

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

1. Introduction	5
2. MPDS: Upload Data	6
4. 2.1 Get Data	6
5. 2.2 Draw Molecule	6
2. Data Libraries	7
3.1. Module 1-Literature	7
3.2. Module 2-Target Library	7
3.3. Module 2- Compound Library	8
3.3.1. Database ID Search:	8
3.3.2. Exact Structure Search:	10
3.3.3. Sub-structure Search:	11
3.3.4. Molecular Property-based Search:	12
3.3.5. Fingerprint-based Search:	13
3.3.6. Molecule cloud:	14
3.3.7. Library Generator:	15
4. Data Processing	16
4.1. Module 4- File Format Convertor	16
4.1.1 Converter	16
4.1.2. Generate 3D coordinates	17
4.2. Descriptor Calculator	18
4.2.1. PaDEL Descriptor Calculator	18
4.2.2. CDK Descriptor Calculator	18
5. Data Analysis	20
5.1. QSAR	20
5.1.1. QSAR Model Building using McQSAR	20
5.1.2. QSAR Model Building using Weka	26
5.1.3. QSAR Model Building using SVMlight	26

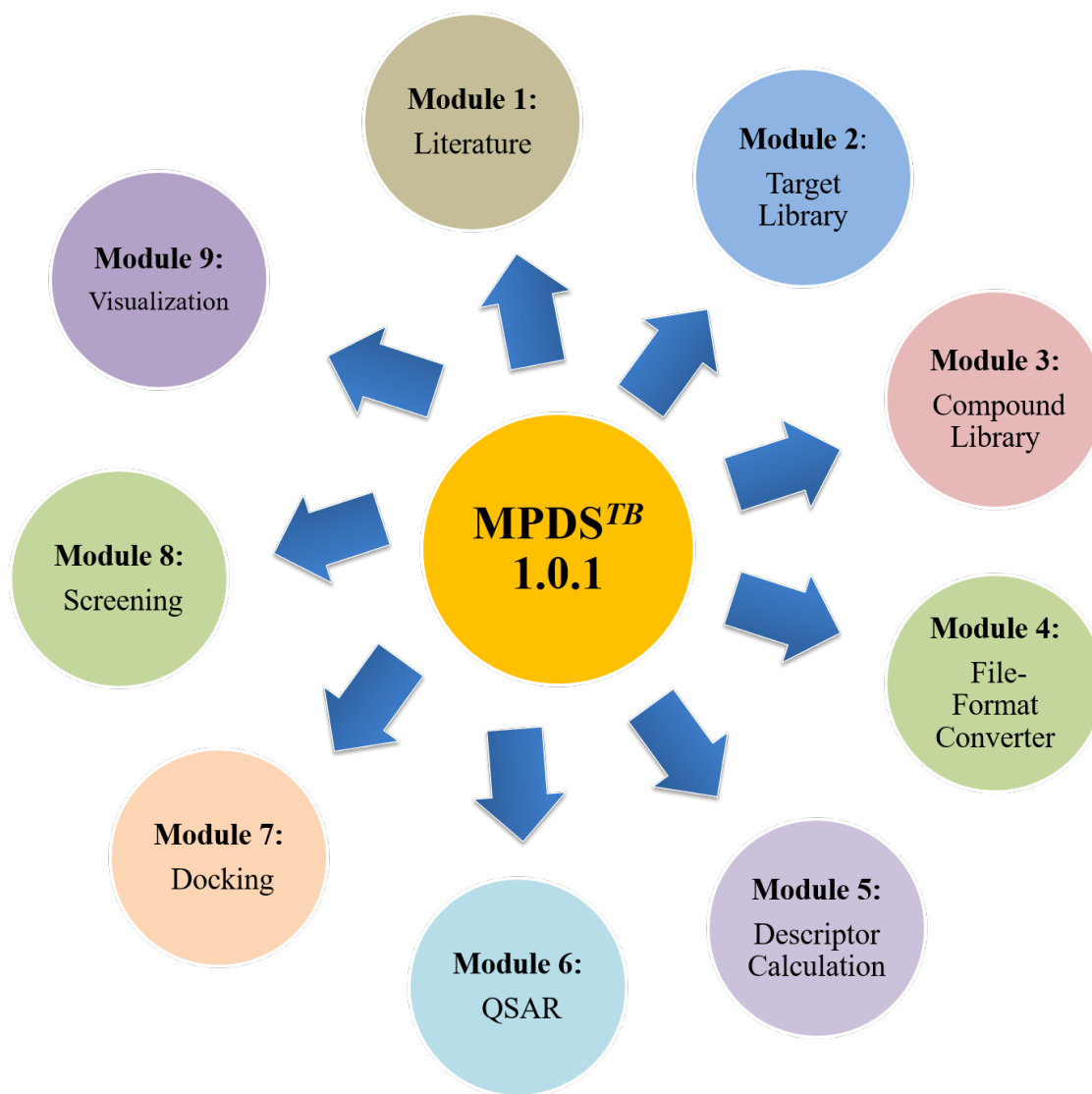
5.2. Module 7- Docking.....	30
5.2.1. Optimize Ligand.....	30
5.2.2. Generate Conformers.....	31
5.2.3. Molecular Docking : Dock your ligand with target protein structure (with inbuilt ligand optimization).....	32
5.2.4. Molecular Docking: Dock your ligand with target protein structure (without ligand optimization)	35
5.2.5. Molecular Docking with advanced features	35
5.3. Screening.....	36
5.4.1. Descriptor Calculator	36
5.4.2. DruLiTo :.....	37
5.4.3. Segregate Molecules	38
5.4.3. BCS Classification	39
5.4.4. Toxicity Filter.....	40_Toc476085384
5.6. Visualization	41
5.6.1. 3D Visualization by Jmol.....	41
5.6.2. Generate Ligplot plots an interaction between protein-ligand.....	43

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-likeness 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).

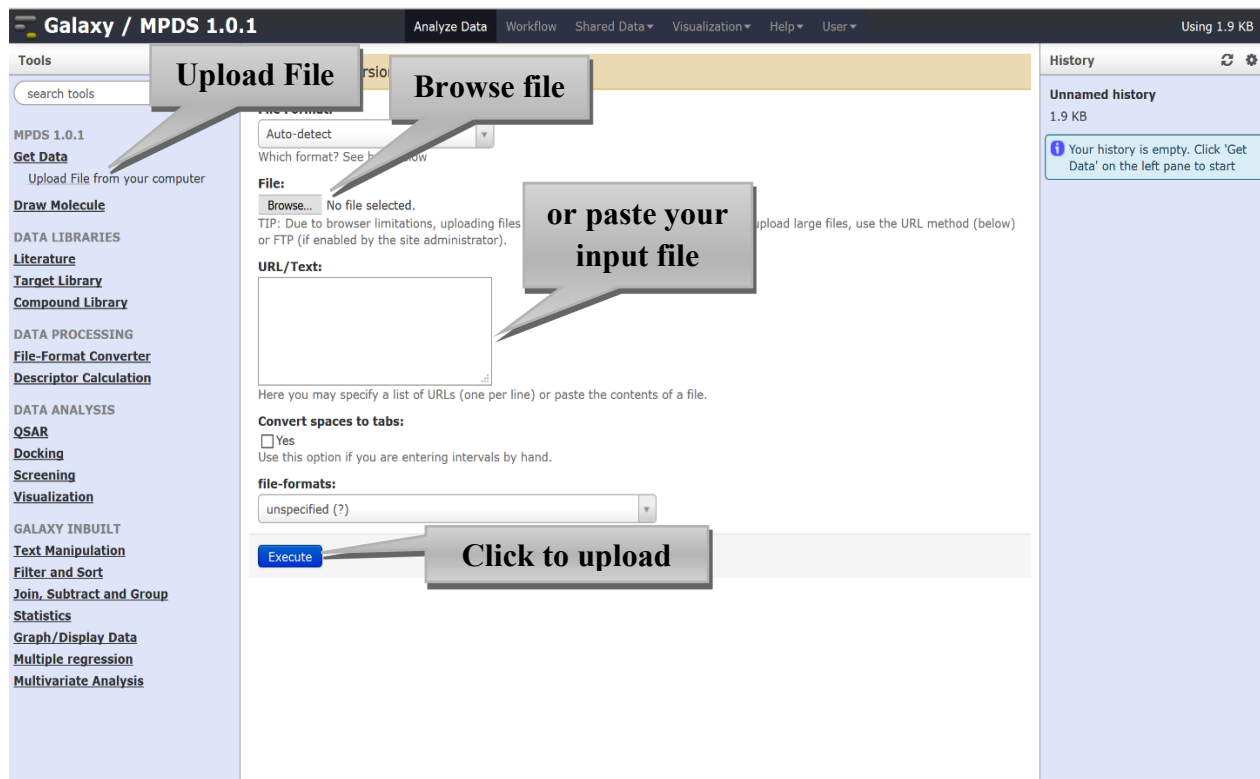


Figure 1a. Upload your input file (ligand, receptor).

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).

Galaxy / MPDS 1.0.1

Analyze Data Workflow Shared Data Visualization Help User

Using 12.3 MB

Tools

search tools

MPDS 1.0.1

Get Data

Draw Molecule

JMol Editor A Chemical Structure Editor

DATA LIBRARIES

Literature

Target Library

Compound Library

DATA PROCESSING

File-Format Converter

Descriptor Calculation

DATA ANALYSIS

QSAR

Docking

Screening

Visualization

Draw your structure and add it to MPDS Galaxy!

Molecule Editor:

H
C
N
O
S
P
F
Cl
Br
...

Export Mofile to Galaxy Export Smiles to Galaxy

Mol. Weight: 177.2423 Formula: C₁₁H₁₅NO

History

Unnamed history

12.3 MB

30: imported SMILES file

1 line

format: smi, database: ?

C(C=CC1) (=C (N=1) CC2O) C (C2) CC

29: imported mol file

30 lines

format: mol, database: ?

JSDraw02281712072D

13	14	0	0	0	0	0	0	0	0	1	V2000
11.	1280		-6.7600		0.0000		C	0	0		
9.	7770		-5.9800		0.0000		C	0	0		

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

Galaxy / MPDS 1.0.1

Analyze Data Workflow Shared Data Visualization Help User

Tools

search tools

MPDS 1.0.1

Get Data

Upload File from your computer

Draw Molecule

DATA LIBRARIES

Literature

Target Library

Mtb Targets Library Search

Download Mtb Targets in PDB format

Compound Library

DATA PROCESSING

File-Format Converter

Descriptor Calculation

DATA ANALYSIS

QSAR

Docking

Screening

Visualization

GALAXY INBUILT

Text Manipulation

Filter and Sort

Join, Subtract and Group

Statistics

Graph / Display Data

Multiple regression

Multivariate Analysis

Mtb Targets Library Search (version 1.0.0)

Enter PDB ID:

Execute

Input example : 1BVR, 4G44 etc. MPDS Database Search Web Page from galaxy Interface

This tools is still under development for automatic upload of files given by user.

