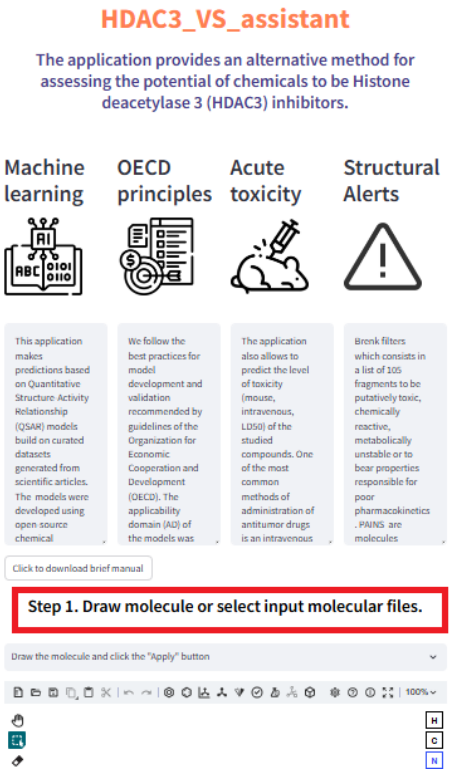
Manual of HDAC3_VS_assistant

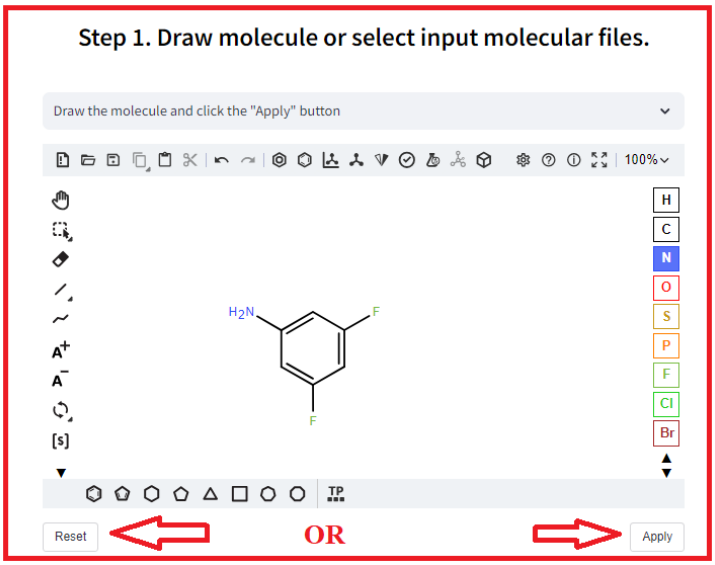
open-source software

Appearance of the main window of the HDAC3_VS_assistant web application:

