

MPDS^{DM} 1.0.1: Manual

Molecular Property Diagnostic Suite (MPDS^{DM})













*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	8
3.3. Module 3-Gene Library	8
3.3. Module 2- Compound Library	10
3.3.1. Database ID Search:	11
3.3.2. Exact Structure Search:	13
3.3.3. Sub-structure Search:	14
3.3.4. Molecular Property-based Search:	15
3.3.5. Fingerprint-based Search:	16
3.3.6. Molecule cloud:	17
3.3.7. Library Generator:	19
4. Data Processing	20
4.1. Module 4- File Format Convertor	20
4.1.1 Converter	20
4.1.2. Generate 3D coordinates	21
4.2.Descriptor Calculator	23
4.2.1.PaDEL Descriptor Calculator	23
4.2.2.CDK Descriptor Calculator	23

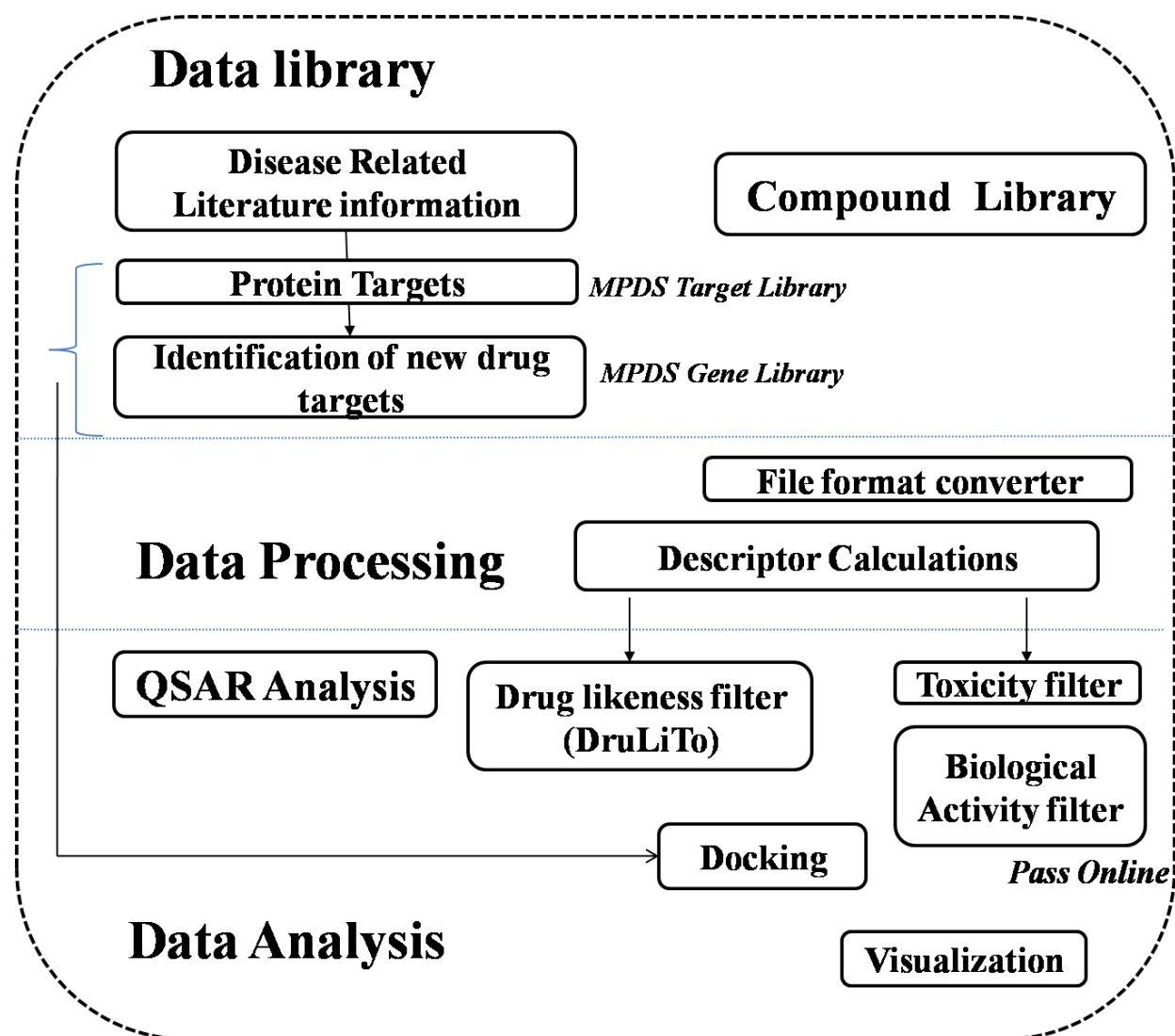
5. Data Analysis	25
5.1. QSAR	25
5.1.1. QSAR Model Building using McQSAR.....	25
5.1.2. QSAR Model Building using Weka	31
5.1.3. QSAR Model Building using SVMlight	31
5.2. Module 7- Docking	36
5.2.1. Optimize Ligand	36
5.2.2. Generate Conformers.....	37
5.2.3. Molecular Docking : Dock your ligand with target protein structure (with inbuilt ligand optimization)	38
5.2.4. Molecular Docking: Dock your ligand with target protein structure (without ligand optimization).....	41
5.2.5. Molecular Docking with advanced features	42
5.3. Screening	42
5.4.1. Descriptor Calculator	42
5.4.2. DruLiTo :	43
5.4.3. Segregate Molecules.....	45
5.4.3. BCS Classification.....	46
5.4.4. Toxicity Filter	47
5.6. Visualization	49
5.6.1. 3D Visualization by Jmol	49
5.6.2. Generate Ligplot plots an interaction between protein-ligand	51
5.7. Drug Repurposing	53

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^{DM} 1.0.1 covers informatics (databases, file format conversion, visualization), structure and analog based drug design approaches (property calculation, QSAR, docking, drug repurposing).



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).

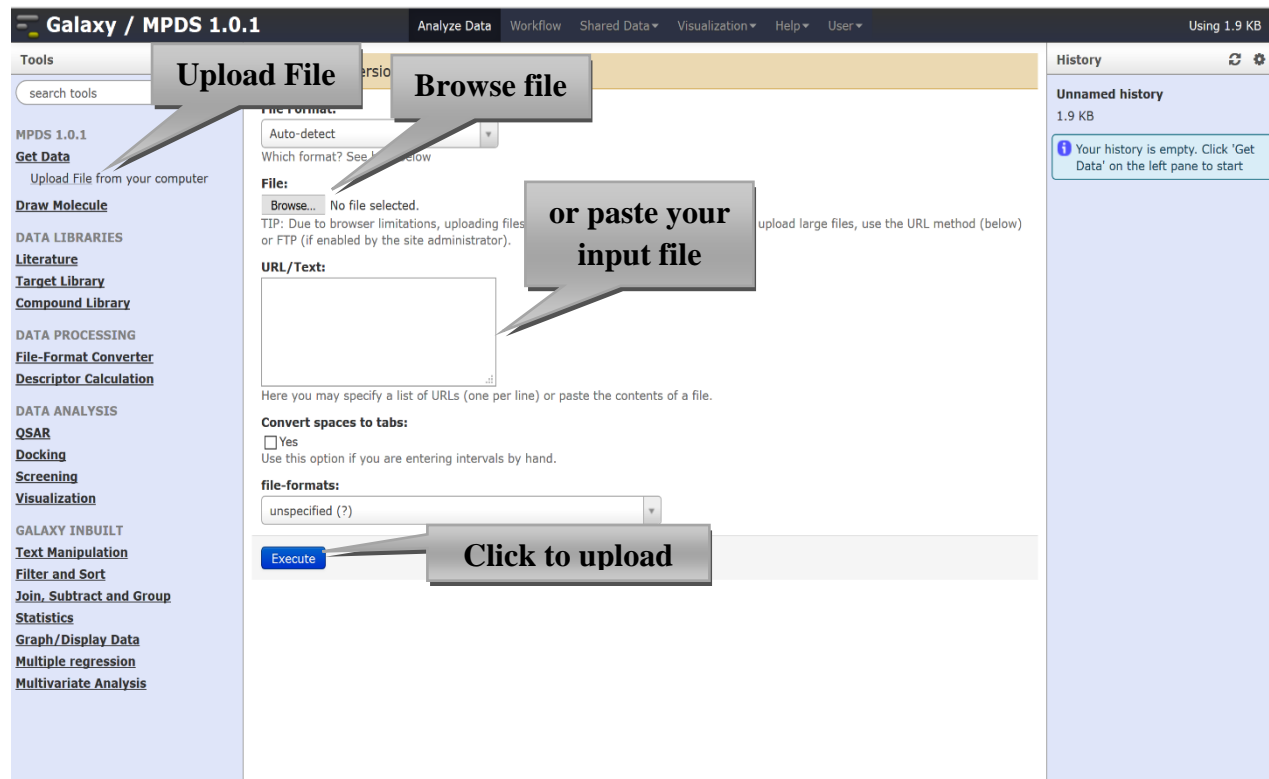


Figure 1a. Upload your input file (ligand, receptor).

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

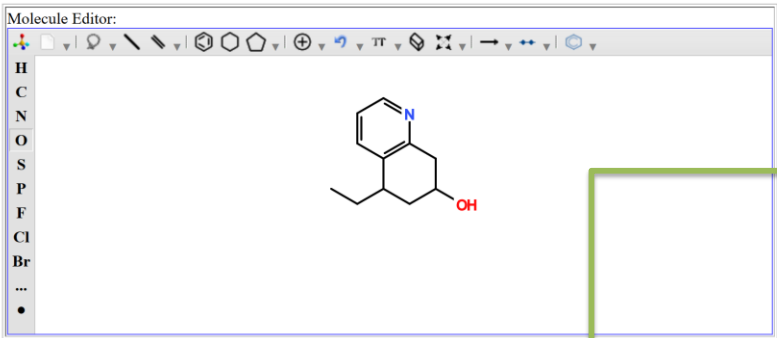
DATA LIBRARIES
Literature
Target Library
Compound Library

DATA PROCESSING
File-Format Converter
Descriptor Calculation

DATA ANALYSIS
QSAR
Docking
Screening
Visualization

Draw your structure and add it to MPDS Galaxy!

