BAPPL+ Web Server Manual

BAPPL+

(Binding Affinity Prediction of Protein-Ligand)

[1]Initially, BAPPL was developed at the Supercomputing Facility for Bioinformatics and Computational Biology (SCFBio) IIT Delhi for the binding affinity prediction of non-metallo protein-ligand complexes.

[2]The new version introduced from the SCFBO lab, named BAPPL-Z, can predict the binding affinity of non-metallo protein-ligand complexes and zinc metallo protein-ligand complexes.

[3]In 2020, the latest version named BAPPL+ was introduced from the SCFBio Lab to predict the binding affinity of non-metallo and metallo protein-ligand complexes using the machine learning method.

1. Jain T, Jayaram B (2005). **An all atom energy based computational protocol for predicting binding affinities of protein-ligand complexes**. FEBS Lett. 5;579(29):6659-66. doi: 10.1016/j.febslet.2005.10.031.
2. Jain T, Jayaram B (2007). **Computational protocol for predicting the binding affinities of zinc containing metalloprotein-ligand complexes**. Proteins. 67(4):1167-78. doi: 10.1002/prot.21332.
3. Soni A, Bhat R, Jayaram B (**2020**). **Improving the binding affinity estimations of protein-ligand complexes using machine-learning facilitated force field method**. J Comput Aided Mol Des. 34(8):817-830. doi: 10.1007/s10822-020-00305-1.

Chapter 1 Introduction

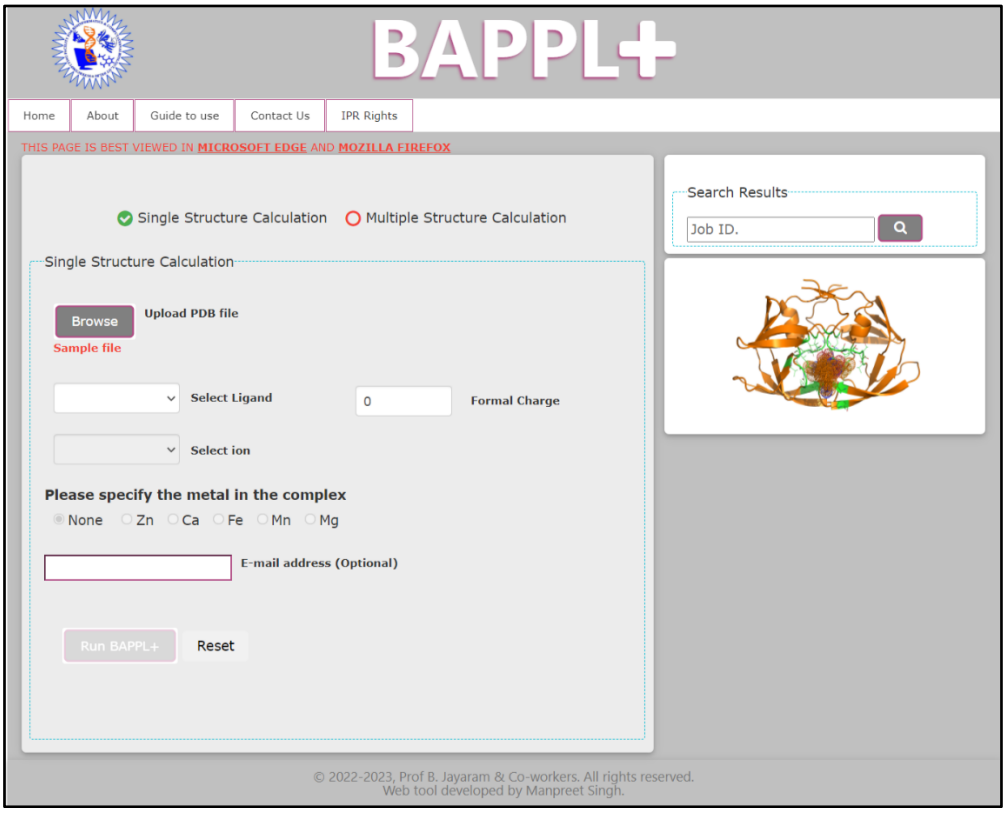
BAPPL+ (Binding Affinity Prediction of Protein-Ligand complex) is a computationally fast protocol for predicting the binding affinity of non-metallo and metallo protein-ligand complexes. The BAPPL+ methodology calculates the binding affinities of non-metallo and Metallo protein-ligand complexes based on electrostatic (ΔE_{ele}), van der Waals (ΔE_{vdw}), hydrophobic (ΔE_{hyp}), and entropic ($-T\Delta S_{conf}$) contributions. Explicit quantum mechanical calculations are adopted to handle commonly occurring metal ions in the binding pockets of proteins. BAPPL+ has parameters for five metal ions named **Zn^{2+}, Mg^{2+}, Ca^{2+}, Mn^{2+} and Fe^{3+}.**