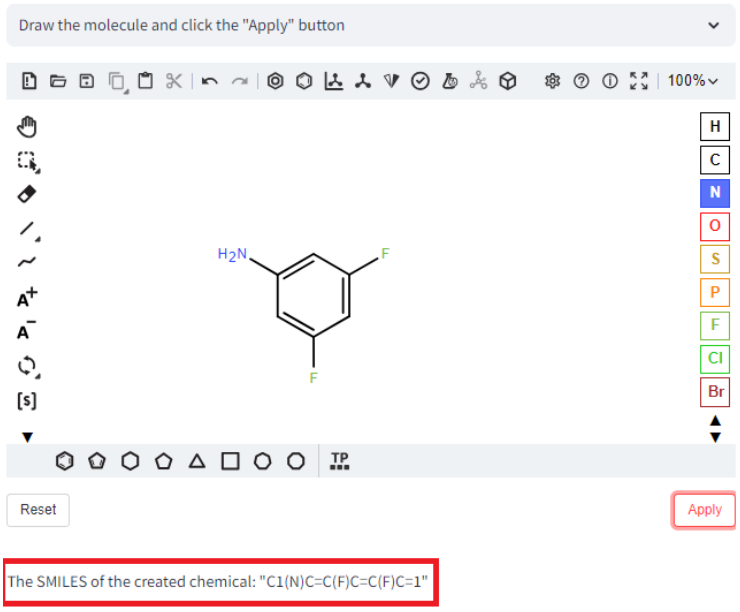


Step 1. *Draw molecule or select input molecular files.*

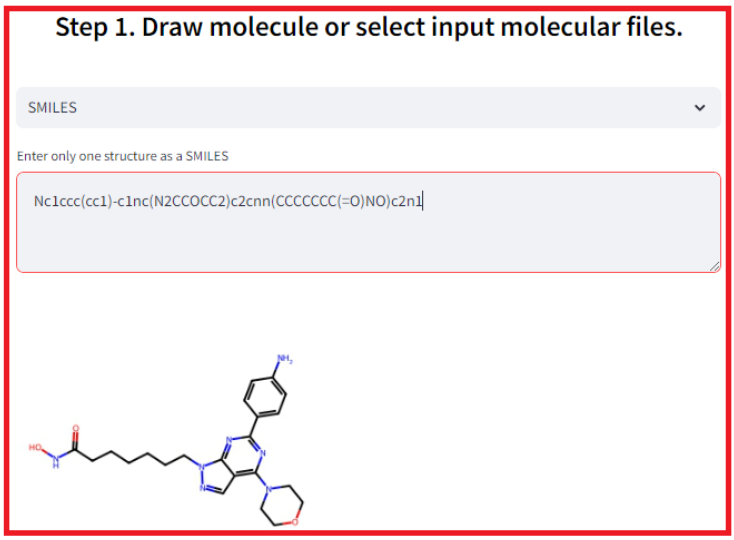
If you want to draw the structure of a chemical compound, you can use the capabilities of the built-in chemical editor Ketcher (https://github.com/epam/ketcher). After creating the structure of a chemical compound, it can be controlled using two buttons: 1)"reset" - deleting the structure to create a new one 2) "Apply" - transferring the structure of the compound for further analysis, forecasting to step 2.



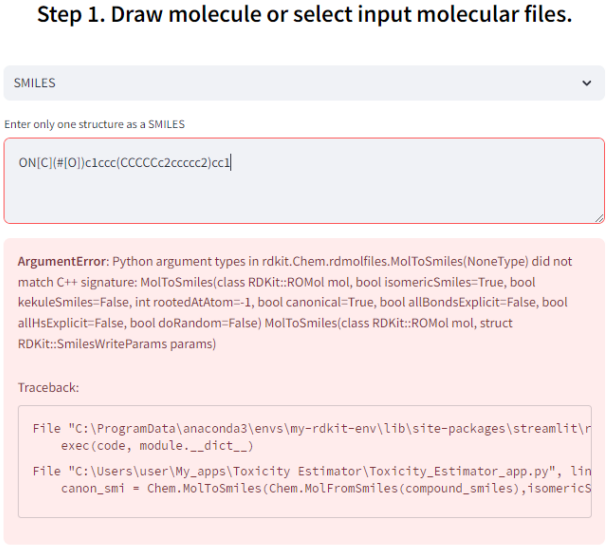
After clicking on the 'Apply' button, make sure that the structure has been created. If the structure is successfully created, its smiles will be displayed under the chemical editor window



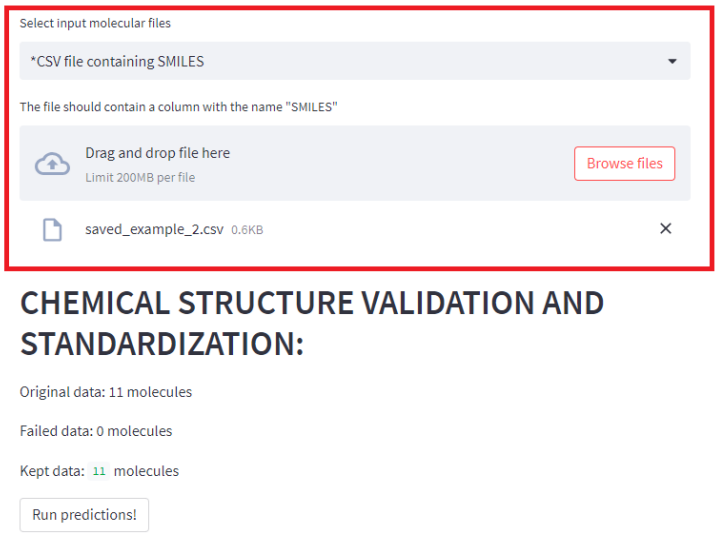
If you choose smiles, please, directly paste the SMILES representation of the desired chemical structure and press Ctrl+Enter. If the entered chemical structure is correct, the application will generate a 2D image of the studied compound. DON'T FORGET TO CLICK THE "APPLY" BUTTON



