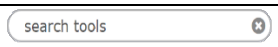
MPDS^{TB} 1.0.1: Manual

Molecular Property Diagnostic Suite (MPDS^{TB}): An Open Source Chemoinformatics Portal

Hosted at
Centre for Molecular Modeling
CSIR-IICT
Tarnaka, Hyderabad-500007
INDIA

Contents

Default symbols used in Galaxy Interface

S. No.	Icon	Description
1		Search box
2		Refresh
3		Settings
4		Edit Tag
5		Edit Annotations
6		View data
7		Edit
8		Delete
9		Download
10		View details
11		Run this job again
12		Run the job

1. Introduction

MPDS^{TB} 1.0.1 consists of nine modules. It covers informatics (databases, file format conversion, visualization), structure and analog based drug design approaches (property calculation, QSAR, docking, fragment library). The Molecular Property Diagnostic Suite (MPDS^{TB}) is an Open Source Chemoinformatics portal; conceptualized to assess and estimate the multifarious aspects of drug-likeliness of any given molecule, in order to diagnose their potential application as drug.

2. MPDS: Upload Data

4. 2.1 Get Data

To upload your input file (ligand, receptor) go to **Get Data** and click upload file (fig.1a).

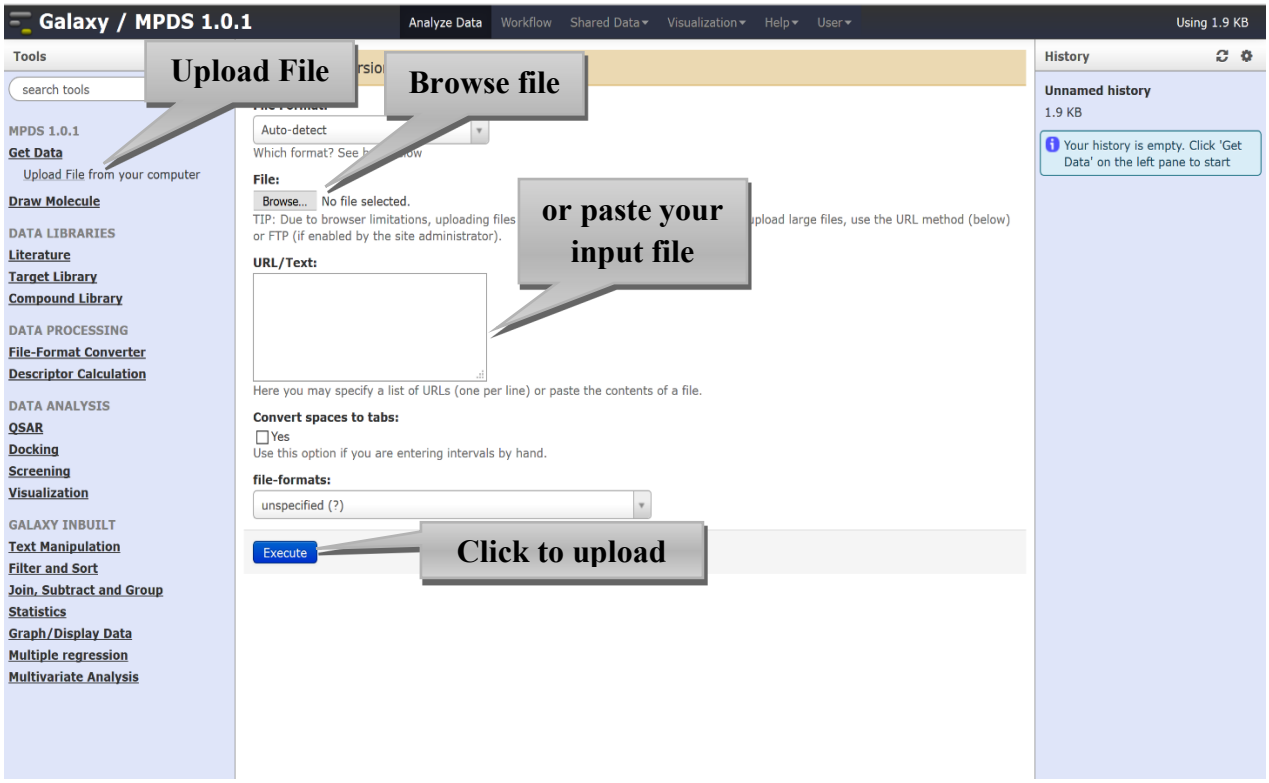


Figure 1a. Upload your input file (ligand, receptor).

5. 2.2 Draw Molecule

To draw your input file (ligand) go to Draw Molecule and sketch your molecule and import it into galaxy either SMILES or mol format. This input can be used for all the compound library searches and other modules in MPDS^{TB} (fig.1b).

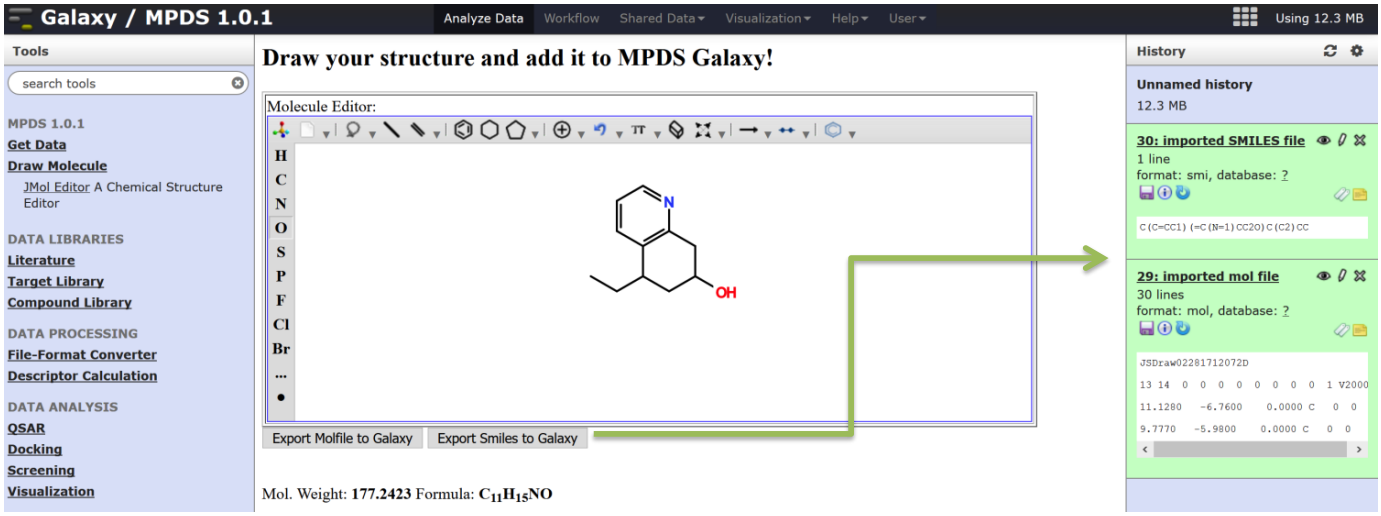
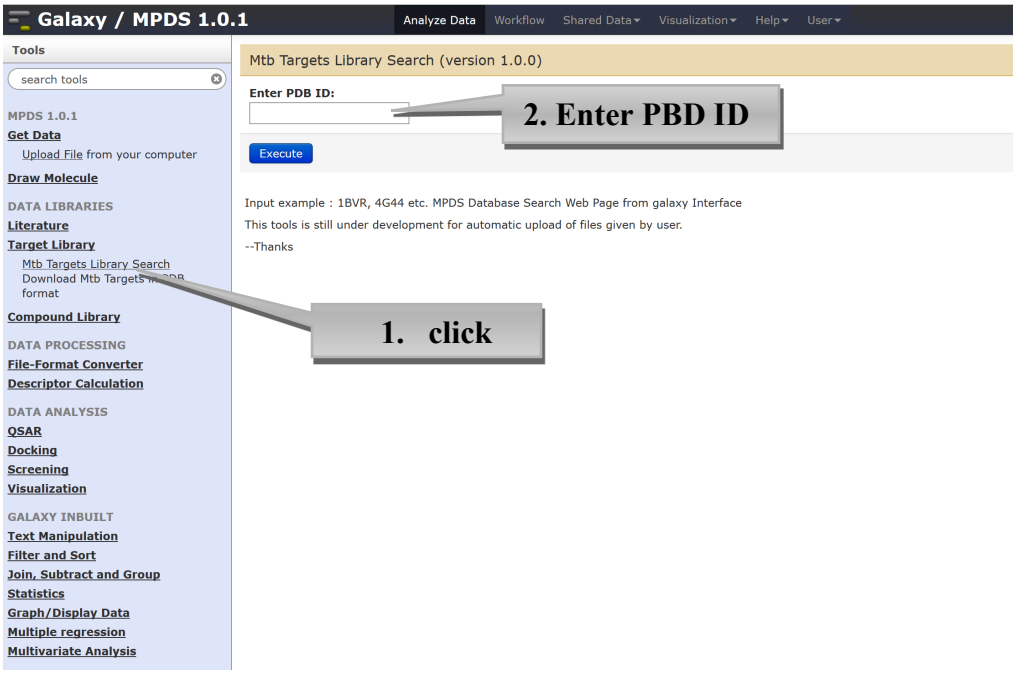


Figure 1b. Draw molecule and import it into galaxy either SMILES or mol format

2. Data Libraries

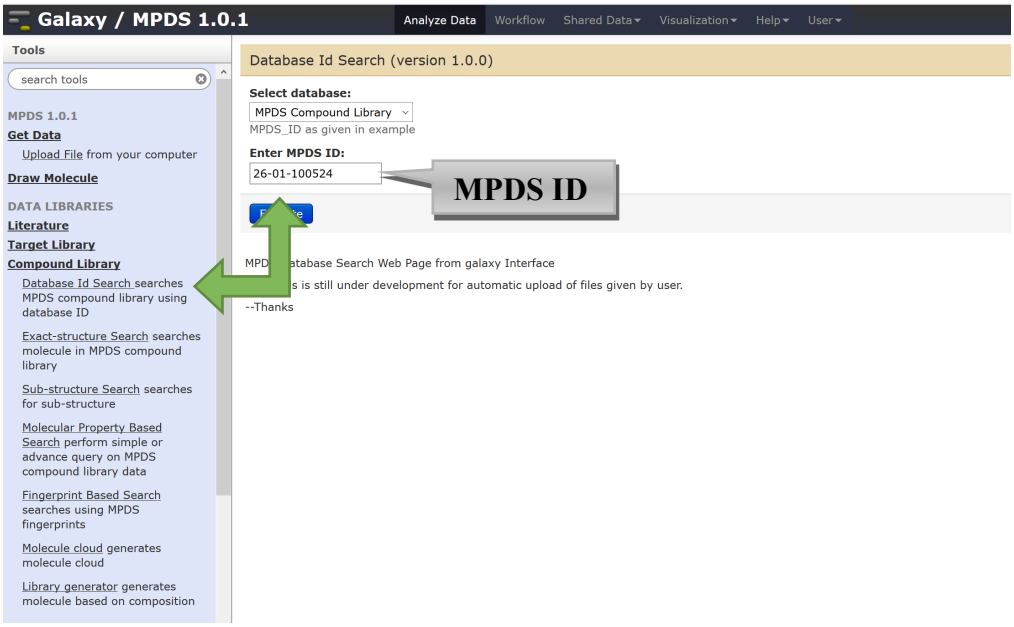
3.1. Module 1-Literature

3.2. Module 2-Target Library



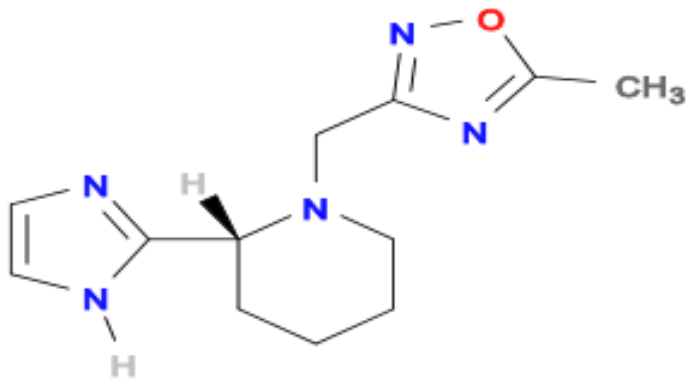
3.3. Module 2- Compound Library

3.3.1. Database ID Search:



Molecular Property Diagnostic Suite

MPDS ID: 26-01-100524



Molecular Formula:
$C_{12}H_{17}N_5O$

IUPAC Name:
N-methyl-2-{[(2R)-morpholin-2-yl]methyl}-3H-imidazo[4,5-b]pyridin-5-amine

Remarks:

Remarks here...

Name/Synonyms:

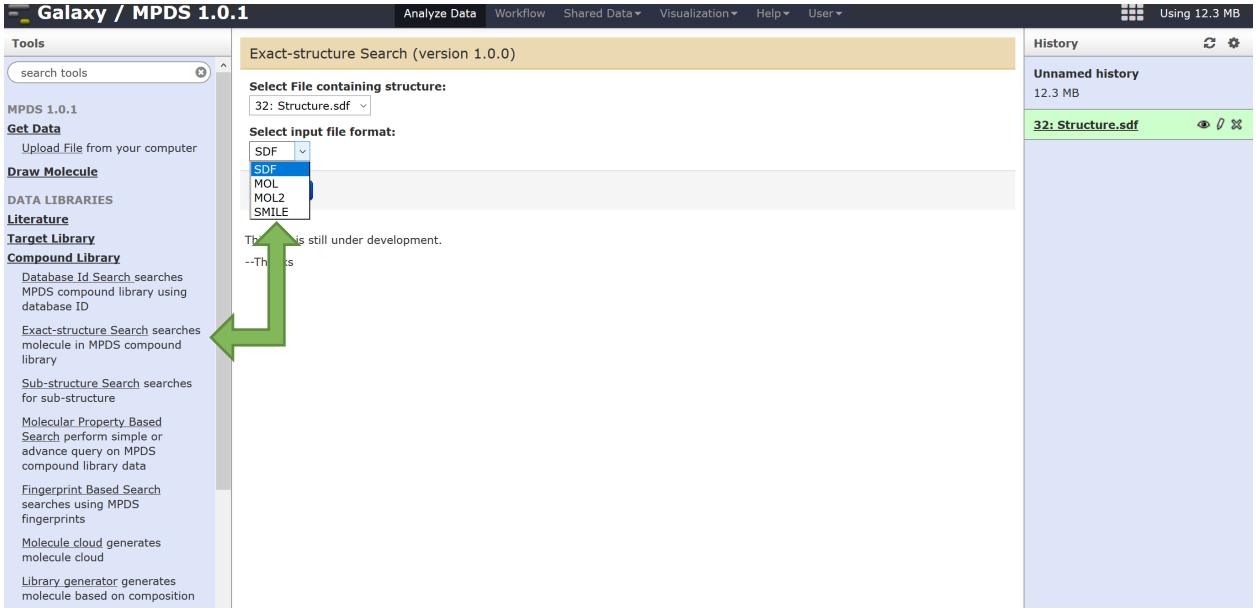
Name/Synonyms here...

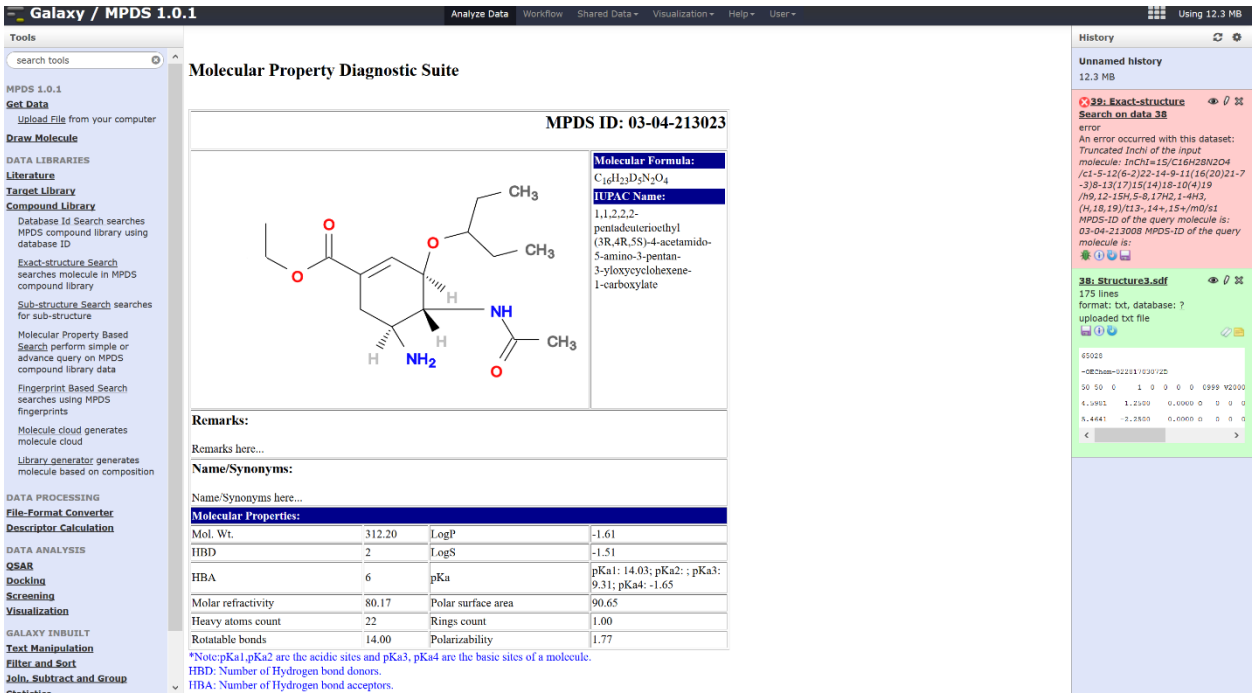
Molecular Properties:

Mol. Wt.	247.14	LogP	-1.40
HBD	1	LogS	-2.33
HBA	4	pKa	pKa1: 12.54; pKa2: ; pKa3: 6.15; pKa4: 2.98
Molar refractivity	35.27	Polar surface area	70.84
Heavy atoms count	18	Rings count	3.00
Rotatable bonds	4.00	Polarizability	1.86

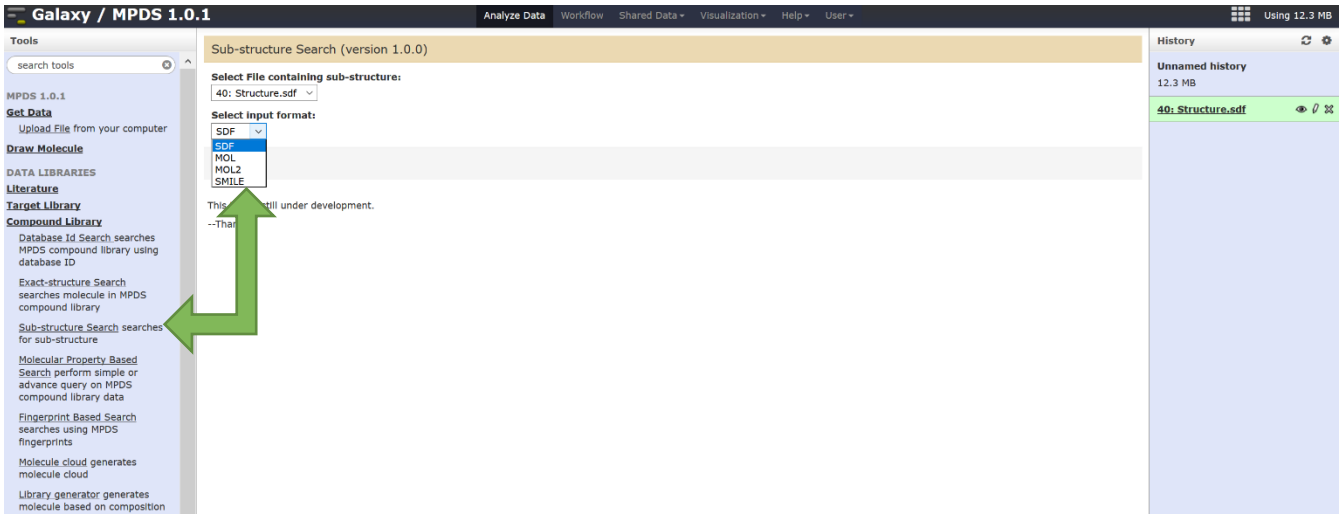
*Note:pKa1,pKa2 are the acidic sites and pKa3, pKa4 are the basic sites of a molecule.
HBD: Number of Hydrogen bond donors.
HBA: Number of Hydrogen bond acceptors.

