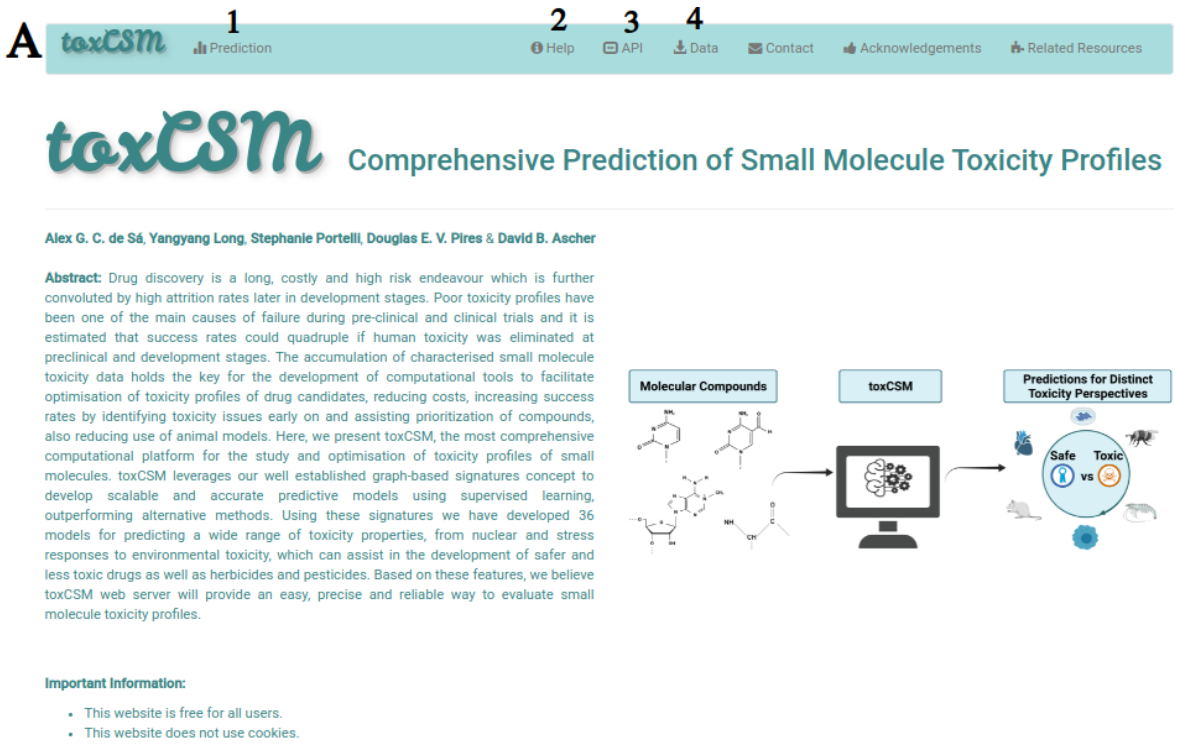
toxCSM: comprehensive prediction of small molecule toxicity profiles

Alex G. C. de Sá[1,2,3,4], Yangyang Long[2,3,5], Stephanie Portelli[1,2,3],
Douglas E. V. Pires[2,3,5], David B. Ascher[1,2,3,4]

alex.desa@unimelb.edu.au; douglas.pires@unimelb.edu.au; david.ascher@unimelb.edu.au

[1]School of Chemistry and Molecular Biosciences, University of Queensland
[2]Systems and Computational Biology, Bio21 Institute, University of Melbourne
[3]Computational Biology and Clinical Informatics, Baker Heart and Diabetes Institute
[4]Baker Department of Cardiometabolic Health, University of Melbourne
[5]School of Computing and Information Systems, University of Melbourne