$$\Delta G^{\circ}_{pred} = \Delta E_{ele} + \Delta E_{vdw} + \Delta E_{hyp} - T\Delta S_{conf}$$

ΔE_{ele} is the electrostatic contribution to the energy, ΔE_{vdw} is the van der Waal term, ΔE_{hyp} is the hydrophobic contribution, and $T\Delta S_{conf}$ is the conformational entropy loss.

We have evaluated the performance of BAPPL+ on a large dataset and found that it outperforms other state-of-the-art scoring functions, achieving a high Pearson Correlation Coefficient of up to ~ 0.76 with low standard deviations. The increased performance results from the machine-learning model and the enlarged training dataset. Here we have used the **"Random Forest (RF)"** machine learning method as a regression model to calculate the binding affinity of protein-ligand complexes.

Availability: http://scfbio-iitd.res.in/bappl+/



Chapter 2 Format Specification

BAPPL+ takes input in the standard PDB format containing **hydrogen added ligand.** The term PDB stands for Protein Data Bank. The PDB formatted file must have a .pdb extension. The formal charge of the ligand will be calculated automatically by the server. However, the user can also enter the value of formal charge using the '**Formal charge**' box for accurate binding affinity prediction. The reference ligand identifier in the complex must be in three-letter codes. Reference ligands are nonstandard residues that bind non-covalently to a macromolecule and are represented as heteroatoms in the PDB file. In the case of metal complexes, please ensure that your input file contains metal ions in addition to protein and ligand and specify the metal ions in the PDB file as per the following formats-

1. Non-metal protein ligand complex format

```
.....
ATOM   5993  O   LEU   390      12.186   6.909  37.018  1.00  0.00           O
ATOM   5994  OXT LEU   390      13.732   6.577  38.568  1.00  0.00           O
TER    5995      LEU   390
HETATM 5995  C   DRG   391      -1.720   0.696  -1.775  1.00  0.00           C
HETATM 5996  H   DRG   391      -2.546   0.335  -2.394  1.00  0.00           H
HETATM 5997  H1  DRG   391      -1.996   0.526  -0.733  1.00  0.00           H
.....
```

2. Metallo-protein ligand complex format (with Zn metal)

```
.....
ATOM   5993  O   LEU   390      12.186   6.909  37.018  1.00  0.00           O
ATOM   5994  OXT LEU   390      13.732   6.577  38.568  1.00  0.00           O
TER
HETATM 3768 ZN   ZN2   488      64.301  50.558  15.004  1.00  9.79          ZN
TER
HETATM 5995  C   DRG   391      -1.720   0.696  -1.775  1.00  0.00           C
HETATM 5996  H   DRG   391      -2.546   0.335  -2.394  1.00  0.00           H
HETATM 5997  H1  DRG   391      -1.996   0.526  -0.733  1.00  0.00           H
.....
```