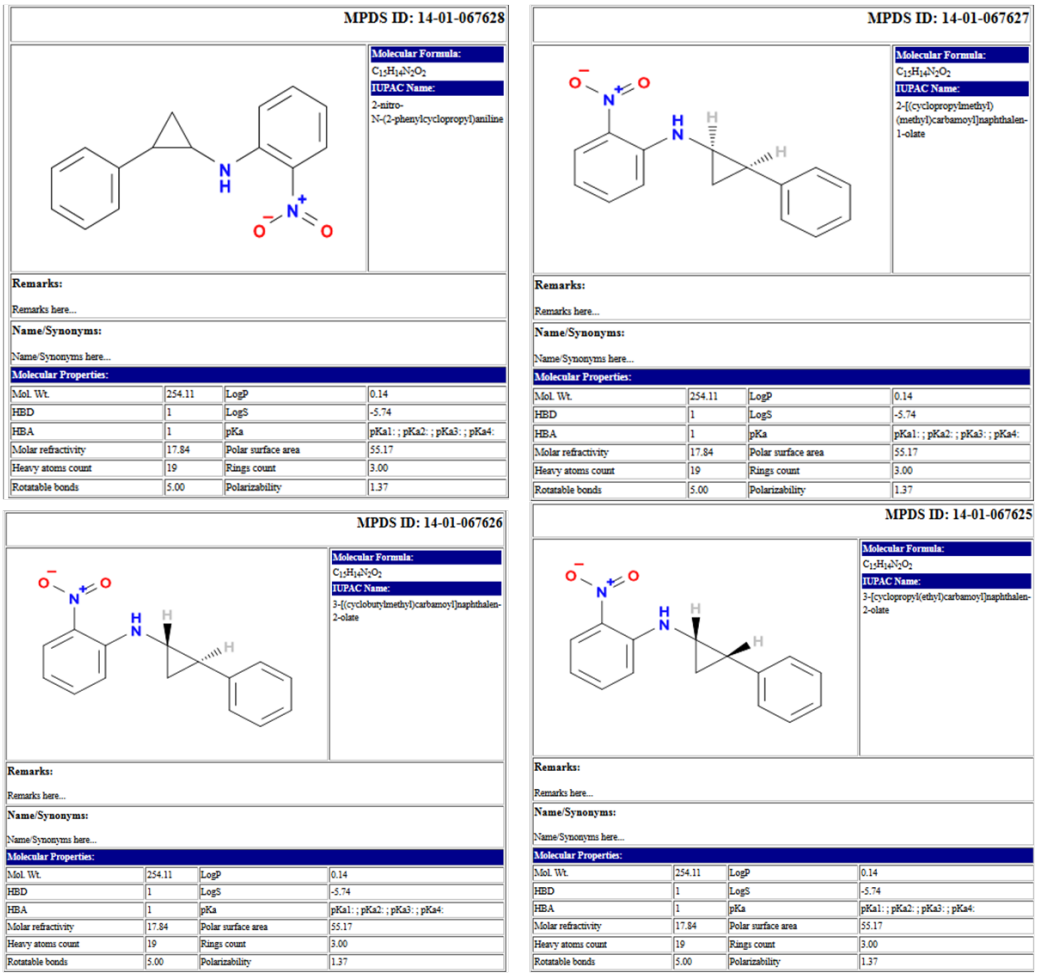
3.3.2. Exact Structure Search:



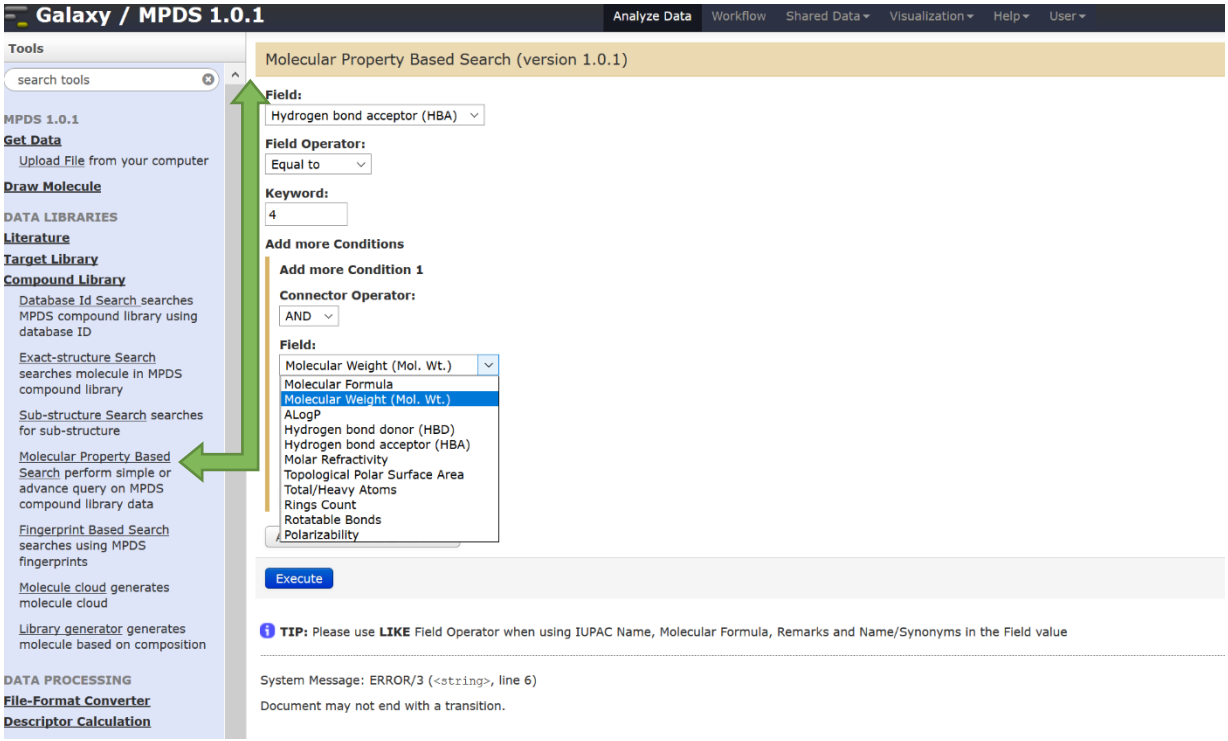


3.3.3. Sub-structure Search:



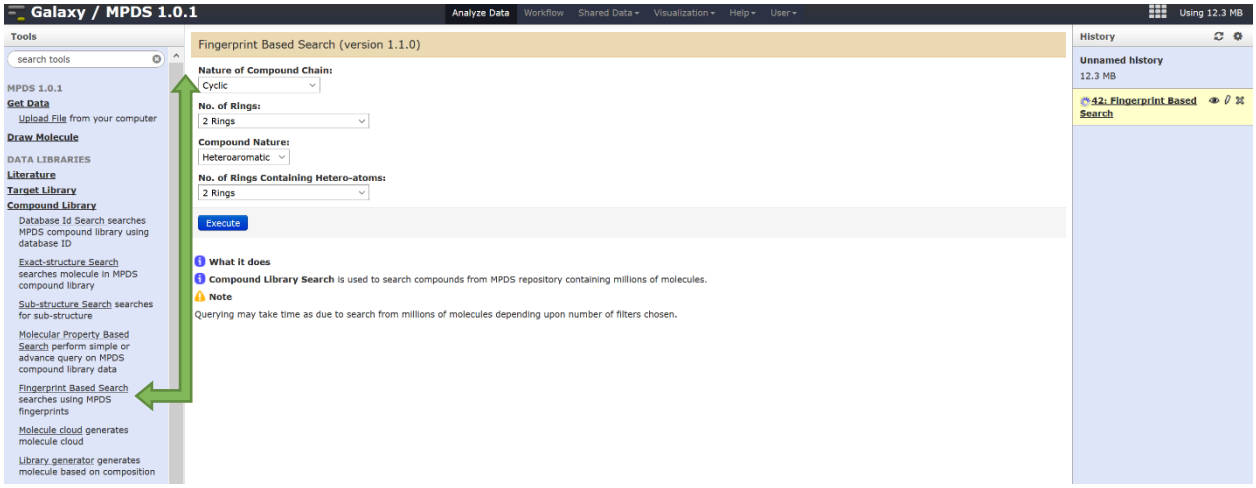


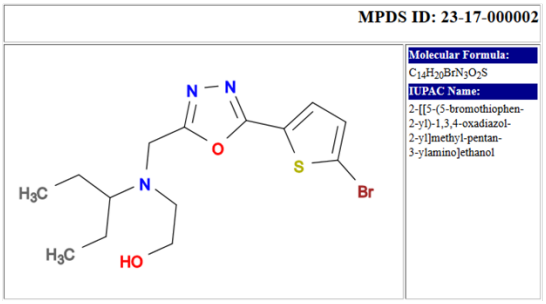
3.3.4. Molecular Property-based Search:

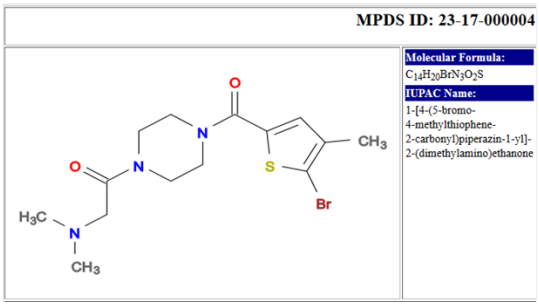


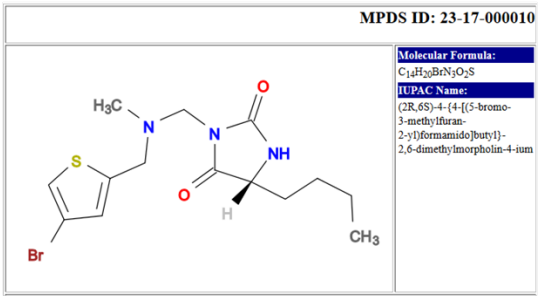
Sr.No.	Source ID	Molecular Formula	Molecular Weight	Total/Heavy Atoms	No. of Rings	No. of Rotatable
1	81254820	C16H24N4	272.20	20	2.00	10.00
2	62781398	C14H28N2O2	256.22	18	1.00	11.00
3	82648576	C12H16N2O3	236.12	17	2.00	6.00
4	ZINC72192526	C9H16N5+	194.14	14	2.00	3.00
5	82474341	C13H14N2O2	230.11	17	2.00	6.00
6	84459040	C16H30N2O2	282.23	20	2.00	8.00
7	19528390	C8H12N4O4	228.09	16	1.00	9.00
8	52349284	C16H27N4O+	291.22	21	2.00	9.00
9	52349282	C16H27N4O+	291.22	21	2.00	9.00
10	52349278	C16H27N4O+	291.22	21	2.00	9.00
11	75831150	C16H27N4O+	291.22	21	2.00	9.00
12	52349280	C16H27N4O+	291.22	21	2.00	9.00
13	79496934	C14H21N3O2	263.16	19	2.00	9.00
14	63027295	C13H24N2O2	240.18	17	2.00	11.00
15	83706632	C12H17N3O	219.14	16	2.00	3.00
16	60654680	C16H19N3O2	285.15	21	1.00	10.00
17	84048566	C13H24N2O2	240.18	17	2.00	8.00
18	20918123	C17H19N3O	281.15	21	4.00	4.00
19	7154525	C11H13N2O2S-	237.07	16	1.00	8.00
20	3485749	C11H14N2O2S	238.08	16	1.00	8.00
21	81340404	C10H9F3N4O3	290.06	20	2.00	9.00
22	85545714	C13H14N2O2	230.11	17	2.00	6.00
23	ZINC32541243	C15H15N5O2	298.13	22	3.00	6.00
24	39782185	C15H15N5O2	298.13	22	3.00	6.00
25	64634347	C11H20N4OS	256.14	17	2.00	6.00
26	61902106	C9H7F3N2O2S	264.02	17	1.00	7.00
27	63791537	C12H19NO3S	257.11	17	2.00	8.00

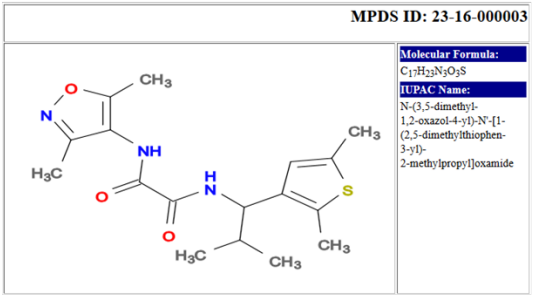
3.3.5. Fingerprint-based Search:

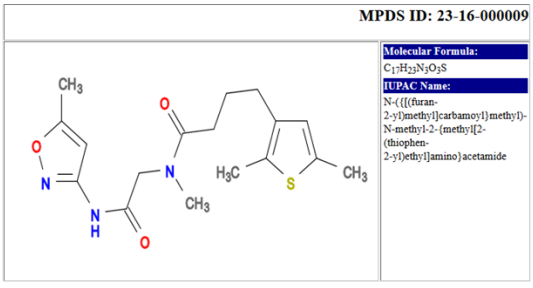


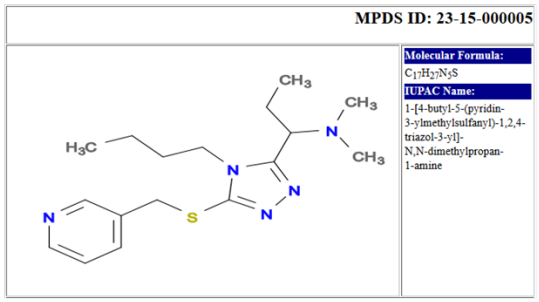




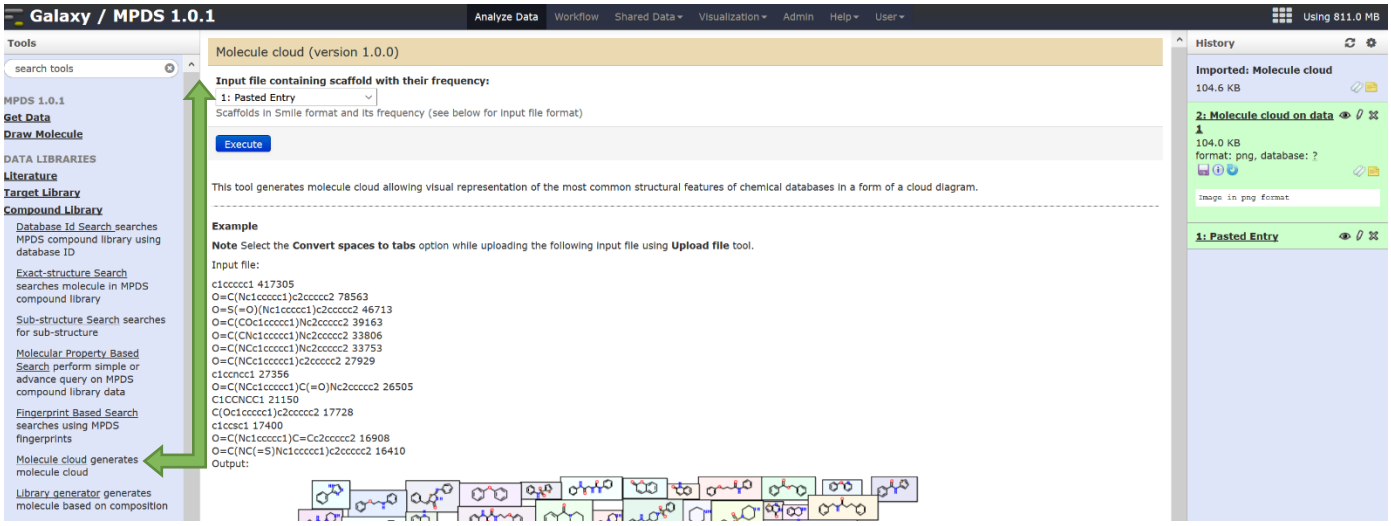


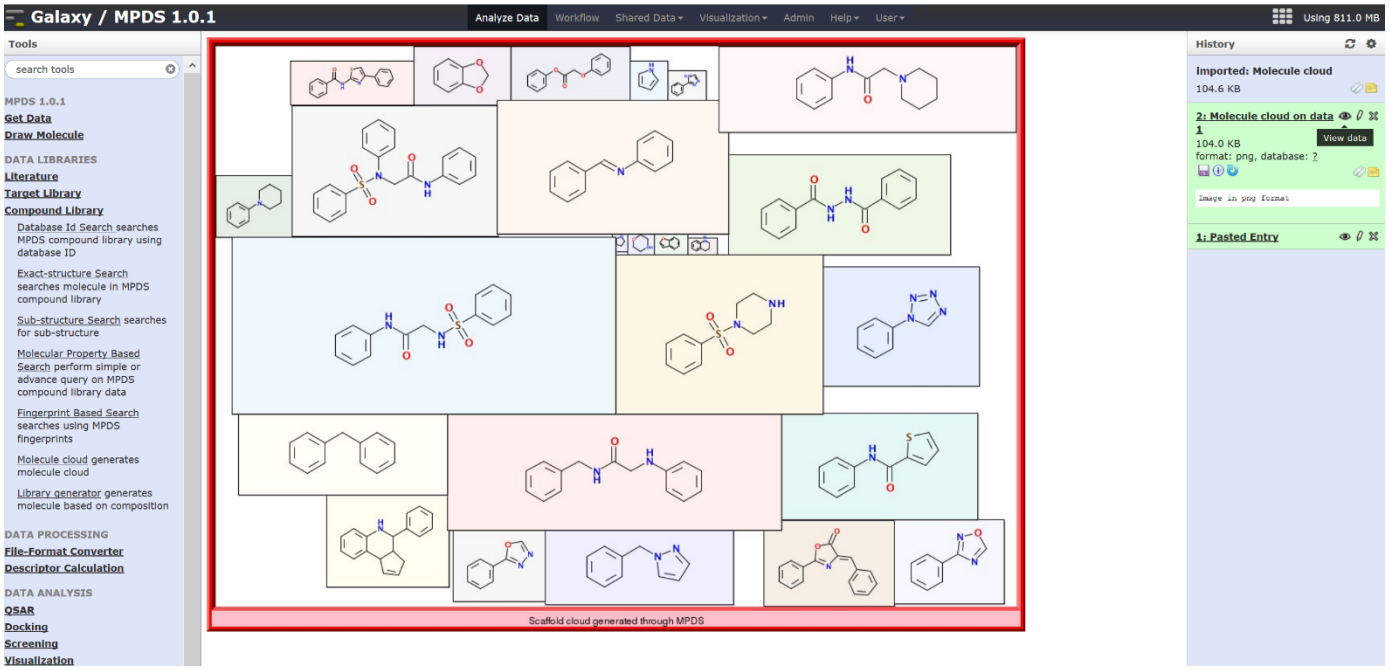




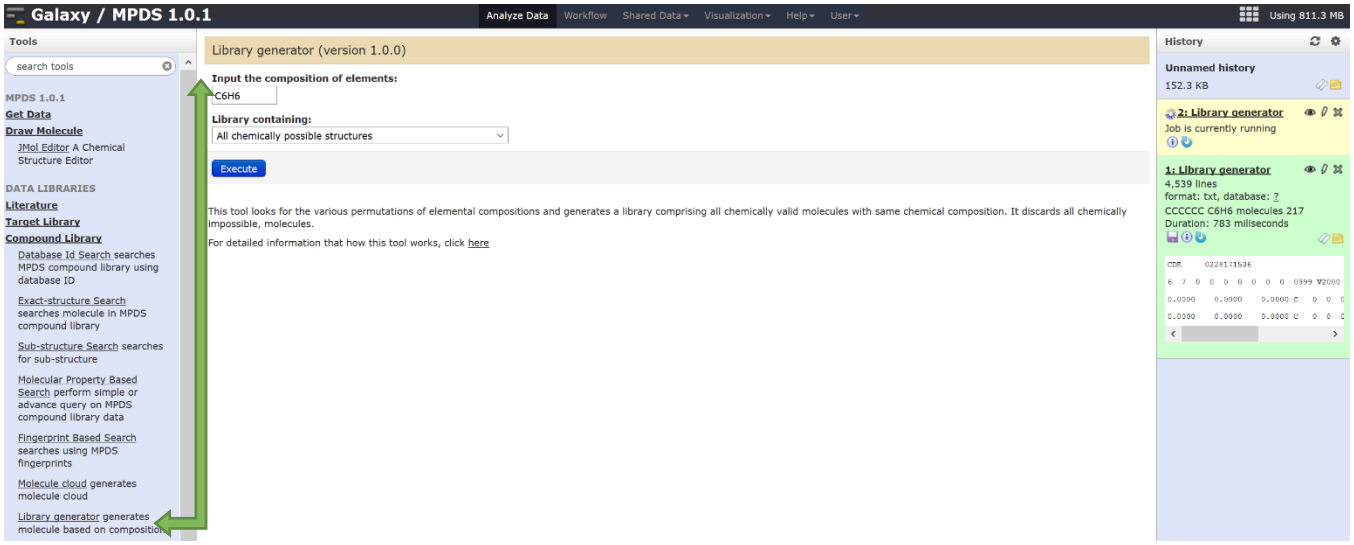


3.3.6. Molecule cloud:

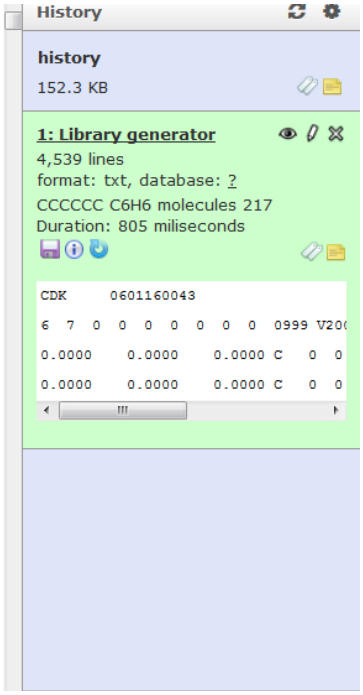




3.3.7. Library Generator:




```
  CDK     0601160043

  6  7  0  0  0  0  0  0  0  0999 V2000
    0.0000    0.0000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.0000    0.0000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.0000    0.0000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.0000    0.0000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.0000    0.0000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.0000    0.0000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
  1  2  1  0  0  0  0
  1  6  1  0  0  0  0
  2  3  1  0  0  0  0
  2  5  1  0  0  0  0
  3  4  2  0  0  0  0
  4  6  1  0  0  0  0
  5  6  2  0  0  0  0
M  END
> <Id>
1

> <can_string>
0100010010100002000000010000020000000

$$$$

  CDK     0601160043

  6  7  0  0  0  0  0  0  0  0999 V2000
    0.0000    0.0000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.0000    0.0000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.0000    0.0000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.0000    0.0000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
```



4. Data Processing

4.1. Module 4- File Format Convertor

4.1.1 Converter

Step 1: Upload your ligand file from **Get Data** (fig.1).