--Thanks

1. click

2. Enter PDB ID

The screenshot shows the Galaxy / MPDS 1.0.1 interface. A green notification box at the top center states: "The following job has been successfully added to the queue: 22: Mtb Targets Library Search. 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." The History pane on the right shows a job named "22: Mtb Targets Library Search" with 6,686 lines of data in tabular format. A callout box points to the job with the text "Download PDB ID or view".

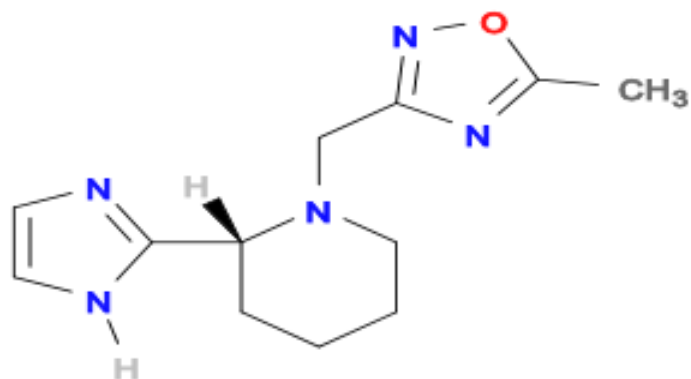
3.3. Module 2- Compound Library

3.3.1. Database ID Search:

The screenshot shows the Galaxy / MPDS 1.0.1 interface for the "Database Id Search (version 1.0.0)" tool. The "Select database:" dropdown is set to "MPDS Compound Library". The "Enter MPDS ID:" field contains the value "26-01-100524". A callout box points to this field with the text "MPDS ID". A green arrow points from the "MPDS ID" callout box to the "MPDS ID" field. The left sidebar shows the "Compound Library" section with various search options.

Molecular Property Diagnostic Suite

MPDS ID: 26-01-100524



Molecular Formula:

C₁₂H₁₇N₅O

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:

Galaxy / MPDS 1.0.1 | Analyze Data | Workflow | Shared Data | Visualization | Help | User | Using 12.3 MB

Tools

search tools

MPDS 1.0.1

Get Data
Upload File from your computer

Draw Molecule

DATA LIBRARIES

Literature

Target Library

Compound Library
Database Id Search searches MPDS compound library using database ID
Exact-structure Search searches molecule in MPDS compound library
Sub-structure Search searches for sub-structure
Molecular Property Based Search perform simple or advance query on MPDS compound library data
Fingerprint Based Search searches using MPDS fingerprints
Molecule cloud generates molecule cloud
Library generator generates molecule based on composition

Exact-structure Search (version 1.0.0)

Select File containing structure:
32: Structure.sdf

Select input file format:
SDF
MOL
MOL2
SMILE

This tool is still under development.
--Thanks

History

Unnamed history
12.3 MB

32: Structure.sdf

Galaxy / MPDS 1.0.1 | Analyze Data | Workflow | Shared Data | Visualization | Help | User | Using 12.3 MB

Tools

search tools

MPDS 1.0.1

Get Data
Upload File from your computer

Draw Molecule

DATA LIBRARIES

Literature

Target Library

Compound Library
Database Id Search searches MPDS compound library using database ID
Exact-structure Search searches molecule in MPDS compound library
Sub-structure Search searches for sub-structure
Molecular Property Based Search perform simple or advance query on MPDS compound library data
Fingerprint Based Search searches using MPDS fingerprints
Molecule cloud generates molecule cloud
Library generator generates molecule based on composition

Molecular Property Diagnostic Suite

MPDS ID: 03-04-213023

Molecular Formula:
C₁₆H₂₃D₃N₂O₄

IUPAC Name:
1,1,2,2,2-pentadeuterioethyl (3R,4R,5S)-4-acetamido-5-amino-3-pentan-3-yloxyoxoloxene-1-carboxylate

Remarks:
Remarks here...

Name/Synonyms:
Name/Synonyms here...

Molecular Properties:

Mol. Wt.	312.20	LogP	-1.61
HBD	2	LogS	-1.51
HBA	6	pKa	pKa1: 14.03; pKa2: 9.31; pKa3: 9.31; pKa4: -1.65
Molar refractivity	80.17	Polar surface area	90.65
Heavy atoms count	22	Rings count	1.00
Rotatable bonds	14.00	Polarizability	1.77

*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.

History

Unnamed history
12.3 MB

38: Structure3.sdf
175 lines
format: txt, database: 7
uploaded txt file

3.3.3. Sub-structure Search:

Galaxy / MPDS 1.0.1 Analyze Data Workflow Shared Data Visualization Help User

Tools History

Sub-structure Search (version 1.0.0)

Select File containing sub-structure: 40: Structure.sdf

Select input format: SDF (selected), MOL, MOL2, SMILE

This tool is still under development.

MPDS 1.0.1 search tools

Set Data Upload File from your computer

Draw Molecule

DATA LIBRARIES

Literature

Target Library

Compound Library

Database Id Search searches MPDS compound library using database ID

Exact-structure Search searches molecule in MPDS compound library

Sub-structure Search searches for sub-structure

Molecular Property Based Search perform simple or advance query on MPDS compound library data

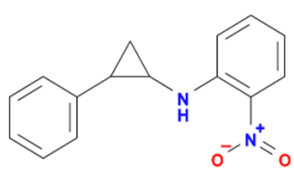
Fingerprint Based Search searches using MPDS fingerprints

Molecule cloud generates molecule cloud

Library generator generates molecule based on composition

History: Unnamed history (12.3 MB), 40: Structure.sdf

MPDS ID: 14-01-067628



Molecular Formula:
C₁₅H₁₄N₂O₂

IUPAC Name:
2-nitro-N-(2-phenylcyclopropyl)aniline

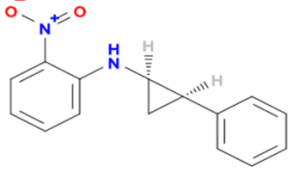
Remarks: Remarks here...

Name/Synonyms: Name/Synonyms here...

Molecular Properties:

Mol. Wt.	254.11	LogP	0.14
HBD	1	LogS	-5.74
HBA	1	pKa	pKa1: ; pKa2: ; pKa3: ; pKa4:
Molar refractivity	17.84	Polar surface area	55.17
Heavy atoms count	19	Rings count	3.00
Rotatable bonds	5.00	Polarizability	1.37

MPDS ID: 14-01-067627



Molecular Formula:
C₁₅H₁₄N₂O₂

IUPAC Name:
2-((cyclopropylmethyl)(methyl)carbamoyl)naphthalen-1-olate

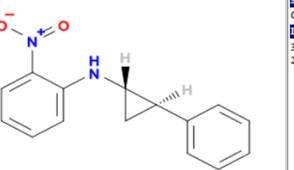
Remarks: Remarks here...

Name/Synonyms: Name/Synonyms here...

Molecular Properties:

Mol. Wt.	254.11	LogP	0.14
HBD	1	LogS	-5.74
HBA	1	pKa	pKa1: ; pKa2: ; pKa3: ; pKa4:
Molar refractivity	17.84	Polar surface area	55.17
Heavy atoms count	19	Rings count	3.00
Rotatable bonds	5.00	Polarizability	1.37

MPDS ID: 14-01-067626



Molecular Formula:
C₁₅H₁₄N₂O₂

IUPAC Name:
3-((cyclobutylmethyl)(methyl)carbamoyl)naphthalen-2-olate

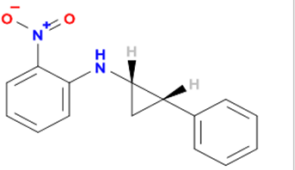
Remarks: Remarks here...

Name/Synonyms: Name/Synonyms here...

Molecular Properties:

Mol. Wt.	254.11	LogP	0.14
HBD	1	LogS	-5.74
HBA	1	pKa	pKa1: ; pKa2: ; pKa3: ; pKa4:
Molar refractivity	17.84	Polar surface area	55.17
Heavy atoms count	19	Rings count	3.00
Rotatable bonds	5.00	Polarizability	1.37

MPDS ID: 14-01-067625



Molecular Formula:
C₁₅H₁₄N₂O₂

IUPAC Name:
3-((cyclopropyl(ethyl)(methyl)carbamoyl)naphthalen-2-olate

Remarks: Remarks here...

Name/Synonyms: Name/Synonyms here...

Molecular Properties:

Mol. Wt.	254.11	LogP	0.14
HBD	1	LogS	-5.74
HBA	1	pKa	pKa1: ; pKa2: ; pKa3: ; pKa4:
Molar refractivity	17.84	Polar surface area	55.17
Heavy atoms count	19	Rings count	3.00
Rotatable bonds	5.00	Polarizability	1.37

3.3.4. Molecular Property-based Search:

Galaxy / MPDS 1.0.1 Analyze Data Workflow Shared Data Visualization 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

Database Id Search searches MPDS compound library using database ID

Exact-structure Search searches molecule in MPDS compound library

Sub-structure Search searches for sub-structure

Molecular Property Based Search perform simple or advance query on MPDS compound library data

Fingerprint Based Search searches using MPDS fingerprints

Molecule cloud generates molecule cloud

Library generator generates molecule based on composition

DATA PROCESSING

File-Format Converter
Descriptor Calculation

Molecular Property Based Search (version 1.0.1)

Field: Hydrogen bond acceptor (HBA) ▾

Field Operator: Equal to ▾

Keyword: 4

Add more Conditions

Add more Condition 1

Connector Operator: AND ▾

Field: Molecular Weight (Mol. Wt.) ▾

Molecular Formula

Molecular Weight (Mol. Wt.)

ALogP

Hydrogen bond donor (HBD)

Hydrogen bond acceptor (HBA)

Molar Refractivity

Topological Polar Surface Area

Total/Heavy Atoms

Rings Count

Rotatable Bonds

Polarizability

Execute

TIP: Please use LIKE Field Operator when using IUPAC Name, Molecular Formula, Remarks and Name/Synonyms in the Field value

System Message: ERROR/3 (<string>, line 6)

Document may not end with a transition.

