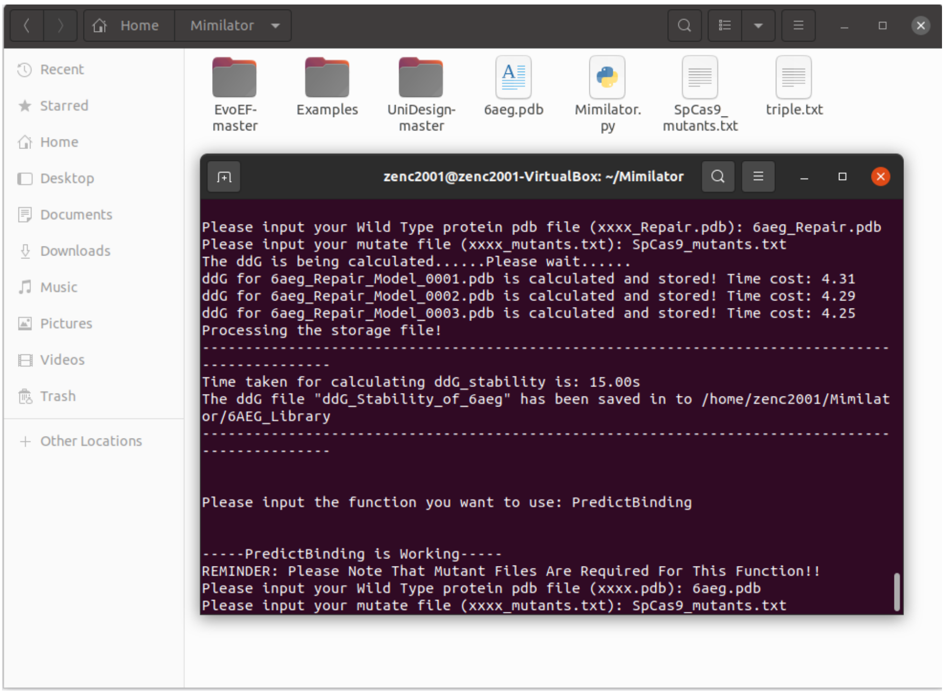


```
Please input the function you want to use: PredictStability
-----PredictStability is Working-----
REMINDER: This Function Can Only Be Used After The "MutatePDB" Function
Please Note That Mutant Files Are Required For This Function!!


Please input your Wild Type protein pdb file (xxxx_Repair.pdb): 6aeg_Repair.pdb
Please input your mutate file (xxxx_mutants.txt): SpCas9_mutants.txt
The ddG is being calculated......Please wait......
ddG for 6aeg_Repair_Model_0001.pdb is calculated and stored! Time cost: 4.31
ddG for 6aeg_Repair_Model_0002.pdb is calculated and stored! Time cost: 4.29
ddG for 6aeg_Repair_Model_0003.pdb is calculated and stored! Time cost: 4.25
Processing the storage file!
-----------------------------------------------------------------------------
--------------
Time taken for calculating ddG_stability is: 15.00s
The ddG file "ddG_Stability_of_6aeg" has been saved in to /home/zenc2001/Mimilat
or/6AEG_Library
-----------------------------------------------------------------------------
--------------


Please input the function you want to use:
```

Note that there is no format for this file, when you want to move it to Windows/Mac system for downstream analysis, please add a suffix *.csv*. Otherwise, the Windows/Mac system cannot open it.

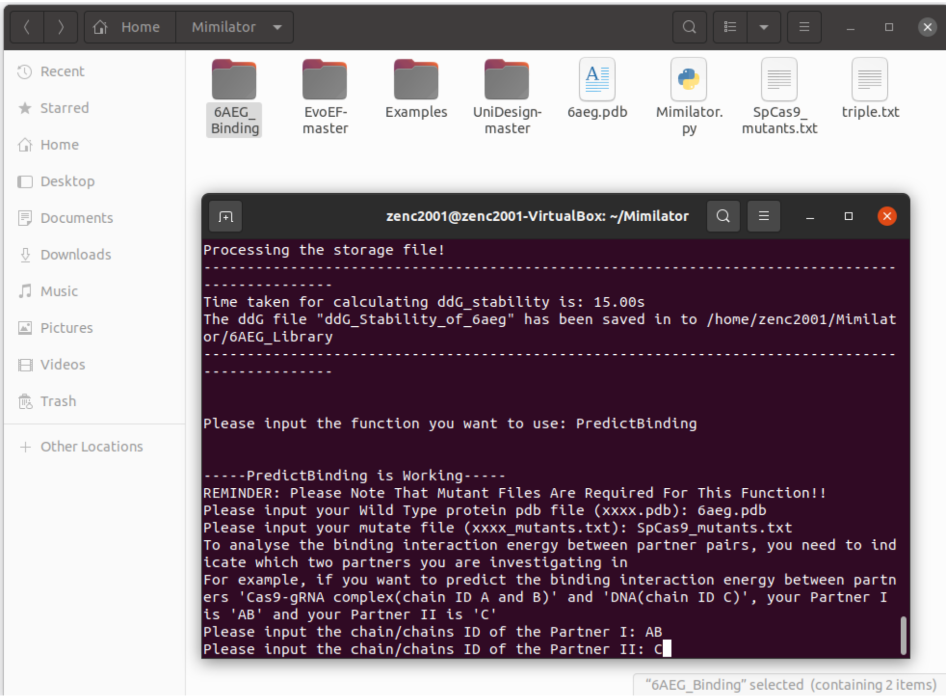
d. The function *PredictBinding* serves for ΔΔG_Binding calculation, it can be used as following:

   i. Input the ***PredictBinding*** command to the API, it will require the protein model and mutant file of the target protein (PDBID.pdb). Note that the mutant file here follows the same format as previous section:



