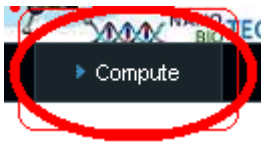
Compute molecular descriptors over the web!

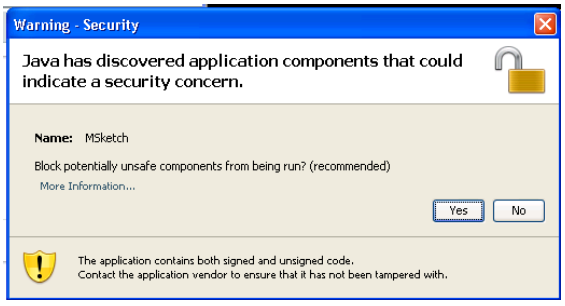
Molecular descriptors are usually used for QSAR, QSPR and QSTR related studies. The property or bio-activity of a molecule depends on the type of functional groups, rings or scaffolds present in the molecule. There are some functional groups like Methoxy (-OCH3) enhance the bio activity. There are certain functional groups which are called 'Toxicophores' which are very risky in drug discovery program, where their presence in the molecule would produce toxic or adverse effects. Chemists usually think of Natural Products associated with their bio-activity. For example class of natural products such as Taxols, Flavones etc., with distinct scaffolds modified extensively by the chemist to identify potent compounds as drugs.

As part of on-going cloud computing infrastructure or facility here we propose to build a Web-based chemical property / toxicity prediction module where the researchers can upload their molecules (create dynamically using the molecular structure drawing tools (e.g., ChemAxon's Marvin Sketch)) and apply opensource based chemoinformatics modules, algorithms to generate molecular descriptors. As per the principle of Structure Activity Relationship studies, similar molecules expected to have similar activity. First step in this initiative is to produce molecular descriptors for the set of given molecules. Later learning from these experiments, building dynamic / automatic QSAR models using machine learning techniques, the molecular data and bioactivity/toxicity data input could be used to predict bio-activity or toxicity of molecules which are not yet synthesized.

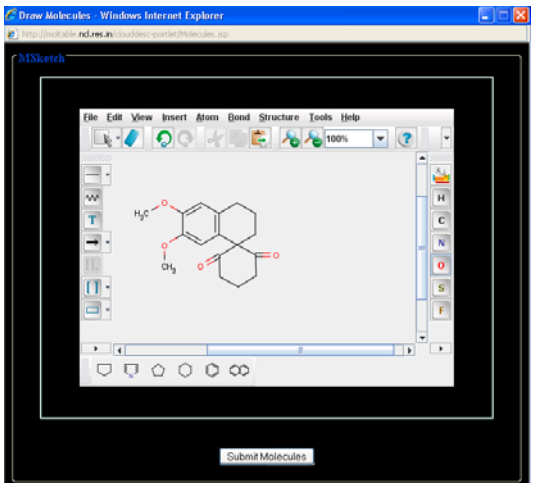
How to use the web based 'Moltable's chemoinformatics tool (Compute) to calculate molecular descriptors?

1) Visit http://moltable.ncl.res.in/
2) No registration required for evaluation

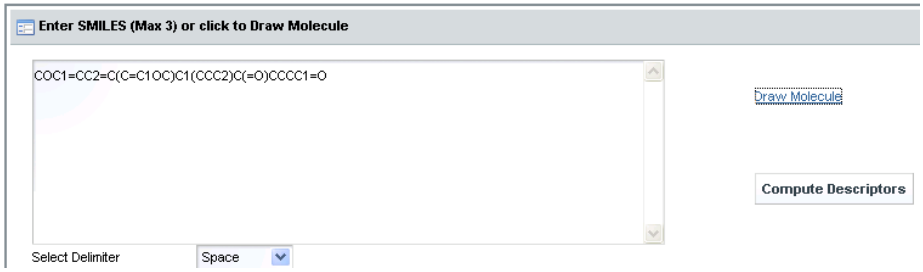


3) Click 'Compute' tab from the Moltable Portal Menu
4) Enter molecules in SMILES format (up to maximum of three SMILES strings)
5) Alternatively you can use Drawing tool available in that portal (Marvin Sketch)
6) Click "Draw Molecule" you will be asked to 'accept' the applet (as per the browser's requirement on Java Applets). Select 'Always Trust' the 'ChemAxon' Applet.
7) 