Main Page

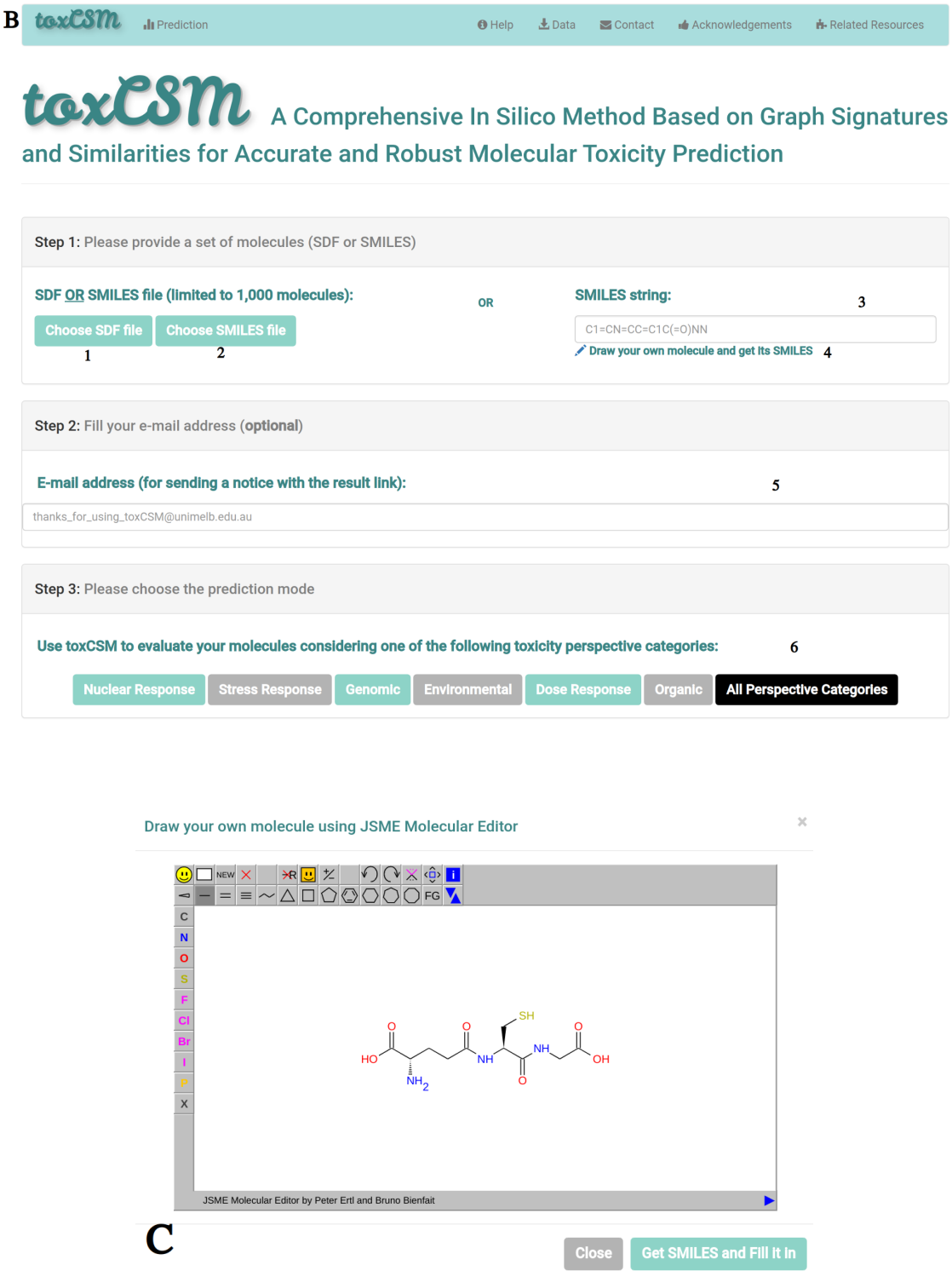


About toxCSM

toxCSM is a robust machine learning method that relies on graph-based signatures, molecular descriptors and toxic-active molecular similarities to predict small molecule toxicity profiles. Currently, toxCSM is the most comprehensive method in the literature, encompassing 36 different endpoint properties, which vary from nuclear response to environmental chemical activity.

