

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

¹School of Chemistry and Molecular Biosciences, University of Queensland

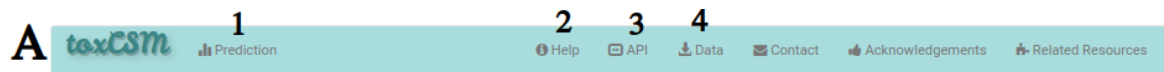
²Systems and Computational Biology, Bio21 Institute, University of Melbourne

³Computational Biology and Clinical Informatics, Baker Heart and Diabetes Institute

⁴Baker Department of Cardiometabolic Health, University of Melbourne

⁵School of Computing and Information Systems, University of Melbourne

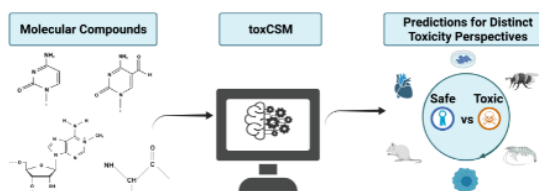
Main Page



toxCSM Comprehensive Prediction of Small Molecule Toxicity Profiles

Alex G. C. de Sá, Yangyang Long, Stephanie Portelli, Douglas E. V. Pires & David B. Ascher

Abstract: Drug discovery is a long, costly and high risk endeavour which is further convoluted by high attrition rates later in development stages. Poor toxicity profiles have been one of the main causes of failure during pre-clinical and clinical trials and it is estimated that success rates could quadruple if human toxicity was eliminated at preclinical and development stages. The accumulation of characterised small molecule toxicity data holds the key for the development of computational tools to facilitate optimisation of toxicity profiles of drug candidates, reducing costs, increasing success rates by identifying toxicity issues early on and assisting prioritization of compounds, also reducing use of animal models. Here, we present toxCSM, the most comprehensive computational platform for the study and optimisation of toxicity profiles of small molecules. toxCSM leverages our well established graph-based signatures concept to develop scalable and accurate predictive models using supervised learning, outperforming alternative methods. Using these signatures we have developed 36 models for predicting a wide range of toxicity properties, from nuclear and stress responses to environmental toxicity, which can assist in the development of safer and less toxic drugs as well as herbicides and pesticides. Based on these features, we believe toxCSM web server will provide an easy, precise and reliable way to evaluate small molecule toxicity profiles.



Important Information:

- This website is free for all users.
- This website does not use cookies.

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

toxCSM

A Comprehensive In Silico Method Based on Graph Signatures and Similarities for Accurate and Robust Molecular Toxicity Prediction

Step 1: Please provide a set of molecules (SDF or SMILES)

SDF OR SMILES file (limited to 1,000 molecules): OR SMILES string: 3

1 2 [Draw your own molecule and get its SMILES](#) 4

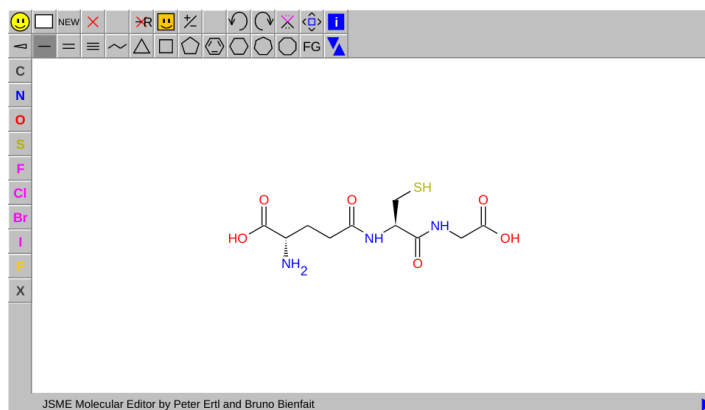
Step 2: Fill your e-mail address (optional)

E-mail address (for sending a notice with the result link): 5

Step 3: Please choose the prediction mode

Use toxCSM to evaluate your molecules considering one of the following toxicity perspective categories: 6

Draw your own molecule using JSME Molecular Editor x



C

Close

Get SMILES and Fill it In

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.

D

toxCSM Prediction Help API Data Contact Acknowledgements Related Resources

toxCSM is processing your submission...

We are processing your submission.

You can bookmark (★) this page and come back later to check the results.

If you have filled your e-mail in the previous page, you will receive the a link by email to access your results.