Step 2: Click on file format convertor (Module 3) then go to **convertor to interconvert molecule file format**. Select desired output file format. Click on "Execute" button (fig. 2).

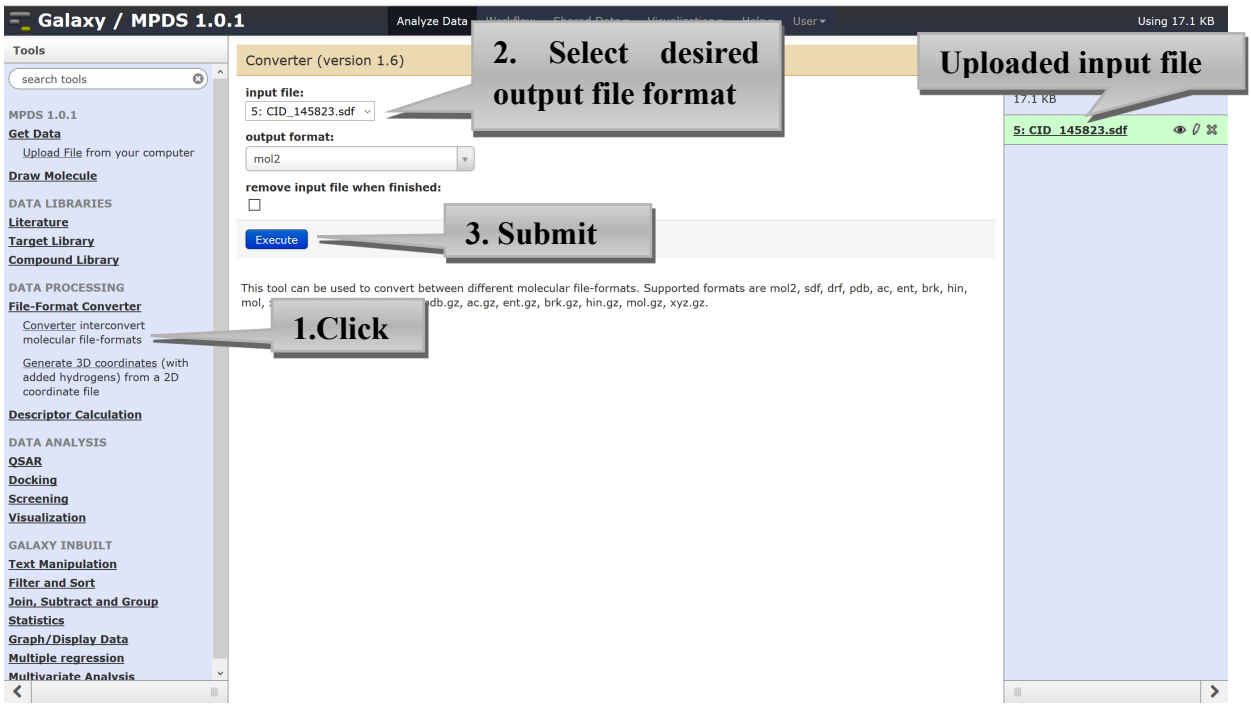


Figure 1

Step 3: The converted output file appears in the history which can be).

Figure 2

4.1.2. Generate 3D coordinates

Step 1: Select input file from local computer and click on “Execute” button to upload (fig. 1).

Step 2: To generate 3D coordinates of input file go to **generate 3D coordinate** sub module (fig 4).

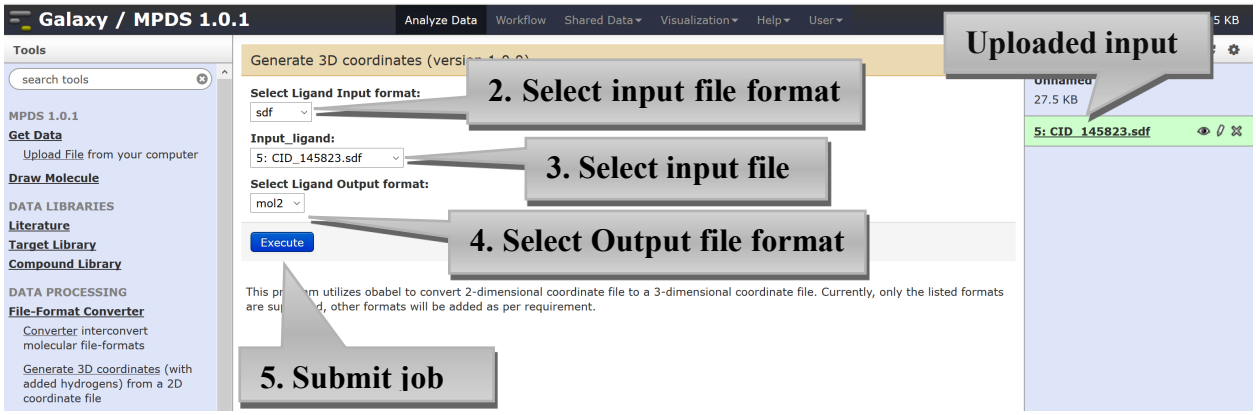


Figure1

Step 4: Results: the 3D coordinates file can be seen and downloaded from history (fig.5).

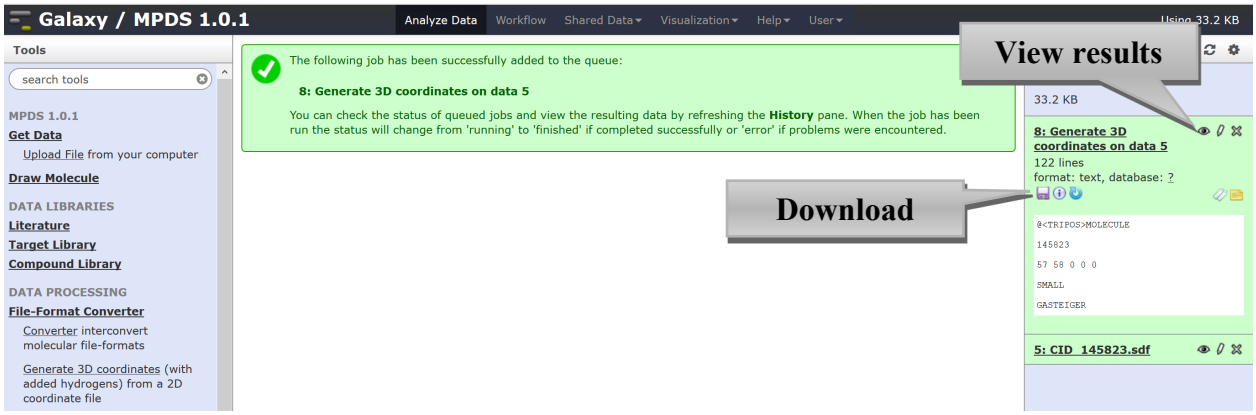


Figure 2

4.2. Descriptor Calculator

4.2.1. PaDEL Descriptor Calculator

- **Step 1:** Upload your ligand file from **Get Data** (fig.1). Select Smile file from local computer and click on "Execute" button.
- **Step 2:** Submitting descriptor calculation job: (Fig. 2).
- **Step 3: Results**: In the history panel of MPDS home page user can see the jobs completed and can download results (same as CDK descriptor results).

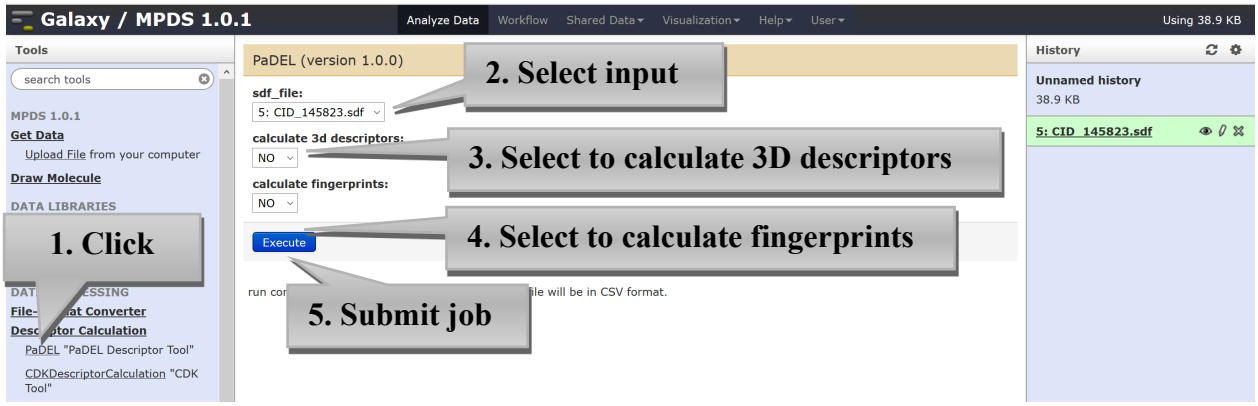


Figure 1

4.2.2.CDK Descriptor Calculator

- **Step 1:** Upload your ligand file from Get Data (fig.1). Select .sdf file from local computer and click on "Execute" button.
- **Step 2:** Submitting descriptor calculation job: For descriptor calculation user may choose all types of descriptors available or may use geometrical, constitutional, electronic, topological or hybrid descriptors for calculation. For fingerprint calculation user have choice for various fingerprints like standard, extended, PubChem etc. (Fig. 6). The recent version (i.e.CDK-1.4.2) do not supports descriptor and fingerprint calculation simultaneously.
- **Step 3:** Results: In the history panel of MPDS home page user can see the jobs completed and can download results (fig 7).

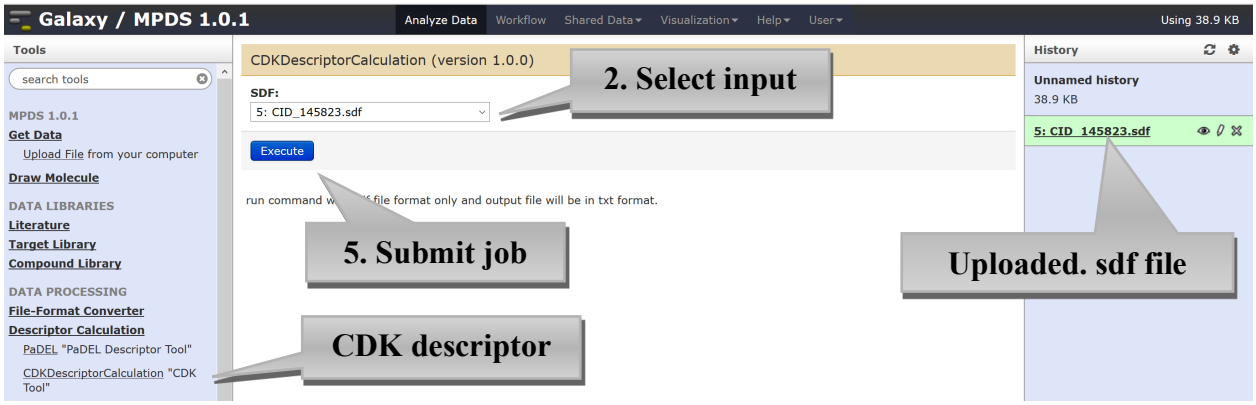


Figure 2

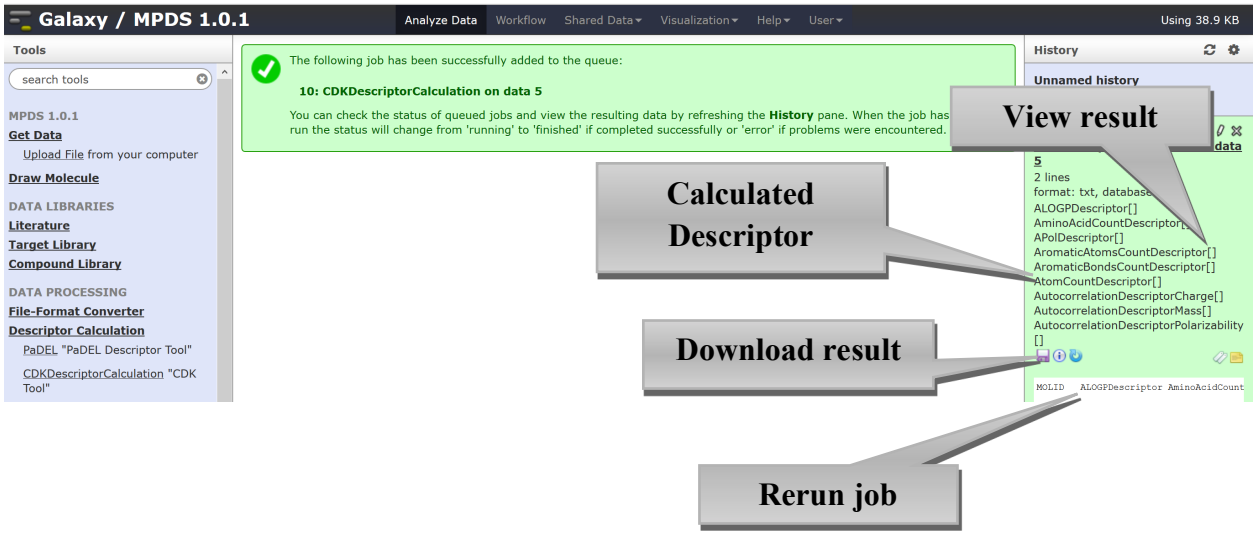


Figure 7

5. Data Analysis

5.1. QSAR

QSAR Model Building: In MPDS^{TB} data mining there are three tools for QSAR model building.

1. SVM light
2. McQSAR
3. Weka

5.1.1. QSAR Model Building using McQSAR

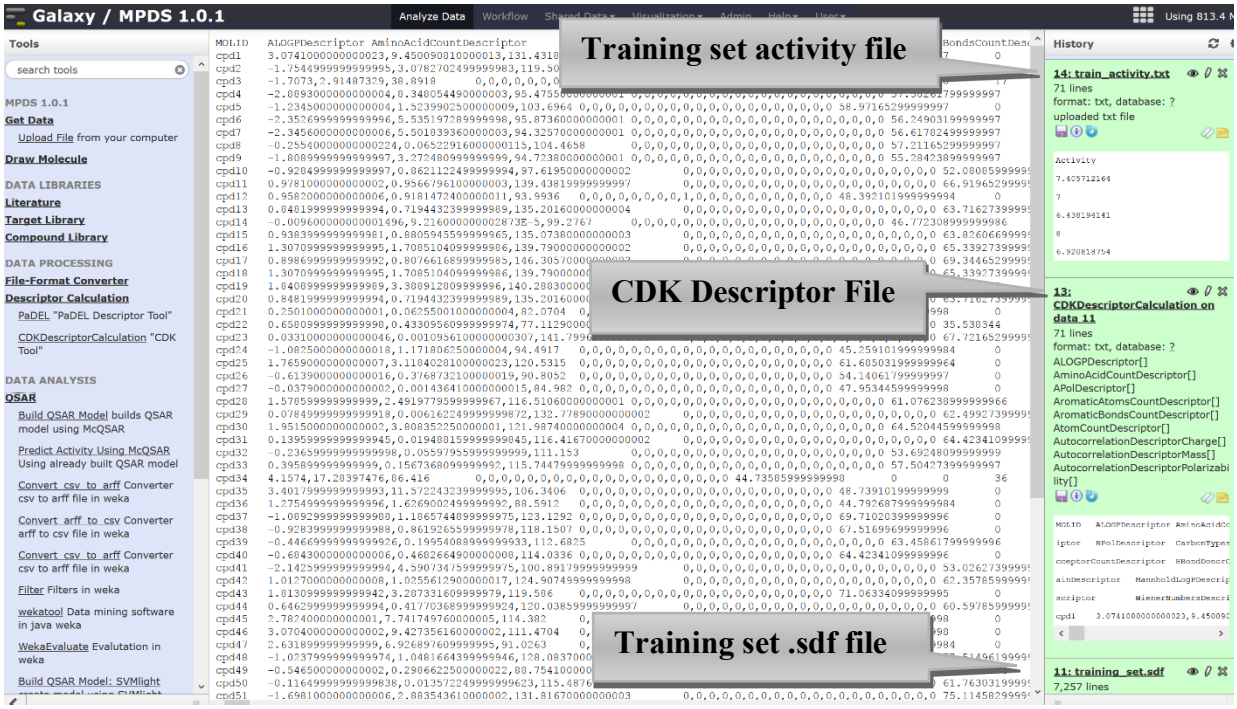
McQSAR builds regression model which can be further used for predicting activity values (in terms of IC50, LD50 or EC50 values or as per requirement, user may choose appropriate field of interest from file). McQSAR model building requires preparation of appropriate descriptor files as it needs activity (or any appropriate field) column.

Input: .sdf

- **Step 1:** Upload your input file (.sdf)from Get Data (fig.1).
- **Step 2:** Calculate its CDK descriptors as mentioned earlier in Cdk descriptor calculation (fig 15).
- **Step 3:** After descriptor calculation user need to add Activity information (e.g. Activity, IC50, mIC or EC50 etc.) to the descriptor file. To add activity information first click on Text Manipulation → Paste and select descriptor file and uploaded activity file from history in appropriate manner (activity at last is preferable) and valid separator (fig.16). On execution resultant file will be input for McQSAR model building.
- **Step 4:** The resultant file obtained from previous step is input for McQSAR model building. There are many options are available for Pre Processing and selecting parameters (fig.17) in tool, '*Build QSAR Model'* in Data mining section. It is advisable to select all preprocessing methods to avoid any complication for model building and also for obtaining significant model(s). Click on "Execution" button.
- **Step 5:** Results of McQSAR build model obtained in logs file and model file (fig.18).

Activity of already built model can be predicted by using one of the sub modules of Data mining module i.e. Predict activity using McQSAR.

- **Step 6:** For prediction, calculate CDK descriptor for sdf file (compounds whose response values need to calculate). Here we used a sdf file whose activity need to be predicted (we have prior information about its activity values as we need to check reliability and significance of result) (fig.19).
- **Step 7:** On execution user will get result of prediction ('Prediction Result') and log file (Standard Output) (fig.20).



6. Figure 15

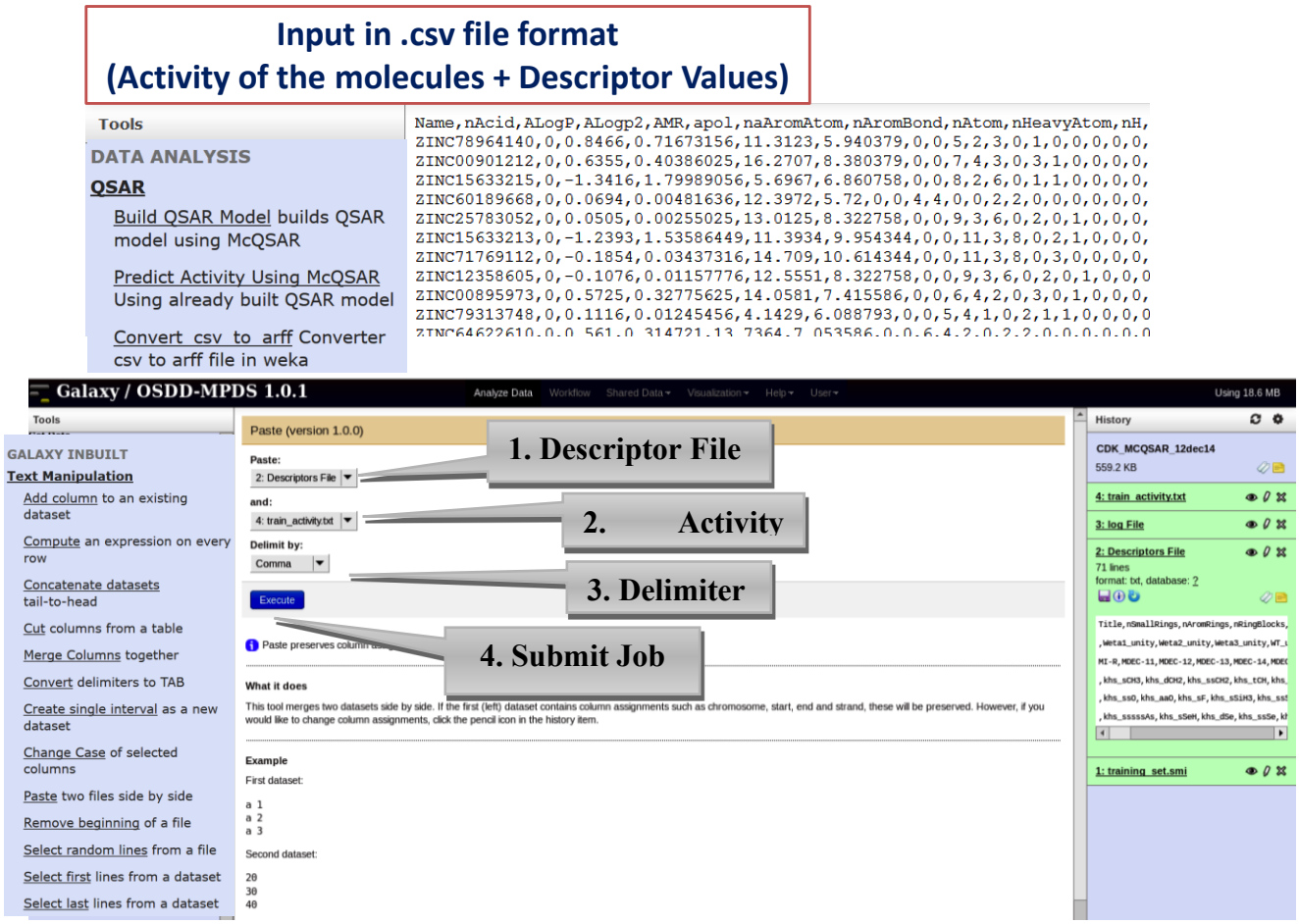

Input in .csv file format
(Activity of the molecules + Descriptor Values)
Tools
DATA ANALYSIS
QSAR
Build QSAR Model builds QSAR model using McQSAR
Predict Activity Using McQSAR Using already built QSAR model
Convert csv to arff Converter csv to arff file in weka
Name,nAcid,ALogP,ALogp2,AMR,apol,naAromAtom,nAromBond,nAtom,nHeavyAtom,nH,
ZINC78964140,0,0.8466,0.71673156,11.3123,5.940379,0,0,5,2,3,0,1,0,0,0,0,0,
ZINC00901212,0,0.6355,0.40386025,16.2707,8.380379,0,0,7,4,3,0,3,1,0,0,0,0,
ZINC15633215,0,-1.3416,1.79989056,5.6967,6.860758,0,0,8,2,6,0,1,1,0,0,0,0,
ZINC60189668,0,0.0694,0.00481636,12.3972,5.72,0,0,4,4,0,0,2,2,0,0,0,0,0,0,
ZINC25783052,0,0.0505,0.00255025,13.0125,8.322758,0,0,9,3,6,0,2,0,1,0,0,0,
ZINC15633213,0,-1.2393,1.53586449,11.3934,9.954344,0,0,11,3,8,0,2,1,0,0,0,
ZINC71769112,0,-0.1854,0.03437316,14.709,10.614344,0,0,11,3,8,0,3,0,0,0,0,
ZINC12358605,0,-0.1076,0.01157776,12.5551,8.322758,0,0,9,3,6,0,2,0,1,0,0,0
ZINC00895973,0,0.5725,0.32775625,14.0581,7.415586,0,0,6,4,2,0,3,0,1,0,0,0,
ZINC79313748,0,0.1116,0.01245456,4.1429,6.088793,0,0,5,4,1,0,2,1,1,0,0,0,0
Galaxy / OSDD-MPDS 1.0.1
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Using 18.6 MB
Tools
GALAXY INBUILT
Text Manipulation
Add column to an existing dataset
Compute an expression on every row
Concatenate datasets tail-to-head
Cut columns from a table
Merge Columns together
Convert delimiters to TAB
Create single interval as a new dataset
Change Case of selected columns
Paste two files side by side
Remove beginning of a file
Select random lines from a file
Select first lines from a dataset
Select last lines from a dataset
Paste (version 1.0.0)
Paste:
2: Descriptors File
and:
4: train_activity.txt
Delimit by:
Comma
Execute
1. Descriptor File
2. Activity
3. Delimiter
4. Submit Job
What it does
This tool merges two datasets side by side. If the first (left) dataset contains column assignments such as chromosome, start, end and strand, these will be preserved. However, if you would like to change column assignments, click the pencil icon in the history item.
Example
First dataset:
a 1
a 2
a 3
Second dataset:
20
30
40
History
CDK_MCQSAR_12dec14
559.2 KB
4: train_activity.txt
3: log File
2: Descriptors File
71 lines
format: txt, database: ?
1: training_set.smi