A depicts the main page of toxCSM. Users are directed to the job submission page by clicking on "**Prediction**" at the top menu (**1**). Users can also access this help page on "**Help" (2),** how to use the **"API" (3),** and all data used in toxCSM experiments by clicking on "**Data**" at the top menu (**4).**

Job Submission Page



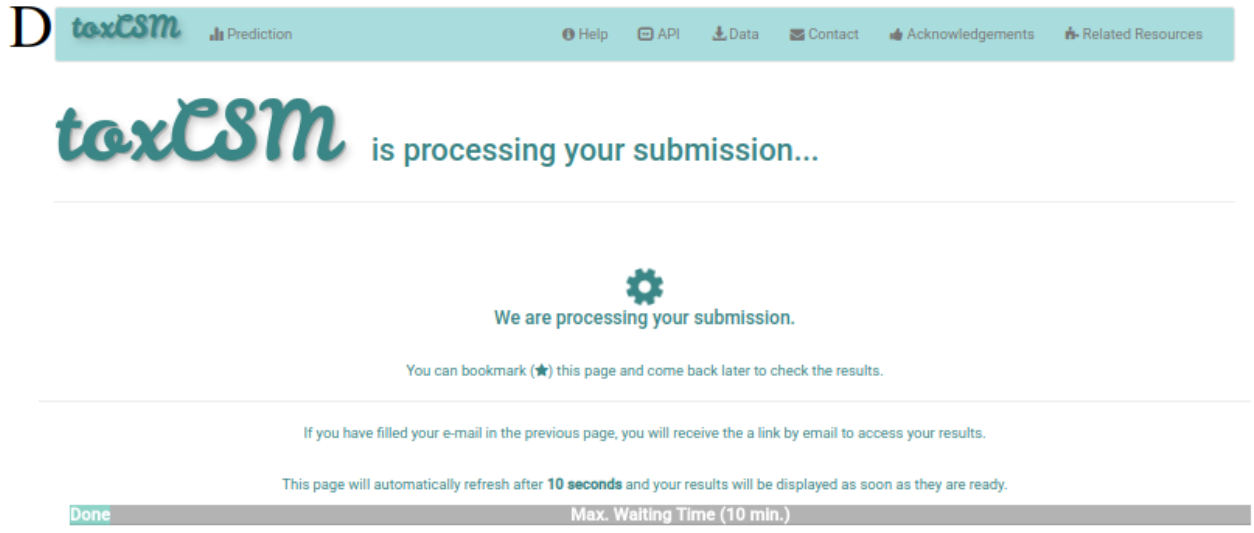
B describes the job submission page, where users can either submit an SDF (**1**), a SMILES file **(2)** or an individual SMILES string **(3).** For the SMILES string, users can also draw their molecule using JSME molecular editor and retrieve its SMILES, as shown at **C.** It is worth noting when the user clicks on **"Get SMILES and Fill It In"** at **C,** the SMILES string **(3)** at **B** will be filled with the respective SMILES of the drawn molecule.

Besides, if users want to, they can provide an email address **(5)** at the submission form.

After these steps, the user can choose among several toxicity categories to run predictions on **(6)**, including Nuclear Response, Stress Response, Genomic, Environmental, Dose Response, Organic and All Perspective Categories. Users can also run an example by clicking in "**Run Example" (7).**