8) Draw a molecule and click "Submit Molecule"

9) 

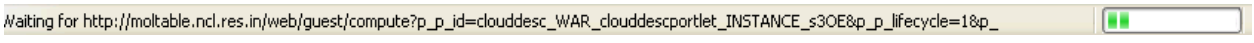
10) You will notice that the molecule is transformed into 'SMILES' format and copied to Text Area in the portlet.

11) 

12) You have the option of selecting delimiter (if you are calculating descriptors for more than one molecule) The following delimiters are available (Comma, Tab, Semicolon and Space)

13) Now you click "Compute Descriptors" button.

14) Please wait till the computation is done (dynamically over the web)

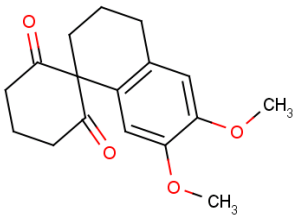
15) 

16) Please note that the time taken to compute the descriptors directly proportional to the complexity associated with the molecule.

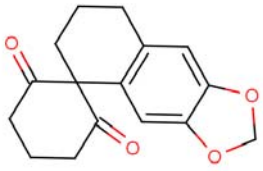
17) The list of descriptors (NCL_1 to NCL_194) now available in the text area which can be copied into Excel or similar applications for further processing.

NCL_0 0.0
NCL_1 0.5594
NCL_2 -0.2077
NCL_3 -0.1044
NCL_4 0.09
NCL_5 -0.0344
NCL_6 24.0979
NCL_7 24.9926
NCL_8 36.6567
NCL_9 46.7596
NCL_10 44.4238
NCL_11 1517.0646
NCL_12 1855.7521
NCL_13 2802.7612
NCL_14 3149.8148
NCL_15 2613.7152
NCL_16 11.9885
NCL_17 16.0063
NCL_18 0.79
NCL_19 1.2102

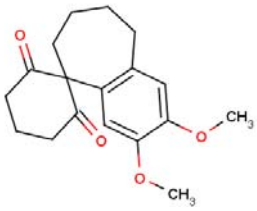
18) The details of descriptors are available on request. The descriptors can be directly used for web-based statistical/machine learning tools. Currently we are implementing PLS, PCR, kNN, ANN, SVM, Clustering, Random Forest and other similar statistical methods to handle molecular data associated with bio-activity, property and toxicity attributes. Any suggestions and collaboration welcome to build a cloud based property prediction / simulation studies related to molecular informatics.



1



2



3

For example if you want to calculate the molecular descriptors for the above molecules

Copy the following SMILES strings and paste directly to the input text area and click Compute Descriptor button.