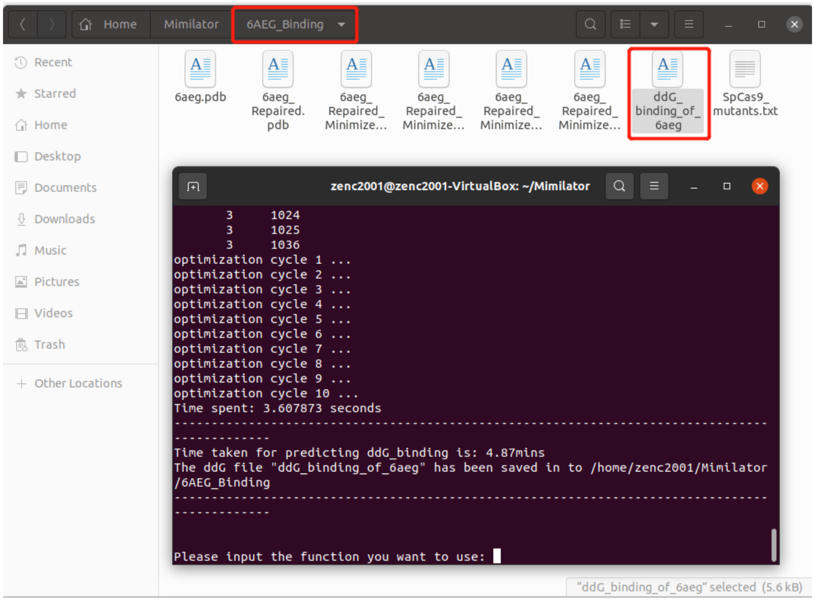
```
Please input your Wild Type protein pdb file (xxxx_Repair.pdb): 6aeg_Repair.pdb
Please input your mutate file (xxxx_mutants.txt): SpCas9_mutants.txt
The ddG is being calculated......Please wait......
ddG for 6aeg_Repair_Model_0001.pdb is calculated and stored! Time cost: 4.31
ddG for 6aeg_Repair_Model_0002.pdb is calculated and stored! Time cost: 4.29
ddG for 6aeg_Repair_Model_0003.pdb is calculated and stored! Time cost: 4.25
Processing the storage file!
--------------------------------------------------------------------------------
--------------
Time taken for calculating ddG_stability is: 15.00s
The ddG file "ddG_Stability_of_6aeg" has been saved in to /home/zenc2001/Mimilat
or/6AEG_Library
--------------------------------------------------------------------------------
--------------


Please input the function you want to use: PredictBinding


-----PredictBinding is Working-----
REMINDER: Please Note That Mutant Files Are Required For This Function!!
Please input your Wild Type protein pdb file (xxxx.pdb): 6aeg.pdb
Please input your mutate file (xxxx_mutants.txt): SpCas9_mutants.txt
```

ii. Similarly, a library will be formed in a name of *PDBID_Binding*, and the downstream mutant files will be stored in it. After that, the binding partners should be indicated in the API respectively. More details can be refered to the figure below:



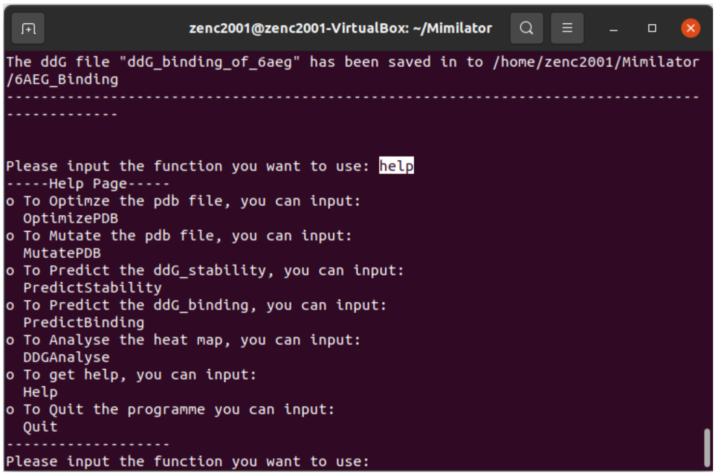
```
Processing the storage file!
--------------------------------------------------------------------------------
--------------
Time taken for calculating ddG_stability is: 15.00s
The ddG file "ddG_Stability_of_6aeg" has been saved in to /home/zenc2001/Mimilat
or/6AEG_Library
--------------------------------------------------------------------------------
--------------


Please input the function you want to use: PredictBinding


-----PredictBinding is Working-----
REMINDER: Please Note That Mutant Files Are Required For This Function!!
Please input your Wild Type protein pdb file (xxxx.pdb): 6aeg.pdb
Please input your mutate file (xxxx_mutants.txt): SpCas9_mutants.txt
To analyse the binding interaction energy between partner pairs, you need to ind
icate which two partners you are investigating in
For example, if you want to predict the binding interaction energy between partn
ers 'Cas9-gRNA complex(chain ID A and B)' and 'DNA(chain ID C)', your Partner I
is 'AB' and your Partner II is 'C'
Please input the chain/chains ID of the Partner I: AB
Please input the chain/chains ID of the Partner II: C
```