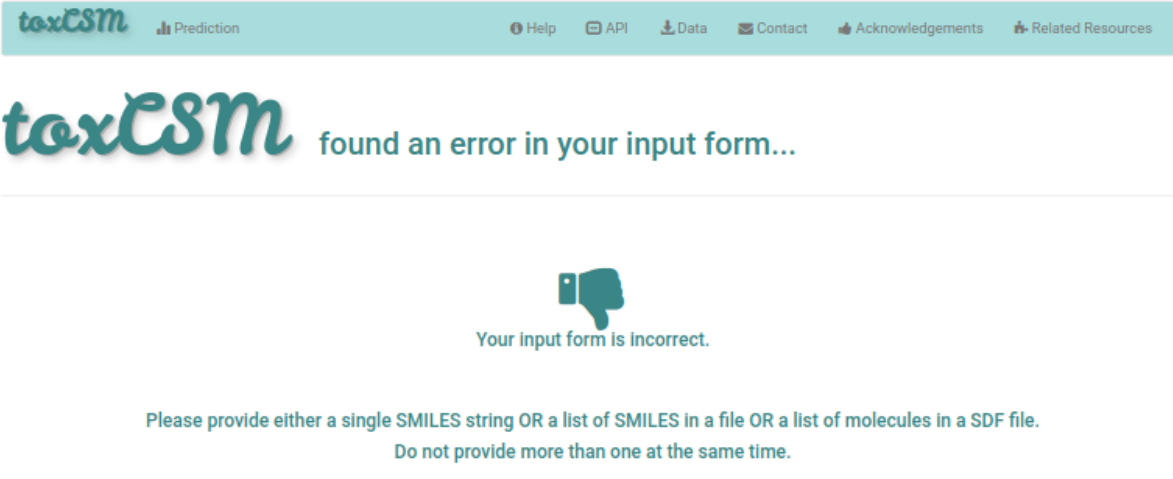
This page will automatically refresh after **10 seconds** and your results will be displayed as soon as they are ready.

Done Max. Waiting Time (10 min.)

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.

E



toxCSM Prediction Help API Data Contact Acknowledgements Related Resources

toxCSM found an error in your input form...

Your input form is incorrect.

Please provide either a single SMILES string OR a list of SMILES in a file OR a list of molecules in a SDF file.
Do not provide more than one at the same time.

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	Interpretation	Organic	Organic	Organic	
		Skin Sensitisation	hERG I Inhibitor	hERG II Inhibitor	
<chem>CC(=O)O.Cc1cc(C(=C2C=CC(=N)C=C2)c2ccc(N)cc2)ccc1N</chem>	View Details 3	Medium Toxicity 5	High Safety	Medium Safety	
<chem>CC[Sn](Cl)(CC)CC</chem>	View Details	Low Toxicity 4	High Safety	Low Safety	
<chem>CN(C)c1ccc(C=Cc2cc[n+](C)c3ccccc23)cc1.[-]</chem>	View Details	Low Toxicity	Medium Safety	Low Safety	
<chem>O=[N+](O)c1ccc(Br)c([N+](=O)[O-])c1</chem>	View Details	High Toxicity	High Safety	High Safety	
<chem>Oc1ccc(Cl)cc1C(c1cc(Cl)ccc1O)C(Cl)(Cl)Cl</chem>	View Details	High Toxicity 6	High Safety	Low Safety	

Showing 1 to 5 of 5 entries

Previous 1 Next

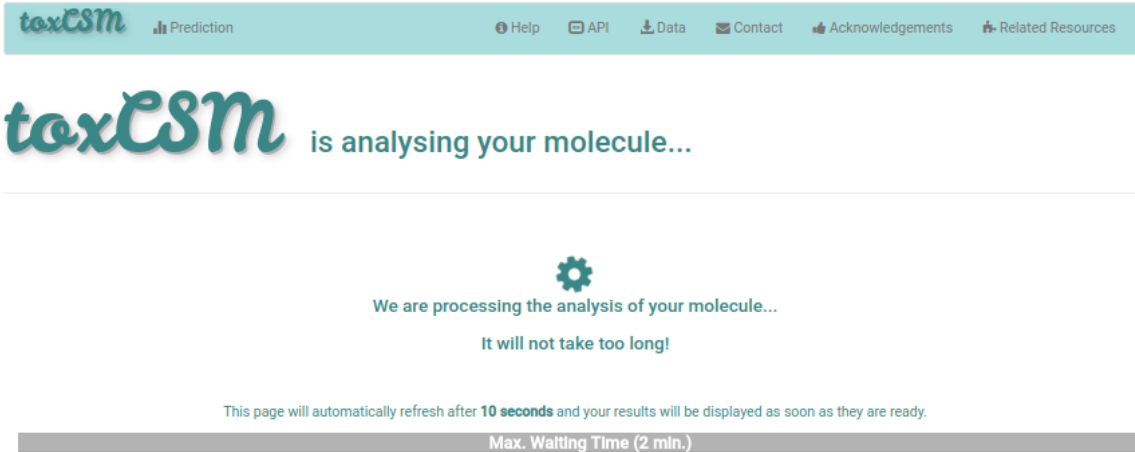
7 8 9

[Run Another Prediction](#)[Download Results \(csv\)](#)[Download Invalid SMILES \(smi\)](#)

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.

G



toxESM Prediction Help API Data Contact Acknowledgements Related Resources

toxESM is analysing your molecule...