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:

Galaxy / MPDS 1.0.1

Tools

MPDS 1.0.1

Get Data

Upload File from your computer

Draw Molecule

DATA LIBRARIES

Literature

Target Library

Compound Library

Database Id Search searches MPDS compound library using database id

Exact-structure Search searches molecule in MPDS compound library

Sub-structure Search searches for sub-structure

Molecular Property Based Search perform simple or advance query on MPDS compound library data

Fingerprint Based Search searches using MPDS fingerprints

Molecule cloud generates molecule cloud

Library generator generates molecule based on composition

Fingerprint Based Search (version 1.1.0)

Nature of Compound Chain: Cyclic

No. of Rings: 2 Rings

Compound Nature: Heteroaromatic

No. of Rings Containing Hetero-atoms: 2 Rings

Execute

What it does

Compound Library Search is used to search compounds from MPDS repository containing millions of molecules.

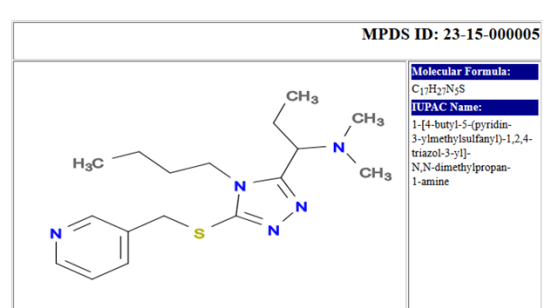
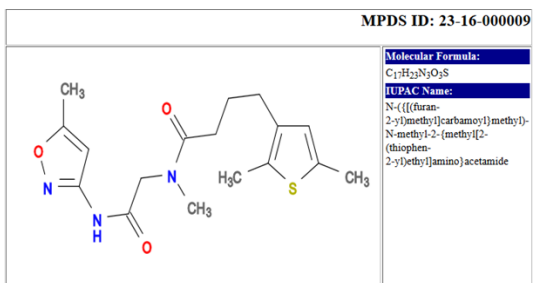
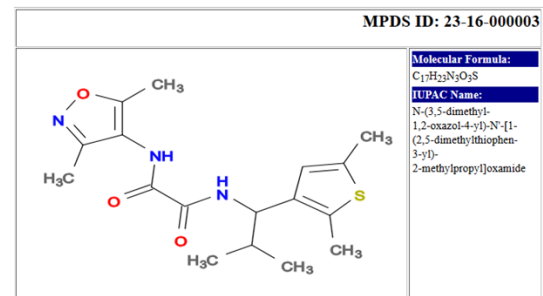
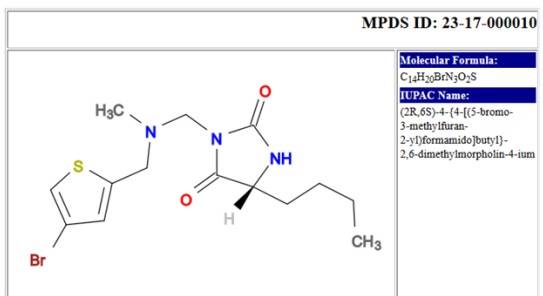
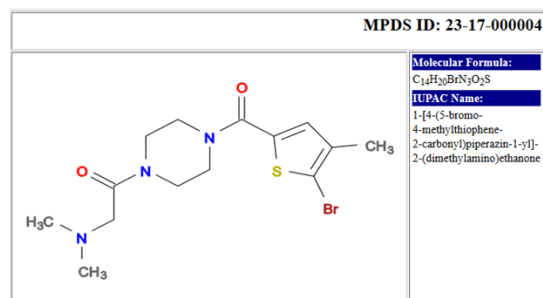
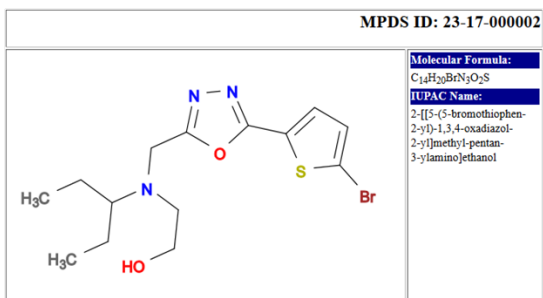
Note

Querying may take time as due to search from millions of molecules depending upon number of filters chosen.

History

12.3 MB

42: Fingerprint Based Search



3.3.6. Molecule cloud:

Galaxy / MPDS 1.0.1 Analyze Data Workflow Shared Data Visualization Admin Help User Using 811.0 MB

Tools
search tools

MPDS 1.0.1
Get Data
Draw Molecule

DATA LIBRARIES
Literature
Target Library
Compound Library

Database Id Search searches molecule in MPDS compound library using database ID
Exact-structure Search searches molecule in MPDS compound library
Sub-structure Search searches for sub-structure
Molecular Property Based Search perform simple or advance query on MPDS compound library data
Fingerprint Based Search searches using MPDS fingerprints
Molecule cloud generates molecule cloud
Library generator generates molecule based on composition

Molecule cloud (version 1.0.0)

Input file containing scaffold with their frequency:
1: Pasted Entry
Scaffolds in Smiles format and its frequency (see below for input file format)

Execute

This tool generates molecule cloud allowing visual representation of the most common structural features of chemical databases in a form of a cloud diagram.

Example
Note Select the **Convert spaces to tabs** option while uploading the following input file using **Upload file** tool.

Input file:
c1ccccc1 417205
O=C(Nc1ccccc1)c2ccccc2 78563
O=S(=O)(Nc1ccccc1)c2ccccc2 46713
O=C(COc1ccccc1)Nc2ccccc2 39163
O=C(CNc1ccccc1)Nc2ccccc2 33806
O=C(NC(=O)c1ccccc1)c2ccccc2 33759
O=C(NC(=O)c1ccccc1)c2ccccc2 27929
c1ccccc1 27356
O=C(NC(=O)c1ccccc1)C(=O)Nc2ccccc2 26505
C1CCNCC1 21150
C(Oc1ccccc1)c2ccccc2 17728
c1ccccc1 17400
O=C(Nc1ccccc1)C=Cc2ccccc2 16908
O=C(NC(=S)Nc1ccccc1)c2ccccc2 16410

Output:

Galaxy / MPDS 1.0.1 Analyze Data Workflow Shared Data Visualization Admin Help User Using 811.0 MB

Tools
search tools

MPDS 1.0.1
Get Data
Draw Molecule

DATA LIBRARIES
Literature
Target Library
Compound Library

Database Id Search searches molecule in MPDS compound library using database ID
Exact-structure Search searches molecule in MPDS compound library
Sub-structure Search searches for sub-structure
Molecular Property Based Search perform simple or advance query on MPDS compound library data
Fingerprint Based Search searches using MPDS fingerprints
Molecule cloud generates molecule cloud
Library generator generates molecule based on composition

DATA PROCESSING
File-Format Converter
Descriptor Calculation

DATA ANALYSIS
QSAR
Docking
Screening
Visualization

Molecule cloud (version 1.0.0)

Input file containing scaffold with their frequency:
1: Pasted Entry
Scaffolds in Smiles format and its frequency (see below for input file format)

Execute

This tool generates molecule cloud allowing visual representation of the most common structural features of chemical databases in a form of a cloud diagram.

Example
Note Select the **Convert spaces to tabs** option while uploading the following input file using **Upload file** tool.

Input file:
c1ccccc1 417205
O=C(Nc1ccccc1)c2ccccc2 78563
O=S(=O)(Nc1ccccc1)c2ccccc2 46713
O=C(COc1ccccc1)Nc2ccccc2 39163
O=C(CNc1ccccc1)Nc2ccccc2 33806
O=C(NC(=O)c1ccccc1)c2ccccc2 33759
O=C(NC(=O)c1ccccc1)c2ccccc2 27929
c1ccccc1 27356
O=C(NC(=O)c1ccccc1)C(=O)Nc2ccccc2 26505
C1CCNCC1 21150
C(Oc1ccccc1)c2ccccc2 17728
c1ccccc1 17400
O=C(Nc1ccccc1)C=Cc2ccccc2 16908
O=C(NC(=S)Nc1ccccc1)c2ccccc2 16410

Output:

3.3.7. Library Generator:

Galaxy / MPDS 1.0.1 | Analyze Data | Workflow | Shared Data | Visualization | Help | User | Using 811.3 MB

Tools

search tools

MPDS 1.0.1

Get Data

Draw Molecule

JMol Editor A Chemical Structure Editor

DATA LIBRARIES

Literature

Target Library

Compound Library

Database ID Search searches MPDS compound library using database ID

Exact-structure Search searches molecule in MPDS compound library

Sub-structure Search searches for sub-structure

Molecular Property Based Search perform simple or advance query on MPDS compound library data

Fingerprint Based Search searches using MPDS fingerprints

Molecule cloud generates molecule cloud

Library generator generates molecule based on composition

Library generator (version 1.0.0)

Input the composition of elements:

C6H6

Library containing:

All chemically possible structures

Execute

This tool looks for the various permutations of elemental compositions and generates a library comprising all chemically valid molecules with same chemical composition. It discards all chemically impossible molecules.

For detailed information that how this tool works, click [here](#)

History

Unnamed history
152.3 KB

2: Library generator
Job is currently running

1: Library generator
4,539 lines
format: txt, database: 2
CCCCC C6H6 molecules 217
Duration: 783 milliseconds

```
CDK      0228171036
6 7 0 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 0 0 0
0.0000 0.0000 0.0000 C 0 0 0 0 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 0 0 0
0.0000 0.0000 0.0000 C 0 0 0 0 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 0 0 0
0.0000 0.0000 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

```
CDK      0601160043

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

> <can_string>
01000100101000002000000010000002000000

$$$$

CDK      0601160043

6 7 0 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 0 0 0
0.0000 0.0000 0.0000 C 0 0 0 0 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 0 0 0
0.0000 0.0000 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

