Molecular standardization

June 7, 2016

The molecular standardization processes of ChemAxon's Standardizer and the Open Source library MolVS have been evaluated on a test suit of 35 compounds containing tautomers, mesomers and functional groups that are often represented differently. The 35 compounds are represented in multiple forms that, from a chemoinformatic perspective, should ideally be transformed into the same representation by the standardization process. However, the standardization process in Standardizer is not intended to transform into one canonical tautomer, but produces so called normal canonical tautomers where the normal canonical tautomer is selected from from subsets of tautomers considered interchangeable under physiological conditions.

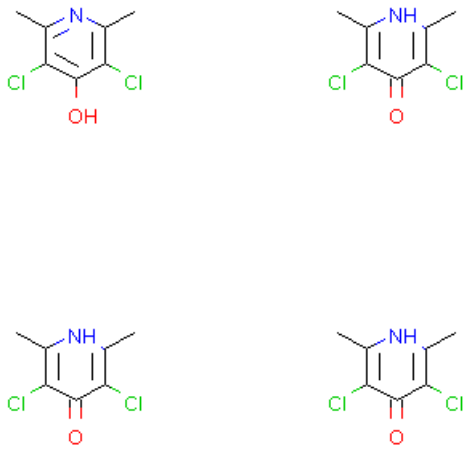
For each method a table is produced showing the input and output structures as drawn by RDKit for the different forms. The table also displays two fingerprint similarities between all pairs of output structures. If the standardization is successful the fingerprints should be identical. Furthermore, the set of 177 RDKit descriptors is calculated for each output structure and the number of differing descriptors is reported for each pair of output structures.

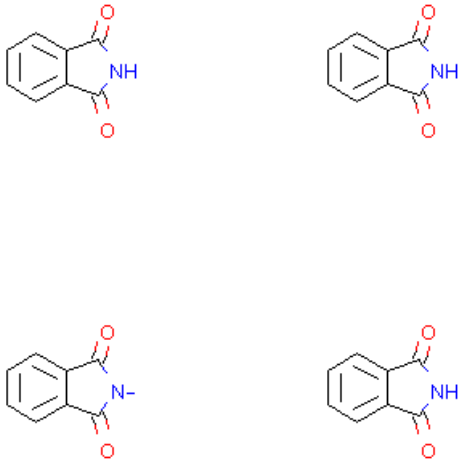
1 Standardization with ChemAxon Standardizer

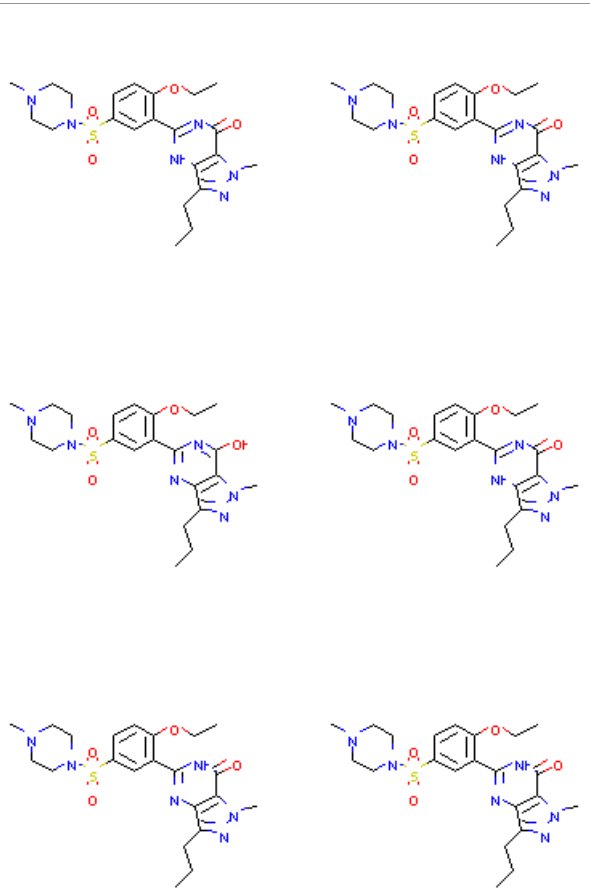
The ChemAxon standardizer was used with 7 actions, applied in the following order, together with the set of transformations defined in the GUI:

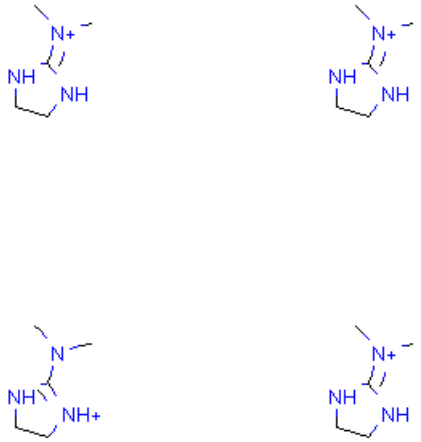
- removesolvents
- stripsalts
- removefragment:method=keeplargest
- mesomerize
- tautomerize
- neutralize

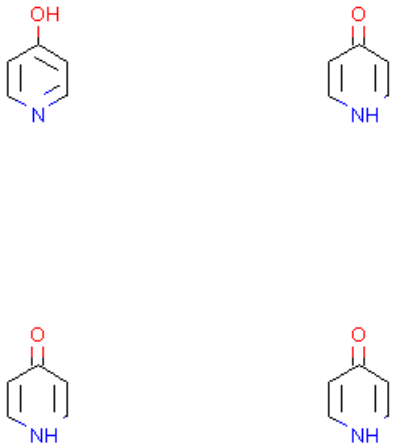
- transformations
- dearomatize

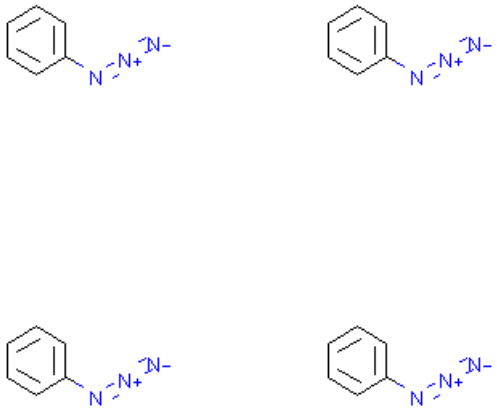
Idx	Name	Structure In and Out	Comment	TanTop	DiceMorgan	No. of DescDiff
1	clopidol			[1.0]	[1.0]	[0]

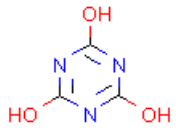
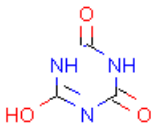
2	phthalimide			[1.0]	[1.0]	[0]

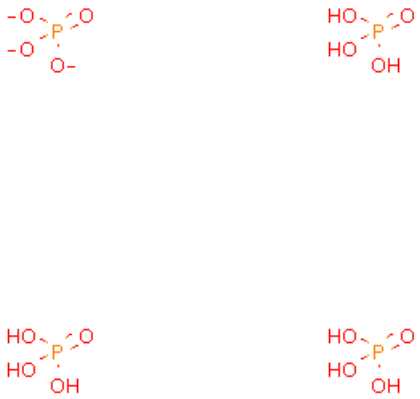
3	Viagra			[1.0, 1.0, 1.0]	[1.0, 0.91, 0.91]	[1, 21, 21]

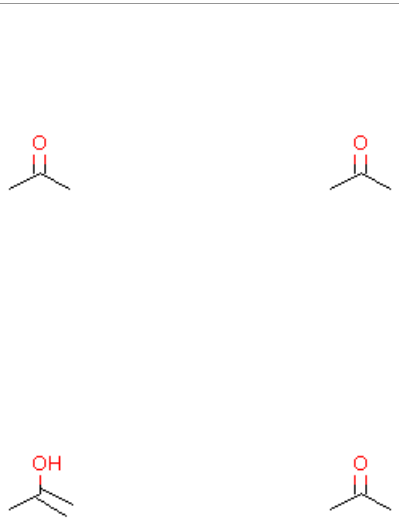
4	quatAmMesomer			[1.0]	[1.0]	[0]

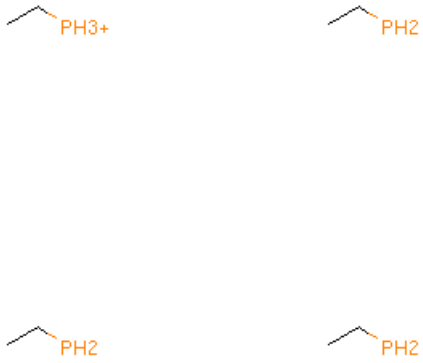
5	pyridinol			[1.0]	[1.0]	[0]