Molecule Editor:



Export Molfine 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: 2
C(C=CC1) (=C (N=1) CC2O) C (C2) CC

29: imported mol file
30 lines
format: mol, database: 2
JSDraw02281712072D
13 14 0 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. Literature

Target, biomarker, DM drugs, literature information

Galaxy

Tools search tools

Get Data

DATA LIBRARIES
Literature
T2D_Target_Library Type 2 Diabetes Druggable Gene
T1D_Target_Library Type 1 Diabetes Druggable Gene Details
T2D_Biomarker_Library Type 2 Diabetes Biomarker Gene Details
T1D_Biomarker_Library Type 1 Diabetes Biomarker Gene Details
DM_DrugInfo Anti-diabetic Drugs Information
Literature References, Web links
Target Library
DM Targets Library Search Download DM Targets in PDB file format
Gene Library
Compound Library

DATA PROCESSING
File-Format Converter

MPDS-DM consists of the approaches.

Text Mining (1145)

Type-1 (236) Unassigned (288) Type-2 (621)

First Page

GENERAL INFORMATIONS

MPDS ID	90-01-043491
GENE	TNFRSF9
DESCRIPTION	TNF receptor superfamily member 9
DIABETES TYPE	T1D
PROTEIN	>sp P11717 MPRI_HUMAN Cation-independent mannose-6-phosphate receptor OS=Homo sapiens GN=IGF2R PE=1 SV=3 SNEHDDCOVTPNSTGHFLDLSLSGRAGTAA YSEKGLVYMSICGENENCPRGVGACFGOTRIS VGGANKRLRYVDQVLQVYKDGSPCTPSGSL SYKSVISFYCRPEARPTNRPMLISLDKQTCLTF FSWHTPLACEQATFCSVRNGSSIVDI.SPLIHR

Second Page

GENE INFORMATION

NCBI GENE ID	3604
CHROMOSOME	1
MAP LOCATION	1p36.23
ORIENTATION	Minus
EXON-COUNT	10

Third Page

PROTEIN INFORMATION

CRYSTAL STRUCTURE	NA
UNIPROT	Q07011.1
PROT_NCBI_ID	NP_001552.2
AA LENGTH	255
STRING LINK	http://bit.ly/2s00qtp
DOMAIN INFORMATION	TNFR domain
KEGG	K05146
OMIM	602250
PubMed ID	24797972

Centre for Molecular Modeling, CSIR-

3.2. Target Library

Galaxy / MPDS-DM

Analyze Data Workflow Shared Data Visualization Help User

Tools

search tools

MPDS 1.0.1

Get Data

DATA LIBRARIES

Literature

Target Library

DM Targets Library Search

Download DM Targets in PDB file format

Gene Library

DM Targets Library Search Download DM Targets in PDB file format (Galaxy Version 1.0.0) Options

Enter PDB ID

Execute

Input example : 2ZVT, 3A... MPDS Target Library Search Web Page from Galaxy Interface

--Thanks

2. Enter PDB ID

1. click

Galaxy / MPDS-DM

Analyze Data Workflow Shared Data Visualization Help User

Using 404.4 KB

Tools

search tools

MPDS 1.0.1

Get Data

DATA LIBRARIES

Literature

Target Library

DM Targets Library Search

Download DM Targets in PDB file format

Gene Library

Compound Library

DATA PROCESSING

File-Format Converter

Descriptor Calculation

DATA ANALYSIS

QSAR

Docking

Screening

Drug Repurposing Tool

Visualization

Workflows

All workflows

1 job has been successfully added to the queue - resulting in the following datasets:

1: DM 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.

History

search datasets

Unnamed history

1 shown

404.39 KB

1: DM Targets Library Search

5,113 lines

format: tabular, database: 2

1

HEADER TRANSCRIPTION

TITLE CYS285SER MUTANT PPARGAMMA L

TITLE 2 WITH 15-DEOXY-DELTA12,14-PR

COMPND MOL_ID: 1;

COMPND 2 MOLECULE: PEROXISOME PROLIF

Download PDB ID or view

3.3. Gene Library

3.3.1. Gene name based search

Galaxy / MPDS-DM

Analyze Data Workflow Shared Data Visualization Help User

Tools

search tools

MPDS 1.0.1

[Get Data](#)

DATA LIBRARIES

[Literature](#)

[Target Library](#)

[Gene Library](#)

[Gene Library Search](#)
searches MPDS gene library using gene name

[MPDS Gene ID Search](#)
searches MPDS gene library using unique MPDS gene identifier

Gene Library Search searches MPDS gene library using gene name (Galaxy Version 1.0.0) Options

Select library

MPDS_Gene_Library

Gene name as given in examples: characterized genes:- BCL2, FGF1, AGRP and uncharacterized genes:- LOC101060024,MGC16275

Enter Gene name

bcl2

Execute

1. Enter gene name

2. click

MPDS gene library search web page from galaxy interface. MPDS gene library is comprised of 60,118 Homo Sapiens genes. These genes includes: Characterized genes: 47,774 (includes mitochondrial genes=1,069 and pseudo genes: 15,012) Uncharacterized genes: 12,3... --Thanks

Galaxy / MPDS-DM

Analyze Data Workflow Shared Data Visualization Help User

Using 599 bytes

Tools

search tools

MPDS 1.0.1

[Get Data](#)

DATA LIBRARIES

[Literature](#)

[Target Library](#)

[Gene Library](#)

[Gene Library Search](#)
searches MPDS gene library using gene name

[MPDS Gene ID Search](#)
searches MPDS gene library using unique MPDS gene identifier

Molecular Property Diagnostic Suite:Gene Database

MPDS Gene Identifier:	90-01-002325
Gene Name:	BCL2
Gene Description:	BCL2, apoptosis regulator [Homo sapiens (human)]
Characterization:	Characterized gene

4. Output

3. Database search completed

History

search datasets

Unnamed history

1 shown

599 b

1: Gene Library Search

599 bytes

format: html, database: 2

mpds gene identifier: 90-01-002325 Gene Name: BCL2

HTML file

3.3.2. MPDS ID based search

The image shows two screenshots of the Galaxy / MPDS-DM interface. The top screenshot shows the 'MPDS Gene ID Search' tool with the 'MPDS_Gene_Library' selected and the gene ID '90-01-000524' entered. Callout boxes indicate '1. Enter MPDS gene ID' pointing to the input field and '2. click' pointing to the 'Execute' button. The bottom screenshot shows the 'Molecular Property Diagnostic Suite:Gene Database' output table with callout boxes indicating '3. Database search completed' pointing to the history panel and '4. Output' pointing to the table.

1. Enter MPDS gene ID

2. click

3. Database search completed

4. Output

MPDS Gene Identifier:	90-01-000524
Gene Name:	ADCY10
Gene Description:	adenylate cyclase 10, soluble [Homo sapiens (human)]
Characterization:	Characterized gene

History: 2: MPDS Gene ID Search
605 bytes
format: html, database: Z
mpds gene identifier: 90-01-000524 Gene Name: ADCY10

3.4. Compound Library

3.4.1. Database ID 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

Database Id Search (version 1.0.0)

Select database:

MPDS Compound Library

MPDS_ID as given in example

Enter MPDS ID:

26-01-100524

MPDS ID

Find

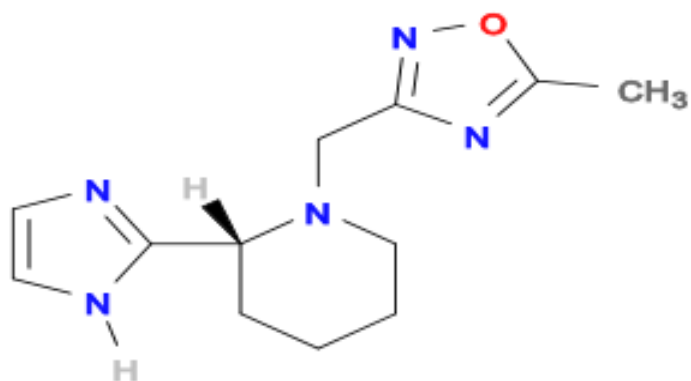
MPDS Database Search Web Page from galaxy Interface

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

--Thanks

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.

3.4.2.Exact Structure Search:

Galaxy / MPDS 1.0.1

Tools

Exact-structure Search (version 1.0.0)

Select File containing structure:
32: Structure.sdf

Select input file format:
SDF
MOL
MOL2
SMILE

History

Unnamed history
12.3 MB

32: Structure.sdf

Galaxy / MPDS 1.0.1

Molecular Property Diagnostic Suite

MPDS ID: 03-04-213023

CC(=O)Nc1ccc(O[C@@H](C)C)cc1C(=O)OCC

Molecular Formula:
C16H23D3N2O4

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

Remarks:

Name/Synonyms:

Molecular Properties:			
Mol. Wt.	312.20	LogP	-1.61
HBD	2	LogS	-1.51
HBA	6	pKa	pKa1: 14.03; pKa2: ; 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

39: Exact-structure Search on data 38
error
An error occurred with this dataset:
Truncated Inchi of the input molecule: InChI=1S/C16H28N2O4/1:1-5:12(6-2):22:14-9-11(16:20):21-7-318-13(17)15(14)18-10(4)19/h9,12-15H,5-8,17H2,1-4H3,(14,18,19)(13-,14+,15+/m0/s1
MPDS-ID of the query molecule is: 03-04-213008
MPDS-ID of the query molecule is:

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

3.4.3. Sub-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 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

Sub-structure Search (version 1.0.0)

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

Select input format:
SDF
MOL
MOL2
SMILE

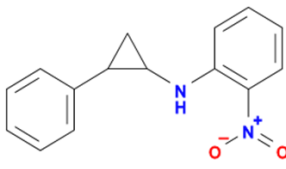
This tool is still under development.

