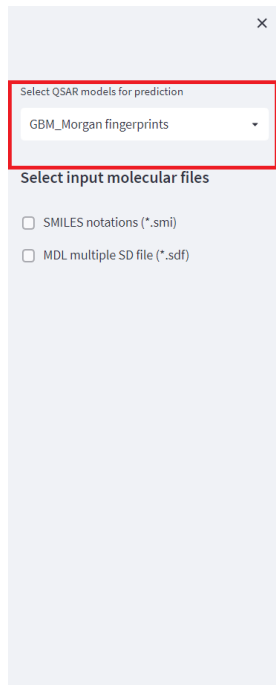


Manual of HDAC6 DETECTOR

open-source software

Step 1. *Select QSAR models for prediction.*

You can choose a model developed using the Morgan fingerprints and gradient boosting method (GBM) or a model developed using topological fingerprints, and multilayer perceptron (MLP) classifier.



Select QSAR models for prediction

GBM_Morgan fingerprints

Select input molecular files

☐ SMILES notations (*.smi)

☐ MDL multiple SD file (*.sdf)

HDAC6 DETECTOR

The application provides an alternative method for assessing the potential of chemicals to be Histone deacetylase 6 (HDAC6) inhibitors.

Machine learning



The models were developed using open-source chemical descriptors based on Morgan and topological fingerprints, along with the gradient boosting method.

OECD principles



We follow the best practices for model development and validation recommended by guidelines of the Organization for Economic

Predictive power



The QSAR models showed high predictive power with vigorous validation metrics for test sets, achieving accuracy, sensitivity, and specificity ranging

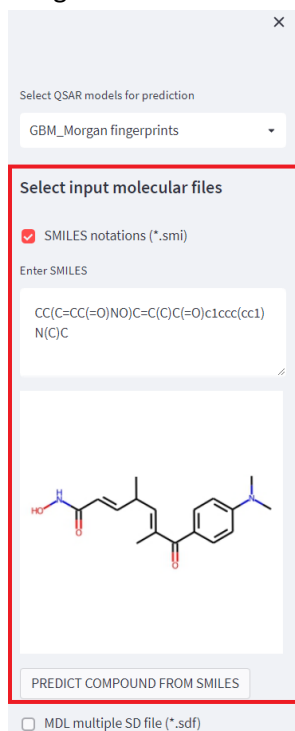
Lipinski's rule



Estimating the bioavailability of a compound is an important factor in drug development. Lipinski's rule of five was introduced to

Step 2. *Select input molecular files.*

If you choose smiles, please, directly paste the SMILES representation of the desired chemical structure and press Ctrl+Enter. If the entered chemical structure is correct, the application will generate a 2D image of the studied compound. Click on the "PREDICT COMPOUND FROM SMILES" button.



Select QSAR models for prediction

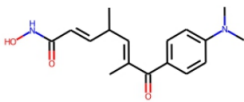
GBM_Morgan fingerprints

Select input molecular files

☒ SMILES notations (*.smi)

Enter SMILES

CC(C=CC(=O)NO)C=C(C)C(=O)c1ccc(cc1)N(C)C



PREDICT COMPOUND FROM SMILES

☐ MDL multiple SD file (*.sdf)

HDAC6 DETECTOR

The application provides an alternative method for assessing the potential of chemicals to be Histone deacetylase 6 (HDAC6) inhibitors.

Machine learning



The models were developed using open-source chemical descriptors based on Morgan and topological fingerprints, along with the gradient boosting method.

OECD principles



We follow the best practices for model development and validation recommended by guidelines of the Organization for Economic

Predictive power



The QSAR models showed high predictive power with vigorous validation metrics for test sets, achieving accuracy, sensitivity, and specificity ranging

Lipinski's rule



Estimating the bioavailability of a compound is an important factor in drug development. Lipinski's rule of five was introduced to

If the entered structure is incorrect, the application reports an error that has occurred.