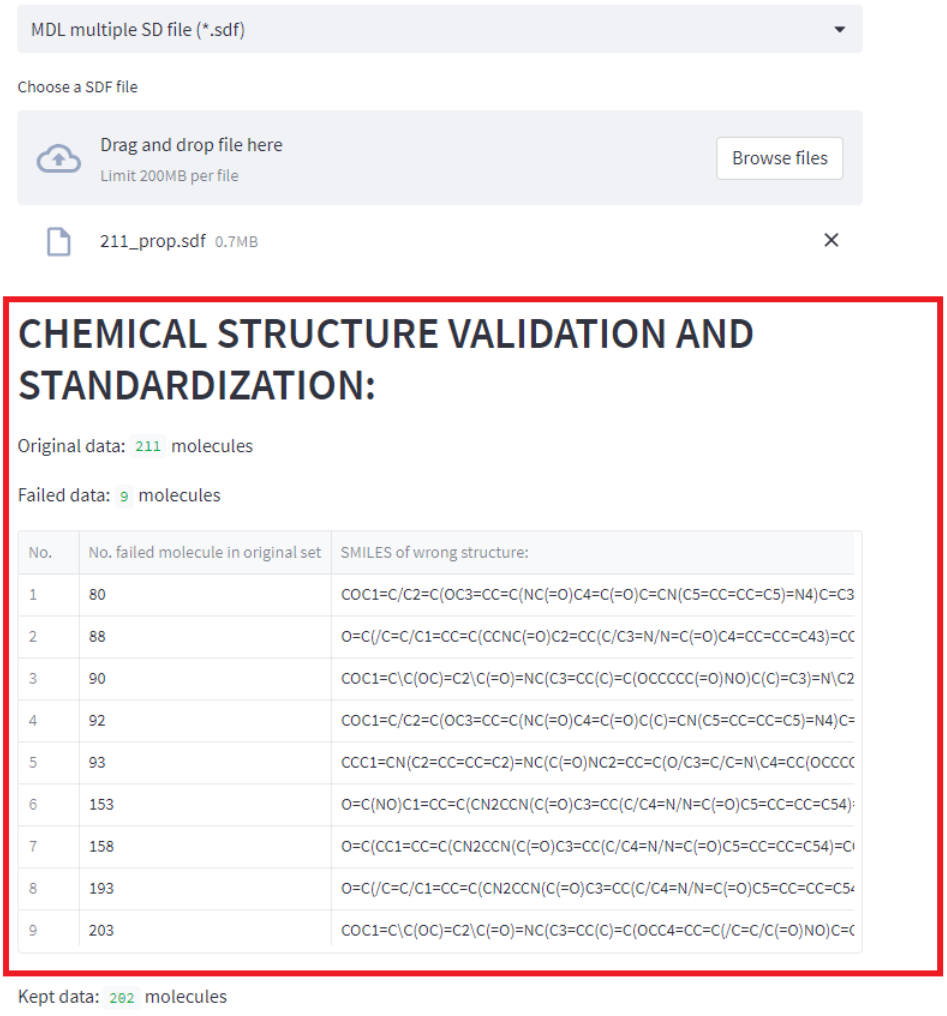
If the entered structure is incorrect, the application reports an error.



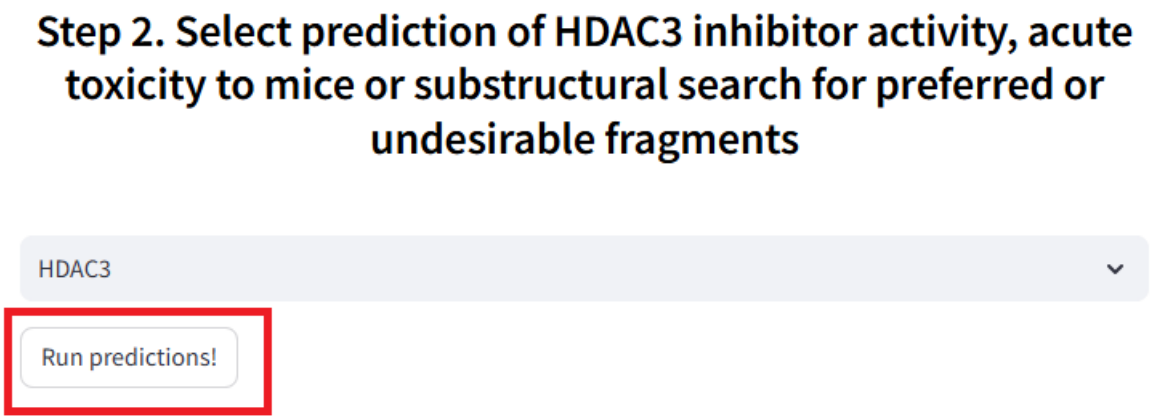
If you choose a file *sdf or *csv, that may contain a different number of chemical structures, please specify the path to this file on your computer's hard drive. In this case, you need to click the "Browse files" button. It is important to note that if you choose a file with the *csv extension, the file should contain a column with the name "SMILES"