History

Unnamed history
12.3 MB

40: Structure.sdf

MPDS ID: 14-01-067628



Molecular Formula:
C₁₇H₁₇N₂O₂

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

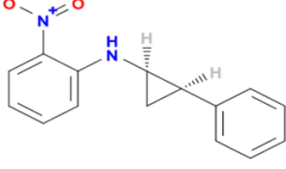
Remarks:
Remarks here...

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:
C₁₇H₁₇N₂O₂

IUPAC Name:
2-[(cyclopropylmethyl)(methoxycarbonyl)asphaltalene-1-olate

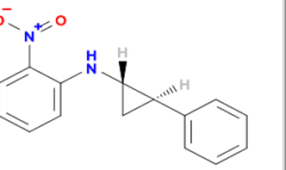
Remarks:
Remarks here...

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:
C₁₇H₁₇N₂O₂

IUPAC Name:
3-[(cyclopropylmethyl)(methoxycarbonyl)asphaltalene-2-olate

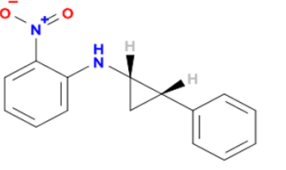
Remarks:
Remarks here...

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:
C₁₇H₁₇N₂O₂

IUPAC Name:
3-[(cyclopropyl)(methyl)(methoxycarbonyl)asphaltalene-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

3.4.4. Molecular Property-based Search:

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 contains a 'Tools' section with a search box and a list of tools under 'MPDS 1.0.1'. The 'Molecular Property Based Search' tool is highlighted in the sidebar, with a green arrow pointing to the main tool interface. The main interface shows the following fields:

- Field:** Hydrogen bond acceptor (HBA)
- Field Operator:** Equal to
- Keyword:** 4

Below these fields is the 'Add more Conditions' section, which includes:

- Connector Operator:** AND
- Field:** A dropdown menu is open, showing options: 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, and Polarizability.

At the bottom of the search area is an 'Execute' button. Below the search area, a tip states: 'TIP: Please use LIKE Field Operator when using IUPAC Name, Molecular Formula, Remarks and Name/Synonyms in the Field value'. A system message at the bottom reads: '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	C11H20N4O5	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

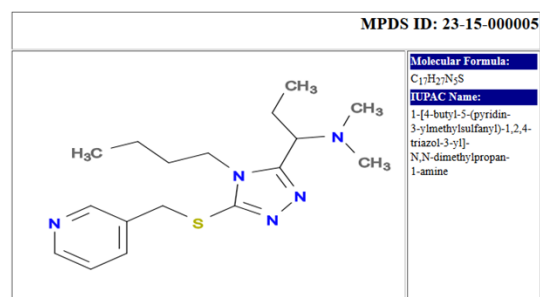
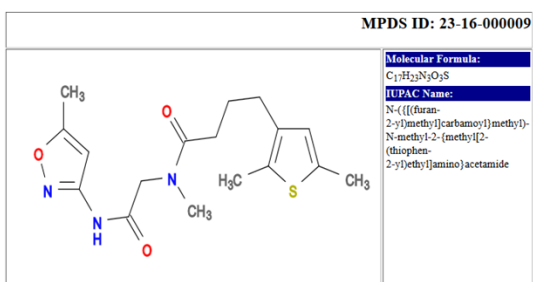
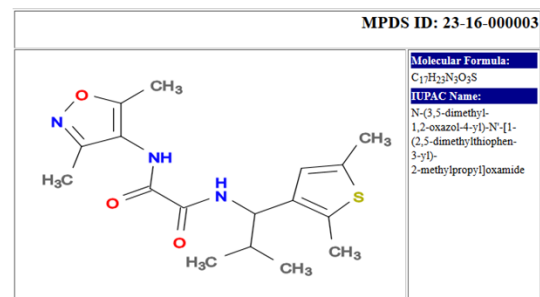
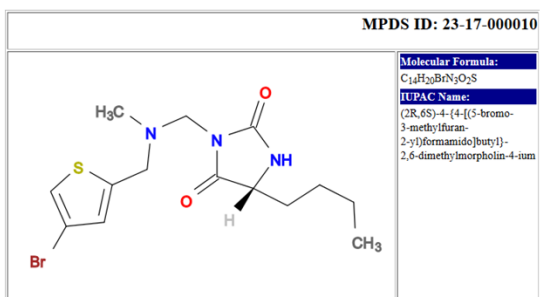
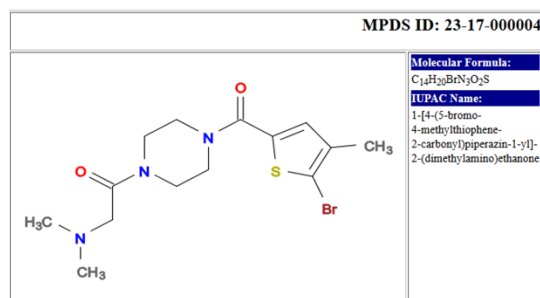
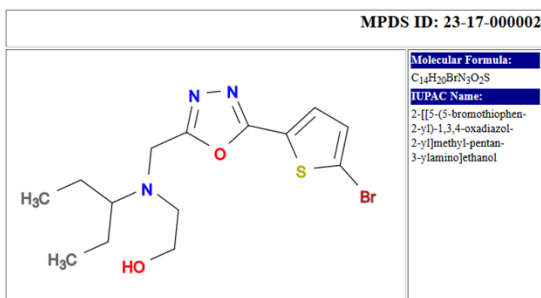
3.4.5. Fingerprint-based Search:

The screenshot displays the Galaxy/MPDS 1.0.1 web interface. The main panel shows the configuration for the 'Fingerprint Based Search (version 1.1.0)' tool. The configuration includes:

- Nature of Compound Chain:** Cyclic
- No. of Rings:** 2 Rings
- Compound Nature:** Heteroaromatic
- No. of Rings Containing Hetero-atoms:** 2 Rings

Below the configuration, there is an 'Execute' button and a 'What it does' section. A note states: 'Compound Library Search is used to search compounds from MPDS repository containing millions of molecules. Querying may take time as due to search from millions of molecules depending upon number of filters chosen.'

The left sidebar contains a 'Tools' section with a search bar and a list of tools. The 'Fingerprint Based Search' tool is highlighted in blue, and a green arrow points from this tool name to the main configuration panel.



3.4.6. Molecule cloud:

Galaxy / MPDS 1.0.1

Tools

search tools

MPDS 1.0.1

Get Data

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

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 417305
O=C(Nc1ccccc1)c2ccccc2 78563
O=S(=O)(Nc1ccccc1)c2ccccc2 46713
O=C(Oc1ccccc1)Nc2ccccc2 39163
O=C(CNc1ccccc1)Nc2ccccc2 33806
O=C(NC1ccccc1)Nc2ccccc2 33753
O=C(NC1ccccc1)c2ccccc2 27929
c1ccccc1 27356
O=C(NC1ccccc1)C(=O)Nc2ccccc2 26505
C1CN(C)C1 21159
C(=O)c1ccccc1c2ccccc2 17728
c1ccccc1 17400
O=C(Nc1ccccc1)C=Cc2ccccc2 16908
O=C(NC(=S)Nc1ccccc1)c2ccccc2 16410
```

Output:

History

Imported: Molecule cloud

104.6 KB

1

2: Molecule cloud on data

104.0 KB

format: png, database: ?

Image in png format

1: Pasted Entry

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

Set Data

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

DATA ANALYSIS

QSAR

Docking

Screening

Visualization

Scaffold cloud generated through MPDS

History

Imported: Molecule cloud
104.6 KB

2: Molecule cloud on data
1
104.0 KB
format: png, database: ?
View data

Image in png format

1: Pasted Entry

3.4.7. 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:

Library containing:

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: ?
 CCCCC C6H6 molecules 217
 Duration: 783 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
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>
010001001010000200000001000002000000

$$$$

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: ?
 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
0.0000 0.0000 0.0000 C 0 0

```

4. Data Processing

4.1. 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 then go to **converter** to **interconvert moleculefile format**. Select desired output file format. Click on "Execute" button (fig. 2).

The screenshot displays the Galaxy / MPDS 1.0.1 web interface. The left sidebar contains a search bar and a list of tools under categories like 'DATA LIBRARIES', 'DATA PROCESSING', 'DATA ANALYSIS', and 'GALAXY INBUILT'. The 'File-Format Converter' tool is highlighted in the sidebar. The main content area shows the configuration for the 'Converter (version 1.6)' tool. The 'input file' is set to '5: CID_145823.sdf' and the 'output format' is set to 'mol2'. There is an 'Execute' button and a checkbox for 'remove input file when finished:'. Below the tool configuration, a note states: 'This tool can be used to convert between different molecular file-formats. Supported formats are mol2, sdf, drf, pdb, ac, ent, brk, hin, mol, ...'. On the right side, a file upload area shows the file '5: CID_145823.sdf' (17.1 KB) has been uploaded. Three callout boxes with arrows point to specific elements: '1. Click' points to the 'File-Format Converter' tool in the sidebar; '2. Select desired output file format' points to the 'output format' dropdown menu; and '3. Submit' points to the 'Execute' button.

Figure 1

Step 3: The converted output file appears in the history which **View results** and (fig. 3).