Select QSAR models for prediction
GBM_Morgan fingerprints

Select input molecular files

☒ SMILES notations (*.smi)

Enter SMILES

COC1=C/C2=C(OC3=CC=C(NC(=O)C4=C(O)C=CN(C5=CC=CC=C5)=N4)C=C3)C=CN=C2/C=C\1OCCCCCCC(=O)NO

build on curated datasets generated from scientific articles. The models were developed using open-source chemical

Organization for Economic Cooperation and Development (OECD). For assessment of the applicability domain, we used

accuracy, sensitivity, and specificity ranging from 88 to 100%. The predictive ability of the developed QSAR models is further

radar is displayed for a quick assessment of the compliance of the tested compound with the Lipinski rules.

Click to download brief manual

rdkit.Chem.rdchem.AtomValenceException: This app has encountered an error. The original error message is redacted to prevent data leaks. Full error details have been recorded in the logs (if you're on Streamlit Cloud, click on 'Manage app' in the lower right of your app).

Traceback:

```

File ~/home/appuser/venv/lib/python3.9/site-packages/streamlit/runtime/script_runner.py, line 441, in exec(code, module.__dict__)
File ~/app/hdac6_detector/HDAC6_detector_app.py, line 183, in <module>
    smiles=standardize_smiles(compound_smiles)
File ~/home/appuser/venv/lib/python3.9/site-packages/molvs/standardize.py, line 11, in standardize
    mol = Standardizer().standardize(mol)
File ~/home/appuser/venv/lib/python3.9/site-packages/molvs/standardize.py, line 11, in standardize
    Chem.SanitizeMol(mol)

```

If you choose a file *.sdf, that may contain a different number of chemical structures, please specify the path to this file on your computer's hard drive. In this case, you need to click the " Browse files" button.

Select QSAR models for prediction
GBM_Morgan fingerprints

Select input molecular files

☐ SMILES notations (*.smi)

☒ MDL multiple SD file (*.sdf)

Choose a file

Drag and drop file here
Limit 200MB per file

Browse files

example.sdf
13.9KB

HDAC6 DETECTOR

The application provides an alternative method for assessing the potential of chemicals to be Histone deacetylase 6 (HDAC6) inhibitors.

Machine learning



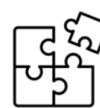
OECD principles



Predictive power



Lipinski's rule



The models were developed using open-source chemical descriptors based on Morgan and topological fingerprints, along with the gradient

We follow the best practices for model development and validation recommended by guidelines of the Organization for

ability of the developed QSAR models is further confirmed by our synthesis of new HDAC6 inhibitors and in vivo studies. The application HDAC6

development. Lipinski's rule of five was introduced to estimate the oral bioavailability of a compound. Our bioavailability radar is displayed

Step 3. *Prediction results.*

Select QSAR models for prediction

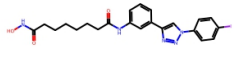
GBM_Morgan fingerprints

Select input molecular files

☒ SMILES notations (*.smi)

Enter SMILES

ONC(=O)CCCCCCC(=O)Nc1cccc(c1)-c1cn(nn1)-c1ccc(l)cc1



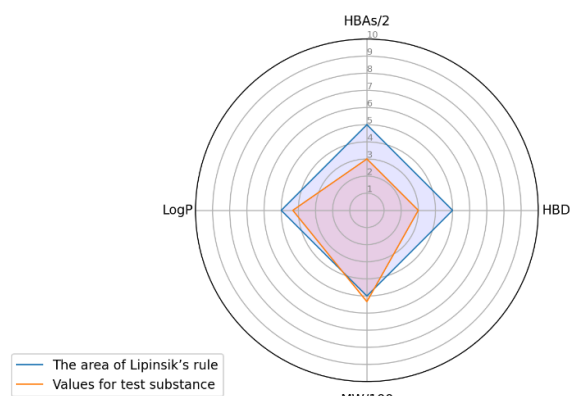
PREDICT COMPOUND FROM SMILES

☐ MDL multiple SD file (*.sdf)

Prediction results:

	SMILES	Predicted value, pIC ₅₀	Applicability domain	Experimental value, pIC ₅₀	Standard deviation	ChEMBL ID
1	<chem>O=C(CCCCCC(=O)Nc1cccc(-c2cn(-c3ccc(l)cc3)nn2)c1)NO</chem>	see experimental value	-	8.66	0.0	ChEMBL1091475

The Bioavailability Radar: compliance the Lipinski's rule of five



The final table contains the following columns:

- 1) **SMILES** - the chemical structure is displayed in the SMILES notation
- 2) **Predicted value, pIC₅₀** - the predicted value of the activity to inhibit the HDAC6 enzyme, expressed in pIC₅₀, where pIC₅₀ is the negative logarithm of IC₅₀ in molar concentration. If experimental data is available in the ChEMBL database, the label "see experimental value" is displayed in this cell.
- 3) **Applicability domain** - compliance of the chemical compound with Applicability domain. If experimental data is available in the ChEMBL database, the label "-" is displayed in this cell.
- 4) **Experimental value, pIC₅₀** - experimental data presented in the ChEMBL database. Where pIC₅₀ is the negative logarithm of IC₅₀ in molar concentration. If there is more than one value in the database, then the average value is given in the cell.
- 5) **Standard deviation** - the cell indicates the standard deviation of the experimental activity values presented in the ChEMBL database.
- 6) **ChEMBL ID** - identifier from the ChEMBL chemical database of molecule

The form of presentation of the results depends on the type of descriptors selected, as well as the format of the input chemical data. For example, when selecting SMILES, the results will be displayed for a single molecule. The chemical fragments are colored in green (predicted to reduce inhibitory activity) or magenta (predicted to increase activity HDAC6 inhibitors). The gray isolines separate positive and negative contributions.

Select QSAR models for prediction

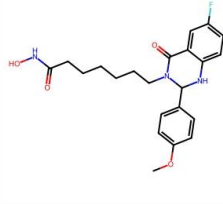
GBM_Morgan fingerprints

Select input molecular files

☒ SMILES notations (*.smi)

Enter SMILES

COc1ccc(C2Nc3ccc(F)cc3C(=O)N2CCCCC.CC(=O)NO)cc1



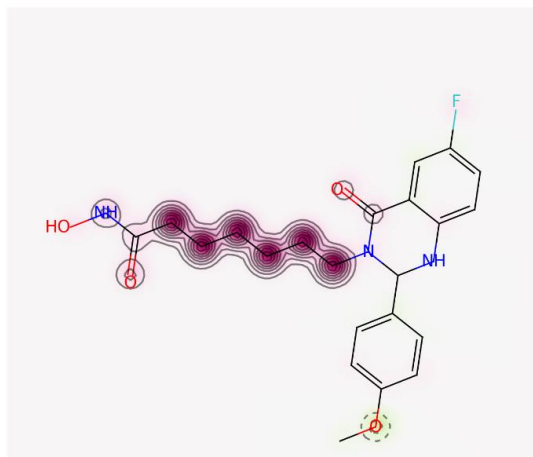
PREDICT COMPOUND FROM SMILES

☐ MDL multiple SD file (*.sdf)

Prediction results:

	SMILES	Predicted value, pIC50	Applicability domain	Experimental value, pIC50	Standard deviation	ChEMBL ID
1	<chem>COc1ccc(C2Nc3ccc(F)cc3C(=O)N2CCCCC.CC(=O)NO)cc1</chem>	Active	Inside AD	-	-	not detected

Predicted fragments contribution:



If you select a file *.sdf, the results will be presented in tabular form. If incorrect structures are detected in the file *.sdf, the corresponding information will appear in the section "CHEMICAL STRUCTURE VALIDATION AND STANDARDIZATION"