Figure 16


1. Input descriptor + Activity File

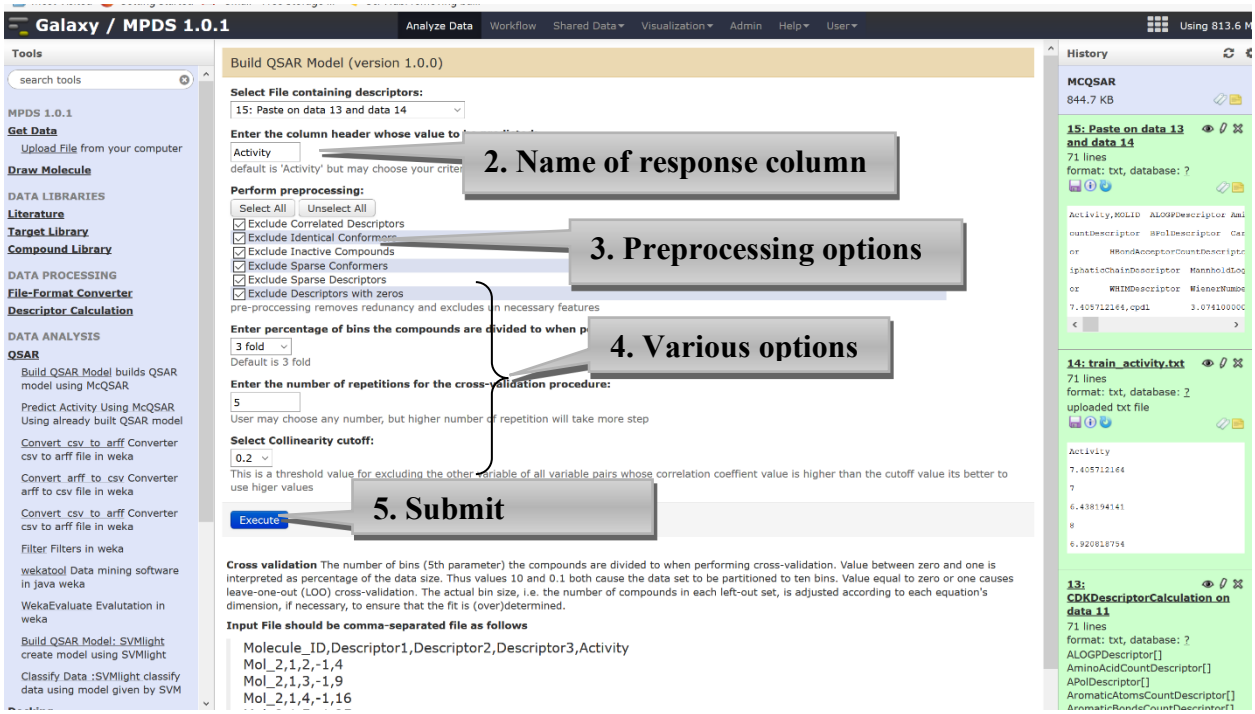


Figure 17

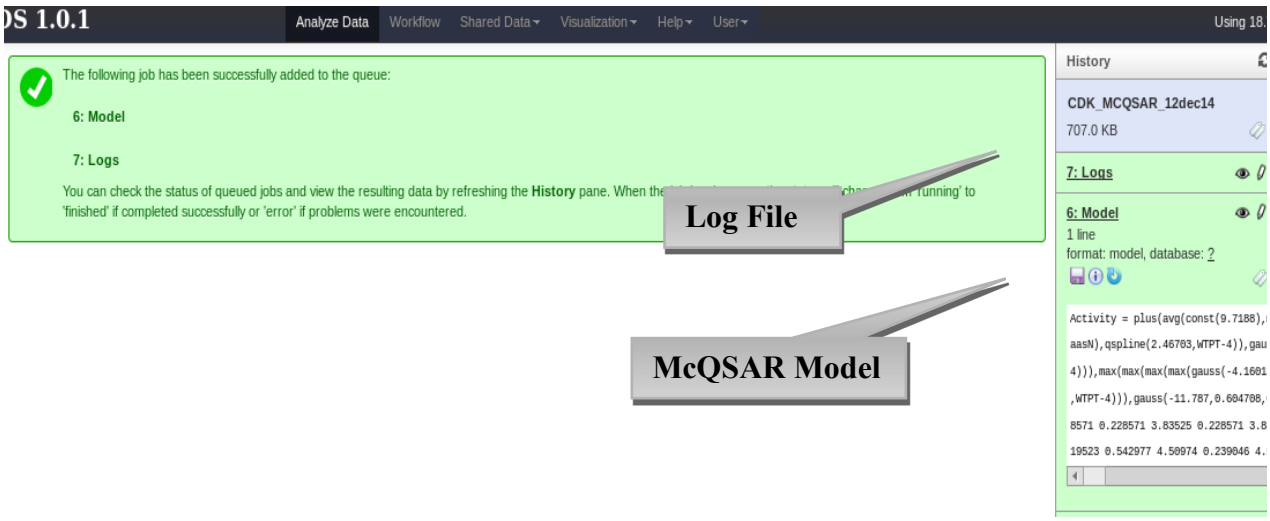


Figure 18

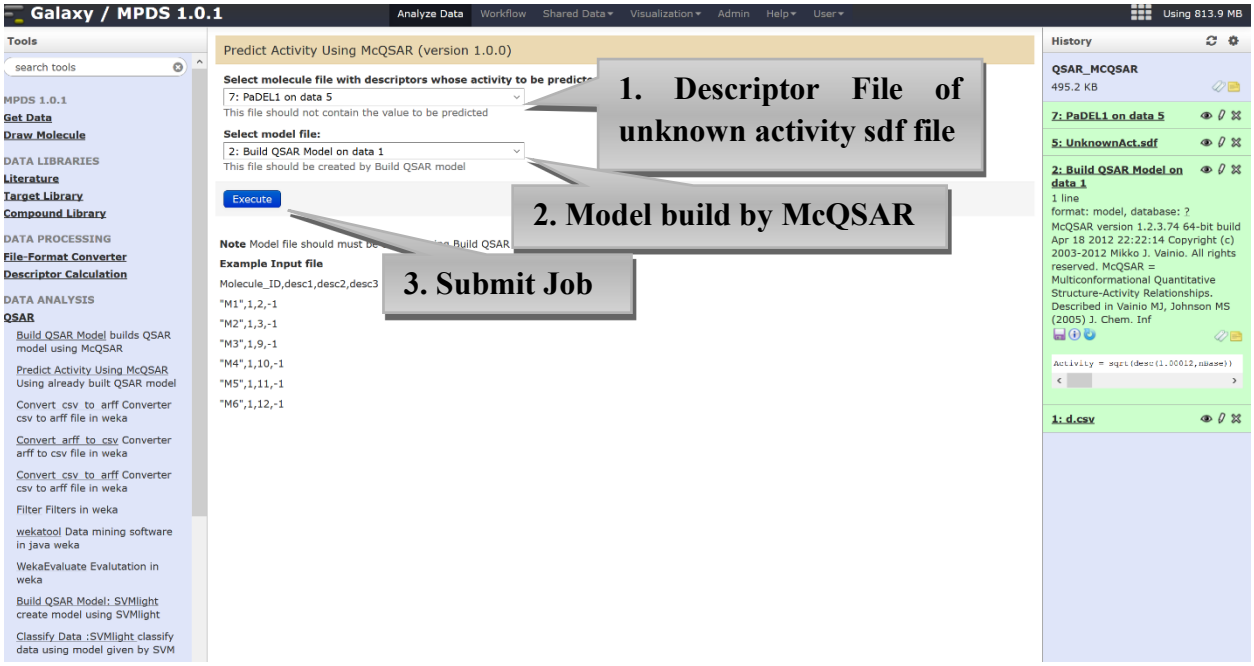
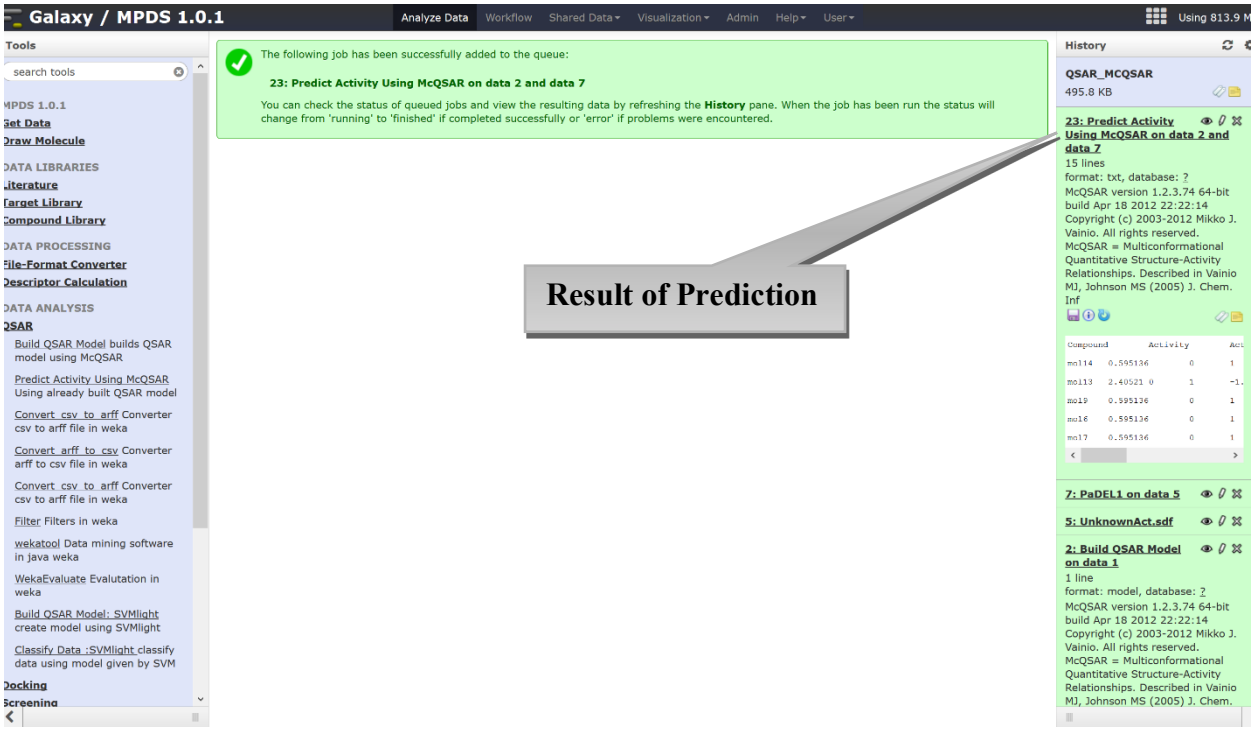


Figure 19



We compared the predicted value with those of actual values and result is shown in table 1.

Compound	Activity	Actual Values
mol1	6.49483	6.443697499
mol2	7.08538	6.397940009
mol3	7.08538	6.337242168
mol4	7.08538	6.145693958
mol5	7.08538	6.124938737
mol6	7.08538	6.004364805
mol7	7.08538	6.663540266
mol8	7.08538	6.13667714
mol9	7.08538	6.823908741
mol10	7.08538	6.425968732
mol11	7.08538	6.045757491
mol12	6.49483	6.420216403
mol13	7.08538	6.193820026
mol14	7.08538	7.096910013

Table1: Actual Vs Predicted values

NOTE: *For demo purpose, User may download the data set (convert .smi to sdf) used for this case study at link given below:*
https://drive.google.com/file/d/0B3c9isKbTnxtZmpzYVc0VVNpWmM/view?usp=sharing

7.

Complete workflow of Cdk-McQSAR model building and activity prediction.

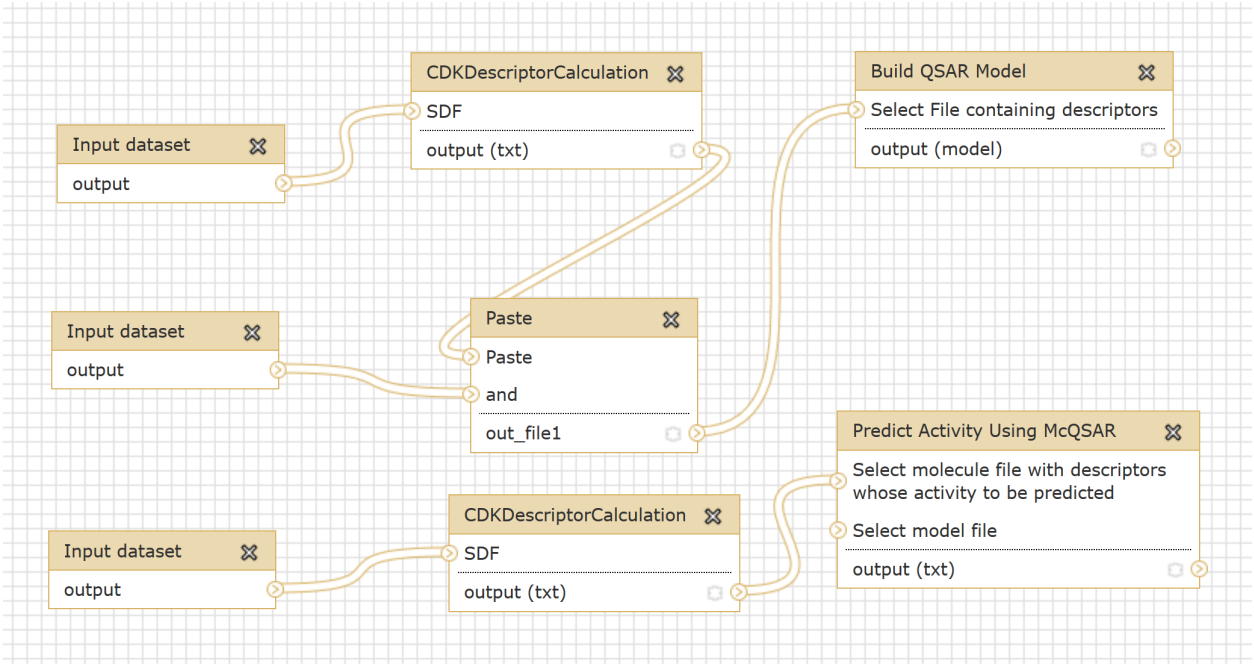


Figure 21

5.1.2. QSAR Model Building using Weka

5.1.3. QSAR Model Building using SVMlight

Input: one descriptor file for known active and another for known inactive.

Step 1: Upload your file from Get Data (fig.1). Select inactive sdf or active sdf files one by one from local computer and click on "Execute" button.

Step 2: Descriptor calculation (as described in module 5 help).

User may choose Classification or Regression as methodology and various kernel functions (linear, radial basis, Polynomial etc.) and its parameters (Fig. 9)

Step 3: Select appropriate options for QSAR model building (fig 10). Here, we used default options (e.g. Methodology: Classification, kernel method: linear and other parameters as default value).

Step 4: Results of SVMlight QSAR model are obtained as statistics file and model file, which can be seen in the history panel of MPDS home page (fig 11).

Step 5: Select descriptor file having unknown activity and model created in previous step (fig.12).

Step 6: Classification of data obtained from SVMlight can be by done using one of the sub module of Data mining module i.e. Classify data. Results of classification of SVM (fig. 13).

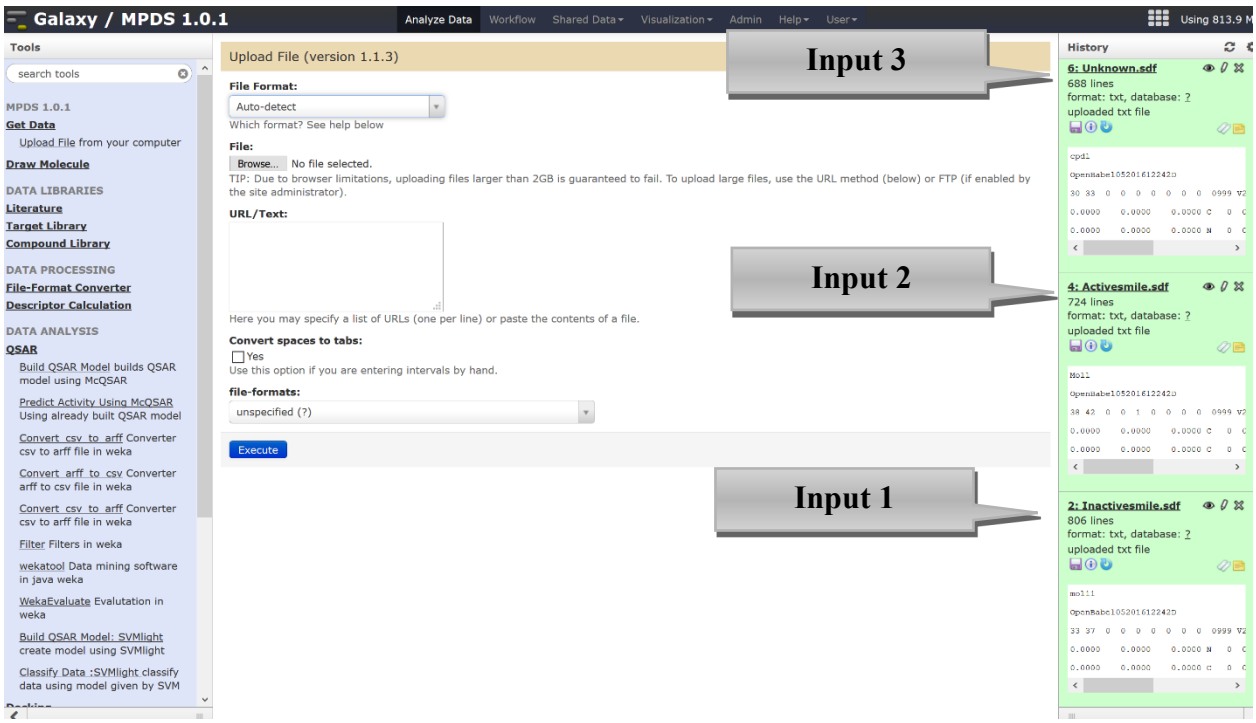

Galaxy / MPDS 1.0.1
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Tools
search tools
MPDS 1.0.1
Get Data
Upload File from your computer
Draw Molecule
DATA LIBRARIES
Literature
Target Library
Compound Library
DATA PROCESSING
File-Format Converter
Descriptor Calculation
DATA ANALYSIS
QSAR
Build QSAR Model builds QSAR model using McQSAR
Predict Activity Using McQSAR Using already built QSAR model
Convert csv to arff Converter csv to arff file in weka
Convert arff to csv Converter arff to csv file in weka
Convert csv to arff Converter csv to arff file in weka
Filter Filters in weka
wekatool Data mining software in java weka
WekaEvaluate Evalutation in weka
Build QSAR Model: SVMlight create model using SVMlight
Classify Data :SVMlight classify data using model given by SVM
Upload File (version 1.1.3)
File Format:
Auto-detect
Which format? See help below
File:
Browse...
No file selected.
TIP: Due to browser limitations, uploading files larger than 2GB is guaranteed to fail. To upload large files, use the URL method (below) or FTP (if enabled by the site administrator).
URL/Text:
Here you may specify a list of URLs (one per line) or paste the contents of a file.
Convert spaces to tabs:
Yes
Use this option if you are entering intervals by hand.
file-formats:
unspecified (?)
Execute
Input 3
Input 2
Input 1
History
6: Unknown.sdf
688 lines
format: txt, database: ?
uploaded txt file
4: Activesmile.sdf
724 lines
format: txt, database: ?
uploaded txt file
2: Inactivesmile.sdf
806 lines
format: txt, database: ?
uploaded txt file