3. Metallo-protein ligand complex format (with Ca metal)

```
.....
ATOM   5993  O   LEU   390      12.186   6.909  37.018  1.00  0.00           O
ATOM   5994  OXT LEU   390      13.732   6.577  38.568  1.00  0.00           O
TER
HETATM 3768  CA  CA2   488      64.301  50.558  15.004  1.00  9.79          CA
TER
HETATM 5995  C   DRG   391      -1.720   0.696  -1.775  1.00  0.00           C
HETATM 5996  H   DRG   391      -2.546   0.335  -2.394  1.00  0.00           H
HETATM 5997  H1  DRG   391      -1.996   0.526  -0.733  1.00  0.00           H
.....
```

4. Metallo-protein ligand complex format (with Mg metal)

```
.....
ATOM   5993  O   LEU   390      12.186   6.909  37.018  1.00  0.00           O
ATOM   5994  OXT LEU   390      13.732   6.577  38.568  1.00  0.00           O
TER
HETATM 3768  MG  MG2   488      64.301  50.558  15.004  1.00  9.79          MG
TER
HETATM 5995  C   DRG   391      -1.720   0.696  -1.775  1.00  0.00           C
HETATM 5996  H   DRG   391      -2.546   0.335  -2.394  1.00  0.00           H
HETATM 5997  H1  DRG   391      -1.996   0.526  -0.733  1.00  0.00           H
.....
```

5. **Metallo-protein ligand complex format (with Mn metal)**

```
.....
ATOM   5993  O   LEU   390      12.186   6.909  37.018  1.00  0.00           O
ATOM   5994  OXT LEU   390      13.732   6.577  38.568  1.00  0.00           O
TER
HETATM 3768  MN  MN2   488      64.301  50.558  15.004  1.00  9.79          MN
TER
HETATM 5995  C   DRG   391      -1.720   0.696  -1.775  1.00  0.00           C
HETATM 5996  H   DRG   391      -2.546   0.335  -2.394  1.00  0.00           H
HETATM 5997  H1  DRG   391      -1.996   0.526  -0.733  1.00  0.00           H
.....
```

6. **Metallo-protein ligand complex format (with Fe metal)**

```
.....
ATOM   5993  O   LEU   390      12.186   6.909  37.018  1.00  0.00           O
ATOM   5994  OXT LEU   390      13.732   6.577  38.568  1.00  0.00           O
TER
HETATM 3768  FE  FE2   488      64.301  50.558  15.004  1.00  9.79          FE
TER
HETATM 5995  C   DRG   391      -1.720   0.696  -1.775  1.00  0.00           C
HETATM 5996  H   DRG   391      -2.546   0.335  -2.394  1.00  0.00           H
HETATM 5997  H1  DRG   391      -1.996   0.526  -0.733  1.00  0.00           H
.....
```

While submitting a job for multiple structures, submit a compressed file (.zip format) containing all the PDB files to be executed. Additionally, the user needs to provide a text file (.txt) (as shown below) containing the name of the complex, a formal charge of ligand, metal, or non-metal information (for non-metal XX and metal could be Zn, Ca, Mg, Mn, Fe) and ligand identifier. The text file contains space-separated information of protein name(s) (without extension .pdb), then the charge of ligand(s) followed by metal information, and finally ligand identifier as shown in template.txt below.