History

history
152.3 KB

1: Library generator
4,539 lines
format: txt, database: 2
CCCCC C6H6 molecules 217
Duration: 805 milliseconds

```
CDK      0601160043

6 7 0 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 0 0 0
0.0000 0.0000 0.0000 C 0 0 0 0 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).

The screenshot shows the Galaxy File-Format Converter tool interface. The left sidebar contains a 'Tools' menu with 'File-Format Converter' selected. The main panel shows the tool configuration: 'Input file' is '5: CID_145823.sdf', 'output format' is 'mol2', and the 'Execute' button is visible. A callout '1. Click' points to the 'File-Format Converter' tool in the sidebar. Another callout '2. Select desired output file format' points to the 'output format' dropdown menu. A third callout '3. Submit' points to the 'Execute' button. On the right, a file upload area shows '5: CID_145823.sdf' (17.1 KB) with a callout 'Uploaded input file'.

Figure 1

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

The screenshot shows the Galaxy interface after the conversion job is complete. A green notification box at the top states: 'The following job has been successfully added to the queue: 6: Converter on data 5. You can check the status of queued jobs and view the resulting data by refreshing. The status will change from 'running' to 'finished' if completed successfully.' A callout 'Download' points to the notification. On the right, the 'History' panel shows a job entry: '6: Converter on data 5' (21.8 KB) with a callout 'View results'. Below the history, the output file '5: CID_145823.sdf' is visible.

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).

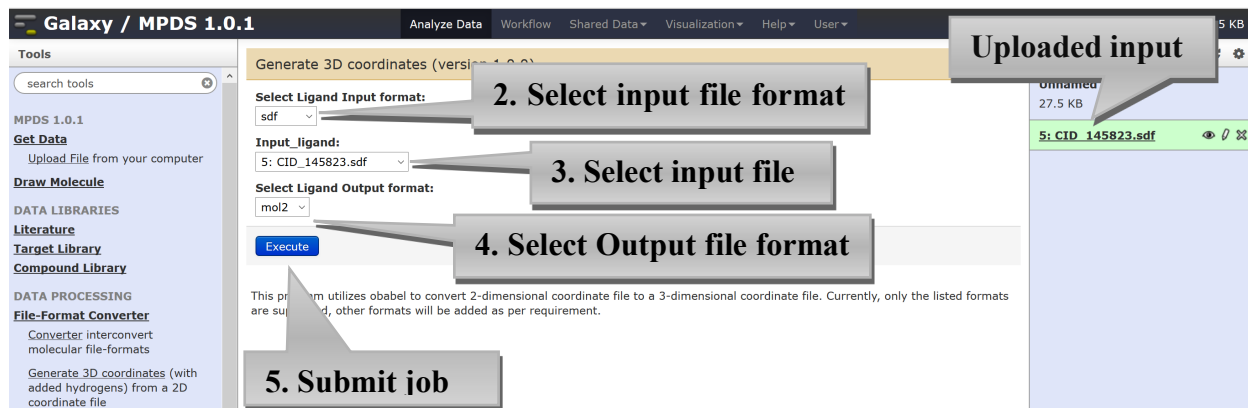


Figure 1

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

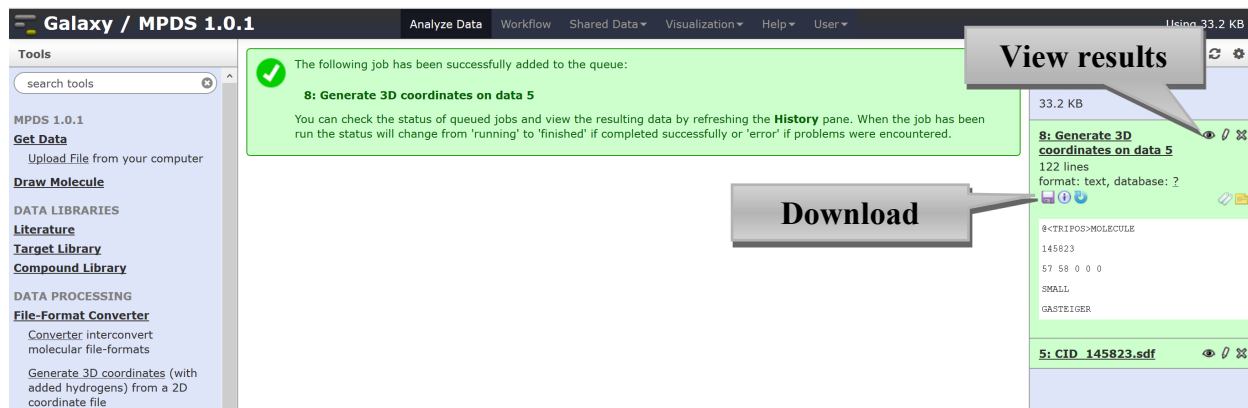


Figure 2

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).

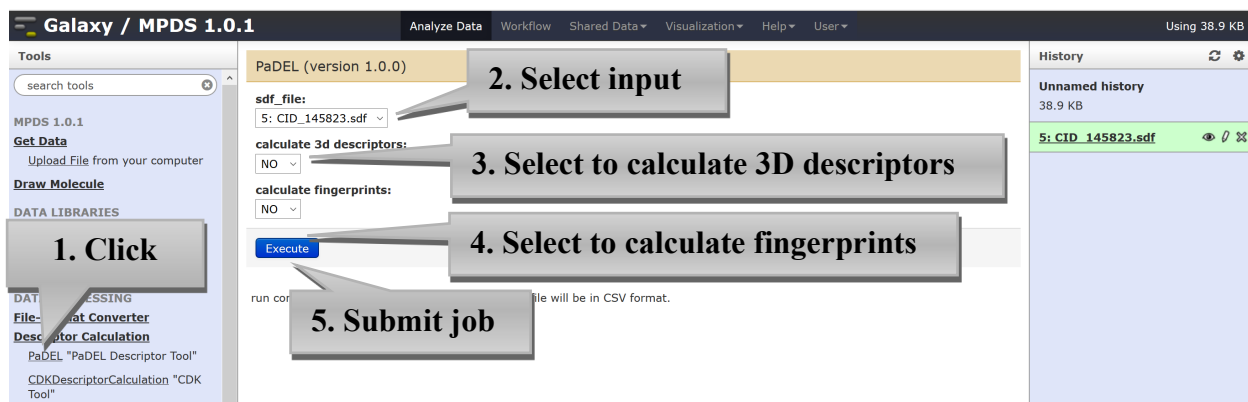


Figure 1

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).

Galaxy / MPDS 1.0.1

Analyze Data Workflow Shared Data Visualization Help User Using 38.9 KB

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"

CDKDescriptorCalculation (version 1.0.0)

SDF:
5: CID_145823.sdf

Execute

run command with file format only and output file will be in txt format.

History

Unnamed history
38.9 KB

5: CID_145823.sdf

2. Select input

5. Submit job

CDK descriptor

Uploaded. sdf file

Figure 2

Galaxy / MPDS 1.0.1

Analyze Data Workflow Shared Data Visualization Help User Using 38.9 KB

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"

The following job has been successfully added to the queue:

10: CDKDescriptorCalculation on data 5

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History