If incorrect structures are detected in the file *sdf or *csv, the corresponding information will appear in the section "CHEMICAL STRUCTURE VALIDATION AND STANDARDIZATION"



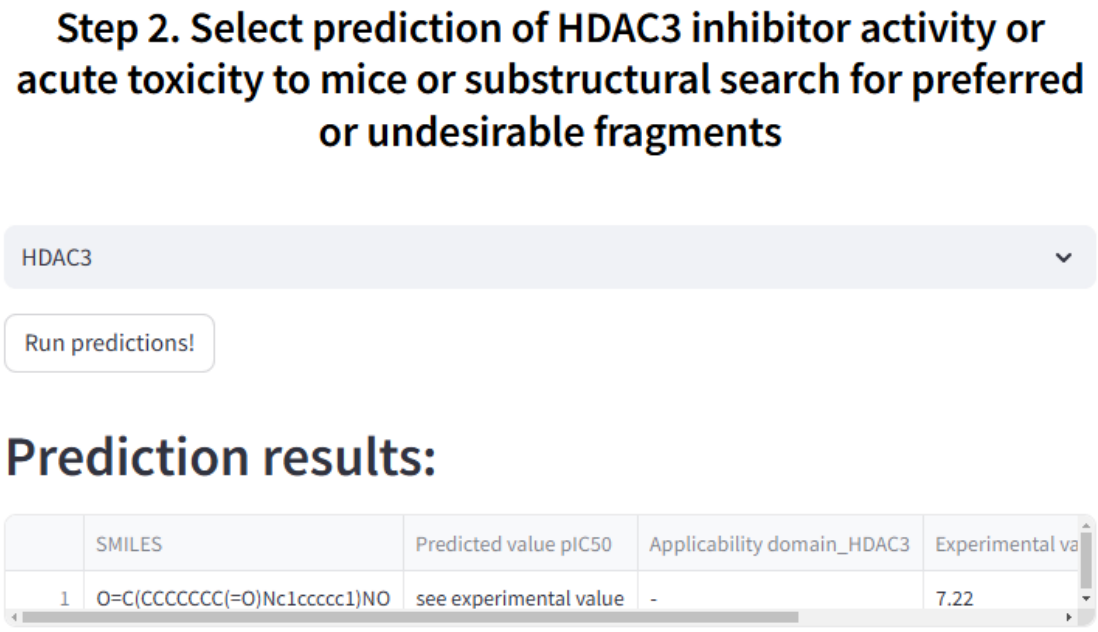
Step 2. *Select prediction of HDAC3 inhibitor activity or acute toxicity to mice or substructural search for preferred or undesirable fragments.*



Click on the “Run predictions!” button for prediction.

Step 3. *Prediction results*.

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. When displaying the results on the screen, it is taken into account whether there are experimental values of activity and toxicity for the studied compound.



The final table ‘Prediction results’ 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 HDAC3 enzyme, expressed in pIC_{50}, where pIC_{50} is the negative logarithm of IC_{50} 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_HDAC3*** - 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 pIC_{50} is the negative logarithm of IC_{50} in molar concentration. If there is more than one value in the database, then the average value is given in the cell.

5) ***STD*** - ehe cell indicates the standard deviation of the experimental activity values presented in the ChEMBL database.

6) ***chemble_ID*** - identifier from the ChEMBL chemical database of molecule

7) **Predicted value toxicity, mice, intravenous, Ld50, *mg/kg*** - predicted value of acute toxicity when administered intravenous to mouse. If experimental data is available in the PubChem database, the label "see experimental value" is displayed in this cell.

8) ***Applicability domain_tox*** - compliance of the chemical compound with Applicability domain. If experimental data is available in the PubChem database, the label "-" is displayed in this cell.

9) ***Experimental* value toxicity, mice, intravenous, Ld50, *mg/kg*** - experimental data presented in the PubChem database. Toxicity was measured by a dose of LD_{50} when administered orally to rats

10) ***CAS number*** - a unique identification number assigned by the Chemical Abstracts Service (CAS)

If you choose a file *sdf or *csv, the prediction results for correct chemical structures are displayed in a table that can be downloaded.

Original data: 11 molecules

Failed data: 0 molecules

Kept data: 11 molecules

Run predictions!