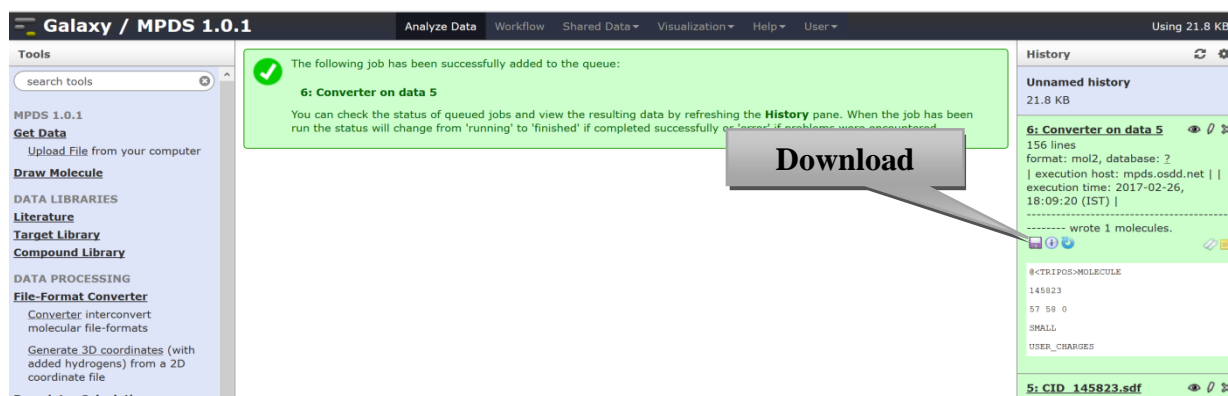


Figure 2

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).

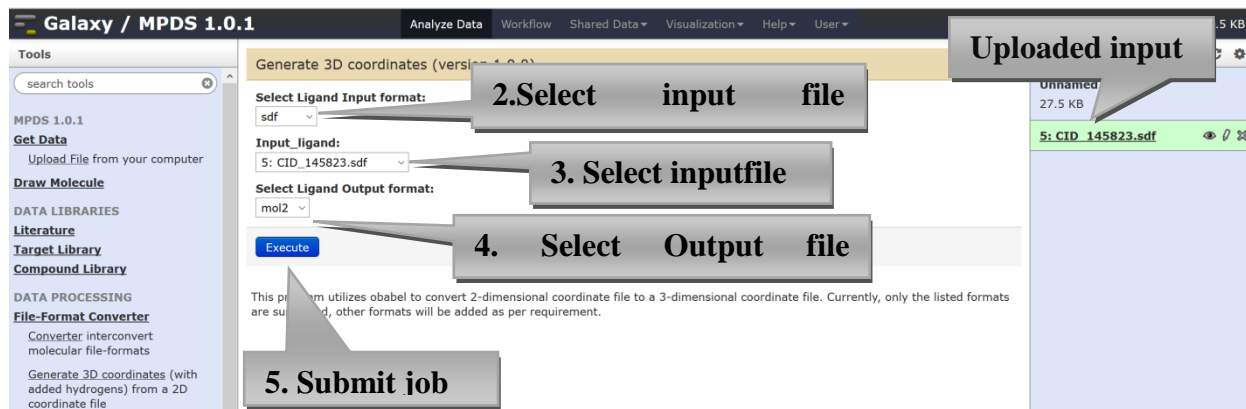


Figure1

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

The screenshot displays the Galaxy / MPDS 1.0.1 web interface. At the top, a navigation bar includes 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. A status bar on the right indicates 'Using 33.2 KB'. On the left, a 'Tools' sidebar lists various categories: 'MPDS 1.0.1', 'Get Data', 'Draw Molecule', 'DATA LIBRARIES', 'Literature', 'Target Library', 'Compound Library', 'DATA PROCESSING', and 'File-Format Converter'. The main workspace features a green notification box with a checkmark icon, stating: 'The following job has been successfully added to the queue: 8: Generate 3D coordinates 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 been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.' To the right, a file entry is shown: '8: Generate 3D coordinates on data 5' (33.2 KB, 122 lines, text format). Below this, a preview of the file content is visible: '4<TRIPPOS>MOLECULE', '145823', '57 58 0 0 0', 'SMALL', and 'GASTRIGER'. A 'Download' button is positioned over the file entry. A 'View results' button is located above the file entry.

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).

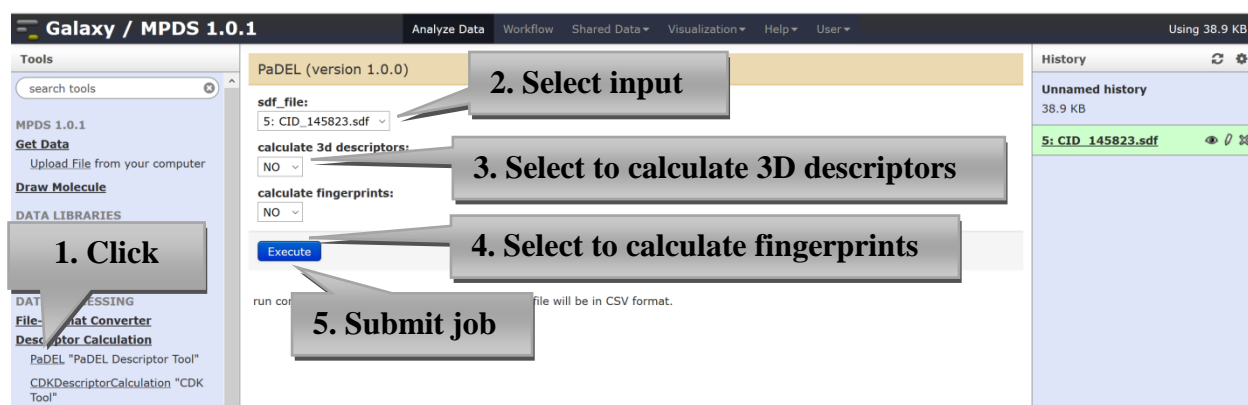


Figure 1

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 sdf 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

0 data

5

2 lines

format: txt, database

ALOGPDescriptor[]

AminoAcidCountDescriptor[]

APolDescriptor[]

AromaticAtomsCountDescriptor[]

AromaticBondsCountDescriptor[]

AtomCountDescriptor[]

AutocorrelationDescriptorCharge[]

AutocorrelationDescriptorMass[]

AutocorrelationDescriptorPolarizability

[]

MOLID ALOGPDescriptor AminoAcidCount

View result

Calculated Descriptor

Download result

Rerun job

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 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).

The screenshot shows the Galaxy / MPDS 1.0.1 interface. The main window displays a table of molecular descriptors for a training set. The columns include MolID, ALogP, nAcid, and a long list of binary descriptor values. Three callout boxes highlight specific files: 'Training set activity file', 'CDK Descriptor File', and 'Training set .sdf file'. The sidebar on the left contains various tools and options, including 'PaDEL Descriptor Tool', 'CDKDescriptorCalculation', and 'Build QSAR Model'.

6. Figure 15

**Input in .csv file format
(Activity of the molecules + Descriptor Values)**

Tools	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, 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, 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, 1, 0, 0, 0, 0, ZINC15633213, 0, -1.2393, 1.53586449, 11.3934, 9.954344, 0, 0, 11, 3, 8, 0, 2, 1, 0, 0, 0, 0, ZINC71769112, 0, -0.1854, 0.03437316, 14.709, 10.614344, 0, 0, 11, 3, 8, 0, 3, 0, 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, 2, 1, 1, 0, 0, 0, 0, ZINC64622610, 0.0 561.0 314721.13 7364.7 053586.0.0.6.4.2.0.2.0.0.0.0.0.0
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	

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: 2
- 1: training_set.smi

Figure 16

Galaxy / MPDS 1.0.1

Analyze Data Workflow Shared Data Visualization Help User Using 813.6 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

- 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

Build QSAR Model (version 1.0.0)

Select File containing descriptors: 15: Paste on data 13 and data 14

Enter the column header whose value to use as response: Activity

default is 'Activity' but may choose your criteria

Perform preprocessing:

Select All Unselect All

- Exclude Correlated Descriptors
- Exclude Identical Conformers
- Exclude Inactive Conformers
- Exclude Sparse Conformers
- Exclude Sparse Descriptors
- Exclude Descriptors with zeros

pre-processing removes redundancy and excludes unnecessary features

Enter percentage of bins the compounds are divided to when performing cross-validation: 3 fold

Default is 3 fold

Enter the number of repetitions for the cross-validation procedure: 5

User may choose any number, but higher number of repetition will take more step

Select Collinearity cutoff: 0.2

This is a threshold value for excluding the other variable of all variable pairs whose correlation coefficient value is higher than the cutoff value its better to use higher values

Execute

5. Submit

1. Input descriptor + Activity File

2. Name of response column

3. Preprocessing options

4. Various options

History

- MCQSAR
844.7 KB
- 15: Paste on data 13 and data 14
71 lines
format: txt, database: ?
- 14: train_activity.txt
71 lines
format: txt, database: 2
uploaded txt file
- Activity
- 7: 405712164
- 6: 438194141
- 8
- 6: 920818754
- 13: CDKDescriptorCalculation on data 13
71 lines
format: txt, database: 2

Cross validation The number of bins (5th parameter) the compounds are divided to when performing cross-validation. Value between zero and one is interpreted as percentage of the data size. Thus values 10 and 0.1 both cause the data set to be partitioned to ten bins. Value equal to zero or one causes leave-one-out (LOO) cross-validation. The actual bin size, i.e. the number of compounds in each left-out set, is adjusted according to each equation's dimension, if necessary, to ensure that the fit is (over)determined.

Input File should be comma-separated file as follows

```
Molecule_ID,Descriptor1,Descriptor2,Descriptor3,Activity
Mol_2,1,2,-1,4
Mol_2,1,3,-1,9
Mol_2,1,4,-1,16
```

Figure 17

MPDS 1.0.1 Analyze Data Workflow Shared Data Visualization Help User Using 18

The following job has been successfully added to the queue:

6: Model

7: Logs

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job is in the state 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

Log File

McQSAR Model

History

CDK_MCQSAR_12dec14
707.0 KB

7: Logs

6: Model

1 line
format: model, database: ?

```
Activity = plus(avg(const(0.7188),
aasn), qspLine(2.46783, WPT-4)), gau
4)), max(max(max(gauss(-4.1661
,WPT-4))), gauss(-11.787, 0.694798,
8571 0.228571 3.83525 0.228571 3.8
19523 0.542977 4.50974 0.239046 4.
```

Figure 18

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

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

1. Descriptor File of unknown activity sdf file

2. Model build by McQSAR

3. Submit Job

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

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 = eqcl(desc(1.00012, nbase))

1: d.csv

Figure 19

The screenshot shows the Galaxy / MPDS 1.0.1 interface. A green notification box at the top states: "The following job has been successfully added to the queue: 23: Predict Activity Using McQSAR on data 2 and data 7". Below this, a message explains that the job status will change from 'running' to 'finished' or 'error' upon completion. On the right, the 'History' panel displays job details for '23: Predict Activity Using McQSAR on data 2 and data 7', including a table of predicted vs actual activity values for compounds mol14 through mol1.