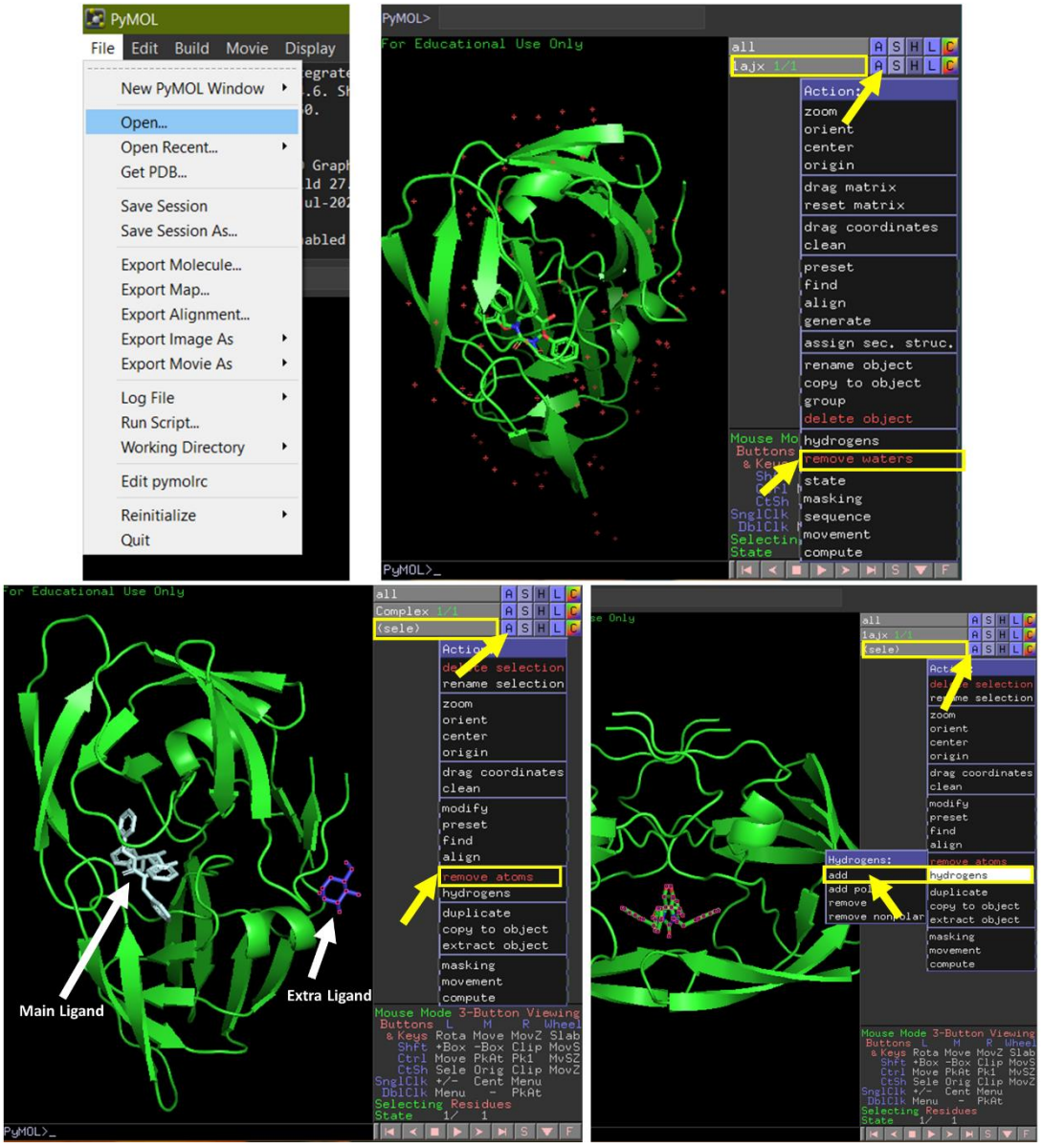
template.txt:

sample 0 XX DRG
sample_zn 0 Zn DRG
sample_ca 0 Ca PEG
sample_mg +1 Mg BIF
sample_mn +2 Mn DRG
sample_fe 0 Fe DRG

Chapter 3 File Preparation

It is advisable to clean the input file before submitting it to the BAPPL+. There are several ways to clean a PDB file. One is given below using PyMol (a molecular visualization tool)-