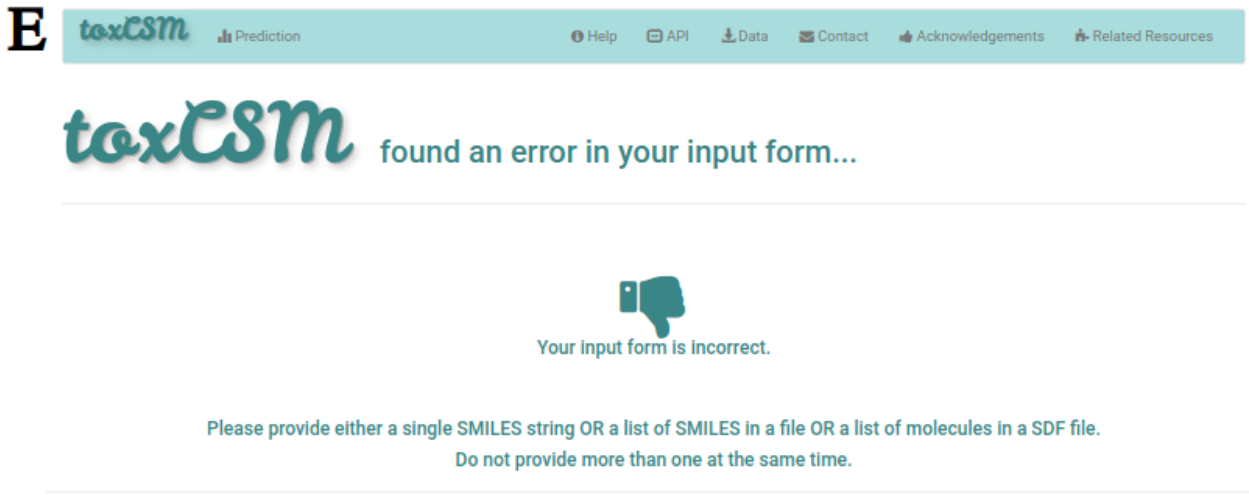
Waiting Page

When the user clicks on any perspective category option at **B (6 or 7),** he/she will be redirected to a waiting page while the job is being run **(D)**. If the email address was filled on the previous page, the user can wait for an email with the link to access the results.



Error Page

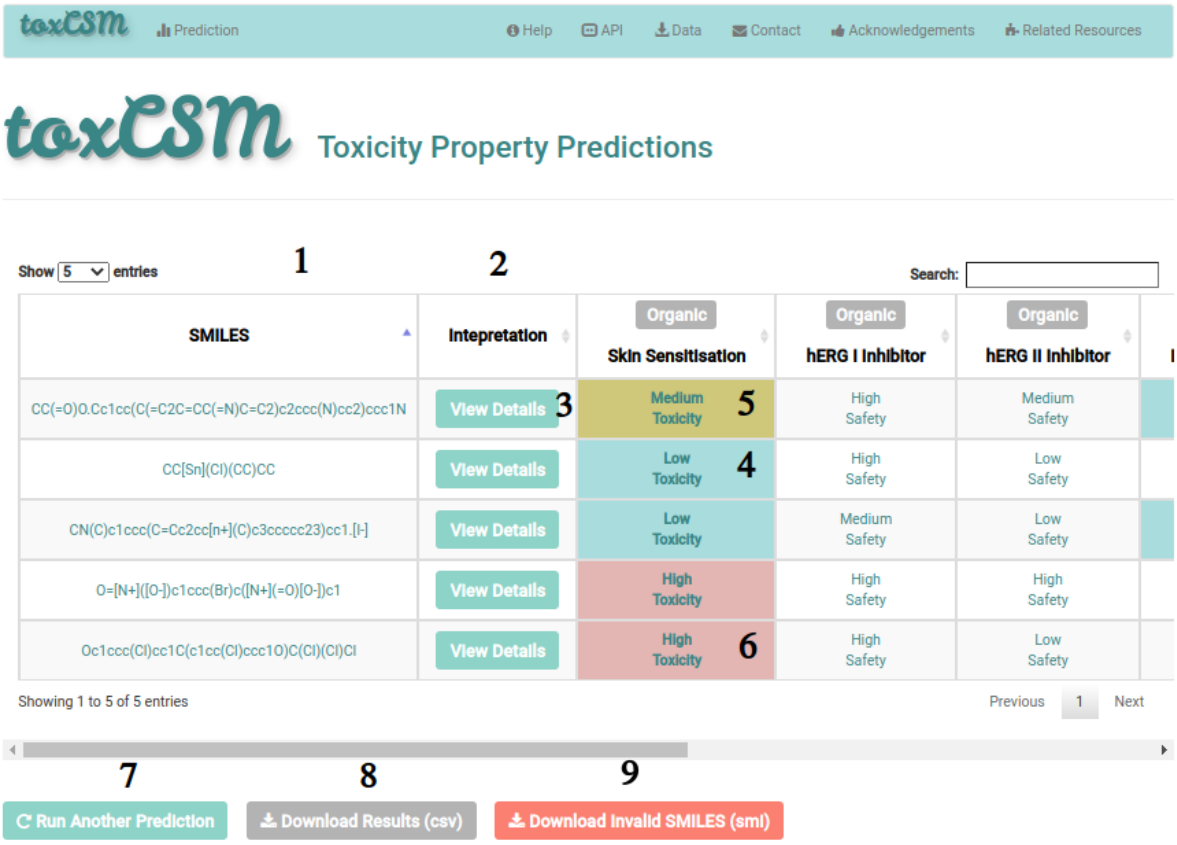
If anything goes wrong in the submission or job processing, the user will visualise an error page such as the one presented at **E.** If the email address was filled on the previous page and the job already started its processing, the user will also receive an email mentioning the occurred issue.