Figure 9

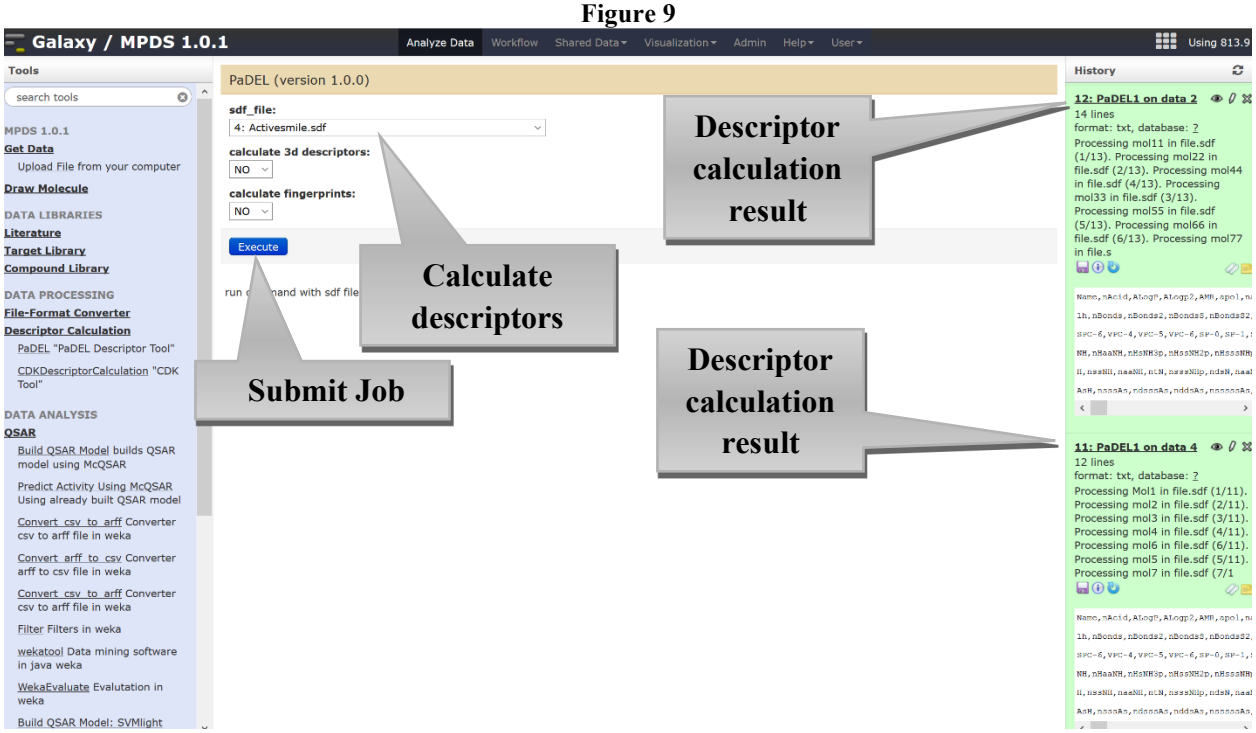
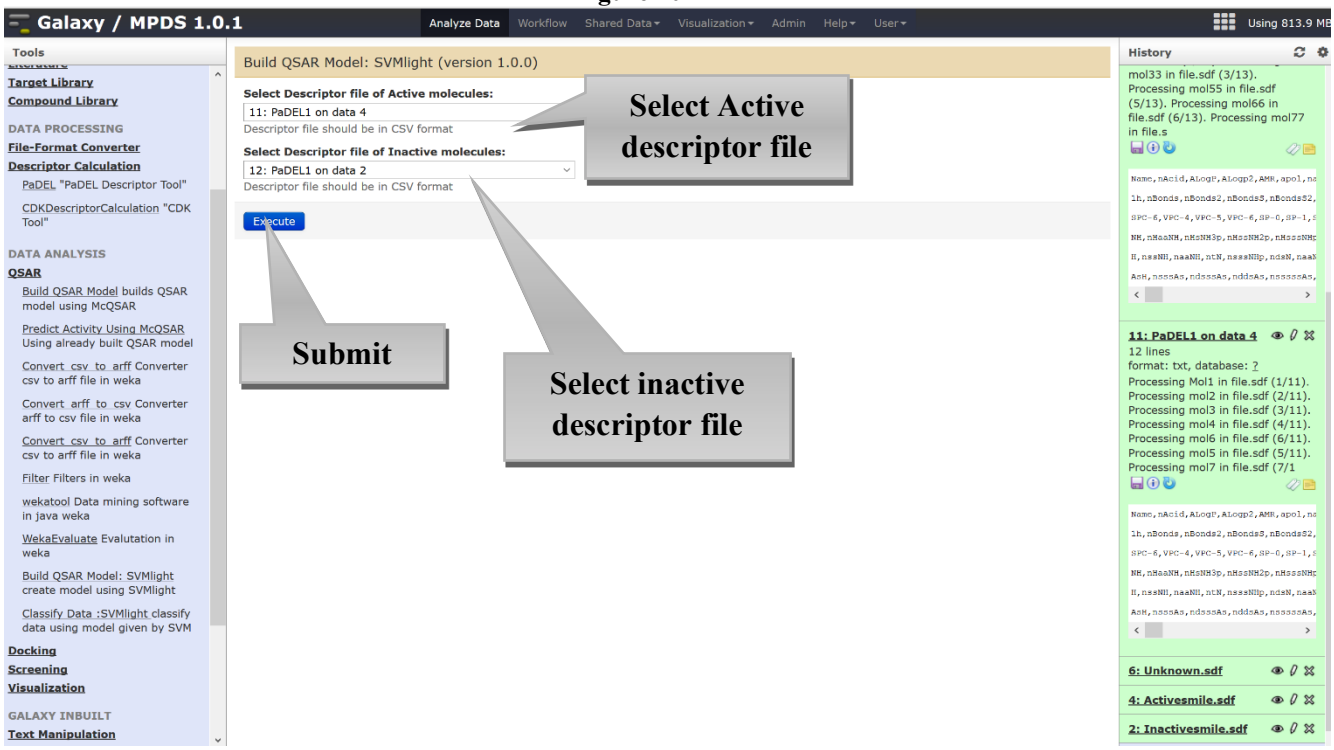
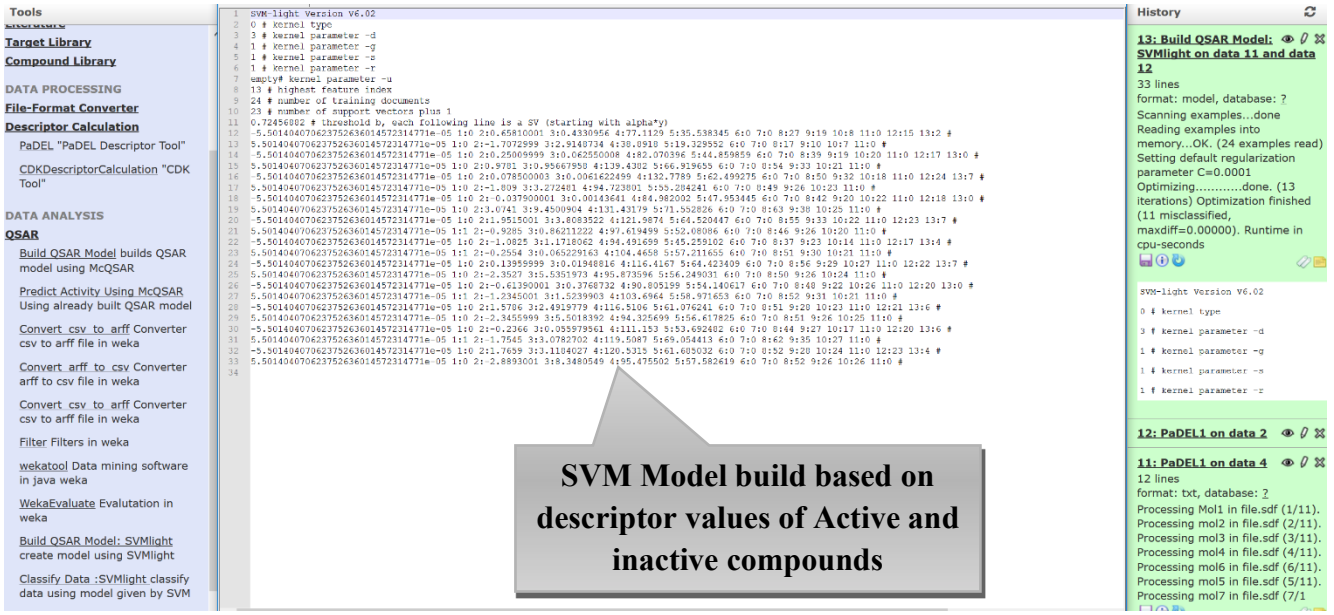
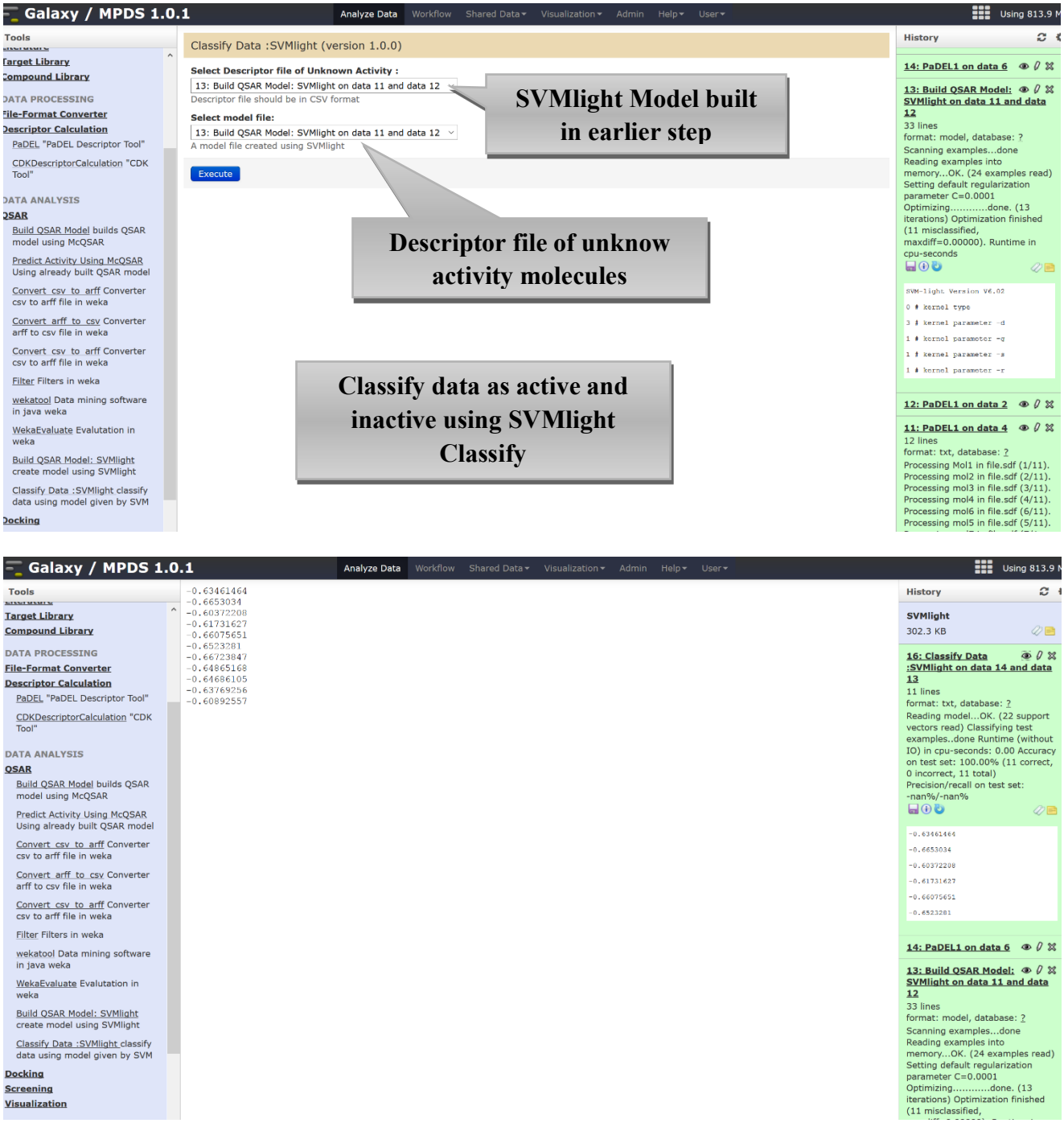

Galaxy / MPDS 1.0.1
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Tools
search tools
MPDS 1.0.1
Get Data
Upload File from your computer
Draw Molecule
DATA LIBRARIES
Literature
Target Library
Compound Library
DATA PROCESSING
File-Format Converter
Descriptor Calculation
PaDEL "PaDEL Descriptor Tool"
CDKDescriptorCalculation "CDK Tool"
DATA ANALYSIS
QSAR
Build QSAR Model builds QSAR model using McQSAR
Predict Activity Using McQSAR Using already built QSAR model
Convert csv to arff Converter csv to arff file in weka
Convert arff to csv Converter arff to csv file in weka
Convert csv to arff Converter csv to arff file in weka
Filter Filters in weka
wekatool Data mining software in java weka
WekaEvaluate Evalutation in weka
Build QSAR Model: SVMlight
PaDEL (version 1.0.0)
sdf_file:
4: Activesmile.sdf
calculate 3d descriptors:
NO
calculate fingerprints:
NO
Execute
Descriptor calculation result
Calculate descriptors
Submit Job
Descriptor calculation result
History
12: PaDEL1 on data 2
14 lines
format: txt, database: ?
11: PaDEL1 on data 4
12 lines
format: txt, database: ?

Figure 10


Galaxy / MPDS 1.0.1
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 813.9 MB
Tools
Target Library
Compound Library
DATA PROCESSING
File-Format Converter
Descriptor Calculation
PaDEL "PaDEL Descriptor Tool"
CDKDescriptorCalculation "CDK Tool"
DATA ANALYSIS
QSAR
Build QSAR Model builds QSAR model using McQSAR
Predict Activity Using McQSAR Using already built QSAR model
Convert csv to arff Converter csv to arff file in weka
Convert arff to csv Converter arff to csv file in weka
Convert csv to arff Converter csv to arff file in weka
Filter Filters in weka
wekatool Data mining software in java weka
WekaEvaluate Evalutation in weka
Build QSAR Model: SVMlight create model using SVMlight
Classify Data :SVMlight classify data using model given by SVM
Docking
Screening
Visualization
GALAXY INBUILT
Text Manipulation
Build QSAR Model: SVMlight (version 1.0.0)
Select Descriptor file of Active molecules:
11: PaDEL1 on data 4
Descriptor file should be in CSV format
Select Descriptor file of Inactive molecules:
12: PaDEL1 on data 2
Descriptor file should be in CSV format
Execute
Select Active descriptor file
Submit
Select inactive descriptor file
History
mol33 in file.sdf (3/13). Processing mol55 in file.sdf (5/13). Processing mol66 in file.sdf (6/13). Processing mol77 in file.s
11: PaDEL1 on data 4
12 lines
format: txt, database: ?
Processing Mol1 in file.sdf (1/11). Processing mol2 in file.sdf (2/11). Processing mol3 in file.sdf (3/11). Processing mol4 in file.sdf (4/11). Processing mol6 in file.sdf (6/11). Processing mol5 in file.sdf (5/11). Processing mol7 in file.sdf (7/1
6: Unknown.sdf
4: Activesmile.sdf
2: Inactivesmile.sdf


Tools
Target Library
Compound Library
DATA PROCESSING
File-Format Converter
Descriptor Calculation
PaDEL "PaDEL Descriptor Tool"
CDKDescriptorCalculation "CDK Tool"
DATA ANALYSIS
QSAR
Build QSAR Model builds QSAR model using McQSAR
Predict Activity Using McQSAR Using already built QSAR model
Convert csv to arff Converter csv to arff file in weka
Convert arff to csv Converter arff to csv file in weka
Convert csv to arff Converter csv to arff file in weka
Filter Filters in weka
wekatool Data mining software in java weka
WekaEvaluate Evalutation in weka
Build QSAR Model: SVMlight create model using SVMlight
Classify Data :SVMlight classify data using model given by SVM
SVM-light Version V6.02
SVM Model build based on descriptor values of Active and inactive compounds
History
13: Build QSAR Model: SVMlight on data 11 and data 12
33 lines
format: model, database: ?
Scanning examples...done
Reading examples into memory...OK. (24 examples read)
Setting default regularization parameter C=0.0001
Optimizing.....done. (13 iterations) Optimization finished (11 misclassified, maxdiff=0.00000). Runtime in cpu-seconds
12: PaDEL1 on data 2
11: PaDEL1 on data 4
12 lines
format: txt, database: ?
Processing Mol1 in file.sdf (1/11). Processing mol2 in file.sdf (2/11). Processing mol3 in file.sdf (3/11). Processing mol4 in file.sdf (4/11). Processing mol6 in file.sdf (6/11). Processing mol5 in file.sdf (5/11). Processing mol7 in file.sdf (7/1


Galaxy / MPDS 1.0.1
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Tools
Classify Data :SVMlight (version 1.0.0)
Select Descriptor file of Unknown Activity :
13: Build QSAR Model: SVMlight on data 11 and data 12
Descriptor file should be in CSV format
Select model file:
13: Build QSAR Model: SVMlight on data 11 and data 12
A model file created using SVMlight
Execute
SVMlight Model built in earlier step
Descriptor file of unknow activity molecules
Classify data as active and inactive using SVMlight Classify
History
14: PaDEL1 on data 6
13: Build QSAR Model: SVMlight on data 11 and data 12
33 lines
format: model, database: ?
Scanning examples...done
Reading examples into memory...OK. (24 examples read)
Setting default regularization parameter C=0.0001
Optimizing.....done. (13 iterations) Optimization finished (11 misclassified, maxdiff=0.00000). Runtime in cpu-seconds
12: PaDEL1 on data 2
11: PaDEL1 on data 4
12 lines
format: txt, database: ?
Processing Mol1 in file.sdf (1/11).
Processing mol2 in file.sdf (2/11).
Processing mol3 in file.sdf (3/11).
Processing mol4 in file.sdf (4/11).
Processing mol6 in file.sdf (6/11).
Processing mol5 in file.sdf (5/11).

Galaxy / MPDS 1.0.1
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Tools
-0.63461464
-0.6653034
-0.60372208
-0.61731627
-0.66075651
-0.6523281
-0.66723847
-0.64865168
-0.64686105
-0.63769256
-0.60892557
History
SVMlight
302.3 KB
16: Classify Data :SVMlight on data 14 and data 13
11 lines
format: txt, database: ?
Reading model...OK. (22 support vectors read) Classifying test examples..done Runtime (without IO) in cpu-seconds: 0.00 Accuracy on test set: 100.00% (11 correct, 0 incorrect, 11 total)
Precision/recall on test set: -nan%/-nan%
14: PaDEL1 on data 6
13: Build QSAR Model: SVMlight on data 11 and data 12
33 lines
format: model, database: ?
Scanning examples...done
Reading examples into memory...OK. (24 examples read)
Setting default regularization parameter C=0.0001
Optimizing.....done. (13 iterations) Optimization finished (11 misclassified,

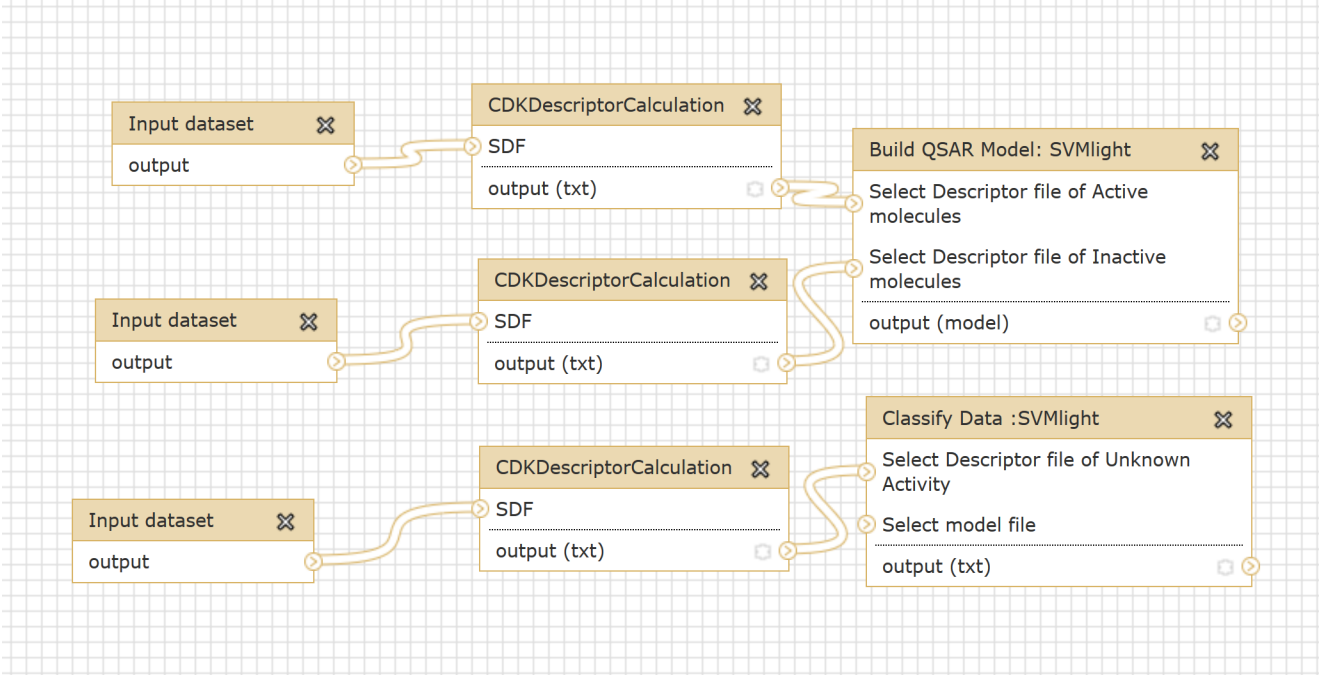
Figure 11

NOTE: *All data used for this demo purpose in this module can be downloaded from link given below.*

Data

https://drive.google.com/file/d/0B3c9isKbTnxtN2l1U1ZwVE03VVU/view?usp=sharing

Complete workflow of Cdk-SVMlight QSAR model building and classification.