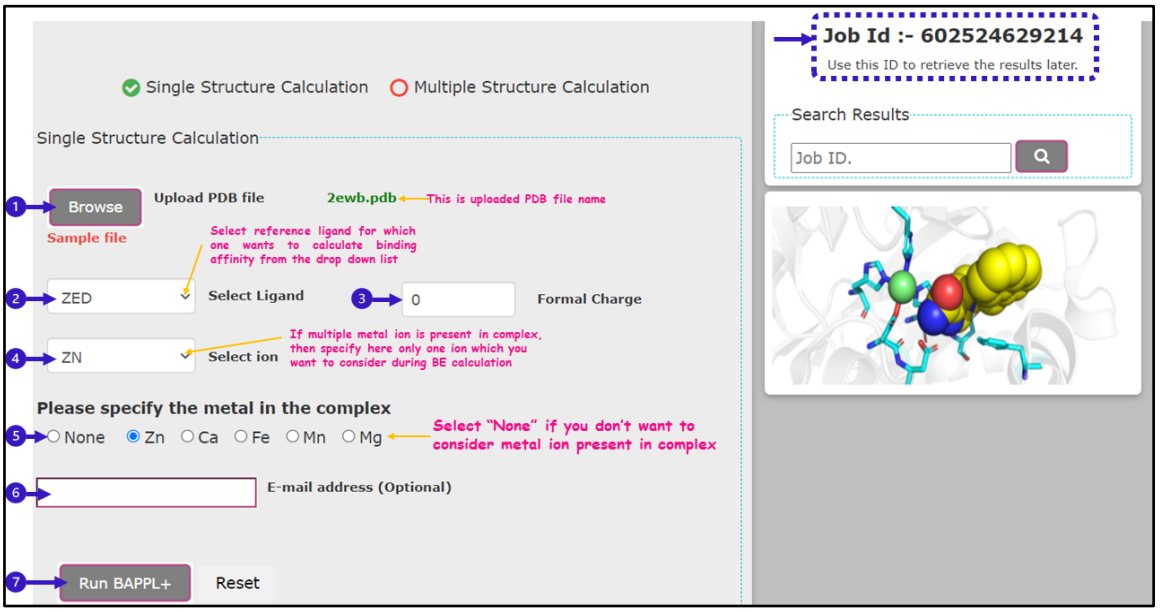
- Open PyMol and click on File tab.
- Select Open.
- Select the path and type the name of your protein file.
- **To remove water molecule from complex**, click on menu Action (A) located at the right corner and select remove waters.
- **To remove extra heteroatoms/ligand from complex**, select the heteroatoms by using left-click button of mouse. Now click on the menu Action (A) of section object and select remove atoms.
- **To add hydrogen molecule in ligand**, click on menu Action (A), select hydrogens, and click on add.



Chapter 4 Job Submission

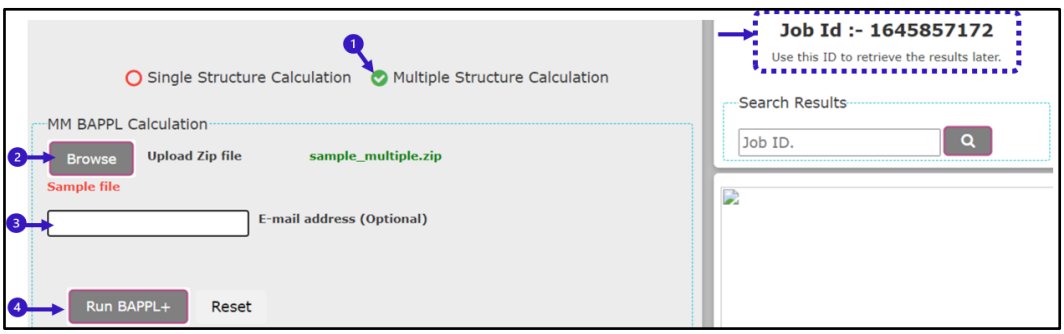
Job submission of single structure run:

First, upload the complex using the browse button in front of **'Upload PDB file',** then select one reference ligand present in the complex from the drop-down menu. Verify the formal charge from the **'Formal charg'** box. BAPPL+ will detect the metal ion present in the complex file and show it on the 'Select Ion' box. Sometimes, multiple metal ions may be present in a complex, in that case, choose a metal ion you want to consider during binding affinity (BA) prediction from the drop-down menu of the 'Select Ion' box. If one wishes to discard the metal ion during the binding affinity prediction, select **'None'**. Now specify email id (optional) to get job completion notification and click on the **'Run BAPPL+'** button to submit your work on the server. Users can also check the job status and result using the Job ID provided for each submitted job.



Job submission of Multiple Structure Run

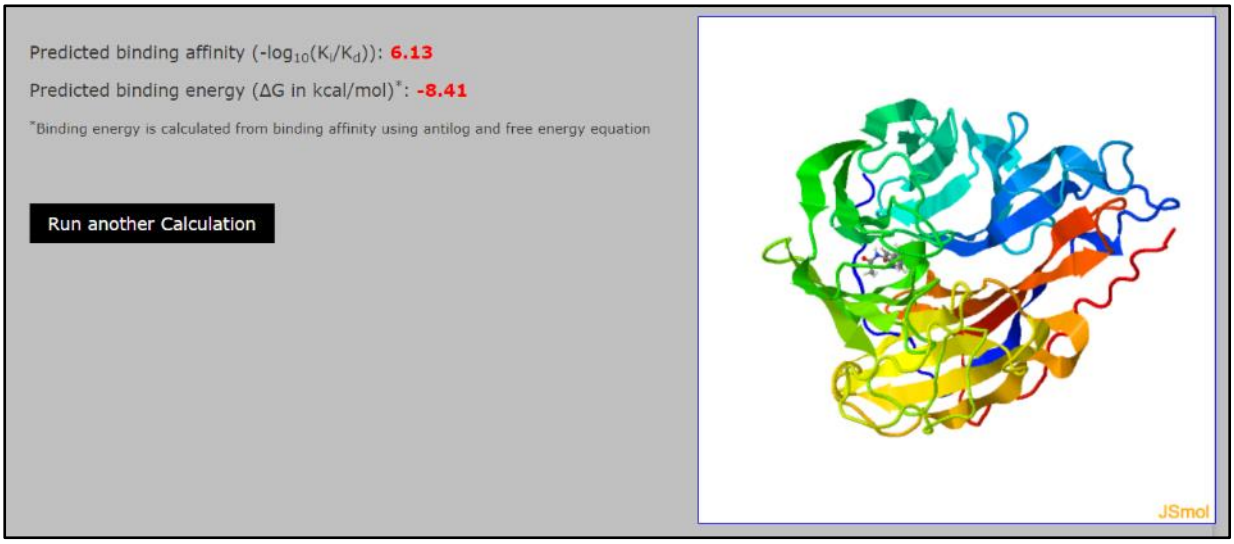
For running BAPPL+ on multiple structures, click on the "**Multiple Structure Calculation**" option, then upload the compressed file (.zip) containing all the PDB files and a text file (.txt) (as described above template.txt) using the browse button in front of **'Upload Zip file.'** Now specify email id to get job completion notification and click on the **'Run BAPPL+'** button to submit your work on the server.
Sample files are also available on the website. Please check the sample for further help.



Chapter 5 Result

It takes about 1-2 minutes for BAPPL+ to run a single structure calculation job. In some cases, the service time may increase to 5 minutes depending on the input file size and queue length. The result will be sent to the email id specified during job submission. If mail is not specified, the result can be retrieved by the **Job Id** provided after each job submission on the top right corner. Job ID can also be used to check your job status. The results are presented as binding affinity and binding energy (in kcal/mol) between the protein-ligand complexes.

Below is shown the result of the single structure calculation job-



Below is shown the result of the multiple structure calculation job-