Results Page

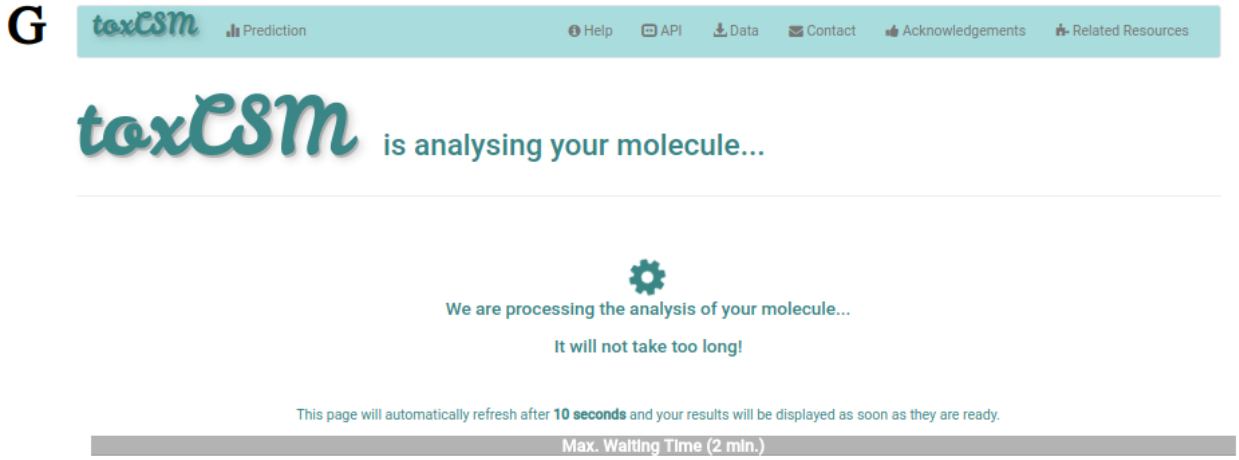
F describes the result page for five molecules. It includes the SMILES of the molecule **(1)**, a set of interpretation buttons **(2, 3)** and the predictions for the selected toxicity categories (**4, 5, 6**). Classification predictions are presented in terms of high, low and medium toxicity or safety as exemplified at **F.** Regression predictions, in turn, are presented as real-valued numbers. In classification tasks, compounds with high **(6)**, medium **(5)** and low (**4**) toxic profiles are coloured in red, mustard and cyan, respectively. On the other hand, safe compounds are kept in grey or white. Users can also run another prediction **(7)**, download the results into a comma-separated value (csv) file **(8)** and, if any SMILES was considered as invalid, the user can download them within a (smi) file **(9)**. The button presented in **(9)** will appear only when invalid compounds are found.