5.2. Module 7- Docking

5.2.1. Optimize Ligand

Ligand: cdx, sdf, mol, mol2, smi, pdb

Step 1: Upload your ligand file from Get Data (fig.1).

Step 2: (a) Go to Module 7: Molecular Docking Protein -Ligand Interaction (fig 22.1) and then click optimize input ligand.

(b) Fill data and select files fig. (22.2)

Step 3: Results: In the history panel of MPDS home page user can see the jobs completed and can download results (fig 23). Output formats for optimization are: sdf, mol, mol2, and pdb.

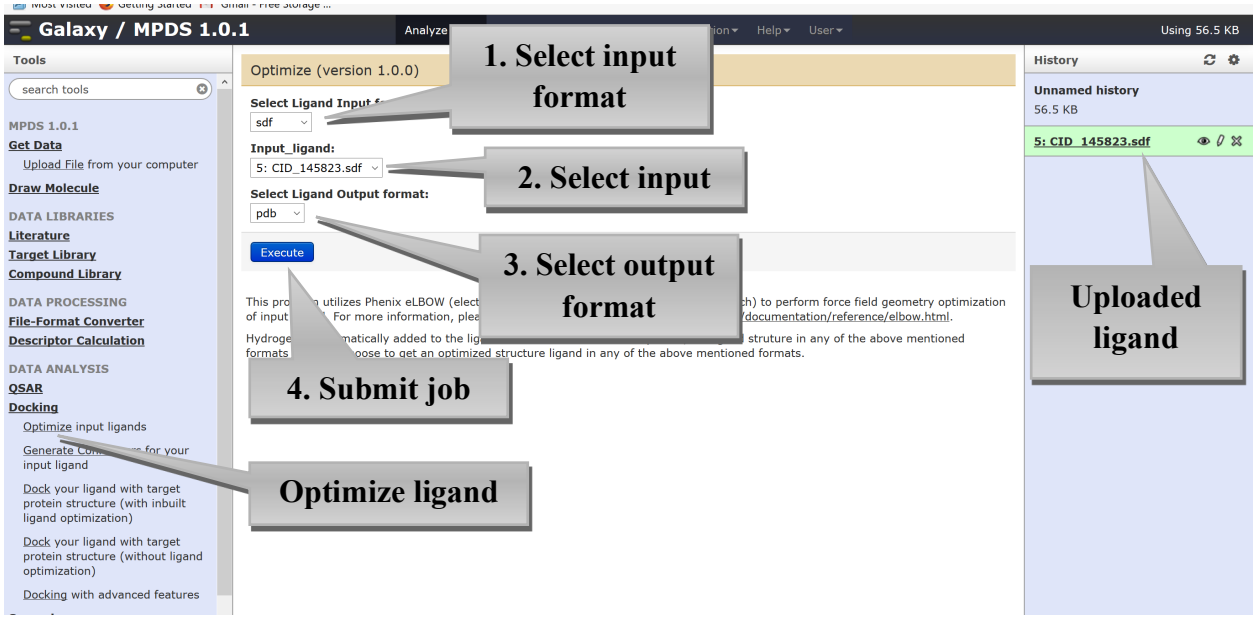
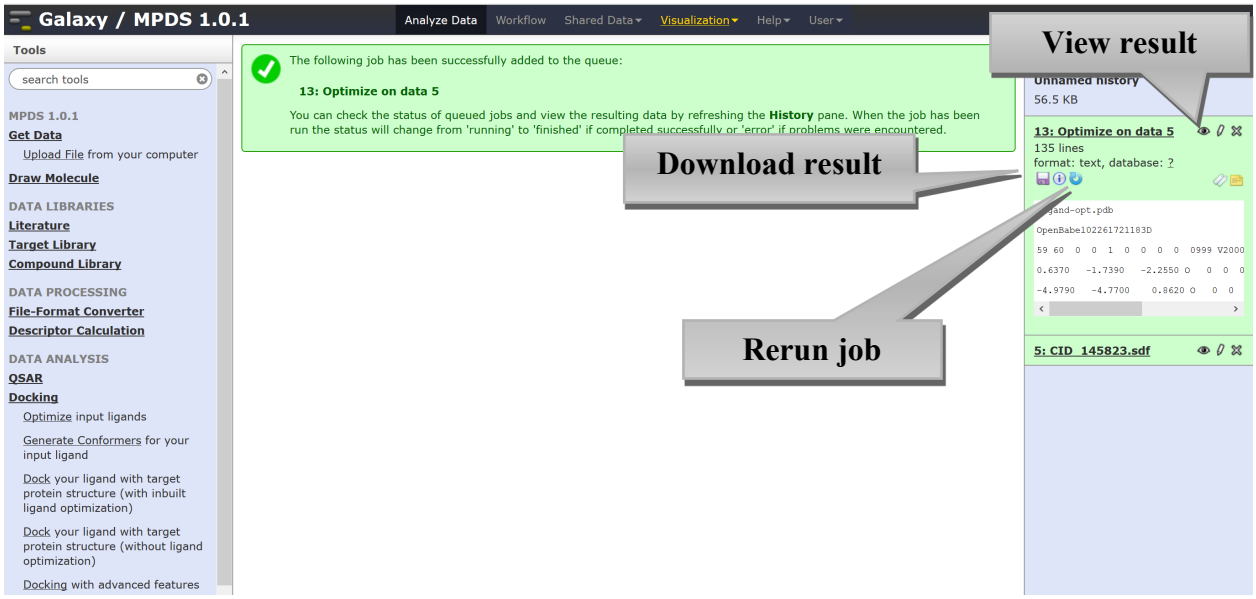


Figure22.2



5.2.2. Generate Conformers

Ligand: sdf

Step 1: Upload your ligand file from Get Data (fig.1).

Step 2: Go to Module 7: Molecular Docking Protein -Ligand Interaction (fig 24) and then click Generate Conformers for input ligand.

Step 3: Results: In the history panel of MPDS home page user can see the jobs completed and can download results.

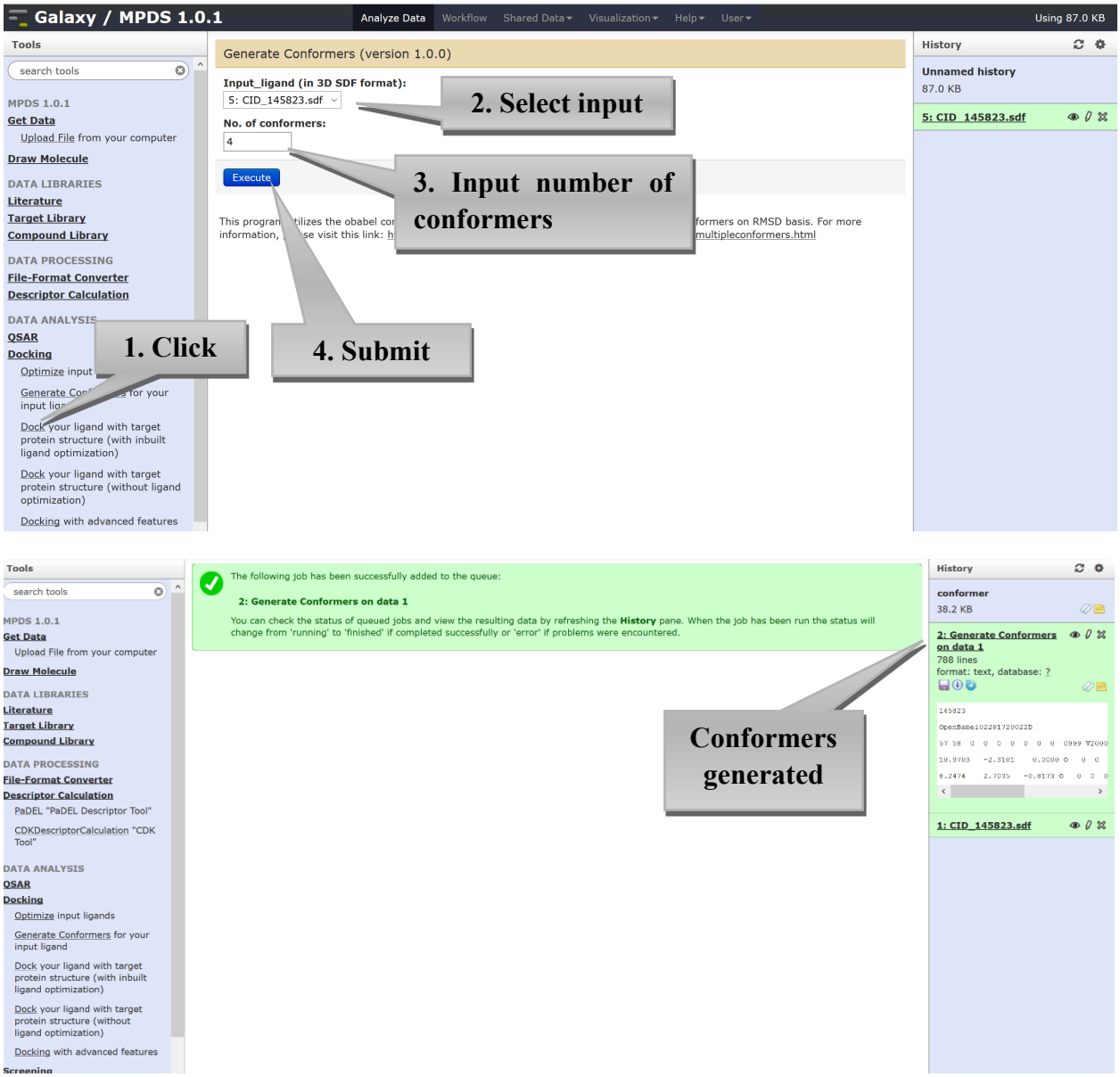


Figure 24

5.2.3. Molecular Docking : Dock your ligand with target protein structure (with inbuilt ligand optimization

The Autodock Vina is used in this program to simulate the complex formation between a receptor protein and a small molecule (ligand). Docking ligand with target protein structure with inbuilt ligand optimization is slower docking as it takes extra time for ligand optimization.

•Receptor: .pdb
•Ligand: .pdb, .sdf

Step 1: Upload your ligand and receptor files one by one from Get Data (fig1).

Step 2:

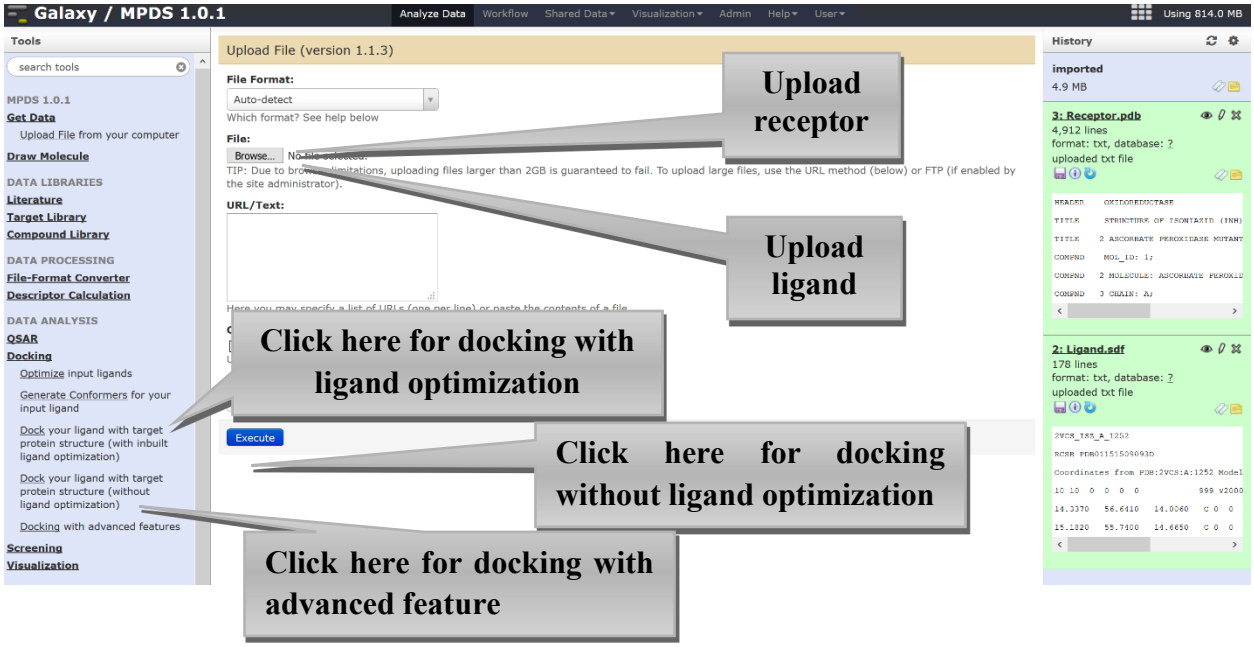


Figure 25

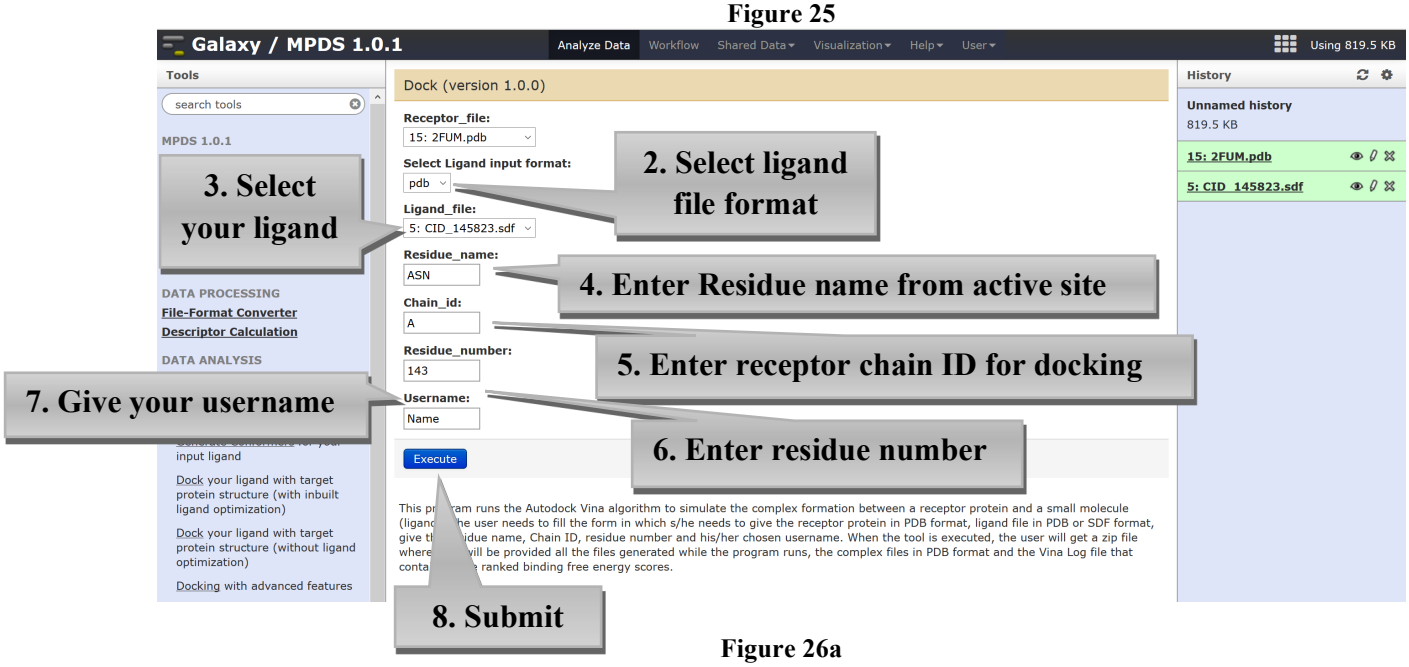


Figure 26a

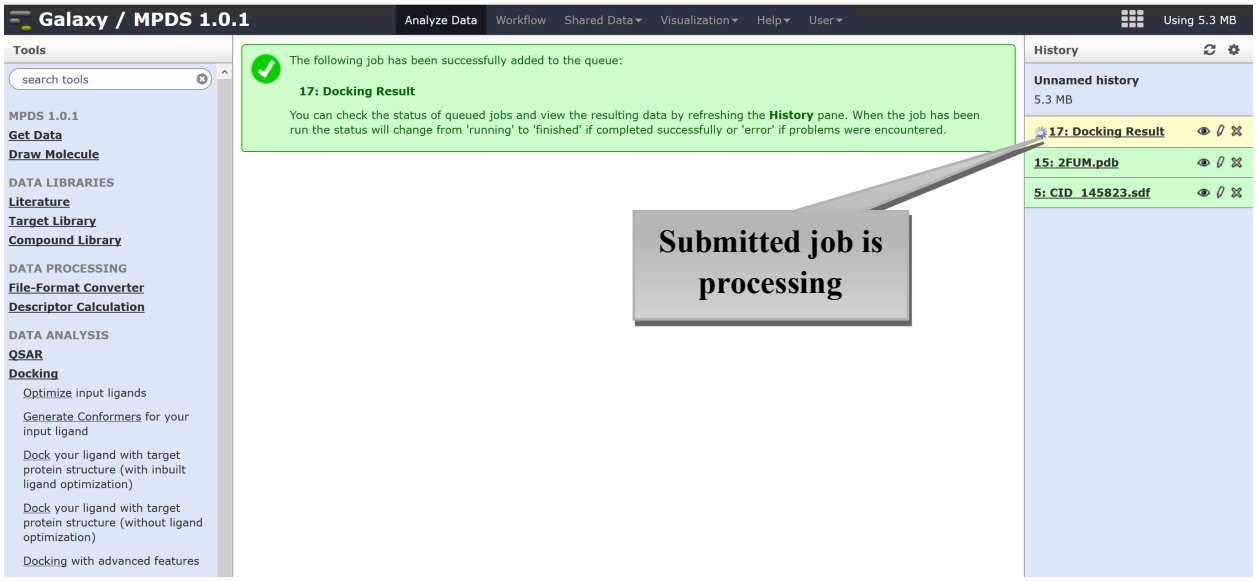


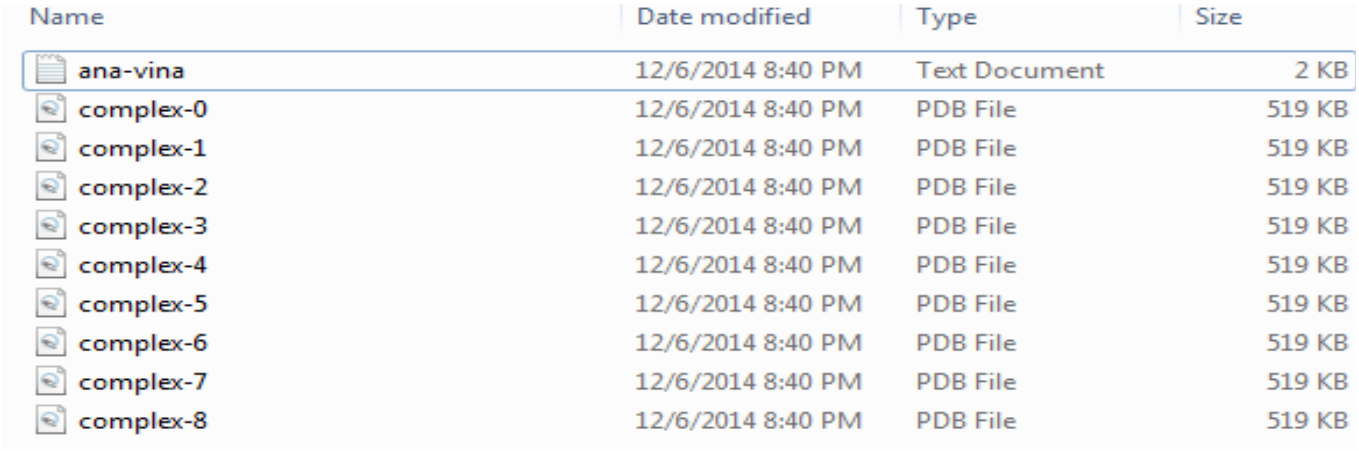
Figure 26b

Step 4: Results: If your job has been successfully completed (fig.7), then the submitted docking result will be in green color or if some error is found then the result will be displayed in red color along with the details of error found.