Compound	Activity	Actual
mol14	0.595136	0
mol13	2.40521	0
mol12	0.595136	0
mol11	0.595136	0
mol10	0.595136	0
mol9	0.595136	0
mol8	0.595136	0
mol7	0.595136	0

Result of Prediction

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/OB3c9isKbTnxtZmpzYVcOVVNpWmM/view?usp=sharing>

7.

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

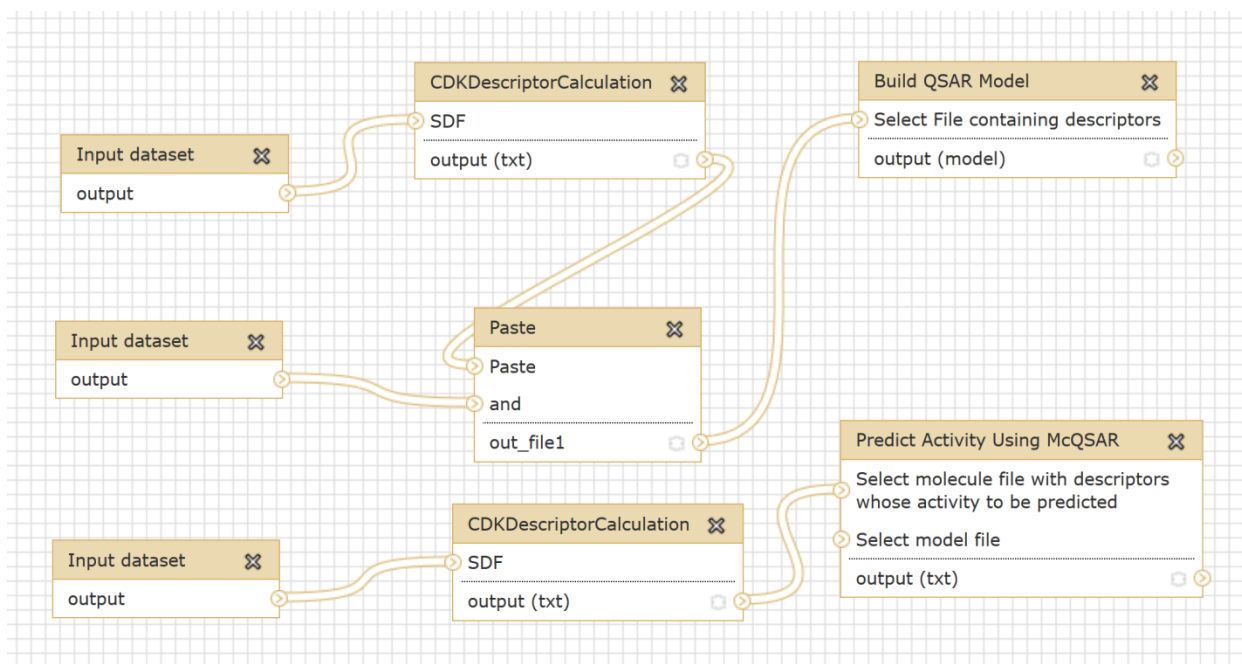


Figure 21

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).

The screenshot displays the Galaxy / MPDS 1.0.1 interface. The main window is titled "Upload File (version 1.1.3)". It features a "File Format" dropdown menu set to "Auto-detect", a "File" field with a "Browse..." button, and a "URL/Text" text area. Below these is a "Convert spaces to tabs" checkbox and a "file-formats" dropdown menu set to "unspecified (?)". An "Execute" button is located at the bottom of the main area. Three callout boxes labeled "Input 1", "Input 2", and "Input 3" point to the "file-formats", "URL/Text", and "File Format" fields respectively. On the right side, a "History" panel shows three entries: "6: Unknown.sdf" (688 lines), "4: Activesmile.sdf" (724 lines), and "2: Inactivesmile.sdf" (806 lines). Each entry includes a preview of a data table with columns for activity and various descriptors.

Figure 9

Galaxy / MPDS 1.0.1 Analyze Data Workflow Shared Data Visualization Admin Help User Using 813.9

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 Evaluation in weka
Build QSAR Model: SVMlight

PaDEL (version 1.0.0)

sdf_file:
4: Activesmile.sdf

calculate 3d descriptors:
NO

calculate fingerprints:
NO

Execute

run command with sdf file

Submit Job

Calculate descriptors

Descriptor calculation result

Descriptor calculation result

History

12: PaDEL1 on data 2
14 lines
format: txt, database: 2
Processing mol11 in file.sdf (1/13). Processing mol22 in file.sdf (2/13). Processing mol44 in file.sdf (4/13). Processing 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

```
Name, nAcid, ALogP, ALogP2, AMR, apol, n
1h, nBonds, nBonds2, nBonds3, nBonds2,
SPC-6, VPC-4, VPC-5, VPC-6, SP-6, SP-1,
NH, nMaAM, nStNH3p, nStNH2p, nStNHNC
H, nSaNH, nSaNH, nTM, nSaNHp, nSaM, nSaA
ASH, nSaSaA, nSaSaA, nSaSaA, nSaSaA,
```

11: PaDEL1 on data 4
12 lines
format: txt, database: 2
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)

```
Name, nAcid, ALogP, ALogP2, AMR, apol, n
1h, nBonds, nBonds2, nBonds3, nBonds2,
SPC-6, VPC-4, VPC-5, VPC-6, SP-6, SP-1,
NH, nMaAM, nStNH3p, nStNH2p, nStNHNC
H, nSaNH, nSaNH, nTM, nSaNHp, nSaM, nSaA
ASH, nSaSaA, nSaSaA, nSaSaA, nSaSaA,
```

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

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

Submit

Select Active descriptor file

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

```
Name, nAcid, ALogP, ALogP2, AMR, apol, n
1h, nBonds, nBonds2, nBonds3, nBonds2,
SPC-6, VPC-4, VPC-5, VPC-6, SP-6, SP-1,
NH, nMaAM, nStNH3p, nStNH2p, nStNHNC
H, nSaNH, nSaNH, nTM, nSaNHp, nSaM, nSaA
ASH, nSaSaA, nSaSaA, nSaSaA, nSaSaA,
```

11: PaDEL1 on data 4
12 lines
format: txt, database: 2
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)

```
Name, nAcid, ALogP, ALogP2, AMR, apol, n
1h, nBonds, nBonds2, nBonds3, nBonds2,
SPC-6, VPC-4, VPC-5, VPC-6, SP-6, SP-1,
NH, nMaAM, nStNH3p, nStNH2p, nStNHNC
H, nSaNH, nSaNH, nTM, nSaNHp, nSaM, nSaA
ASH, nSaSaA, nSaSaA, nSaSaA, nSaSaA,
```

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 Evaluation in weka

Build QSAR Model: SVMlight create model using SVMlight

Classify Data :SVMlight classify data using model given by SVM

```

1 SVM-light version V6.02
2 0 # kernel type
3 # kernel parameter -d
4 # kernel parameter -g
5 # kernel parameter -o
6 # kernel parameter -r
7 empty# kernel parameter -u
8 13 # highest feature index
9 24 # number of training documents
10 23 # number of support vectors plus 1
11 0.7245602 # threshold b, each following line is a SV (starting with alpha*)
12 -5.501404070623752636014572314771e-05 1:0 2:1.04510091 3:0.4330656 4:77.1129 5:35.538345 6:0 7:0 8:27 9:10 10:10 11:0 12:15 13:2 #
13 5.501404070623752636014572314771e-05 1:0 2:-1.7072899 3:2.944734 4:35.0510 5:19.32952 6:0 7:0 8:17 9:10 10:7 11:0 12:17 13:0 #
14 -5.501404070623752636014572314771e-05 1:0 2:0.2500999 3:0.06255008 4:82.070396 5:144.859859 6:0 7:0 8:19 9:19 10:20 11:0 12:17 13:0 #
15 5.501404070623752636014572314771e-05 1:0 2:0.9781 3:0.9567959 4:139.4302 5:166.319653 6:0 7:0 8:14 9:13 10:21 11:0 #
16 -5.501404070623752636014572314771e-05 1:0 2:0.07850003 3:0.03642289 4:133.7789 5:162.489278 6:0 7:0 8:10 9:12 10:18 11:0 12:24 13:7 #
17 5.501404070623752636014572314771e-05 1:0 2:-1.809 3:3.272401 4:94.723091 5:155.284241 6:0 7:0 8:19 9:16 10:23 11:0 #
18 -5.501404070623752636014572314771e-05 1:0 2:0.33793001 3:0.00143641 4:184.382002 5:147.953449 6:0 7:0 8:12 9:20 10:22 11:0 12:18 13:0 #
19 5.501404070623752636014572314771e-05 1:0 2:0.0945 3:9.490904 4:131.43179 5:71.352826 6:0 7:0 8:10 9:20 10:23 11:0 #
20 -5.501404070623752636014572314771e-05 1:0 2:1.9515001 3:3.8083522 4:121.2874 5:164.520447 6:0 7:0 8:15 9:13 10:22 11:0 12:23 13:7 #
21 5.501404070623752636014572314771e-05 1:1 2:0.9285 3:0.8621222 4:97.619499 5:152.08086 6:0 7:0 8:14 9:16 10:20 11:0 #
22 -5.501404070623752636014572314771e-05 1:0 2:1.12625 3:1.113042 4:94.491599 5:145.259102 6:0 7:0 8:17 9:23 10:14 11:0 12:17 13:4 #
23 -5.501404070623752636014572314771e-05 1:0 2:0.1395999 3:0.0196816 4:1116.4167 5:164.423409 6:0 7:0 8:16 9:25 10:27 11:0 12:22 13:7 #
24 5.501404070623752636014572314771e-05 1:0 2:-2.3527 3:15.5851973 4:190.872096 5:156.249031 6:0 7:0 8:15 9:16 10:21 11:0 #
25 -5.501404070623752636014572314771e-05 1:0 2:0.61390001 3:0.3768732 4:190.805199 5:154.140617 6:0 7:0 8:16 9:22 10:26 11:0 12:20 13:0 #
26 5.501404070623752636014572314771e-05 1:1 2:-1.2345001 3:1.5239903 4:103.6964 5:158.971653 6:0 7:0 8:12 9:11 10:21 11:0 #
27 -5.501404070623752636014572314771e-05 1:0 2:1.9786 3:1.4919779 4:116.3506 5:161.076241 6:0 7:0 8:11 9:20 10:23 11:0 12:21 13:6 #
28 5.501404070623752636014572314771e-05 1:0 2:-2.3455999 3:5.5018392 4:94.325699 5:156.417825 6:0 7:0 8:15 9:16 10:21 11:0 #
29 -5.501404070623752636014572314771e-05 1:0 2:0.2366 3:0.95979561 4:111.153 5:133.692402 6:0 7:0 8:14 9:27 10:17 11:0 12:20 13:6 #
30 5.501404070623752636014572314771e-05 1:1 2:1.7543 3:3.0782702 4:119.5087 5:168.354413 6:0 7:0 8:12 9:15 10:27 11:0 #
31 -5.501404070623752636014572314771e-05 1:0 2:1.7639 3:3.1194027 4:1120.3315 5:161.603032 6:0 7:0 8:12 9:20 10:24 11:0 12:23 13:4 #
32 5.501404070623752636014572314771e-05 1:0 2:-2.8893001 3:1.3489549 4:95.475502 5:157.582619 6:0 7:0 8:12 9:16 10:26 11:0 #
33
34