F


toxCSM
Prediction
Help
API
Data
Contact
Acknowledgements
Related Resources
toxCSM
Toxicity Property Predictions
Show 5 entries
1
2
Search:
SMILES
Intepretation
Organic
Skin Sensitisation
Organic
hERG I Inhibitor
Organic
hERG II Inhibitor
CC(=O)O.Cc1cc(C(=C2C=CC(=N)C=C2)c2ccc(N)cc2)ccc1N
View Details
3
Medium Toxicity
5
High Safety
Medium Safety
CC[Sn](Cl)(CC)CC
View Details
Low Toxicity
4
High Safety
Low Safety
CN(C)c1ccc(C=Cc2cc[n+](C)c3ccccc23)cc1.[I-]
View Details
Low Toxicity
Medium Safety
Low Safety
O=[N+]([O-])c1ccc(Br)c([N+](=O)[O-])c1
View Details
High Toxicity
High Safety
High Safety
Oc1ccc(Cl)cc1C(c1cc(Cl)ccc1O)C(Cl)(Cl)Cl
View Details
High Toxicity
6
High Safety
Low Safety
Showing 1 to 5 of 5 entries
Previous
1
Next
7
8
9
Run Another Prediction
Download Results (csv)
Download Invalid SMILES (sml)

Analysis Waiting Page

If the user clicks at any button "**View Details" (3)** at **F**, users will be redirected to a waiting page again while the analysis of that particular molecule is being performed. This waiting page is presented at **G.**



Analysis Page

The analysis page shows information of the molecule the user requested more details about. This page is presented from **H.1** to **H.5,** comprehending the molecule depiction and its respective SMILES **(H.1)**, the toxicity properties (model predictions) sorted by highest toxic confidence level **(H.2)**[1]**,** the general molecular properties **(H.3),** the molecular drug-likeness properties **(H.4),** the molecular drug-likeness radar plots **(H.5)**, the analysis of presence (highlighted in red) or absence of 36 toxicophores **(H.6)**[2]**.**

Users can also be redirected to run another prediction (**H.6/1),** download the analysis of the molecule in a zip file (**H.6/2)** -- which encompases the tables at **H.2**, **H.3, H.4 and H.6** or print the analysis page (**H.6/3).**

In the whole page **H,** tooltips are used to help the interpretation of the analysis.

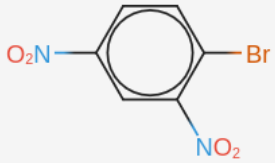
[1] If the molecule is considered to be toxic in a particular endpoint model, the color of the model's row is also highlighted in red, mustard and cyan to indicate high, medium and low toxicity profiles, respectively.
[2] Toxicophores are sorted based on their presence in the analysed molecule. We indicate the appearance of a toxicophore by the color red.

H.1

toxCSM Prediction Help API Data Contact Acknowledgements Related Resources

toxCSM Molecular Properties

Molecule Depiction



Molecule SMILES

Brc1c(cc(cc1)[N+](=O)[O-])[N+](=O)[O-]

H.2

Toxicity Model-Predicted-Based Molecular Properties

Show 9 entries Search:

Toxicity Perspective Category	Endpoint Name	EndPoint Unit	Prediction	Toxicity Confidence	Prediction Interpretation
Genomic	AMES Mutagenesis	Category (Toxic/Safe)	Toxic	1.0	High Toxicity
Environmental	T. Pyriformis	Category (Toxic/Safe)	Toxic	1.0	High Toxicity
Organic	Skin Sensitisation	Category (Toxic/Safe)	Toxic	1.0	High Toxicity
Organic	Eye Irritation	Category (Toxic/Safe)	Toxic	1.0	High Toxicity
Environmental	Fathead Minnow	Category (Toxic/Safe)	Toxic	0.97	High Toxicity
Organic	Eye Corrosion	Category (Toxic/Safe)	Toxic	0.87	High Toxicity
Environmental	Honey Bee	Category (Toxic/Safe)	Toxic	0.67	Medium Toxicity
Stress Response	SR-ARE	Category (Toxic/Safe)	Toxic	0.57	Low Toxicity
Nuclear Response	NR-ER	Category (Toxic/Safe)	Safe	0.5	Low Safety