Figure 27

Output of Docking

The results extracted from zip file (fig. 8) can be analyzed using visualization module of MPDS.



Name	Date modified	Type	Size
ana-vina	12/6/2014 8:40 PM	Text Document	2 KB
complex-0	12/6/2014 8:40 PM	PDB File	519 KB
complex-1	12/6/2014 8:40 PM	PDB File	519 KB
complex-2	12/6/2014 8:40 PM	PDB File	519 KB
complex-3	12/6/2014 8:40 PM	PDB File	519 KB
complex-4	12/6/2014 8:40 PM	PDB File	519 KB
complex-5	12/6/2014 8:40 PM	PDB File	519 KB
complex-6	12/6/2014 8:40 PM	PDB File	519 KB
complex-7	12/6/2014 8:40 PM	PDB File	519 KB
complex-8	12/6/2014 8:40 PM	PDB File	519 KB

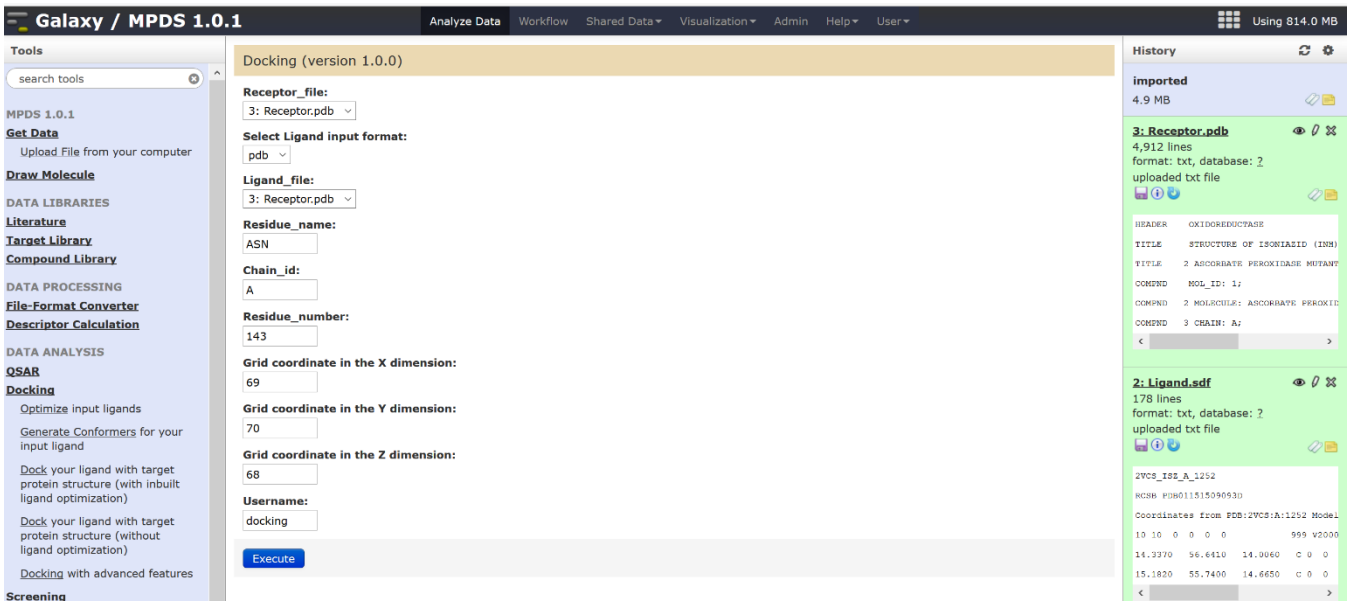
Figure 29

5.2.4. Molecular Docking: Dock your ligand with target protein structure (without ligand optimization)

Refer to 5.2.3. section

5.2.5. Molecular Docking with advanced features:

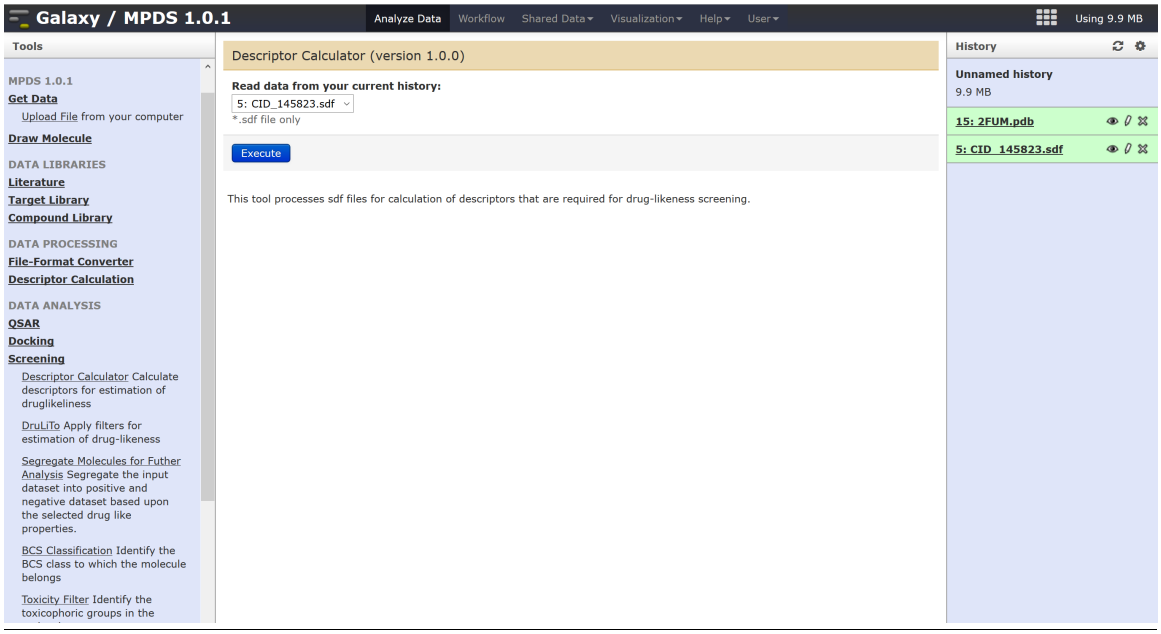
This sub-module of docking provides advanced feature of defining X, Y, Z axis of grid box required in docking. In the previous docking submodules (sections 5.2.3. and 5.2.4) this feature was set to default i.e. not user defined. All other steps can be followed as per given 5.2.3.

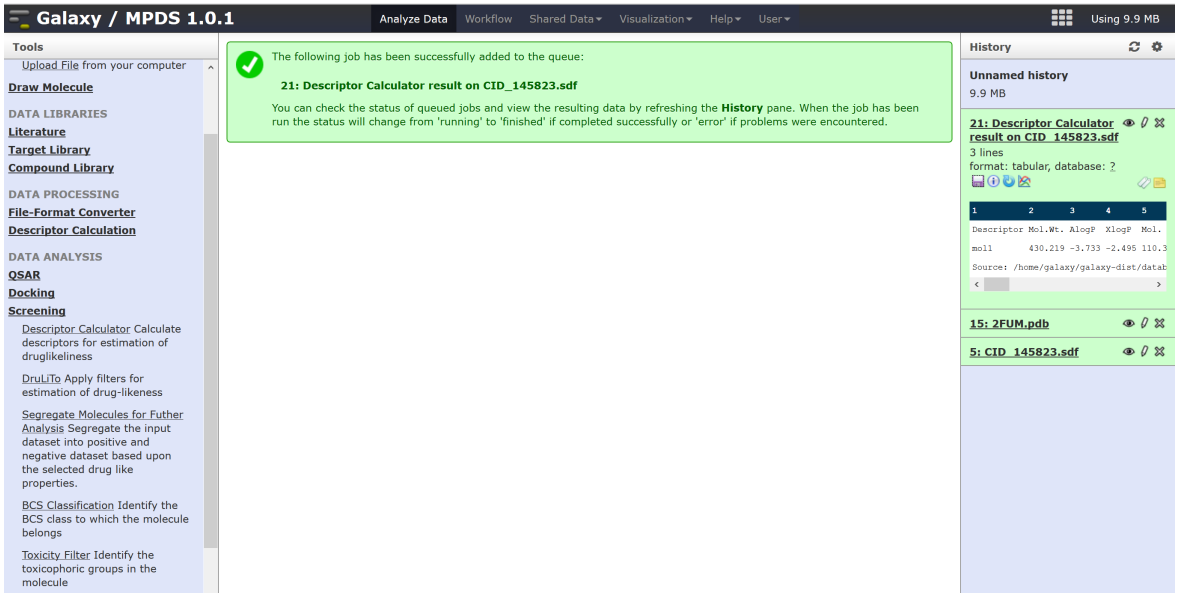


5.3. Screening

5.4.1. Descriptor Calculator

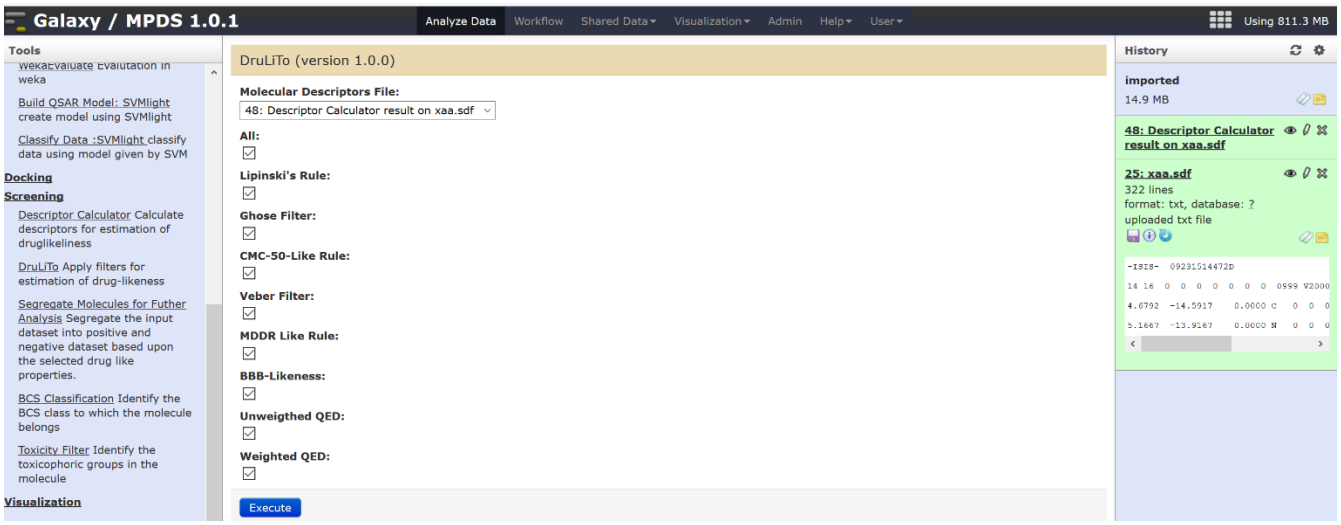
It Calculate descriptors for estimation of drug likeliness

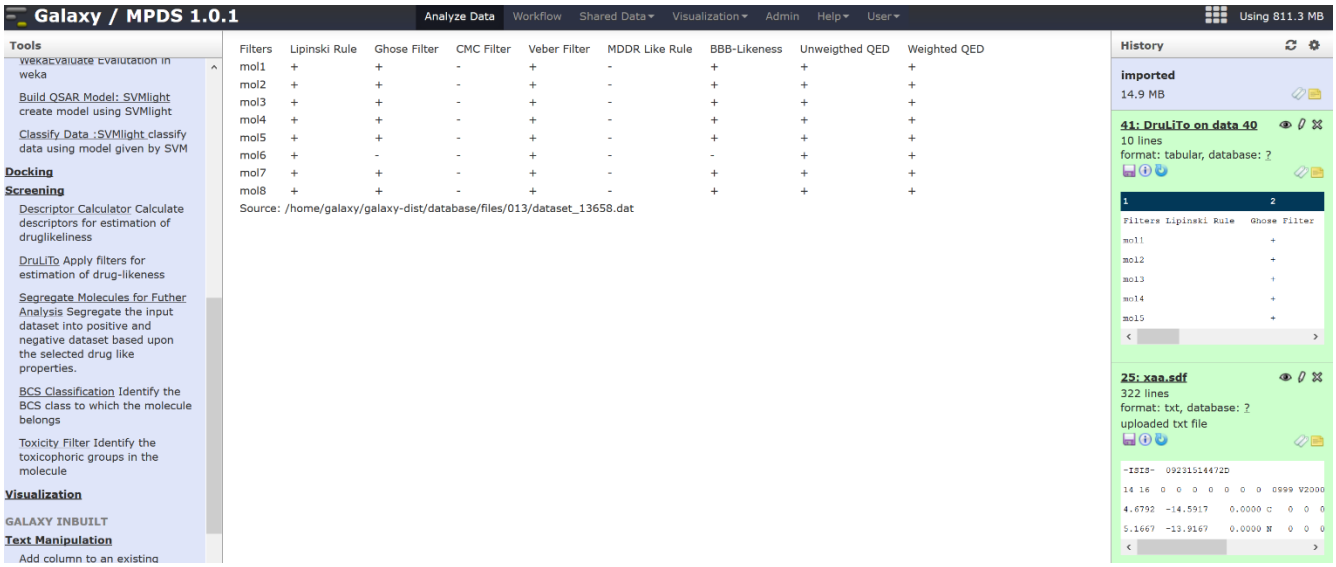




5.4.2. DruLiTo :

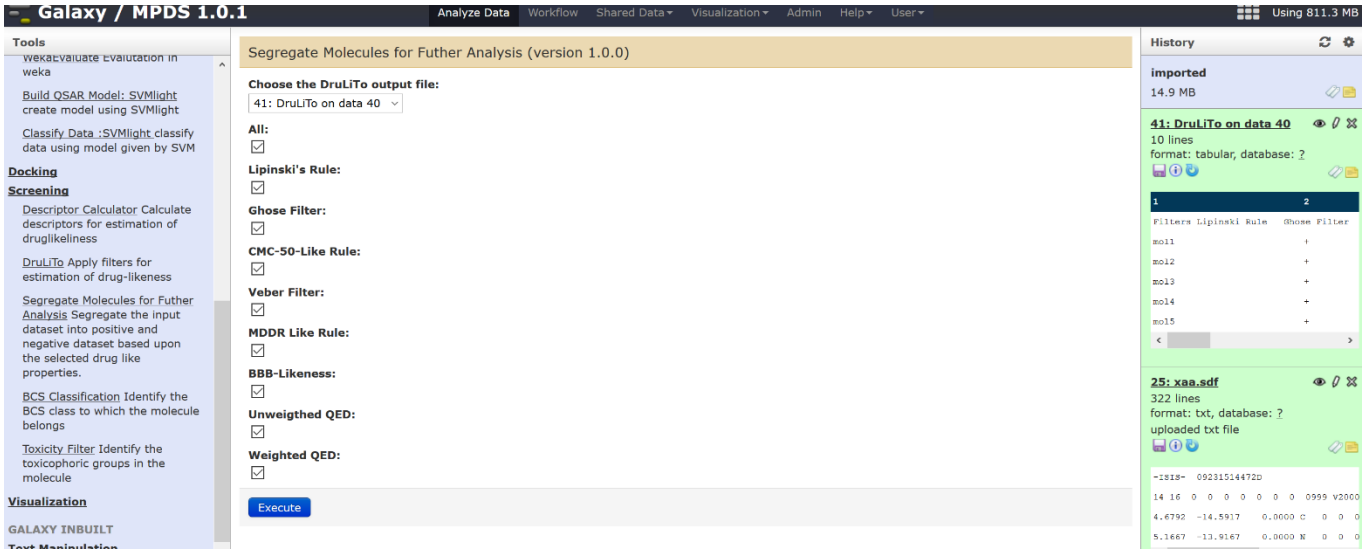
It applies filters for estimation of drug-likeness

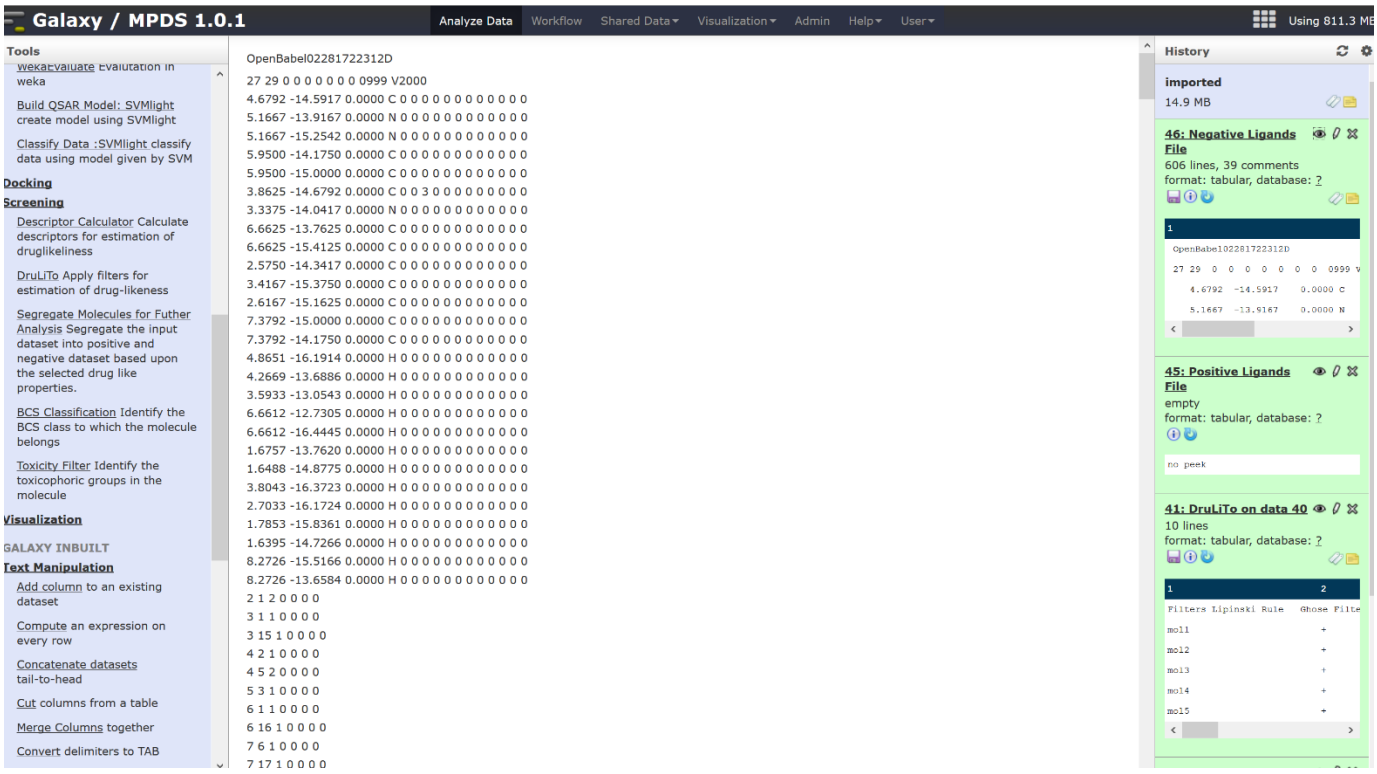




5.4.3. Segregate Molecules

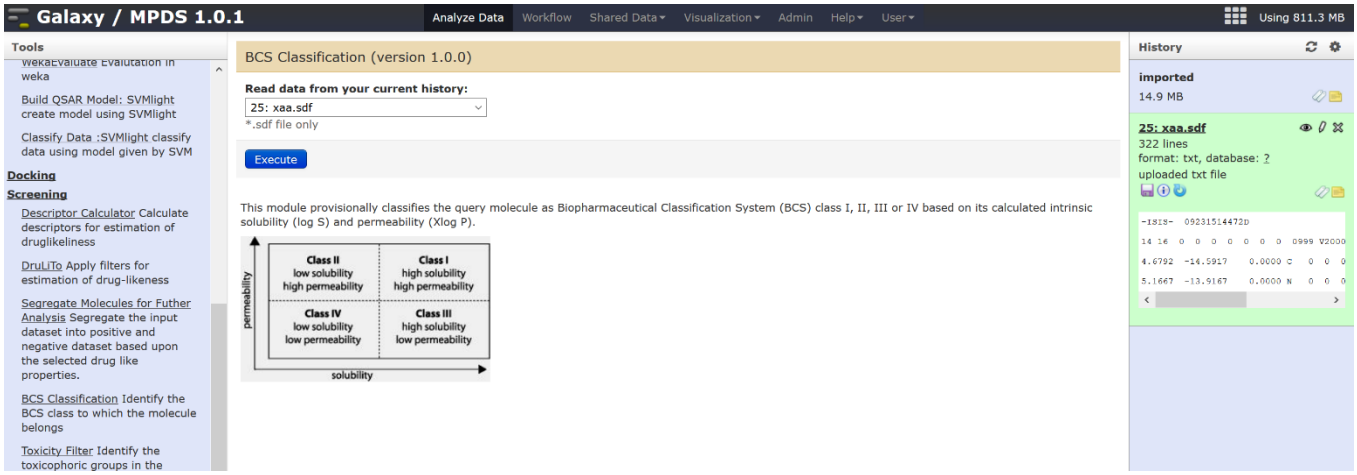
Segregate the input dataset into positive and negative dataset based upon the selected drug like properties.