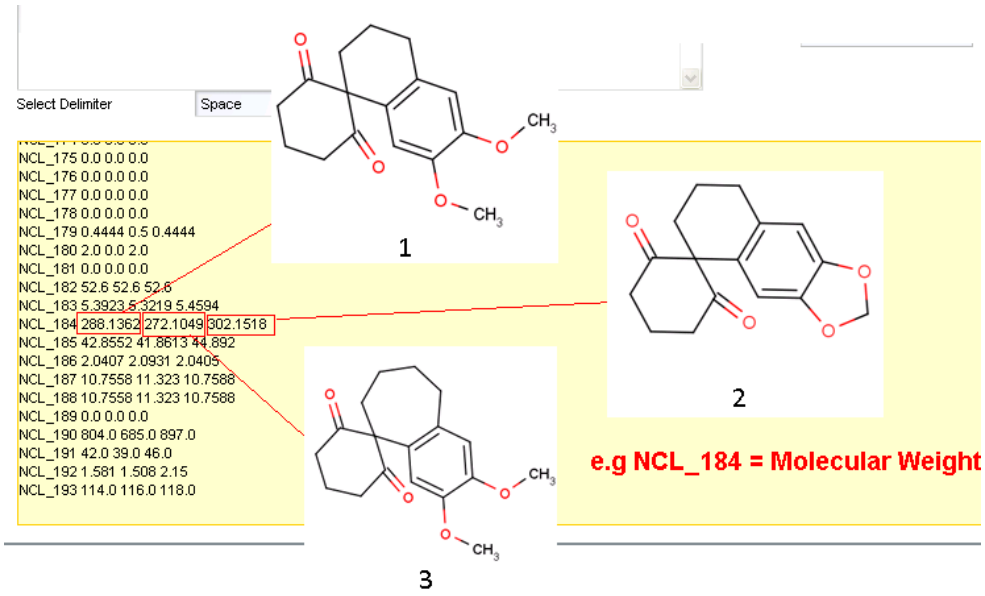
COC1=CC2=C(C=C1OC)C1(CCC2)C(=O)CCCC1=O

C3OC1=CC2=C(C=C1O3)C1(CCC2)C(=O)CCCC1=O

COC1=CC2=C(C=C1OC)C1(CCCC2)C(=O)CCCC1=O

```
NCL_0 0.0 0.0 0.0
NCL_1 0.5594 0.5533 0.5594
NCL_2 -0.2077 -0.2656 -0.2079
NCL_3 -0.1044 0.0207 -0.1053
NCL_4 0.09 -0.0522 0.0897
NCL_5 -0.0344 0.0762 -0.031
NCL_6 24.0979 23.0979 25.0979
NCL_7 24.9926 24.9926 25.9926
NCL_8 36.6567 38.4312 37.6567
NCL_9 46.7596 42.9851 50.7596
NCL_10 44.4238 39.7596 48.4238
NCL_11 1517.0646 1455.1376 1648.7743
NCL_12 1855.7521 1809.1392 2004.7121
NCL_13 2802.7612 2774.8091 2995.418
NCL_14 3149.8148 2948.4638 3661.9611
NCL_15 2613.7152 2397.1234 3024.7823
NCL_16 11.9885 11.9885 11.9884
NCL_17 16.0063 16.0063 16.0063
NCL_18 0.79 0.79 0.79
```

You will be wondering what is the meaning of NCL_XX ? Yes they are all molecular descriptors for example NCL_184 is “Molecular Weight”. The current list includes “Druglike Rule of 5”, Topological, electronic, charge and many more descriptors. This demonstration is just a fraction of list of descriptors available in the system.



Due to anticipated internet requests we restricted to only three molecules and just 193 descriptors at a time now.

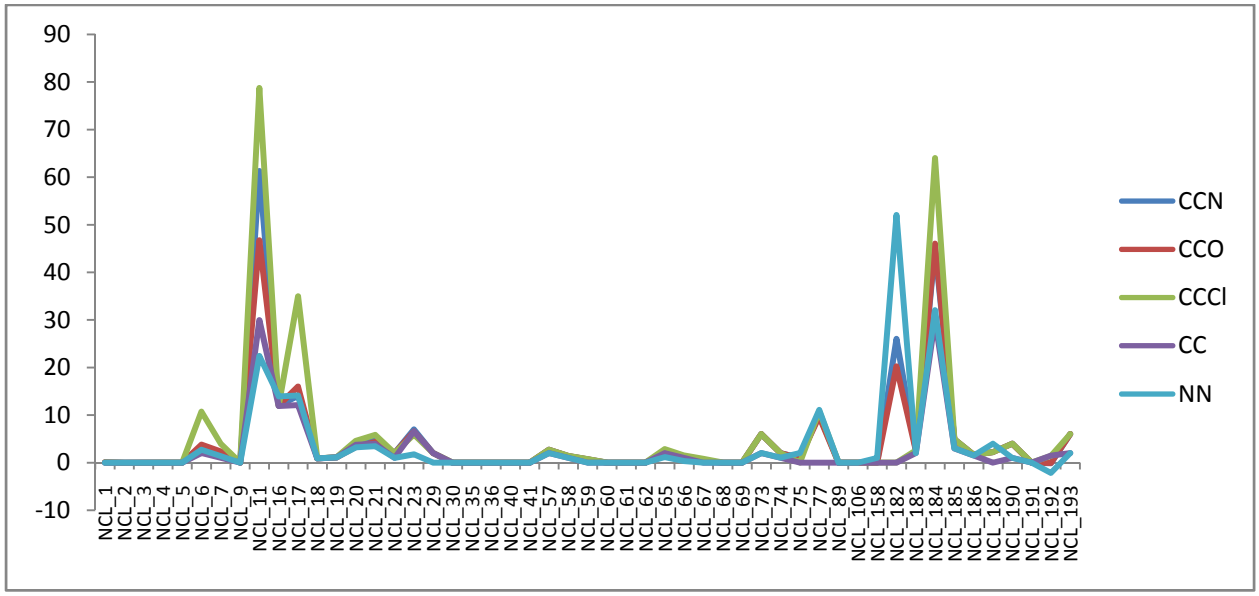
You can enter up to 5 molecules in SMILES format and get the descriptors over the web at a time.