Showing 1 to 9 of 35 entries Previous 1 2 3 4 Next

H.3 General Molecular Properties

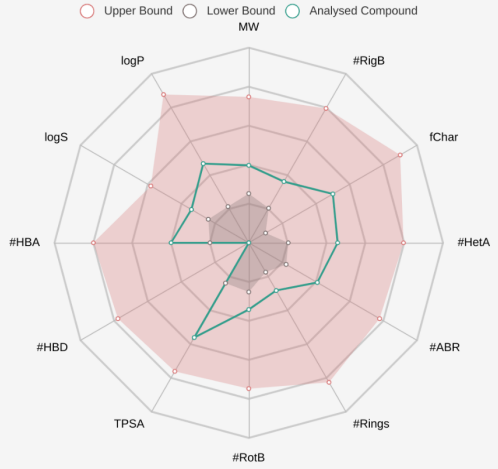
Descriptor	Abbreviation	Value
#Atoms	#Atom	16
#Heavy Atom	#HA	13
Molecular Weight	MW	245.928
Volume	Vol	167.03
Density	Dens	1.472
#Hydrogen Bond Donors	#HBD	0
#Hydrogen Bond Acceptors	#HBA	4
#Rotatable Bonds	#RotB	2
#Rings	#Rings	1
#Atoms in the Biggest Ring	#ABR	6
#Heteroatom in a Molecule	#HetA	7
Formal Charge	fChar	0
#Rigid Bonds	#RigB	8
Flexibility	Flex	0.25
Molar Refractivity	MolRef	47.45
Topological Polar Surface Area	TPSA	86.28
Labute's Approximate Surface Area	Labute_ASA	80.605
Log of the Partition Coefficient Between Octanol and Water	logP	2.266
Logarithm of Compounds Water Solubility	logS	-2.678

H.4 Molecular Drug-likeness Properties

Score/Rule	Value/Decision
Lipinski's Rule of 5 (RO5)	Respect the Lipinski's RO5
Ghose's Rules	Violate the Ghose's Rules (1 Time)
Oprea's Notability Rules	Respect the Oprea's Notability Rules
Pfizer's Rules	Respect the Pfizer's Rules
GSK Rules	Respect the GSK Rules
ADMETLab 2.0 Soft Rules	Respect the ADMETLab 2.0 Rules
QED Score (with unit weights)	0.7
QED Score (with maximal descriptor weights)	0.52
QED Score (with mean descriptor weights)	0.59
SAS Score	2.05
Fsp^3 Score	0.0
NPscore	-1.57

H.5 Molecular Drug-likeness Radars

ADMETLab 2.0 Soft Rules | Lipinski's Rule of 5



H.6 Molecular Toxicophores

Show 9 entries Search:

Toxicophore SMARTS	Molecular Toxicophore Verification
O=N(~O)a	Present
a[NH2]	Absent
a[N;X2]=O	Absent
CO[N;X2]=O	Absent
N[N;X2]=O	Absent
O1[c,C]-[c,C]1	Absent
C1NC1	Absent
N=[N+]=[N-]	Absent
C=[N+]=[N-]	Absent

Showing 1 to 9 of 36 entries Previous 1 2 3 4 Next

1 2 3

Run Another Prediction | Download Analysis (zip) | Print Analysis (pdf)

Data Page

If the user clicks on the data page at the top menu, he/she will be redirected to the data page **I**, where all the endpoints used to train, validate and test toxCSM predictive models. Users can access the name of the endpoint data **(1)** and download it by clicking on their names**.** In addition, information about the perspective category **(2)**, machine learning task **(3)**, number of samples **(4)**, number of generated features **(5),** and source **(6)** are available. Users can also search **(7)** for the respective endpoint data and browse across them **(8)**.

In the whole page **I**, tooltips are available to assist the user.

I