Unnamed history
38.9 KB

5: CID_145823.sdf

View result

Calculated Descriptor

Download result

Rerun job

5
2 lines
format: txt, database
ALOGPDescriptor[]
AminoAcidCountDescriptor[]
APoIDescriptor[]
AromaticAtomsCountDescriptor[]
AromaticBondsCountDescriptor[]
AtomCountDescriptor[]
AutocorrelationDescriptorCharge[]
AutocorrelationDescriptorMass[]
AutocorrelationDescriptorPolarizability[]
MOLID ALOGPDescriptor AminoAcidCount

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)

<p>Tools</p> <p>DATA ANALYSIS</p> <p>QSAR</p> <p>Build QSAR Model builds QSAR model using McQSAR</p> <p>Predict Activity Using McQSAR Using already built QSAR model</p> <p>Convert csv to arff Converter csv to arff file in weka</p>	<pre>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, ZINC00901212, 0, 0.6355, 0.40386025, 16.2707, 8.380379, 0, 0, 7, 4, 3, 0, 3, 1, 0, 0, 0, ZINC15633215, 0, -1.3416, 1.79989056, 5.6967, 6.860758, 0, 0, 8, 2, 6, 0, 1, 1, 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, ZINC25783052, 0, 0.0505, 0.00255025, 13.0125, 8.322758, 0, 0, 9, 3, 6, 0, 2, 0, 1, 0, 0, ZINC15633213, 0, -1.2393, 1.53586449, 11.3934, 9.954344, 0, 0, 11, 3, 8, 0, 2, 1, 0, 0, ZINC71769112, 0, -0.1854, 0.03437316, 14.709, 10.614344, 0, 0, 11, 3, 8, 0, 3, 0, 0, 0, ZINC12358605, 0, -0.1076, 0.01157776, 12.5551, 8.322758, 0, 0, 9, 3, 6, 0, 2, 0, 1, 0, 0, ZINC00895973, 0, 0.5725, 0.32775625, 14.0581, 7.415586, 0, 0, 6, 4, 2, 0, 3, 0, 1, 0, 0, ZINC79313748, 0, 0.1116, 0.01245456, 4.1429, 6.088793, 0, 0, 5, 4, 1, 0, 2, 1, 1, 0, 0, 0, ZINC64622610, 0, 0.561, 0.314721, 13.7364, 7.053586, 0, 0, 6, 4, 2, 0, 2, 0, 0, 0, 0, 0</pre>
---	---

Galaxy / OSDD-MPDS 1.0.1

1. Descriptor File

2. Activity

3. Delimiter

4. Submit Job

1: Input descriptor + Activity File

Figure 16

**1. Input descriptor
+ Activity File**

Galaxy / MPDS 1.0.1

Tools

search tools

MPDS 1.0.1

Get Data

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 Evaluation in weka

Build QSAR Model: SVMlight create model using SVMlight

Classify Data :SVMlight classify data using model given by SVM

Predict Activity Using McQSAR (version 1.0.0)

Select molecule file with descriptors whose activity to be predicted

7: PaDEL1 on data 5
This file should not contain the value to be predicted

Select model file:

2: Build QSAR Model on data 1
This file should be created by Build QSAR model

Execute

Note Model file should must be created by Build QSAR

Example Input file

Molecule_ID,desc1,desc2,desc3

*M1",1,2,-1

*M2",1,3,-1

*M3",1,9,-1

*M4",1,10,-1

*M5",1,11,-1

*M6",1,12,-1

1. Descriptor File of unknown activity sdf file

2. Model build by McQSAR

3. Submit Job

History

QSAR_MCQSAR
495.2 KB

7: PaDEL1 on data 5

