Establishing Moltable an open access initiatives for Molecular Informatics
http://moltable.ncl.res.in/

1) Open Access to Molecular Library of commercially available chemicals.

We made access to over 12 million molecules through MOLTABLE portal which also include all the commercially available chemicals.
http://moltable.ncl.res.in/chemstar/

2) Web based search feature for the molecules

All the molecules are stored in the MySQL database format. User can query the database as his/her criteria for searching the database. Browse, view and print the contents.

3) Access to truly computable and re-usable chemical structures.

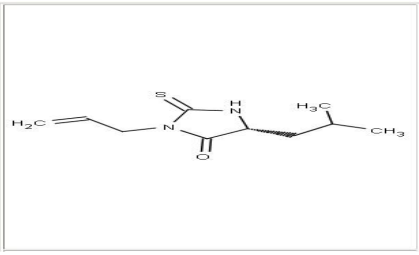
All the molecules in the MOLTABLE portal are stored in SMILES, SDF and MOL formats along with molecular fingerprints. The users can query and download the SMILES string format / SDF format of molecules and use in several computational chemistry or informatics program. The molecules can be directly incorporated into webpages using molecular handling applets.

4) Import and Export molecular data along with computed data.

We collected over 1.2 crore molecules as part of this portal building initiatives from publicly available resources. Every molecule in the database is curated, checked for errors and used for computing molecular properties. Some of the properties include solubility (logP) octanol/water partition co-efficient.

5) 2D and 3D Descriptors for structure activity relationship studies.

There are several molecular descriptors (topological and electronic) calculated describing the 2D and 3D parameters of the molecules which could be used for calculating/predicting other experimental data based on QSAR/QSPR/QSTR related studies.

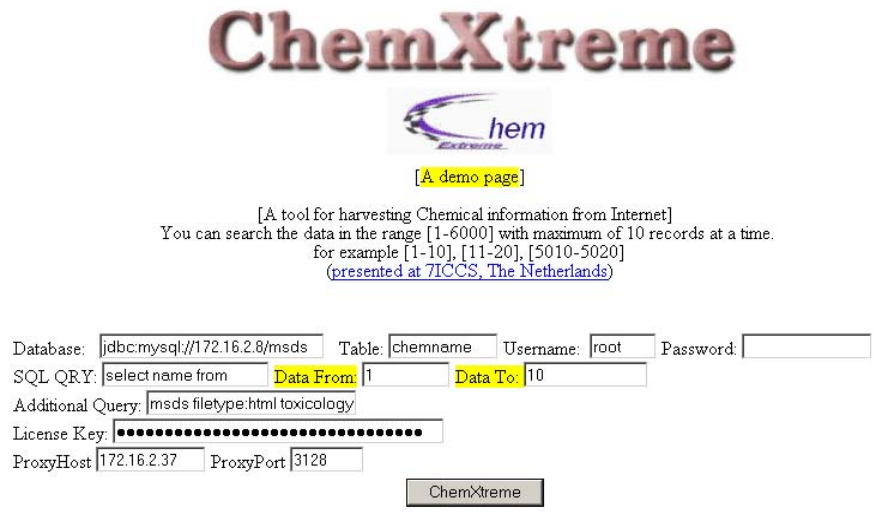


Rec_id: 1
Comp_Name: CC(C)C[C@H]1NC(=S)N(CC=C)C1=O
MolecularForm: C10H16N2OS
ExactMass: 212.0983
AtomCount: 14
HeavyatomCount: 14
ChargeText: -0.063 (0.07),-0.045 (0.03),-0.063 (0.07),-0.026 (0.058),0.088 (0.057),-0.11 (0.131),-0.046,-0.119,0.052,0.03 (0.097),-0.085 (0.058),-0.093 (0.107),0.086,-0.283
Composition: C (56.57%), H (7.6%)
Hpresent: false
Pka:
Logp: 2.1046
LogD: 0 2.105 1 2.105 2 2.105 3 2.105 4 2.105 5 2.105 6 2.105 7 2.104 8 2.103 9 2.089 10 1.977 11 1.659 12 1.491 13 1.465 14 1.463
Tpsa: 64.43
Red: 0
Green: 0
Blue: 0
Mass: 212.3129
OxygenAtomCount: 1
NonisotopeCount: 1
OxygenIstopeCount: 0
IsotopeFormula: C10H16N2OS

6) Access to other chemical risk data from public sources and linking with molecular structures.

We built a local database containing chemical names over 120000 entries along with other parameters like CAS Registry number, synonyms, company

product identifier etc., The chemical names (entries) in the database were used for searching the web using popular search engines like Google. The URLs retrieved using these search engines based some priority criteria were downloaded and analysed for chemical significance. The quality and quantity of the data for every web page is evaluated. The unstructured data from web pages were extracted and transformed into structured format using pattern recognition program, which is built in-house entitled ChemXtreme.