1. CHEMICAL STRUCTURE VALIDATION AND STANDARDIZATION:

Original data: 211 molecules

Failed data: 9 molecules

	No. failed molecule in original set	SMILES of wrong structure:
1	80	<chem>COC1=C/C2=C(OC3=CC=C(NC(=O)C4=C(=O)C=CN(C5=CC=CC=C5)=N4)C=C3)C=C</chem>
2	88	<chem>O=C/C=C/C1=CC=C(C(CNC(=O)C2=CC(C/C3=N/N=C(=O)C4=CC=CC=C43)=CC=C</chem>
3	90	<chem>COC1=C\C(OC)=C2\C(=O)=NC(C3=CC(C)=C(OCCCCC(=O)NO)C(C)=C3)=N\C2=C</chem>
4	92	<chem>COC1=C/C2=C(OC3=CC=C(NC(=O)C4=C(=O)C(C)=CN(C5=CC=CC=C5)=N4)C=C3</chem>
5	93	<chem>CCC1=CN(C2=CC=CC=C2)=NC(C(=O)NC2=CC=C(O/C3=C/C=N\C4=CC(OCCCCC</chem>
6	153	<chem>O=C(NO)C1=CC=C(CN2CCN(C(=O)C3=CC(C/C4=N/N=C(=O)C5=CC=CC=C54)=C</chem>
7	158	<chem>O=C(CC1=CC=C(CN2CCN(C(=O)C3=CC(C/C4=N/N=C(=O)C5=CC=CC=C54)=CC=</chem>
8	193	<chem>O=C/C=C/C1=CC=C(CN2CCN(C(=O)C3=CC(C/C4=N/N=C(=O)C5=CC=CC=C54)=</chem>
9	203	<chem>COC1=C\C(OC)=C2\C(=O)=NC(C3=CC(C)=C(OCCC4=CC=C/C=C/C(=O)NO)C=C4)</chem>

Kept data: 202 molecules

The total number of compounds which have experimental values: 80

Total number of active molecules included in AD: 17

Total number of inactive molecules included in AD: 26

Total number of molecules not included in AD: 79

The prediction results for correct chemical structures are displayed in a table that can be downloaded, or separately for each molecule if you click the " Show results and map of fragments contribution for each molecule separately" button.

GBM_Morgan fingerprints ▾

Select input molecular files


☐ SMILES notations (*.smi)

☒ MDL multiple SD file (*.sdf)

Choose a file

Drag and drop file here
Limit 200MB per file

Browse files

 HDAC6_vs_from_patents.sdf X
33.3KB

Total number of inactive molecules included in AD: 3

Total number of molecules not included in AD: 3

Show results as table

	SMILES	HDAC6 activity	Applicability domain (f
1	<chem>COc1cc2ncn(OCCCCC(=O)NO)c(=O)c2cc1OC</chem>	Inactive	Inside AD
2	<chem>COc1cc2ncn(OCCCCCCC(=O)NO)c(=O)c2cc1OC</chem>	Inactive	Inside AD
3	<chem>O=C(CCCCCOn1cnc2ccc(Br)cc2c1=O)NO</chem>	unreliable prediction	Outside AD
4	<chem>O=C(CCCCCOn1cnc2ccc(Br)cc2c1=O)NO</chem>	unreliable prediction	Outside AD
5	<chem>COc1ccc(C2Nc3cccc3C(=O)N2CCCCC(=O)NO)cc1</chem>	Active	Inside AD
6	<chem>COc1ccc(C2Nc3cccc3C(=O)N2CCCCC(=O)NO)c1</chem>	Active	Inside AD
7	<chem>COc1cc(C2Nc3cccc3C(=O)N2CCCCC(=O)NO)cc(OC)c1OC</chem>	Inactive	Inside AD
8	<chem>O=C(CCCCCN1C(=O)c2cccc2NC1c1ccc(Cl)cc1)NO</chem>	Active	Inside AD
9	<chem>COc1ccc(C2Nc3cccc3C(=O)N2CCCCC(=O)NO)cc1COc1cccc1F</chem>	unreliable prediction	Outside AD
10	<chem>COc1ccc(C2Nc3cccc3C(=O)N2CCCCC(=O)NO)cc1</chem>	Active	Inside AD

Download results of prediction as CSV

Show results, bioavailability radar, map of fragments contribution and for each molecule separately

© Oleg Tinkov, 2023