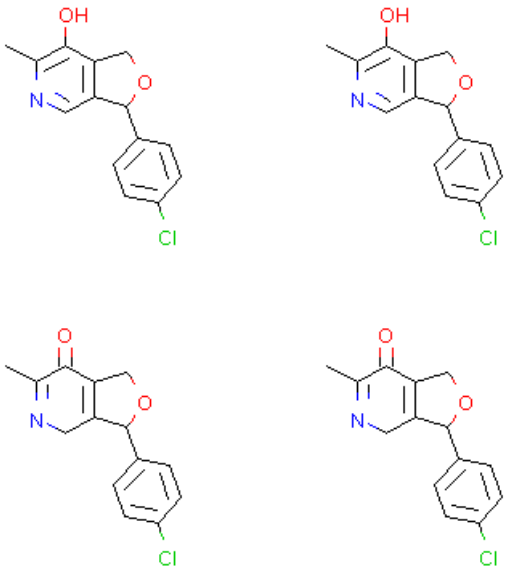
6	azide			[1.0]	[1.0]	[0]

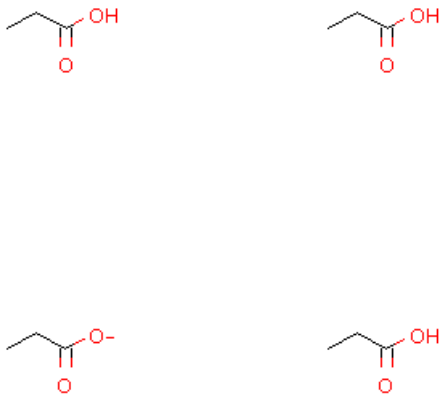
7	triazinol	   		[0.58]	[0.54]	[50]

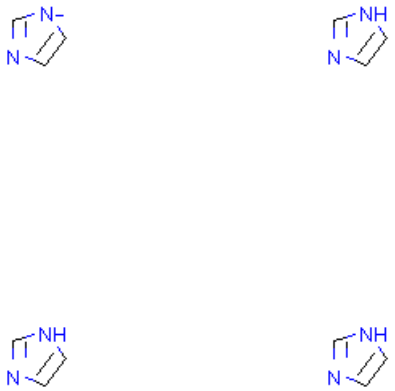
8	phosphate			[1.0]	[1.0]	[0]

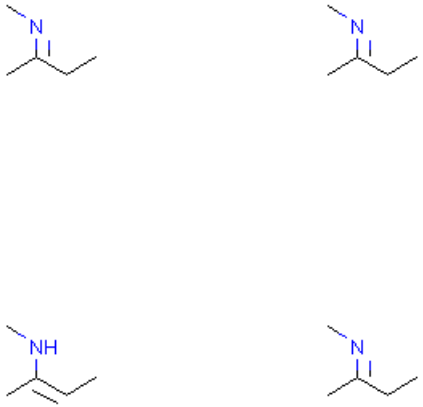
9	Propenol			[1.0]	[1.0]	[0]

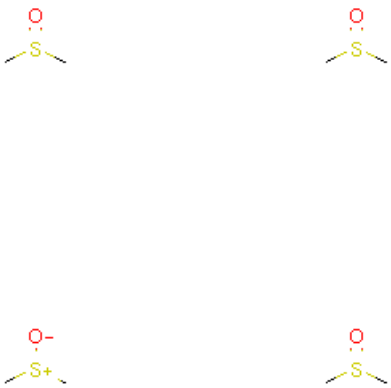
10	phosphine			[1.0]	[1.0]	[0]

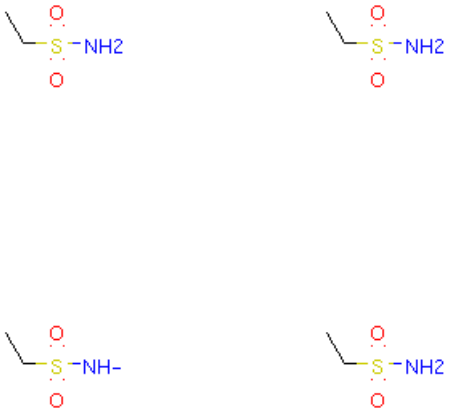
11	cicletanine			[0.3]	[0.67]	[76]

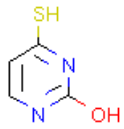
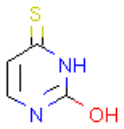
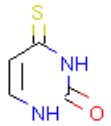
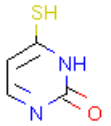
12	propaneAcid			[1.0]	[1.0]	[0]