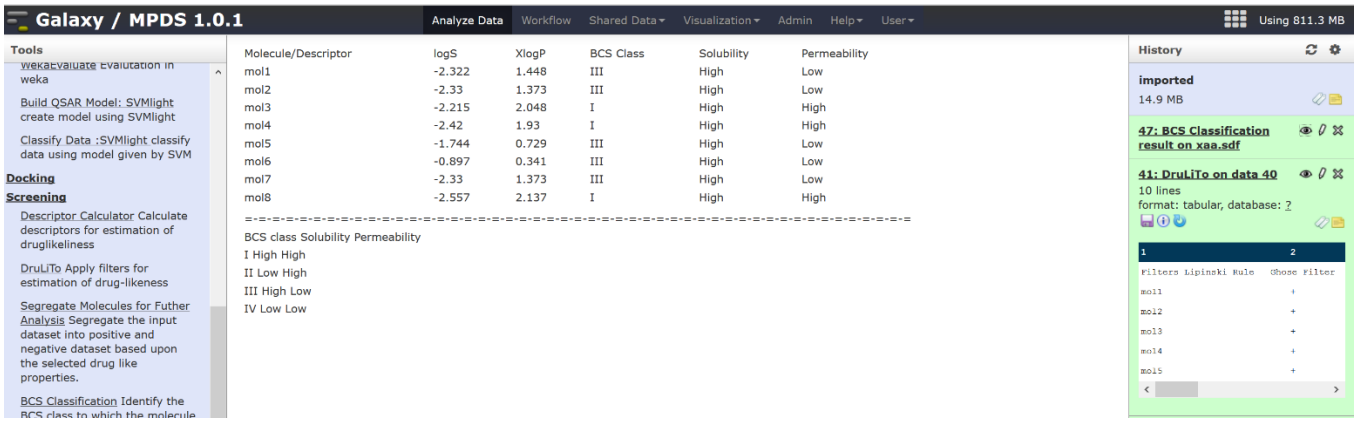




5.4.3. BCS Classification

Identify the BCS class to which the molecule belongs





5.4.4. Toxicity Filter

Identify the toxicophoric groups in the molecule

Toxicity Filter (version 1.0.0)

Read data from your current history:

2: Structure3D_CID_3767.sdf

*.sdf file only

Title for the output file - to remind you what the job was for:

MPDS_ToxFiltResults

Non alphanumeric characters will be trimmed

Execute

MPDS Toxicity Filter - Help Page

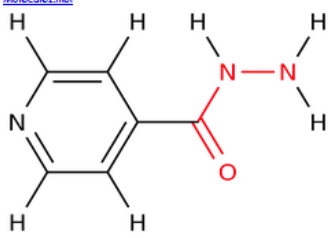
This tool identifies and highlights the structural alerts or unwanted toxicophoric moieties (Brenk, Ruth, et al. ChemMedChem 3.3 (2008)) in the submitted query molecule and renders a downloadable image and summary file.

The complete set of results of the processed dataset can be downloaded as a compressed file using the link (Download All Results Here) on the page. The file named "MPDS_ToxFilterResults_summary.txt" (default output file name) present in the folder provides a summary of results in a text format for all the molecules processed from the input dataset. This file contains the serial number of the molecule, the structural alerts (if present), and the number of times a specific alert occurred in the target molecule ("Occurrence count"). For molecules devoid of any structural alert, "No structural alerts found!" message would be displayed.

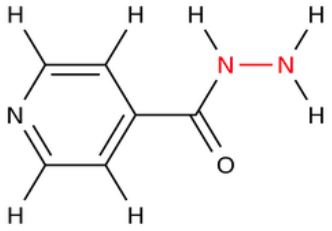
Example

static/images/filters_MPDS/toxicity.png

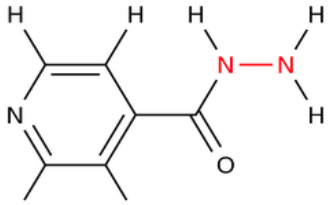
Molecule1.mol



Molecule1.mol



Molecule1.mol



>>Download All Results Here<<

```
################################################
#       Summary of Toxicity Filter results:    #
#       Date: Thu May 26 10:00:46 IST 2016     #
################################################

Molecule 1
Structural Alert found: acyl_hydrazine (C(=O)N[NH2])
Occurrence count: 1
Structural Alert found: hydrazine (N[NH2])
Occurrence count: 2
Structural Alert found: Oxygen-nitrogen_single_bond ([OR0,NR0][OR0,NR0])
Occurrence count: 2
```

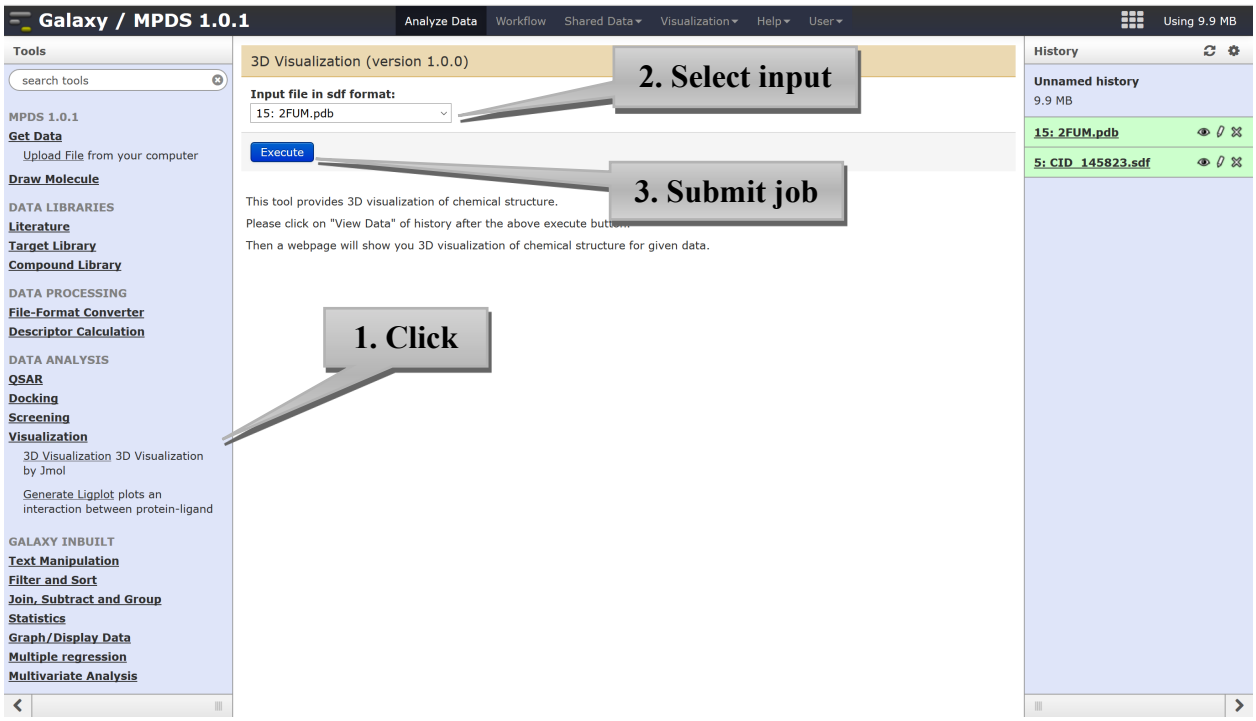
5.6. Visualization

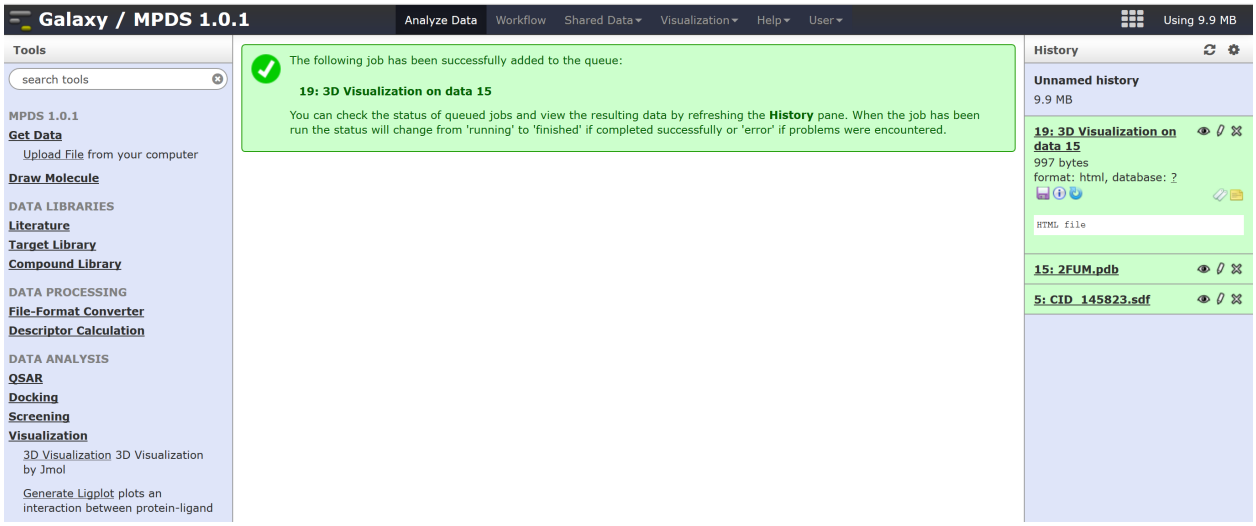
5.6.1. 3D Visualization by Jmol

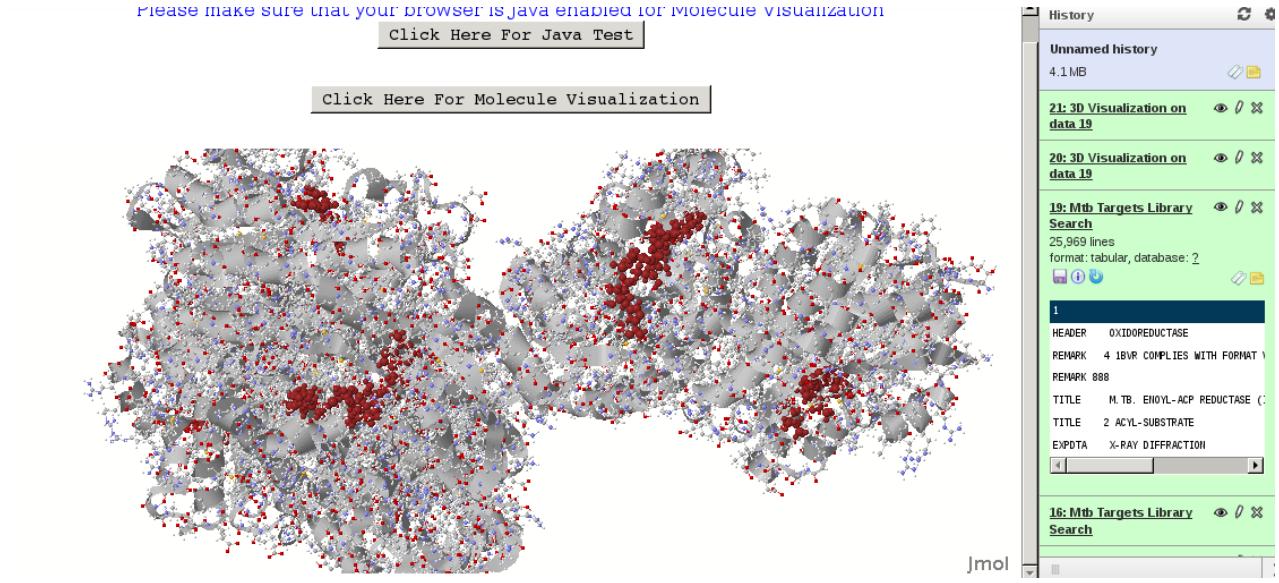
Step1: Upload your ligand file from Get Data (fig.1)

Step 2: Go to Module 9: Visualization and then click 3D Visualization by Jmol (Fig.32).Select the required file and execute.

Step 3: Results: In the history panel of MPDS home page user can see the jobs completed and download results. Generated 3D image is shown below in Fig. 33




Galaxy / MPDS 1.0.1
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Using 9.9 MB
Tools
search tools
MPDS 1.0.1
Get Data
Upload File from your computer
Draw Molecule
DATA LIBRARIES
Literature
Target Library
Compound Library
DATA PROCESSING
File-Format Converter
Descriptor Calculation
DATA ANALYSIS
QSAR
Docking
Screening
Visualization
3D Visualization 3D Visualization by Jmol
Generate Ligplot plots an interaction between protein-ligand
The following job has been successfully added to the queue:
19: 3D Visualization on data 15
You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.
History
Unnamed history
9.9 MB
19: 3D Visualization on data 15
997 bytes
format: html, database: ?
HTML file
15: 2FUM.pdb
5: CID_145823.sdf


Please make sure that your browser is java enabled for Molecule Visualization
Click Here For Java Test
Click Here For Molecule Visualization
Jmol
History
Unnamed history
4.1 MB
21: 3D Visualization on data 19
20: 3D Visualization on data 19
19: Mtb Targets Library Search
25,969 lines
format: tabular, database: ?
1
HEADER OXIDOREDUCTASE
REMARK 4 1BVR COMPLIES WITH FORMAT
REMARK 888
TITLE M.TB. ENOYL-ACP REDUCTASE
TITLE 2 ACYL-SUBSTRATE
EXPDTA X-RAY DIFFRACTION
16: Mtb Targets Library Search

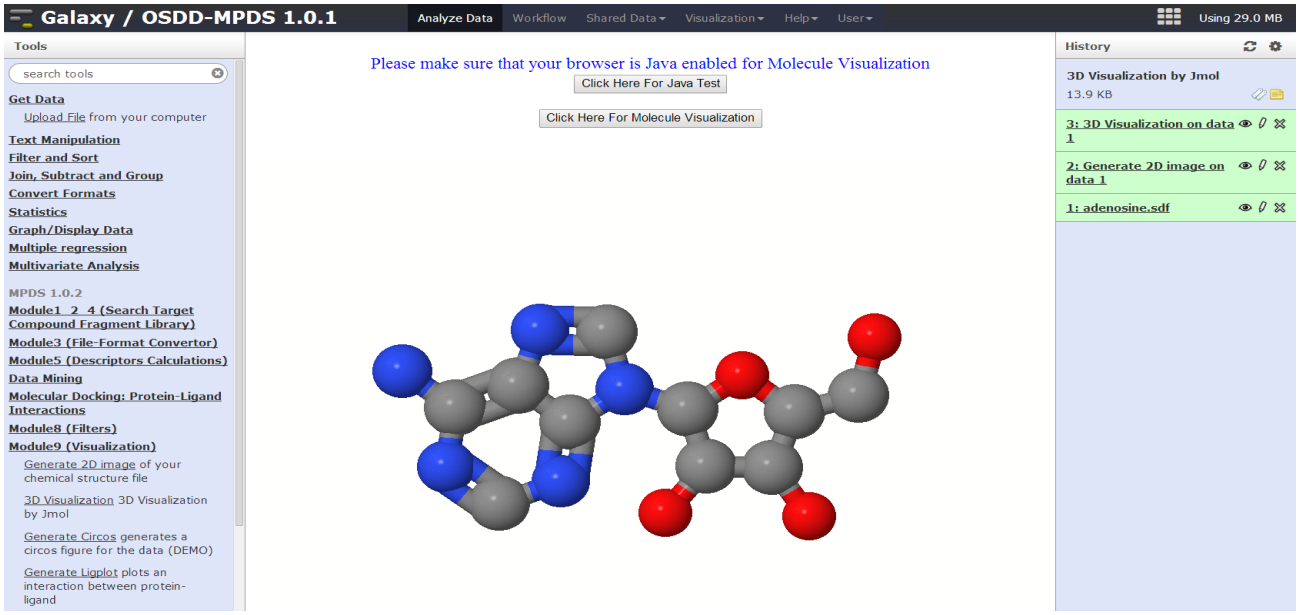


Figure 33

5.6.2. Generate Ligplot plots an interaction between protein-ligand

Input: .pdb file

Step1: Upload your ligand file from Get Data (as in Fig.1).

Step 2: Go to Module 9: Visualization and then click Generate Ligplot. Enter Residue id and Chain id (fig 34).

Step 3: Results: In the history panel of MPDS home page user can see the jobs completed and download results. Generated Ligplot is shown below in fig.35

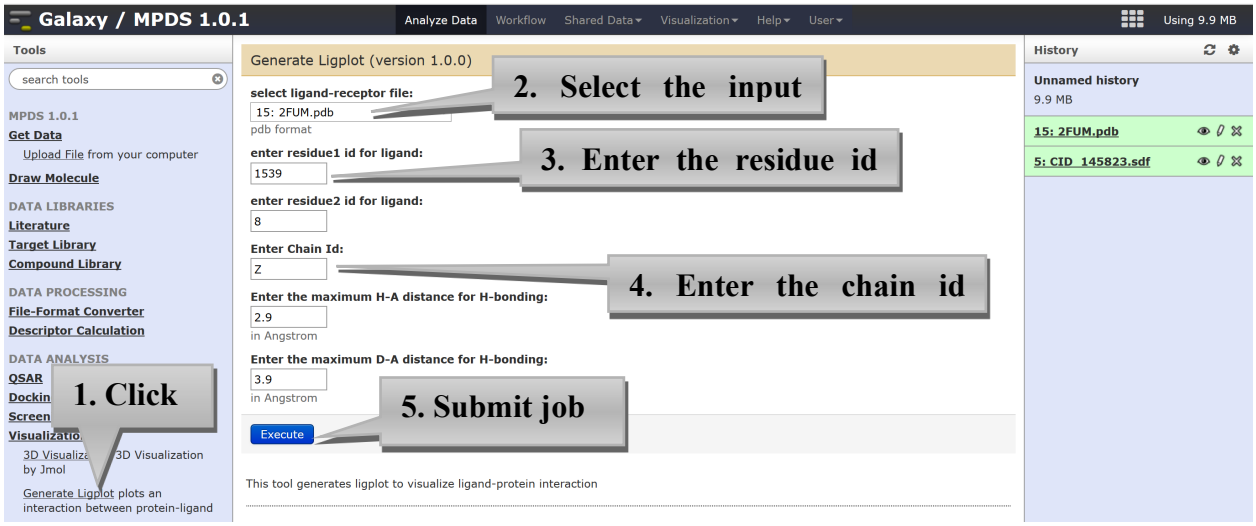


Figure 34

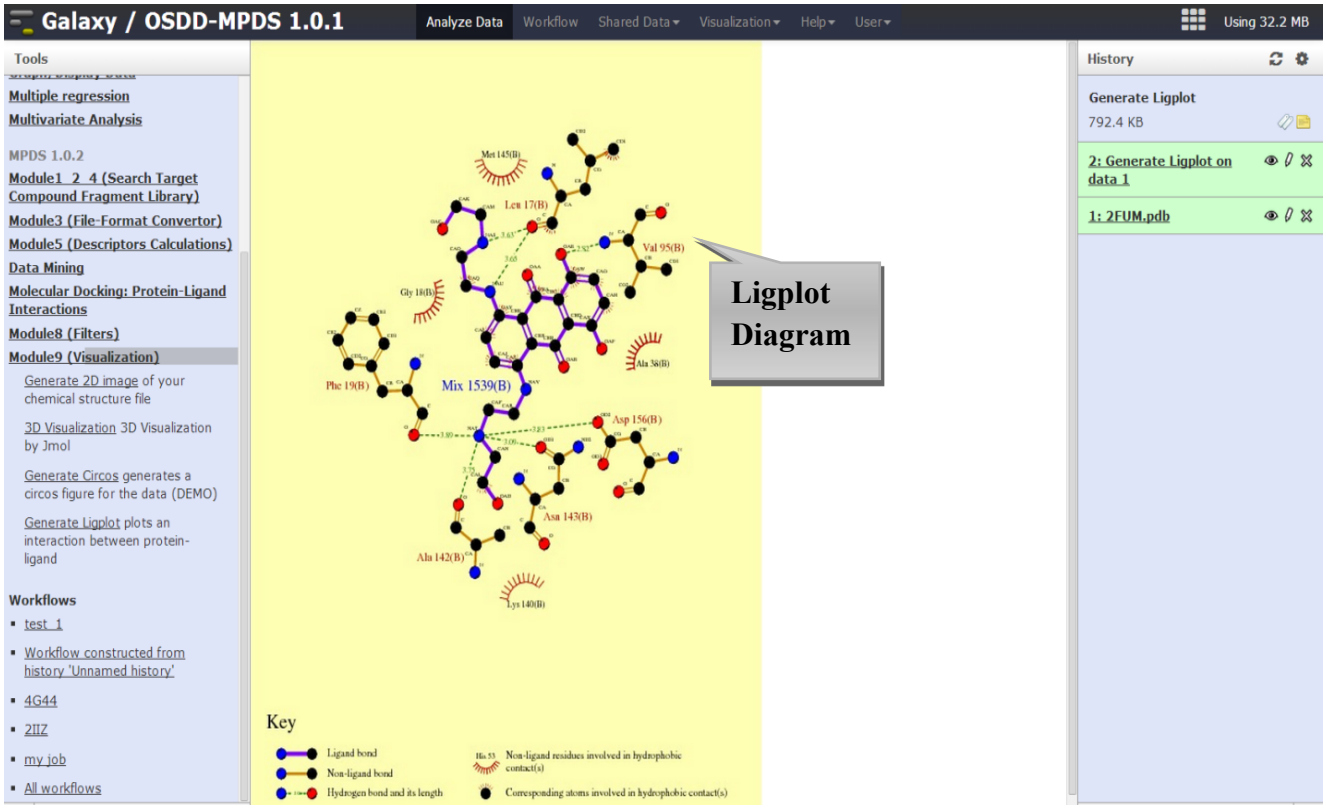


Figure 36