You can copy the descriptors to Microsoft office (Excel) to plot the data to see the variation with in the molecules.

SMILES	CCN	CCO	CCCl	CC	NN
NCL_1	0.0297	0.0836	0.0442	0	0
NCL_2	-0.012	-0.0332	-0.0177	0	0
NCL_3	-0.0029	-0.0086	-0.0044	0	0
NCL_4	0	0	0	0	0
NCL_5	0	0	0	0	0
NCL_6	3.36	3.7745	10.7128	2	2.72
NCL_7	2.1662	2.3321	3.9517	1	1.36
NCL_9	0	0	0	0	0
NCL_11	61.2918	46.7352	78.7445	29.9461	22.4316
NCL_16	11.9938	11.9967	11.9987	11.89	13.8931
NCL_17	14.0091	15.9979	34.9694	12.11	14.1131
NCL_18	0.89	0.89	0.89	0.89	0.89
NCL_19	1.11	1.11	1.11	1.11	1.11
NCL_20	4.1228	3.3125	4.6063	3.7595	3.231
NCL_21	4.974	4.3518	5.8354	3.9795	3.451
NCL_22	2	2	2	1	1
NCL_23	6.9924	6.5592	5.886	6.5592	1.7328
NCL_29	2	2	2	2	0
NCL_30	0	0	0	0	0
NCL_35	0	0	0	0	0
NCL_36	0	0	0	0	0
NCL_40	0	0	0	0	0
NCL_41	0	0	0	0	0
NCL_57	2.7071	2.7071	2.7071	2	2
NCL_58	1.4142	1.4142	1.4142	1	1
NCL_59	0.7071	0.7071	0.7071	0	0
NCL_60	0	0	0	0	0
NCL_61	0	0	0	0	0
NCL_62	0	0	0	0	0
NCL_65	2.2845	2.1543	2.841	2	1.1547
NCL_66	1.1154	1.0233	1.5089	1	0.3333
NCL_67	0.4082	0.3162	0.8018	0	0
NCL_68	0	0	0	0	0
NCL_69	0	0	0	0	0
NCL_73	6	6	6	2	2
NCL_74	2.01	2.01	2.01	1	1.02
NCL_75	1	1	0	0	2
NCL_77	10.0189	9.6823	10.8489	0	11.0679
NCL_89	0	0	0	0	0
NCL_106	0	0	0	0	0

NCL_158	0	0	0	0	1
NCL_182	26.02	20.23	0	0	52.04
NCL_183	2.585	2.585	2.585	2	2
NCL_184	45.0578	46.0419	64.008	30.047	32.0374
NCL_185	4.9142	4.9142	4.9142	3	3
NCL_186	1.6381	1.6381	1.6381	1.5	1.5
NCL_187	2.2071	2.2071	2.2071	0	4
NCL_190	4	4	4	1	1
NCL_191	0	0	0	0	0
NCL_192	-0.143	-0.076	1.398	1.478	-2.164
NCL_193	6	6	6	2	2

Plot the values of descriptors



If you need additional information please send your queries to ‘contacts’ of http://moltable.ncl.res.in