We are processing the analysis of your molecule...
It will not take too long!

This page will automatically refresh after 10 seconds and your results will be displayed as soon as they are ready.

Max. Waiting Time (2 min.)

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 encompasses 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.

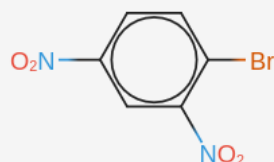
² Toxicophores are sorted based on their presence in the analysed molecule. We indicate the appearance of a toxicophore by the color red.

H.1

toxCSM

Molecular Properties

Molecule Depiction



Molecule SMILES

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

H.2 Toxicity Model-Predicted-Based Molecular Properties

Show entries Search:

Toxicity Perspective Category	Endpoint Name	EndPoint Unit	Prediction	Toxicity Confidence Θ	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

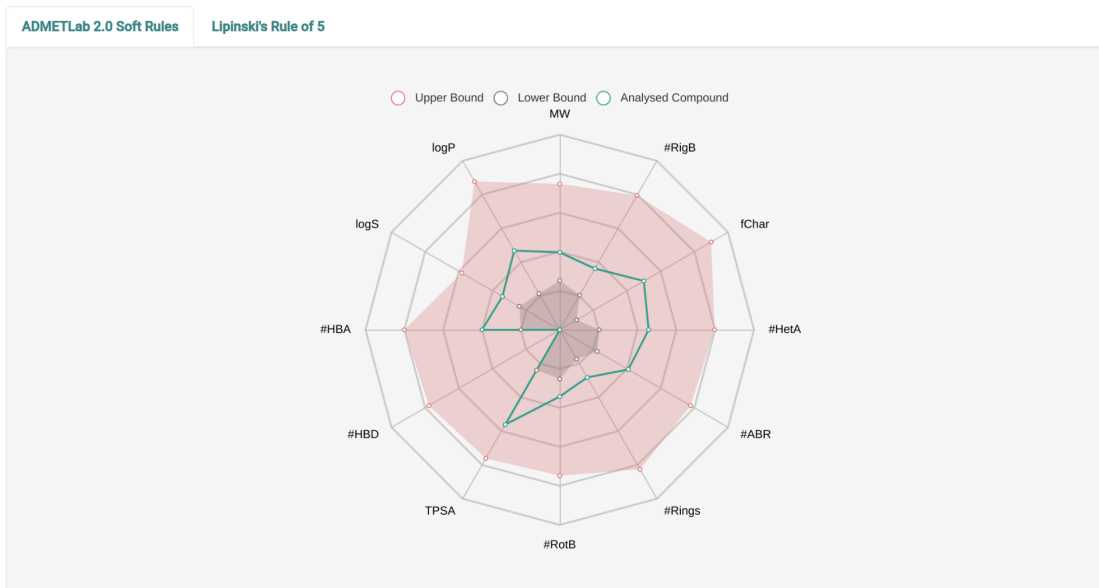
H.3 General Molecular Properties

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



H.6 Molecular Toxicophores

Show entries Search:

Toxicophore SMARTS	Molecular Toxicophore Verification
<chem>O=N(~O)a</chem>	Present
<chem>a[NH2]</chem>	Absent
<chem>a[N;X2]=O</chem>	Absent
<chem>CO[N;X2]=O</chem>	Absent
<chem>N[N;X2]=O</chem>	Absent
<chem>O1[c;C]-[c;C]1</chem>	Absent
<chem>C1NC1</chem>	Absent
<chem>N=[N+]=[N-]</chem>	Absent
<chem>C=[N+]=[N-]</chem>	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  [Help](#) [API](#) [Data](#) [Contact](#) [Acknowledgements](#) [Related Resources](#)

toxCSM Data

36 datasets representing distinct toxicity categories

Show entries **1** **2** **3** **4** **5** Search: **7**

1 Name (Click to Download)	2 Category	3 Learning Task	4 Samples (Negatives/Positives)	5 Generated Features	6 Source
AMES Mutagenesis	Genomic	Classification	8102 (3470/4632)	4431	Xu et al., 2012
Avian	Environmental	Classification	591 (501/90)	344	Zhang et al., 2015
Biodegradation	Environmental	Classification	1592 (1007/585)	789	Cheng et al., 2012
Carcinogenesis	Genomic	Classification	278 (220/58)	315	Li et al., 2015
Crustacean	Environmental	Classification	1020 (487/533)	742	Cao et al., 2018
Eye Corrosion	Organic	Classification	2299 (1412/887)	1061	Wang et al., 2017
Eye Irritation	Organic	Classification	5220 (1346/3874)	3749	Wang et al., 2017
Fathead Minnow	Environmental	Classification	554 (188/366)	592	Yang et al., 2018
Fathead Minnow (pLC50)	Environmental	Regression	554 (-)	364	Yang et al., 2018

Showing 1 to 9 of 36 entries **8** [Previous](#) [1](#) [2](#) [3](#) [4](#) [Next](#)