5: UnknownAct.sdf

2: Build QSAR Model on data 1

1 line
format: model, database: ?
McQSAR version 1.2.3.74 64-bit build
Apr 18 2012 22:22:14 Copyright (c)
2003-2012 Mikko J. Vainio. All rights
reserved. McQSAR =
Multiconformational Quantitative
Structure-Activity Relationships.
Described in Vainio MJ, Johnson MS
(2005) J. Chem. Inf

Activity = sql(desc(1,00012,misses))

1: d.csv

Figure 19

Galaxy / MPDS 1.0.1

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 813.9 MB

Tools

search tools

MPDS 1.0.1

Get Data

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 Evaluation in weka

Build QSAR Model: SVMlight create model using SVMlight

Classify Data :SVMlight classify data using model given by SVM

23: Predict Activity Using McQSAR on data 2 and data 7

The following job has been successfully added to the queue:

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.

Result of Prediction

History

QSAR_MCQSAR
495.8 KB

23: Predict Activity Using McQSAR on data 2 and data 7

15 lines
format: txt, database: ?
McQSAR version 1.2.3.74 64-bit
build Apr 18 2012 22:22:14
Copyright (c) 2003-2012 Mikko J.
Vainio. All rights reserved.
McQSAR = Multiconformational
Quantitative Structure-Activity
Relationships. Described in Vainio
MJ, Johnson MS (2005) J. Chem.
Inf

Compound	Activity	Act
mol14	0.595136	0 1
mol13	2.40521 0	1 -1
mol19	0.995126	0 1
mol6	0.595136	0 1
mol7	0.595136	0 1

7: PaDEL1 on data 5

5: UnknownAct.sdf

2: Build QSAR Model on data 1

1 line
format: model, database: ?
McQSAR version 1.2.3.74 64-bit
build Apr 18 2012 22:22:14
Copyright (c) 2003-2012 Mikko J.
Vainio. All rights reserved.
McQSAR = Multiconformational
Quantitative Structure-Activity
Relationships. Described in Vainio
MJ, Johnson MS (2005) J. Chem.

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.

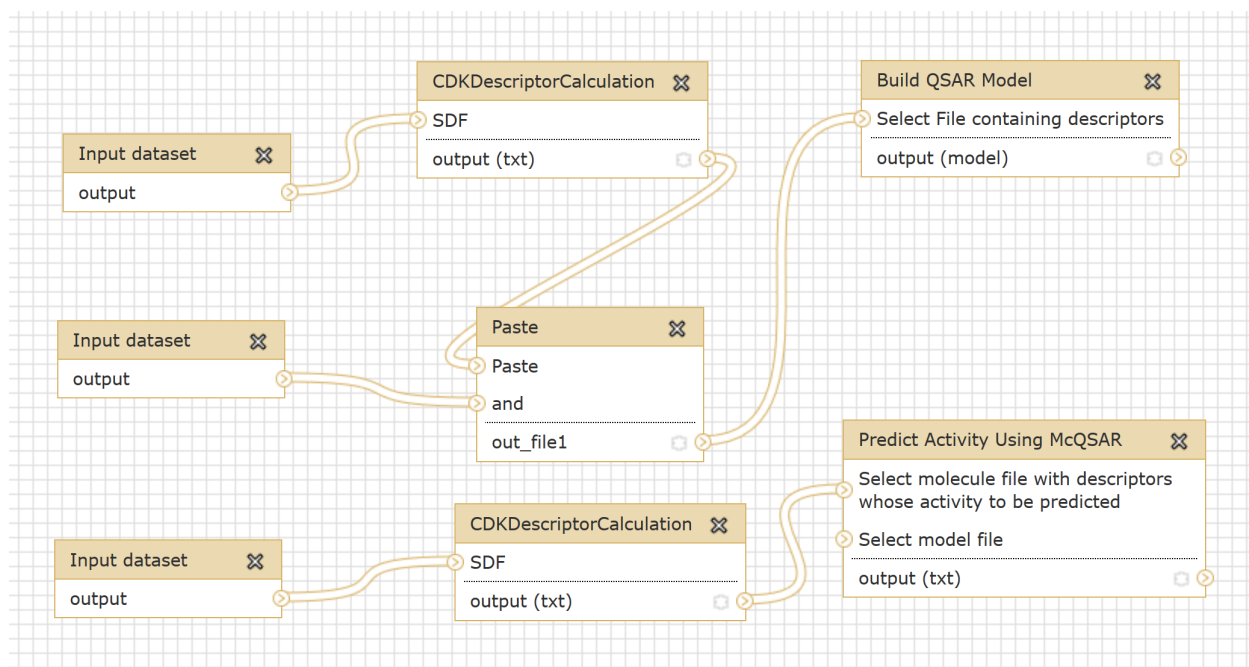


Figure 21

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).

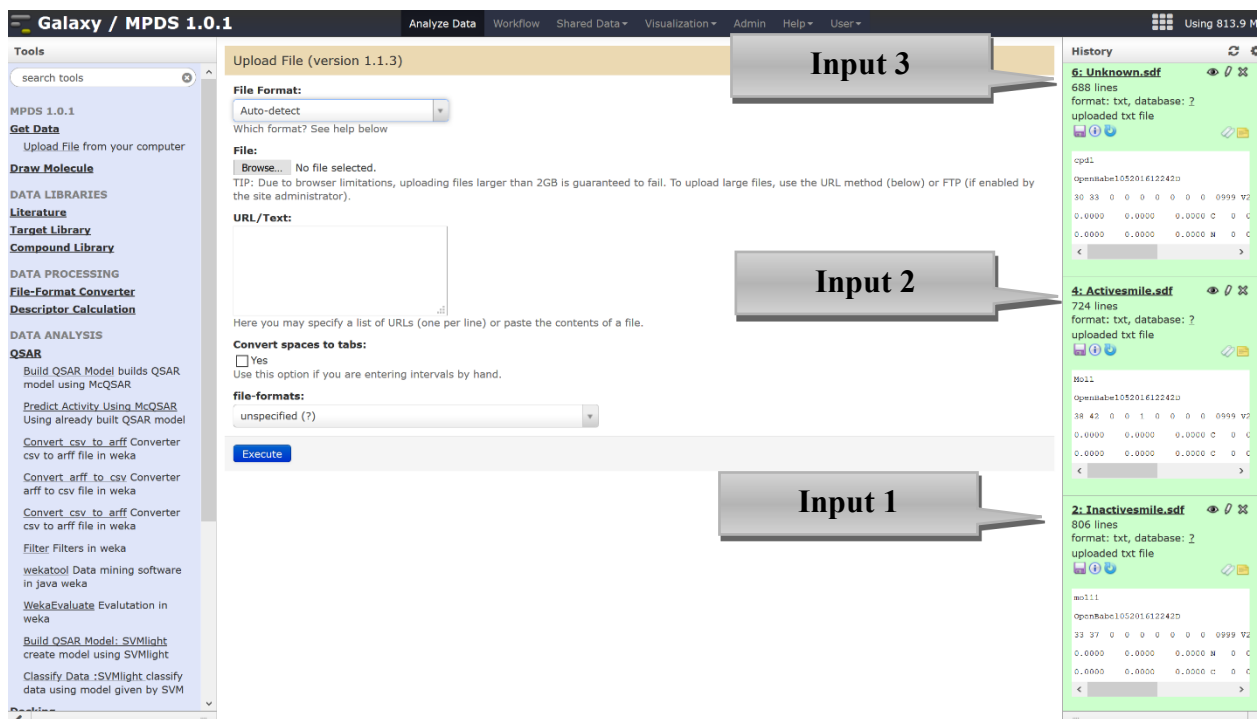


Figure 9

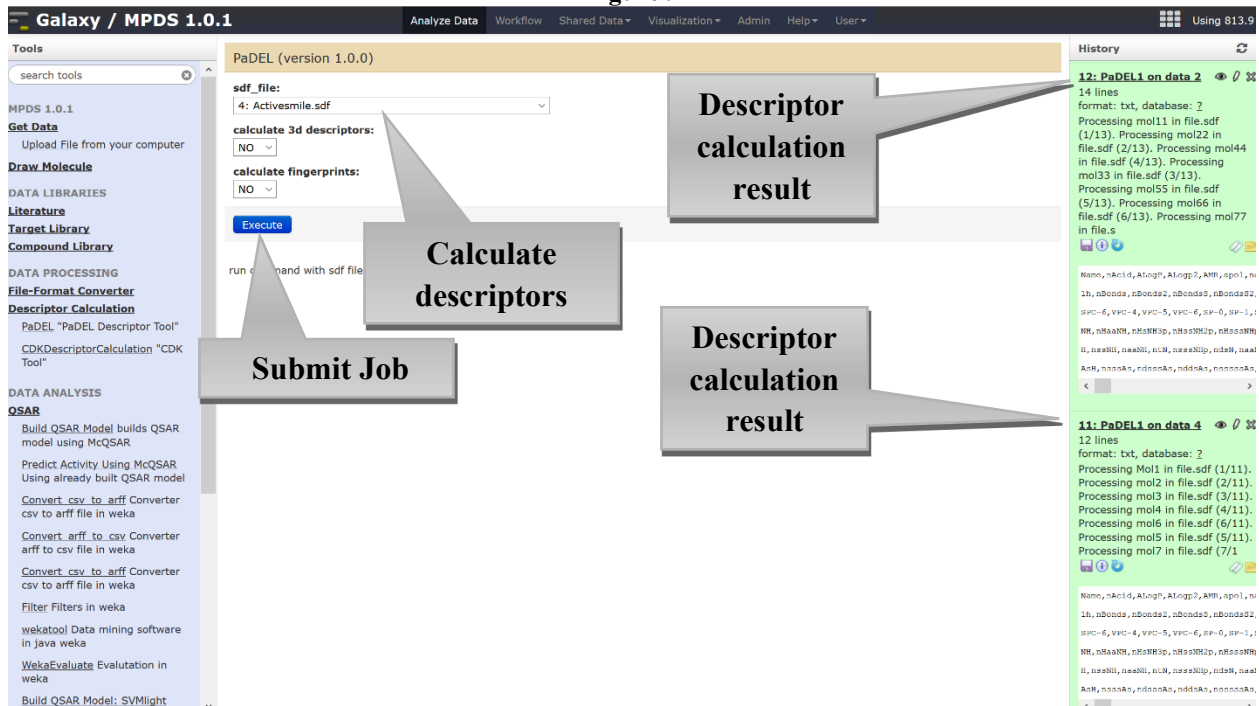


Figure 10

Galaxy / MPDS 1.0.1

Analyze Data Workflow Shared Data Visualization Admin Help User Using 813.9 M

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 Evaluation in weka
Build QSAR Model: SVMlight create model using SVMlight
Classify Data :SVMlight classify data using model given by SVM

Docking

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

SVM-light, Version V6.02
0 # kernel type
3 # kernel parameter -d
1 # kernel parameter -q
1 # kernel parameter -s
1 # kernel parameter -r

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 Using 813.9 M

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 Evaluation in weka
Build QSAR Model: SVMlight create model using SVMlight