iii. The pipeline will automatically construct the mutant models and calculate the corresponding ΔΔG_Binding. The mutant models and the ΔΔG_Binding records will be stored at the PDBID_Binding library. Similar to the previous situation, the ΔΔG_Binding records file has no format, a suffix of ***.csv*** is required for its operation in other systems:


Home
Mimilator
6AEG_Binding
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Other Locations
6aeg.pdb
6aeg_ Repaired. pdb
6aeg_ Repaired_ Minimize...
6aeg_ Repaired_ Minimize...
6aeg_ Repaired_ Minimize...
6aeg_ Repaired_ Minimize...
ddG_ binding_of_ 6aeg
SpCas9_ mutants.txt
zenc2001@zenc2001-VirtualBox: ~/Mimilator
"ddG_binding_of_6aeg" selected (5.6 kB)

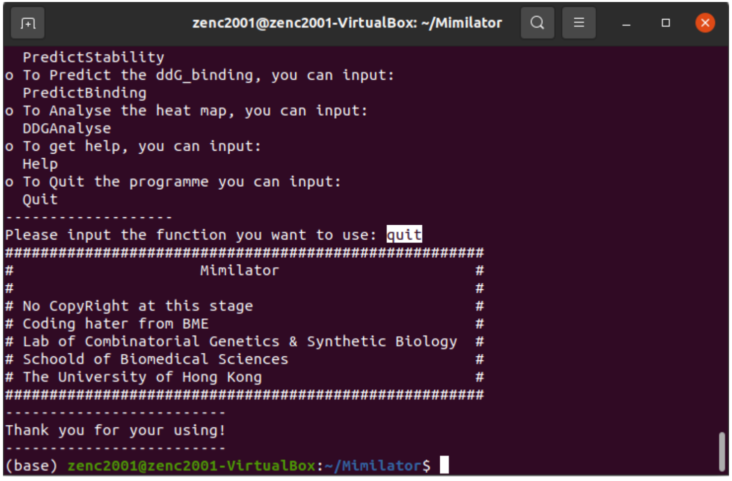
```
       3     1024
       3     1025
       3     1036
optimization cycle 1 ...
optimization cycle 2 ...
optimization cycle 3 ...
optimization cycle 4 ...
optimization cycle 5 ...
optimization cycle 6 ...
optimization cycle 7 ...
optimization cycle 8 ...
optimization cycle 9 ...
optimization cycle 10 ...
Time spent: 3.607873 seconds
--------------------------------------------------------------------------------
------------
Time taken for predicting ddG_binding is: 4.87mins
The ddG file "ddG_binding_of_6aeg" has been saved in to /home/zenc2001/Mimilator
/6AEG_Binding
--------------------------------------------------------------------------------
------------


Please input the function you want to use: 
```

e. If you forget the command of the functions, you can simply type ***Help*** or ***help*** in the API to recall them:



```
The ddG file "ddG_binding_of_6aeg" has been saved in to /home/zenc2001/Mimilator
/6AEG_Binding
--------------------------------------------------------------------------------
------------


Please input the function you want to use: help
-----Help Page-----
o To Optimze the pdb file, you can input:
  OptimizePDB
o To Mutate the pdb file, you can input:
  MutatePDB
o To Predict the ddG_stability, you can input:
  PredictStability
o To Predict the ddG_binding, you can input:
  PredictBinding
o To Analyse the heat map, you can input:
  DDGAnalyse
o To get help, you can input:
  Help
o To Quit the programme you can input:
  Quit
-------------------
Please input the function you want to use:
```

f. If you want to quit the pipeline, you can input ***Quit*** or ***quit***:



```
  PredictStability
o To Predict the ddG_binding, you can input:
  PredictBinding
o To Analyse the heat map, you can input:
  DDGAnalyse
o To get help, you can input:
  Help
o To Quit the programme you can input:
  Quit
-------------------
Please input the function you want to use: quit
######################################################
#                     Mimilator                      #
#                                                    #
# No CopyRight at this stage                         #
# Coding hater from BME                              #
# Lab of Combinatorial Genetics & Synthetic Biology  #
# Schoold of Biomedical Sciences                     #
# The University of Hong Kong                        #
######################################################
-------------------------
Thank you for your using!
-------------------------
(base) zenc2001@zenc2001-VirtualBox:~/Mimilator$ 
```