```

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: PaDEL on data 2

12 lines
format: txt, database: 2
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 (7/11).

SVM Model build based on descriptor values of Active and inactive compounds

Galaxy / MPDS 1.0.1

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

History

14: PaDEL 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: PaDEL on data 2

11: PaDEL on data 4

12 lines
format: txt, database: 2
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).

SVMlight Model built in earlier step

Descriptor file of unknown activity molecules

Classify data as active and inactive using SVMlight Classify

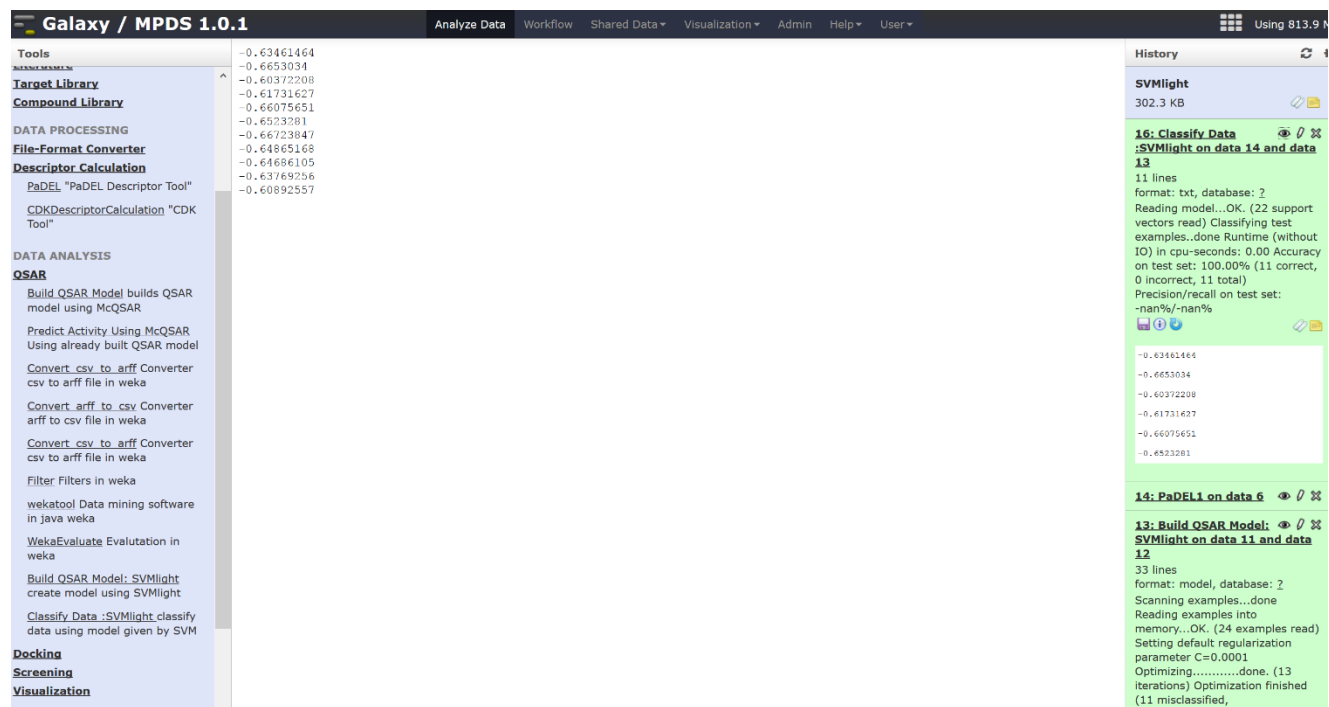


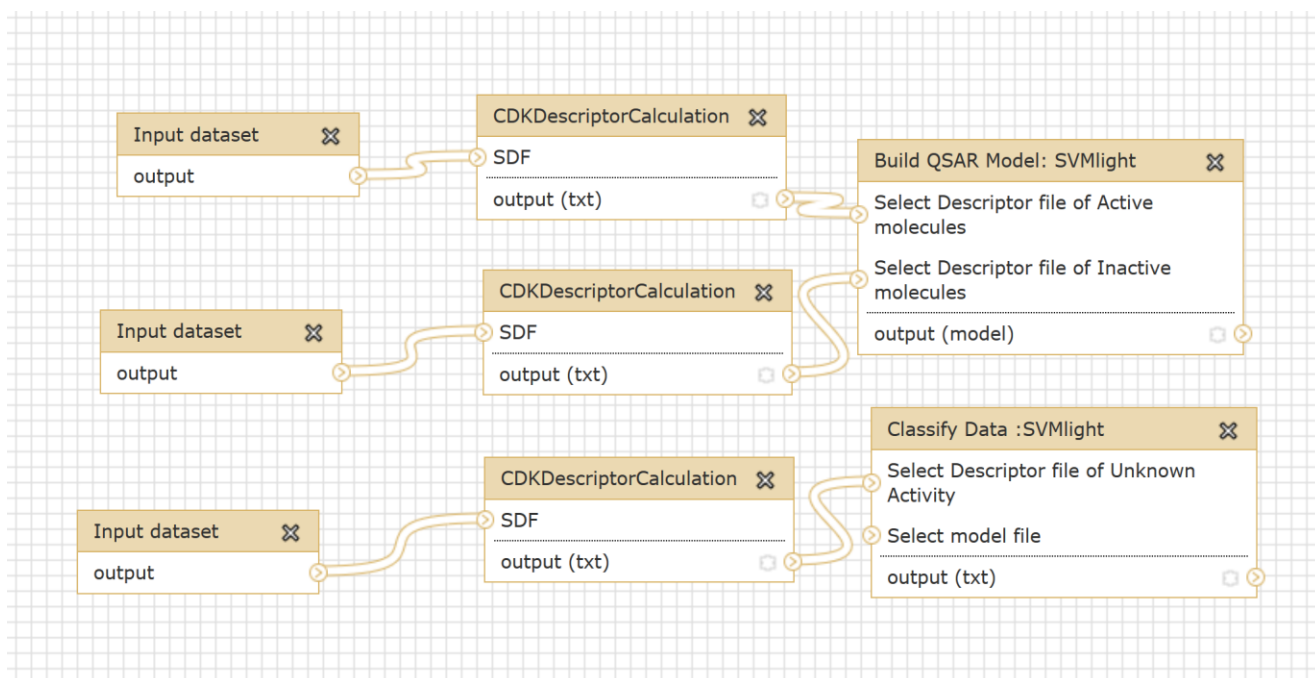
Figure 11

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. 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 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.