Classify Data :SVMlight classify data using model given by SVM

Docking
Screening
Visualization

-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%

-0.63461464
-0.6653034
-0.60372208
-0.61731627
-0.66075651
-0.6523281

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,

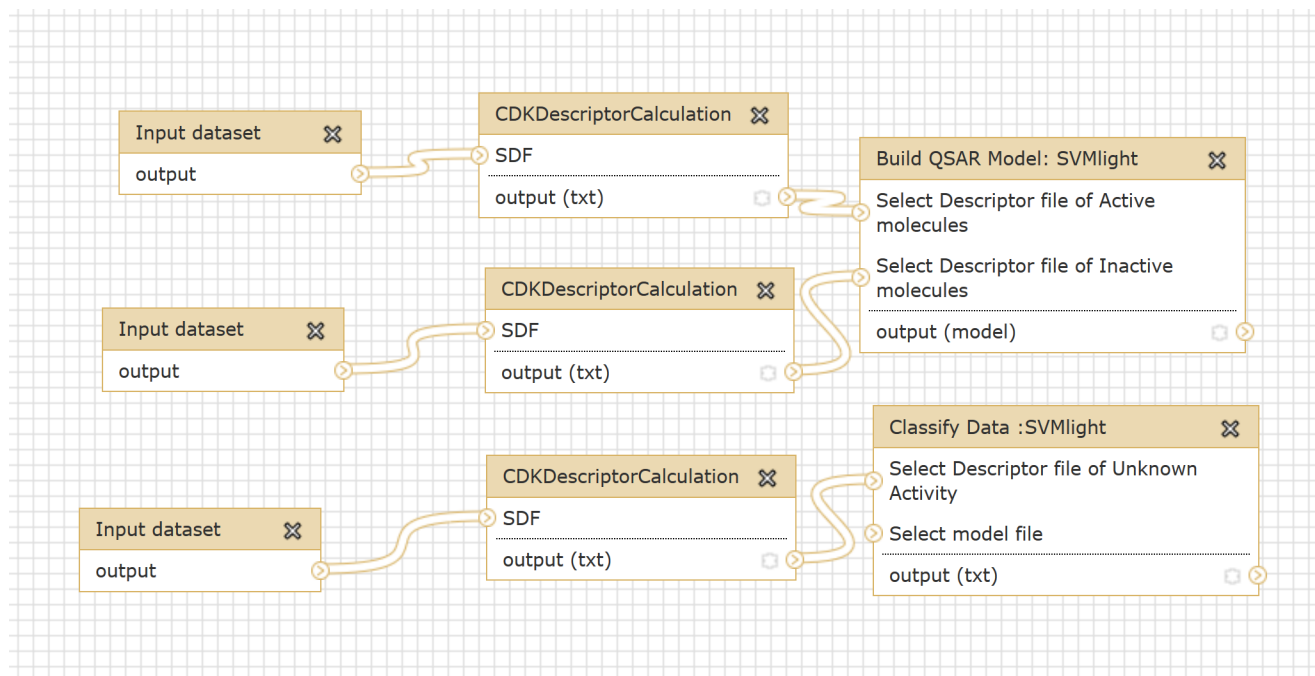
Figure 11

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/0B3c9isKbTnxtN2i1U1ZwVE03VVU/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.

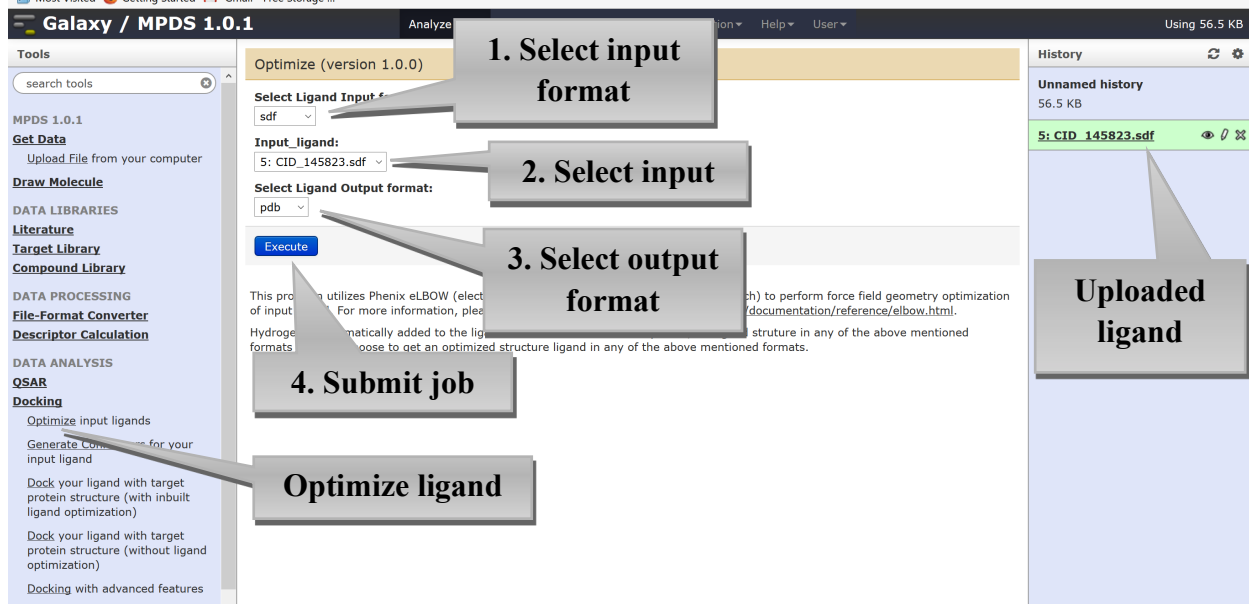


Figure22.2



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.

The screenshot displays the Galaxy / MPDS 1.0.1 interface. The main tool configuration area for 'Generate Conformers (version 1.0.0)' is visible. It includes an 'Input_ligand (in 3D SDF format):' dropdown menu set to '5: CID_145823.sdf', a 'No. of conformers:' input field set to '4', and an 'Execute' button. Four numbered callouts are present: '1. Click' points to the 'Execute' button, '2. Select input' points to the dropdown menu, '3. Input number of conformers' points to the input field, and '4. Submit' points to the 'Execute' button. A green notification box states: 'The following job has been successfully added to the queue: 2: Generate Conformers on data 1. 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.' The 'History' panel on the right shows the job '2: Generate Conformers on data 1' with a preview of the generated conformer data, including a table of coordinates.

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:

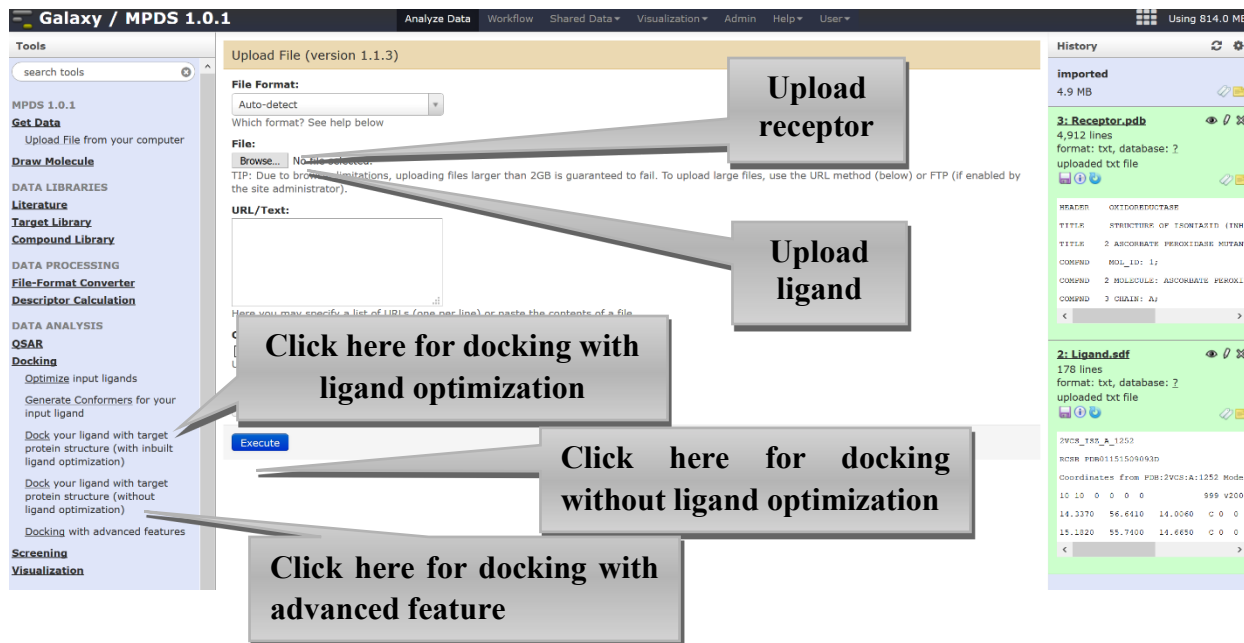


Figure 25

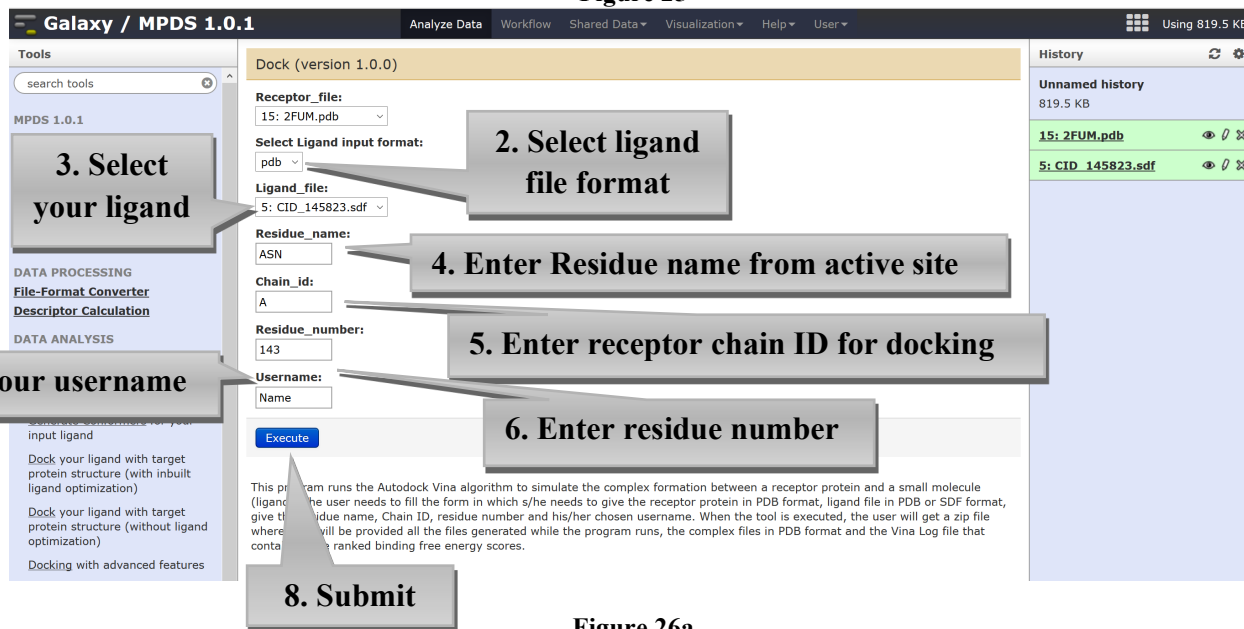


Figure 26a

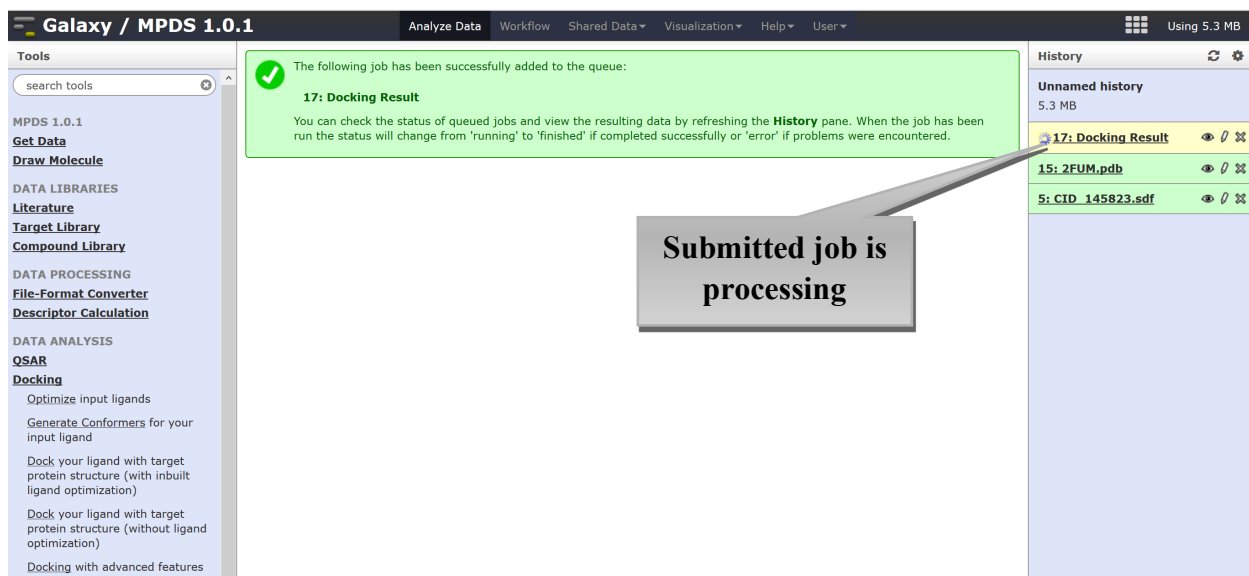


Figure 26b

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.

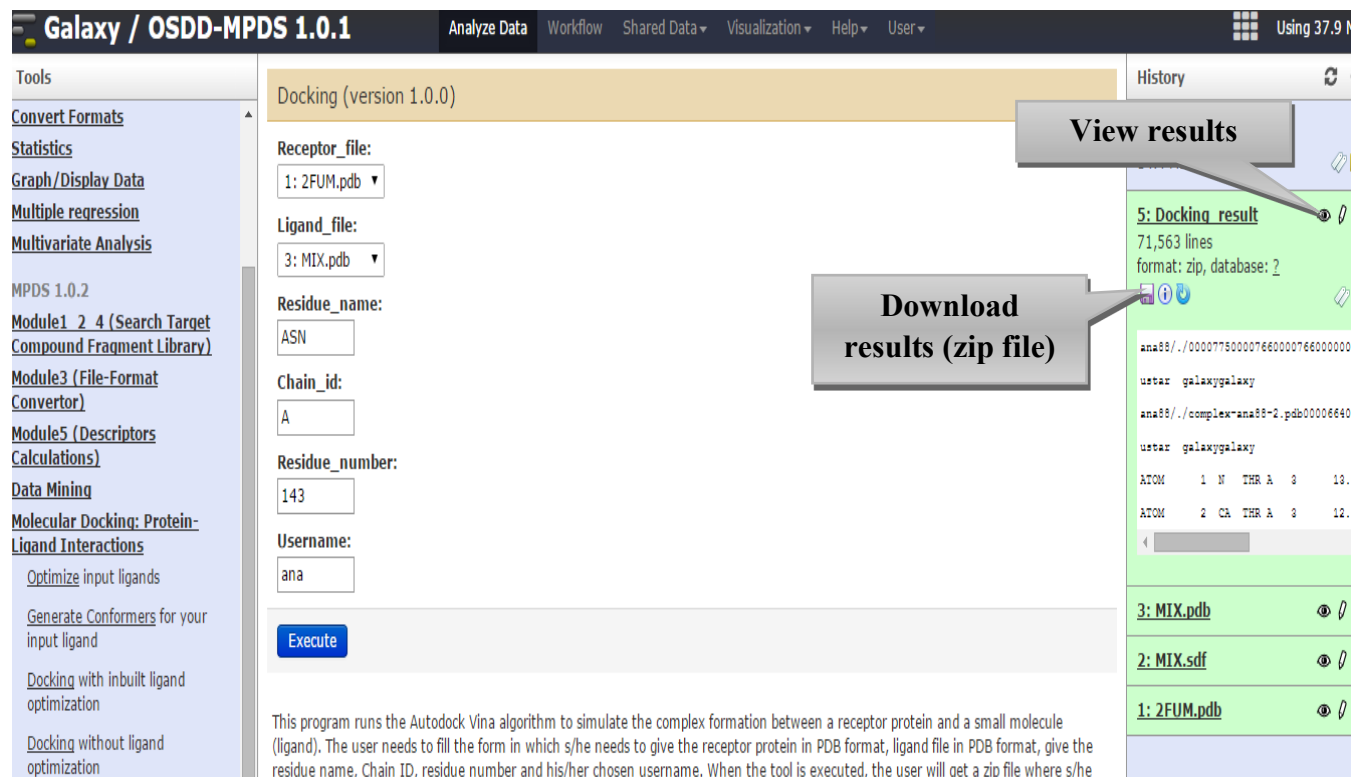


Figure 27

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.

The screenshot displays the Galaxy / MPDS 1.0.1 interface for the 'Docking (version 1.0.0)' workflow. The main configuration area includes the following fields:

- Receptor_file:** 3: Receptor.pdb
- Select Ligand input format:** pdb
- Ligand_file:** 3: Receptor.pdb
- Residue_name:** ASN
- Chain_id:** A
- Residue_number:** 143
- Grid coordinate in the X dimension:** 69
- Grid coordinate in the Y dimension:** 70
- Grid coordinate in the Z dimension:** 68
- Username:** docking

An 'Execute' button is located at the bottom of the configuration area. The right sidebar shows a 'History' panel with two entries:

- imported** (4.9 MB)
- 3: Receptor.pdb** (4,912 lines, format: txt, database: 2, uploaded txt file)

Below the history panel, a preview of the '2: Ligand.sdf' file (178 lines, format: txt, database: 2, uploaded txt file) is shown, including a header and a table of coordinates:

```

HEADER OXIDOREDUCTASE
TITLE STRUCTURE OF ISONIAZID (INH)
TITLE 2 ACCORDATE PEROXIDASE MUTANT
COMPND MOL_ID: 1;
COMPND 2 MOL_SCHL: ACCORDATE PEROXID
COMPND 3 CHAIN: A;
  
```

5.3. Screening

5.4.1. Descriptor Calculator

It Calculate descriptors for estimation of drug likeliness

The screenshot shows the Galaxy / MPDS 1.0.1 interface. The main panel displays the 'Descriptor Calculator (version 1.0.0)' tool. The 'Read data from your current history:' section shows a dropdown menu with '5: CID_145823.sdf' selected. Below this is an 'Execute' button. A descriptive text below the button states: 'This tool processes sdf files for calculation of descriptors that are required for drug-likeness screening.' The left sidebar contains a 'Tools' menu with categories like 'MPDS 1.0.1', 'DATA LIBRARIES', 'DATA PROCESSING', and 'DATA ANALYSIS'. The right sidebar shows a 'History' panel with 'Unnamed history' (9.9 MB) and two entries: '15: 2FUM.pdb' and '5: CID_145823.sdf'.

The screenshot shows the Galaxy / MPDS 1.0.1 interface after the Descriptor Calculator tool has been executed. A green notification box at the top center states: 'The following job has been successfully added to the queue: 21: Descriptor Calculator result on CID_145823.sdf'. Below this, a message reads: '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.' The 'History' panel on the right now shows a new entry: '21: Descriptor Calculator result on CID_145823.sdf' (3 lines, format: tabular, database: ?). Below this entry is a table with 5 columns: Descriptor, Mol.Wt., AlogP, XlogP, Mol. The table contains one row of data: 'mol1 430.219 -3.733 -2.495 110.3'. The source is listed as '/home/galaxy/galaxy-dist/datab'. The left sidebar and other interface elements remain the same as in the previous screenshot.

5.4.2. DruLiTo :

It applies filters for estimation of drug-likeness

The screenshot shows the Galaxy MPDS 1.0.1 interface with the DruLiTo tool (version 1.0.0) selected. The tool configuration is as follows:

- Molecular Descriptors File:** 48: Descriptor Calculator result on xaa.sdf
- All:**
- Lipinski's Rule:**
- Ghose Filter:**
- CMC-50-Like Rule:**
- Veber Filter:**
- MDDR Like Rule:**
- BBB-Likeness:**
- Unweighed QED:**
- Weighted QED:**

The History panel shows the following workflow steps:

- imported (14.9 MB)
- 48: Descriptor Calculator result on xaa.sdf
- 25: xaa.sdf (322 lines, format: txt, database: ?, uploaded txt file)

The screenshot shows the Galaxy MPDS 1.0.1 interface with the DruLiTo tool results displayed. The results are shown in a table format:

Filters	Lipinski Rule	Ghose Filter	CMC Filter	Veber Filter	MDDR Like Rule	BBB-Likeness	Unweighed QED	Weighted QED
mol1	+	+	-	+	-	+	+	+
mol2	+	+	-	+	-	+	+	+
mol3	+	+	-	+	-	+	+	+
mol4	+	+	-	+	-	+	+	+
mol5	+	+	-	+	-	+	+	+
mol6	+	-	-	+	-	-	+	+
mol7	+	+	-	+	-	+	+	+
mol8	+	+	-	+	-	+	+	+

Source: /home/galaxy/galaxy-dist/database/files/013/dataset_13658.dat

The History panel shows the following workflow steps:

- imported (14.9 MB)
- 41: DruLiTo on data 40 (10 lines, format: tabular, database: ?)
- 25: xaa.sdf (322 lines, format: txt, database: ?, uploaded txt file)

5.4.3. Segregate Molecules

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

The screenshot shows the Galaxy MPDS 1.0.1 interface. The main window displays the 'Segregate Molecules for Further Analysis (version 1.0.0)' tool configuration. The 'Choose the DruLiTo output file' dropdown is set to '41: DruLiTo on data 40'. The tool has several filters checked: All, Lipinski's Rule, Ghose Filter, CMC-50-Like Rule, Veber Filter, MDDR Like Rule, BBB-Likeness, Unweighted QED, and Weighted QED. An 'Execute' button is visible at the bottom of the tool configuration. The left sidebar contains various tool categories like 'Tools', 'Docking', 'Screening', 'Visualization', and 'Text Manipulation'. The right sidebar shows the 'History' panel with two entries: 'imported' (14.9 MB) and '41: DruLiTo on data 40' (10 lines, tabular, database: ?). Below the history, a preview of the output data is shown, including a table with columns '1' and '2' and rows for 'mo11' through 'mo15'.

The screenshot shows the Galaxy MPDS 1.0.1 interface displaying the output of the 'Segregate Molecules' tool. The main window shows a large table of output data with columns for molecule ID, various drug-like properties, and a final column with values 'C', 'N', or 'H'. The table is titled 'OpenBabel02281722312D'. The right sidebar shows the 'History' panel with three entries: 'imported' (14.9 MB), '46: Negative Ligands' (606 lines, 39 comments, tabular, database: ?), and '45: Positive Ligands' (empty, tabular, database: ?). Below the history, a preview of the output data is shown, including a table with columns '1' and '2' and rows for 'mo11' through 'mo15'.

5.4.3. BCS Classification

Identify the BCS class to which the molecule belongs

Galaxy / MPDS 1.0.1 Analyze Data Workflow Shared Data Visualization Admin Help User Using 811.3 MB

Tools

- weka.evaluate** Evaluation in weka
- Build QSAR Model: SVMlight** create model using SVMlight
- Classify Data** SVMlight classify data using model given by SVM

Docking

Screening

- Descriptor Calculator** Calculate descriptors for estimation of druglikeliness
- DruLito** Apply filters for estimation of drug-likeness
- Segregate Molecules for Further Analysis** Segregate the input dataset into positive and negative dataset based upon the selected drug like properties.
- BCS Classification** Identify the BCS class to which the molecule belongs
- Toxicity Filter** Identify the toxicophoric groups in the

BCS Classification (version 1.0.0)

Read data from your current history:
25: xaa.sdf
*.sdf file only

Execute

This module provisionally classifies the query molecule as Biopharmaceutical Classification System (BCS) class I, II, III or IV based on its calculated intrinsic solubility (log S) and permeability (Xlog P).

Class II low solubility high permeability	Class I high solubility high permeability
Class IV low solubility low permeability	Class III high solubility low permeability

History

- imported** 14.9 MB
- 25: xaa.sdf** 322 lines
format: txt, database: ?
uploaded txt file

Galaxy / MPDS 1.0.1 Analyze Data Workflow Shared Data Visualization Admin Help User Using 811.3 MB

Molecule/Descriptor	logS	XlogP	BCS Class	Solubility	Permeability
mol1	-2.322	1.448	III	High	Low
mol2	-2.33	1.373	III	High	Low
mol3	-2.215	2.048	I	High	High
mol4	-2.42	1.93	I	High	High
mol5	-1.744	0.729	III	High	Low
mol6	-0.897	0.341	III	High	Low
mol7	-2.33	1.373	III	High	Low
mol8	-2.557	2.137	I	High	High

BCS class Solubility Permeability
I High High
II Low High
III High Low
IV Low Low

History

- imported** 14.9 MB
- 47: BCS Classification result on xaa.sdf** 10 lines
format: tabular, database: ?
- 41: DruLito on data 40** 10 lines
format: tabular, database: ?

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_ToxFilterResults

Non alphanumeric characters will be trimmed

Execute

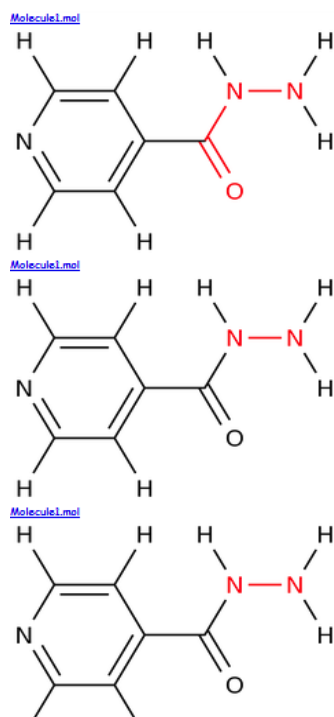
MPDS Toxicity Filter - Help Page

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



[>>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

Step 1: 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

The screenshot displays the Galaxy / MPDS 1.0.1 web interface. The top navigation bar includes 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. The left sidebar lists various tool categories such as 'MPDS 1.0.1', 'DATA LIBRARIES', 'DATA PROCESSING', 'DATA ANALYSIS', and 'GALAXY INBUILT'. The central tool panel is titled '3D Visualization (version 1.0.0)' and features an 'Input file in sdf format' dropdown menu with '15: 2FUM.pdb' selected, and an 'Execute' button. The right sidebar shows a 'History' panel with 'Unnamed history' (9.9 MB) and two entries: '15: 2FUM.pdb' and '5: CID_145823.sdf'. Three callout boxes with arrows indicate the steps: '1. Click' points to the tool in the sidebar, '2. Select input' points to the file dropdown, and '3. Submit job' points to the 'Execute' button.

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

19: 3D Visualization on data 15
The following job has been successfully added to the queue:
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: 2
HTML file

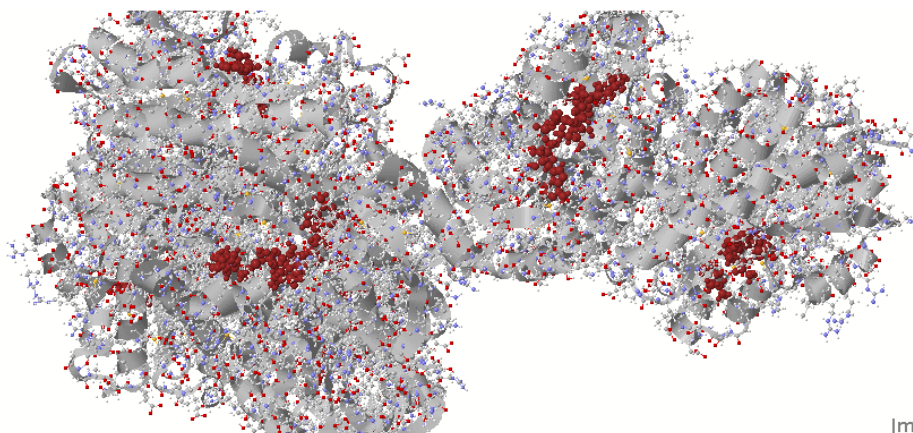
15: 2FUM.pdb

5: CTD_145823.sdf

Please make sure that your browser is java enabled for molecule visualization

[Click Here For Java Test](#)

[Click Here For Molecule Visualization](#)



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: 2

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

The screenshot displays the Galaxy / OSDD-MPDS 1.0.1 web interface. At the top, there is a navigation bar with options like 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. The main content area shows a 3D ball-and-stick model of a protein-ligand complex. Above the model, a blue warning message states: 'Please make sure that your browser is Java enabled for Molecule Visualization'. Below this message are two buttons: 'Click Here For Java Test' and 'Click Here For Molecule Visualization'. On the left side, there is a 'Tools' sidebar with various categories such as 'Get Data', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Convert Formats', 'Statistics', 'Graph/Display Data', 'Multiple regression', 'Multivariate Analysis', 'Data Mining', 'Molecular Docking: Protein-Ligand Interactions', 'Module8 (Filters)', and 'Module9 (Visualization)'. On the right side, there is a 'History' panel showing a list of completed jobs, including '3D Visualization by Jmol' (13.9 KB), '3: 3D Visualization on data_1', '2: Generate 2D image on data_1', and '1: adenosine.sdf'.

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

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

OSAR

Dockin

Screen

Visualization

3D Visualization 3D Visualization by Jmol

Generate Ligplot plots an interaction between protein-ligand

Generate Ligplot (version 1.0.0)

select ligand-receptor file:
15: 2FUM.pdb

enter residue1 id for ligand:
1539

enter residue2 id for ligand:
8

Enter Chain Id:
Z

Enter the maximum H-A distance for H-bonding:
2.9
in Angstrom

Enter the maximum D-A distance for H-bonding:
3.9
in Angstrom

Execute

This tool generates ligplot to visualize ligand-protein interaction

History

Unnamed history
9.9 MB

15: 2FUM.pdb

5: CID_145823.sdf

1. Click

2. Select the input

3. Enter the residue id

4. Enter the chain id

5. Submit job

Figure 34

Galaxy / OSDD-MPDS 1.0.1

Analyze Data Workflow Shared Data Visualization Help User Using 32.2 MB

Tools

Multiple regression

Multivariate Analysis

MPDS 1.0.2

Module1 2 4 (Search Target Compound Fragment Library)

Module3 (File-Format Converter)

Module5 (Descriptors Calculations)

Data Mining

Molecular Docking: Protein-Ligand Interactions

Module8 (Filters)

Module9 (Visualization)

Generate 2D image of your chemical structure file

3D Visualization 3D Visualization by Jmol

Generate Circos generates a circo figure for the data (DEMO)

Generate Ligplot plots an interaction between protein-ligand

Workflows

- test_1
- Workflow constructed from history 'Unnamed history'
- 4G44
- 2IIZ
- my_job
- All workflows

Generate Ligplot

792.4 KB

2: Generate Ligplot on data_1

1: 2FUM.pdb

Ligplot Diagram

Key

- Ligand bond
- Non-ligand bond
- Hydrogen bond and its length
- Non-ligand residues involved in hydrophobic contact(s)
- Corresponding atoms involved in hydrophobic contact(s)

Mix 1539(B)

Me 145(B)

Lew 17(B)

Val 95(B)

Ala 38(B)

Asp 156(B)

Asa 143(B)

Ala 142(B)

Lys 140(B)

Gly 180(B)

Phe 19(B)

Figure 36