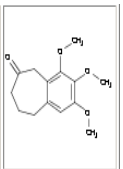
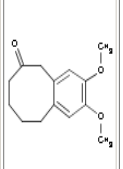


The output of ChemXtreme was stored in the local database. The entire data of chemicals which are very frequently used in the chemical and pharmaceutical sectors along with their risk assessment data extracted from MSDS datasheets were stored in the local database and made web accessible through this MOLTABLE portal.

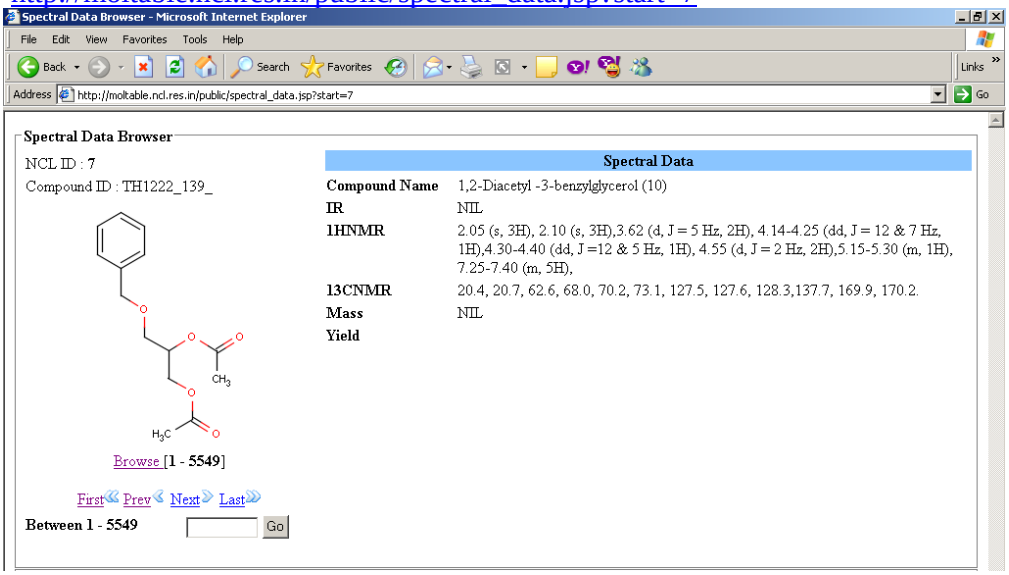
7) Linking Molecular data/crystallographic data with Open Access documents.

There exist several open access based PhD Thesis with high quality of contents. The data from open access documents were extracted and transformed into searchable and re-usable format. One such example is extraction of spectral data (IR, NMR, Mass) for molecular structures represented in PhD thesis.

http://moltable.ncl.res.in/public/thesis_1130.jsp

423	88_1130_159	XXX	2938, 1706, 1492, 1410, 1120 cm-1	1H NMR d 6.5 (s, 1H), 3.90 (s, 6H), 3.85 (s, 3H), 3.75 (s, 2H), 2.85(t, 2H, J = 7.10 Hz), 2.55 (t, 2H, J = 7.10 Hz), 1.95 (m,2H).	13C NMR d 209.84, 152.35, 151.58, 141.21, 136.49, 119.88, 108.96,61.57, 61.05, 56.23, 43.40, 41.47, 33.34, 26.67.	Mass (m/e) 252 (M+), 190 (43), 161 (27), 147 (22), 134 (100), 105 (57),91 (70), 77 (72).	
424	89_1130_185	XXX	3040, 2960, 1700, 1620, 1530, 1450, 1230, 1120 cm-1	1H NMR d 6.70 (s, 1H), 6.65(s, 1H), 3.90 (s, 3H), 3.85 (s, 3H), 3.70 (s, 2H),2.80 (t,J = 6.94 Hz, 2H), 2.35 (t, J = 6.94 Hz, 2H), 1.80 (m, 4H).	13C NMR d 211.76, 148.68, 147.69, 133.13, 125.63, 113.38, 113.15, 56.03,48.20, 41.12, 32.95, 31.33, 24.71.	Mass (m/e) 234 (100% M+), 206 (63), 191 (54), 175 (68), 165 (46), 151 (24),131 (24), 121 (44), 107 (37), 91 (49).	

http://moltable.ncl.res.in/public/spectral_data.jsp?start=7



8) Implementation of search engine compatible chemical identifiers with molecular structures.

The web-based program ChemXtreme implemented in MOLTABLE utilizes the search engine compatible chemical identifiers. We continue to update the local database with all the known chemical information available from Internet. The data collected and stored is periodically verified for authenticity and scientific relevance.

For details, contact:

M Karthikeyan, NCL, Pune at m.karthikeyan@ncl.res.in

Dr S Krishnan, NCL, Pune at krish@ems.ncl.res.in