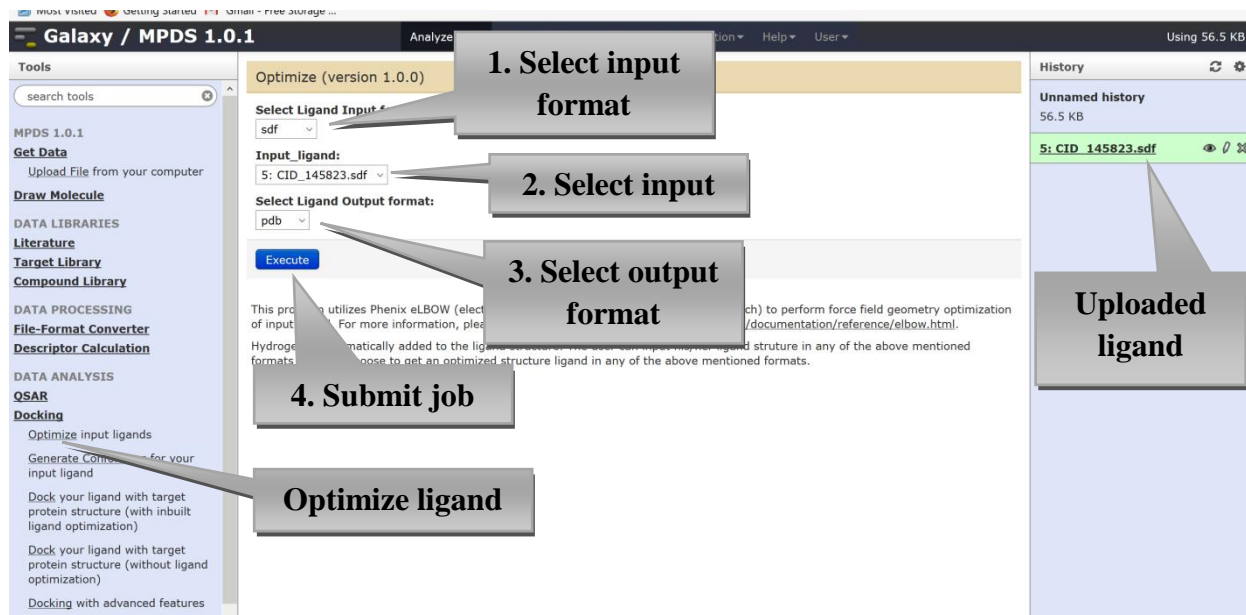
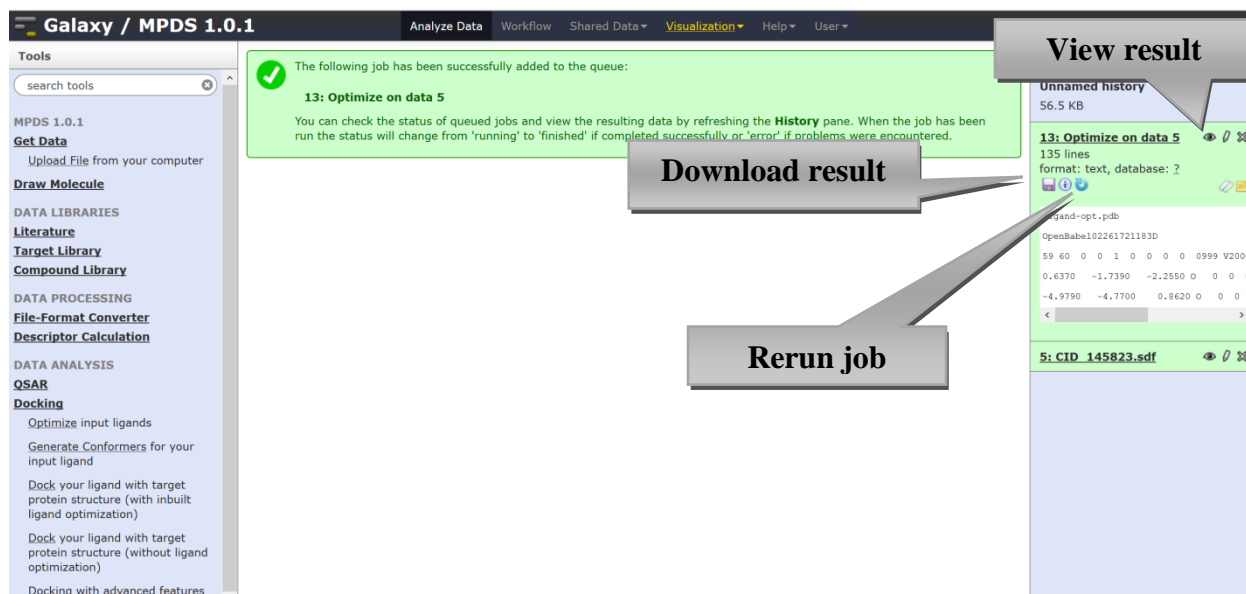


Figure22.2



5.2.2. Generate Conformers

Ligand:sdf

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

Step 2: Go to *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 panel shows the 'Generate Conformers (version 1.0.0)' tool workflow. The workflow includes the following steps:

- 1. Click**: A callout points to the 'Tools' sidebar on the left.
- 2. Select input**: A callout points to the 'Input_ligand (in 3D SDF format):' dropdown menu, which is set to '5: CID_145823.sdf'.
- 3. Input number of conformers**: A callout points to the 'No. of conformers:' input field, which is set to '4'.
- 4. Submit**: A callout points to the 'Execute' button.

The 'History' panel on the right shows the job '5: CID_145823.sdf' completed. Below the main panel, a green notification box states: 'The following job has been successfully added to the queue: 2: Generate Conformers on data 1'. A callout points to this notification with the text 'Conformers generated'. The 'History' panel also shows the job '2: Generate Conformers on data 1' completed, with a file size of 788 lines and a format of 'text, database: ?'. Below this, a table of conformer coordinates is visible:

Atom	X	Y	Z
145923	0.000000	0.000000	0.000000
OpenBabel102281720021D	37.58	0.00	0.00
	10.9703	-2.3101	0.500000
	8.2474	2.7025	-0.8173

Figure 24

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

The AutodockVina 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:

The screenshot displays the Galaxy/MPDS 1.0.1 web interface. The main area is titled 'Upload File (version 1.1.3)'. It features a 'File Format' dropdown set to 'Auto-detect', a 'File' input field with a 'Browse...' button, and a 'URL/Text' input field. A blue 'Execute' button is at the bottom. The left sidebar contains a 'Tools' menu with categories like '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', 'Docking', 'Screening', and 'Visualization'. The 'Docking' section is expanded, showing options: 'Optimize input ligands', 'Generate Conformers for your input ligand', 'Dock your ligand with target protein structure (with inbuilt ligand optimization)', 'Dock your ligand with target protein structure (without ligand optimization)', and 'Docking with advanced features'. The right sidebar shows a 'History' panel with two entries: '3: Receptor.pdb' (4,912 lines, format: txt, database: ?) and '2: Ligand.sdf' (178 lines, format: txt, database: ?). Below the history, a table of coordinates is visible:

2VCH_1RX_A_1252						
RCR: PER0115150403n						
Coordinates from PDB:2VCH:A:1252 Model						
10	10	0	0	0	0	999 V2000
14.3370	56.6310	14.0060	C	0	0	
25.1320	55.7100	14.6650	C	0	0	

Callouts point to the 'Upload receptor' and 'Upload ligand' buttons, the 'Execute' button, and the 'Docking' options in the sidebar.

Figure 25

Galaxy / MPDS 1.0.1 Analyze Data Workflow Shared Data Visualization Help User Using 819.5 KB

Tools search tools

MPDS 1.0.1

3. Select your ligand

2. Select ligand file format

4. Enter Residue name from active site

5. Enter receptor chain ID for docking

7. Give your username

6. Enter residue number

8. Submit

Dock (version 1.0.0)

Receptor_file: 15: 2FUM.pdb

Select Ligand input format: pdb

Ligand_file: 5: CID_145823.sdf

Residue_name: ASN

Chain_id: A

Residue_number: 143

Username: Name

Execute

This program runs the Autodock Vina algorithm to simulate the complex formation between a receptor protein and a small molecule (ligand). The user needs to fill the form in which s/he needs to give the receptor protein in PDB format, ligand file in PDB or SDF format, give the residue name, Chain ID, residue number and his/her chosen username. When the tool is executed, the user will get a zip file where he will be provided all the files generated while the program runs, the complex files in PDB format and the Vina Log file that contains the ranked binding free energy scores.

History Using 819.5 KB

Unnamed history 819.5 KB

15: 2FUM.pdb

5: CID_145823.sdf

Figure 26a

Galaxy / MPDS 1.0.1 Analyze Data Workflow Shared Data Visualization Help User Using 5.3 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

Docking

Optimize input ligands

Generate_Conformers for your input ligand

Dock your ligand with target protein structure (with inbuilt ligand optimization)

Dock your ligand with target protein structure (without ligand optimization)

Docking with advanced features

Submitted job is processing

17: Docking Result

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 Using 5.3 MB

Unnamed history 5.3 MB

17: Docking Result

15: 2FUM.pdb

5: CID_145823.sdf

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.

Galaxy / OSDD-MPDS 1.0.1 Analyze Data Workflow Shared Data Visualization Help User Using 37.9 M

Tools

Convert Formats

Statistics

Graph/Display Data

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

Optimize input ligands

Generate Conformers for your input ligand

Docking with inbuilt ligand optimization

Docking without ligand optimization

Docking (version 1.0.0)

Receptor_file:
1: 2FUM.pdb

Ligand_file:
3: MIX.pdb

Residue_name:
ASN

Chain_id:
A

Residue_number:
143

Username:
ana

Execute

History

5: Docking result
71,563 lines
format: zip, database: 2

3: MIX.pdb

2: MIX.sdf

1: 2FUM.pdb

This program runs the Autodock Vina algorithm to simulate the complex formation between a receptor protein and a small molecule (ligand). The user needs to fill the form in which s/he needs to give the receptor protein in PDB format, ligand file in PDB format, give the residue name, Chain ID, residue number and his/her chosen username. When the tool is executed, the user will get a zip file where s/he

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 web interface. The main panel is titled "Docking (version 1.0.0)". The configuration fields are as follows:

- 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 panel. The right sidebar shows a "History" panel with two entries:

- 3: Receptor.pdb (4,912 lines, format: txt, database: ?) - uploaded txt file
- 2: Ligand.sdf (178 lines, format: txt, database: ?) - uploaded txt file

The "Ligand.sdf" entry shows a table of coordinates:

COORDINATES FROM PDB:2VCS:A:1252 Model
10 10 0 0 0 0
14.2370 56.6410 14.0060 C 0 0
15.1820 55.7400 14.6650 C 0 0

5.3. Screening

5.4.1. Descriptor Calculator

It Calculate descriptors for estimation of drug likeliness

Galaxy / MPDS 1.0.1 Analyze Data Workflow Shared Data Visualization Help User Using 9.9 MB

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

Descriptor Calculator Calculate descriptors for estimation of druglikeness

DruLiTo Apply filters for estimation of drug-likeness

Segregate Molecules for Futher 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

Descriptor Calculator (version 1.0.0)

Read data from your current history:
5: CID_145823.sdf
*.sdf file only

Execute

This tool processes sdf files for calculation of descriptors that are required for drug-likeness screening.