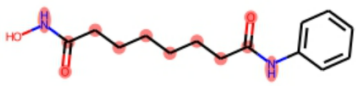
	SMILES	Predicted value, pIC50	Ap
2	CCOc1cc(NC(=O)CCCCCCC(=O)Nc2cc(-c3ccccc3)ccc2N)cc2c1OCC([N+](=O)[O-])=C2	see experimental value	-
3	[N-]=[N+]=NCCC(=O)Nc1ccc(C(=O)Nc2cc(N=[N+]=[N-])ccc2N)cc1	see experimental value	-
4	[N-]=[N+]=Nc1cccc(COC(=O)NCc2ccc(C(=O)Nc3cc(-c4ccccc4)ccc3N)cc2)c1	see experimental value	-
5	[N-]=[N+]=NCc1cc(N=[N+]=[N-])cc(C(=O)Nc2cc(-c3cccs3)ccc2N)c1	see experimental value	-
6	C#CCOc1ccc(C=NNC(=O)c2cccc(C(=O)NO)c2)cc1	see experimental value	-
7	O=C(NO)c1cnc(NC2(c3ccccc3)CCC(F)(F)CC2)nc1	see experimental value	-
8	CN1CCc2c(c3ccccc3n2Cc2ccc(C(=O)NO)cc2)C1	see experimental value	-
9	CCCNNC(=O)c1ccc(CNC(=O)C(Cc2c[nH]c3ccccc23)NC(=O)c2ccc(OC)cc2)cc1	see experimental value	-
10	COc1ccc(C(=O)c2ccc3c(ccn3Cc3ccc(C=CC(=O)NO)cc3)c2)cc1	see experimental value	-
11	CC(C=CC(=O)NO)=CC(C)C(=O)c1ccc(N(C)C)cc1	see experimental value	-

Download results of prediction as CSV

When analyzing single substances, a number of additional options are implemented in the HDAC3_VS_assistant web application. In particular, a set of substructural fragments was formed that steadily increase the activity of HDAC3 inhibitors. When entering structural information through SMILES notations or through the built-in chemical editor, automatic screening is performed for compounds to identify the most substructural fragments that increase the activity to inhibit HDAC3 and the Tanimoto index is calculated, determined by calculating the topological fingerprint of the studied compound and the identified substructural fragment. The results of a substructural search in the HD AC3_VS_assistant web application for Trichostatin (A-F) and Vorinostat (G) are presented below:

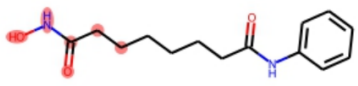
A) The found fragment: Klekota-Roth_1

Tanimoto coefficient: 0.38



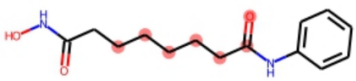
B) The found fragment: Klekota-Roth_2

Tanimoto coefficient: 0.21



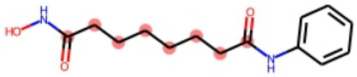
C) The found fragment: Klekota-Roth_3

Tanimoto coefficient: 0.33



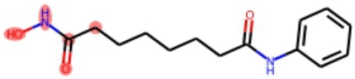
D) The found fragment: Klekota-Roth_4

Tanimoto coefficient: 0.16



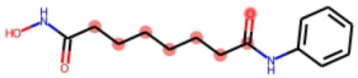
E) The found fragment: Klekota-Roth_8

Tanimoto coefficient: 0.35



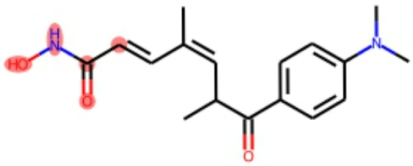
F) The found fragment: Klekota-Roth_9

Tanimoto coefficient: 0.38



G) The found fragment: Klekota-Roth_8

Tanimoto coefficient: 0.34



In addition, the HDAC3_VS_assistant web application integrates detection of such widely used in medicinal chemistry structural alerts as Brenk filters and PAINS when analyzing single structures. Brenk filters which consists in a list of 105 fragments identified by Brenk et al. [Brenk, R. et al. Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. ChemMedChem 3, 435–444 (2008).] to be putatively toxic, chemically reactive, metabolically unstable or to bear properties responsible for poor pharmacokinetics. PAINS (for pan assay interference compounds, a.k.a. frequent

hitters or promiscuous compounds) are molecules containing substructures showing potent response in assays irrespective of the protein target. Such fragments, yielding false positive biological output, have been identified by Baell et al. [Baell, J. B. & Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. J. Med. Chem. 53, 2719–2740 (2010). HDAC3_VS_assistant returns warnings if such moieties are found in the molecule under evaluation.