History

Unnamed history
9.9 MB

15: 2FUM.pdb

5: CID_145823.sdf

Galaxy / MPDS 1.0.1 Analyze Data Workflow Shared Data Visualization Help User Using 9.9 MB

Tools

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

Descriptor Calculator Calculate descriptors for estimation of druglikeness

DruLiTo Apply filters for estimation of drug-likeness

Segregate Molecules for Futher 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 molecule

The following job has been successfully added to the queue:
21: Descriptor Calculator result on CID_145823.sdf
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

21: Descriptor Calculator result on CID_145823.sdf
3 lines
format: tabular, database: 2

1	2	3	4	5
Descriptor Mol.Wt.	AlogP	KlogP	Mol.	
mol1	430.219	-3.733	-2.495	110.3

Source: /home/galaxy/galaxy-dist/datab

15: 2FUM.pdb

5: CID_145823.sdf

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.

The screenshot shows the Galaxy MPDS 1.0.1 interface. The main tool window is titled "Segregate Molecules for Futher Analysis (version 1.0.0)". The configuration includes the following options:

- Choose the DruLiTo output file: 41: DruLiTo on data 40
- All:
- Lipinski's Rule:
- Hose Filter:
- CMC-50-Like Rule:
- Veber Filter:
- MDDR Like Rule:
- BBB-Likeness:
- Unweighted QED:
- Weighted QED:

The "Execute" button is visible at the bottom of the configuration panel. The History panel on the right shows the execution of "41: DruLiTo on data 40" resulting in a tabular dataset with 10 lines. Below it, a table shows filters for Lipinski Rule and Hose Filter for molecules mol1 through mol5.

Filters	Lipinski Rule	Hose Filter
mol1	+	
mol2	+	
mol3	+	
mol4	+	
mol5	+	

The screenshot shows the Galaxy MPDS 1.0.1 interface with the "Segregate Molecules for Futher Analysis" tool configuration. The configuration options are the same as in the previous screenshot. The History panel shows the execution of "46: Negative Ligands" and "45: Positive Ligands". Below the configuration, a table displays the output data for the "OpenBabel02281722312D" dataset.

OpenBabel02281722312D
27 29 0 0 0 0 0 0 0 0 0999 V2000
4.6792 -14.5917 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
5.1667 -13.9167 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0
5.1667 -15.2542 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0
5.9500 -14.1750 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
5.9500 -15.0000 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
3.8625 -14.6792 0.0000 C 0 0 3 0 0 0 0 0 0 0 0 0
3.3375 -14.0417 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0
6.6625 -13.7625 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
6.6625 -15.4125 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
2.5750 -14.3417 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
3.4167 -15.3750 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
2.6167 -15.1625 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
7.3792 -15.0000 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
7.3792 -14.1750 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
4.8651 -16.1914 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0
4.2669 -13.6886 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0
3.5933 -13.0543 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0
6.6612 -12.7305 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0
6.6612 -16.4445 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0
1.6757 -13.7620 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0
1.6488 -14.8775 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0
3.8043 -16.3723 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0
2.7033 -16.1724 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0
1.7853 -15.8361 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0
1.6395 -14.7266 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0
8.2726 -15.5166 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0
8.2726 -13.6584 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0
2 1 2 0 0 0 0
3 1 1 0 0 0 0
3 15 1 0 0 0 0
4 2 1 0 0 0 0
4 5 2 0 0 0 0
5 3 1 0 0 0 0
6 1 1 0 0 0 0
6 16 1 0 0 0 0
7 6 1 0 0 0 0
7 17 1 0 0 0 0

5.4.3. BCS Classification

Identify the BCS class to which the molecule belongs

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).

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

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

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

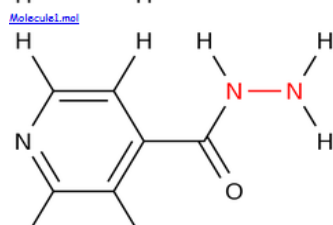
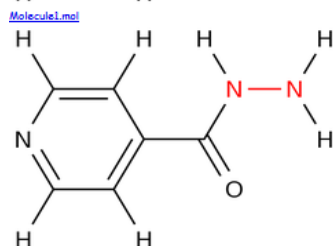
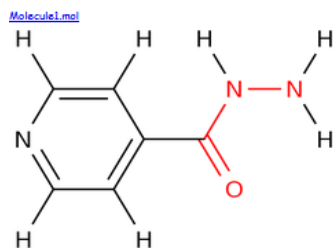
MPDS Toxicity Filter - Help Page

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 tools under categories like 'DATA LIBRARIES', 'DATA PROCESSING', 'DATA ANALYSIS', and 'GALAXY INBUILT'. The 'Visualization' section is expanded, showing '3D Visualization 3D Visualization by Jmol'. The main 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. A 'History' panel on the right shows a list of jobs, including '15: 2FUM.pdb' and '5: CID_145823.sdf'. Three callout boxes with arrows point to specific elements: '1. Click' points to the 'Visualization' option in the sidebar; '2. Select input' points to the dropdown menu; 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

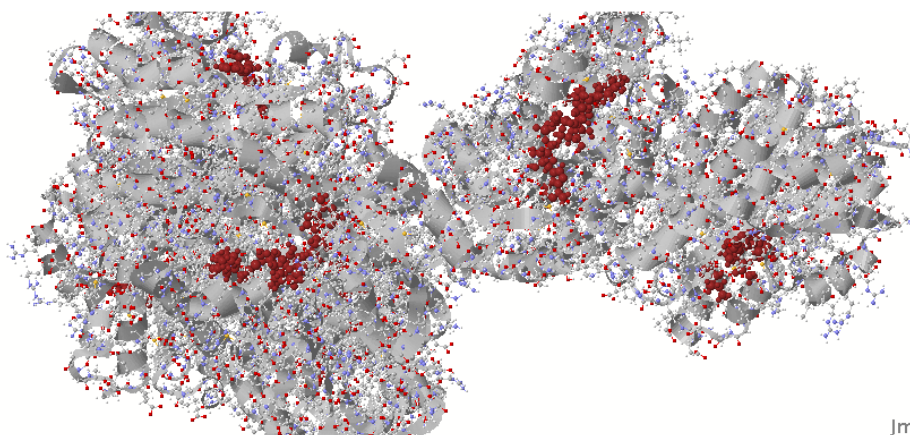
15: 2FUM.pdb

5: 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](#)



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	18VR COMPLIES WITH FORMAT V
REMARK	888	
TITLE	M. TB. ENOYL-ACP REDUCTASE (
TITLE	2	ACVL-SUBSTRATE
EXPOTA	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

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

Docking

Screen

Visualization

3D Visualization 3D Visualization by Jmol

Generate Ligplot plots an interaction between protein-ligand

1. Click

Generate Ligplot (version 1.0.0)

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

2. Select the input

enter residue1 id for ligand:
1539

3. Enter the residue id

enter residue2 id for ligand:
8

Enter Chain Id:
Z

4. Enter the chain id

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

5. Submit job

Execute

This tool generates ligplot to visualize ligand-protein interaction

History

Unnamed history
9.9 MB

15: 2FUM.pdb

5: CID_145823.sdf

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 Circoos generates a circoos 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

Ligplot Diagram

Generate Ligplot

792.4 KB

2: Generate Ligplot on data_1

1: 2FUM.pdb

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)

Figure 36

5.5. Drug repurposing tool

Galaxy / MPDS-DM Analyze Data Workflow Shared Data Visualization Help User

Tools

search tools

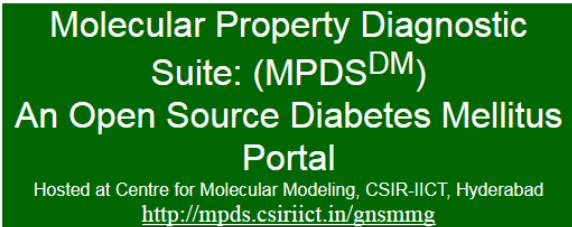
MPDS 1.0.1
[Get Data](#)

DATA LIBRARIES
[Literature](#)
[Target Library](#)
[Gene Library](#)
[Compound Library](#)

DATA PROCESSING
[File-Format Converter](#)
[Descriptor Calculation](#)

DATA ANALYSIS
[QSAR](#)
[Docking](#)
[Screening](#)
[Drug Repurposing Tool](#)

[PASS online](#) server for drug repurposing



[HOME](#) [RELATED LINKS](#) [MANUAL](#) [CONTACT](#)

The disease-specific web portal has the data updated on 15th October 2017
 Next Update: 1st September 2018

PASSonline Nagamani Selvaraman (Log out) Go

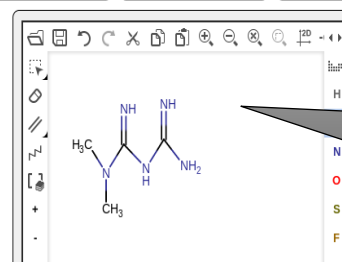
» Home | » Definition | » Products | » Services | » FAQ | » Contacts

Paste
SMILES

[Predict new compound](#) [View old results](#) [View/change profile](#)

[SMILES](#) [MOL file](#) [Molvis JS](#)

Draw the
structure



Way2Drug PREDICTIVE SERVICES
 Understanding Chemical-Biological Interactions

Nagamani Selvaraman (Log out) Go

» Home | » Definition | » Products | » Services | » FAQ | » Contacts

PASS online Better solutions for your research and development
 It is easy to use

[GO for prediction >](#)

Click
here

All Pa>Pi Pa>0,3 Pa>0,7

Pa	Pi	Activity
0,939	0,001	Protein-arginine deiminase inhibitor
0,898	0,010	CDP-glycerol glycerophosphotrar inhibitor
0,847	0,006	NADPH peroxidase inhibitor
0,844	0,009	Pro-opiomelanocortin converting inhibitor
0,837	0,005	Omptin inhibitor
0,836	0,004	Antischematic
0,836	0,006	Arginine 2-monooxygenase inhibi
0,827	0,004	Lipid-lowering factor B inhibitor

Output