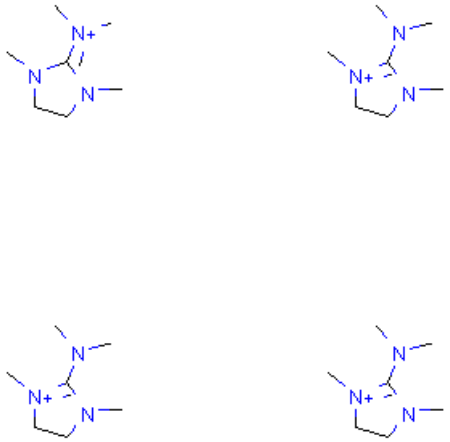
13	imidazole			[1.0]	[1.0]	[0]

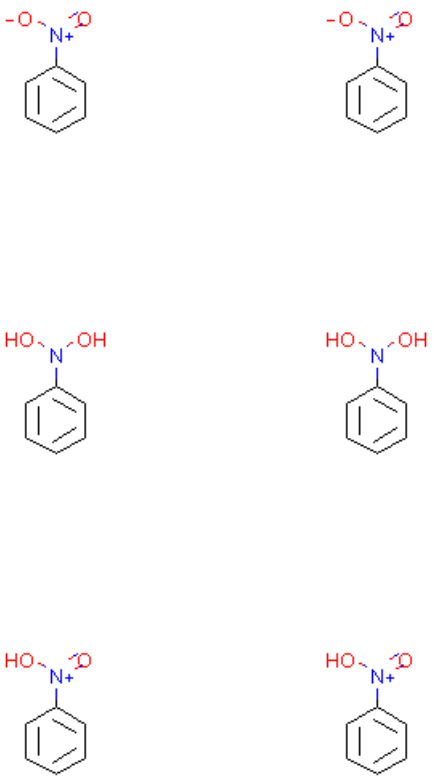
14	enamine			[1.0]	[1.0]	[0]

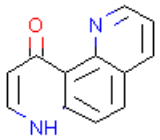
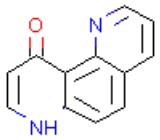
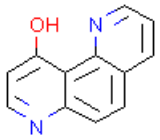
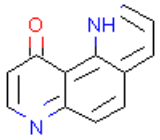
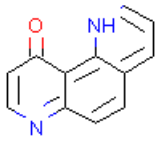
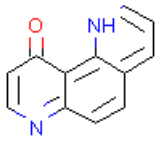
15	sulfoxide			[1.0]	[1.0]	[0]

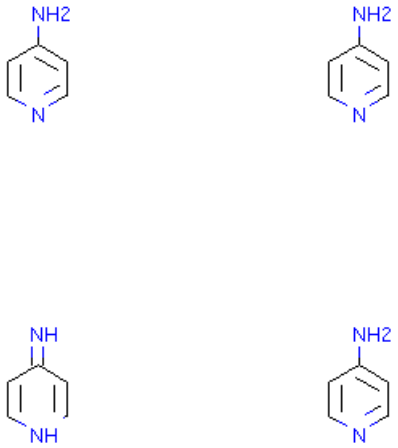
16	sulfonamide			[1.0]	[1.0]	[0]

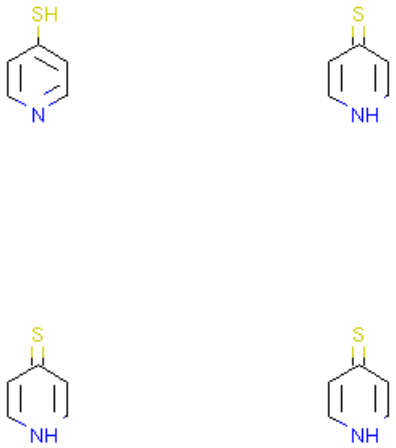
17	tiouracil	   		[0.35]	[0.5]	[54]

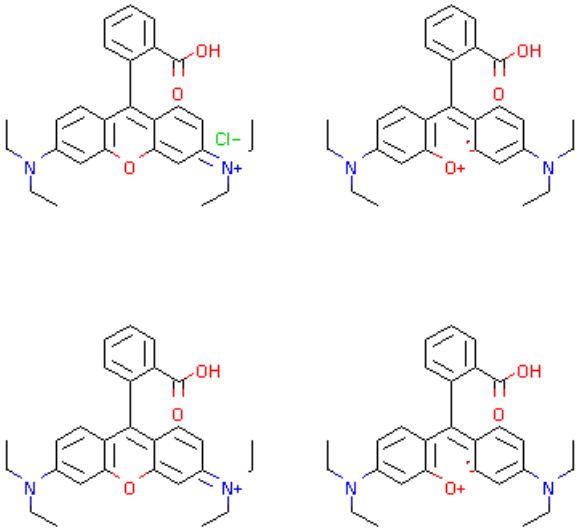
18	quatAmMesomer2			[1.0]	[1.0]	[0]

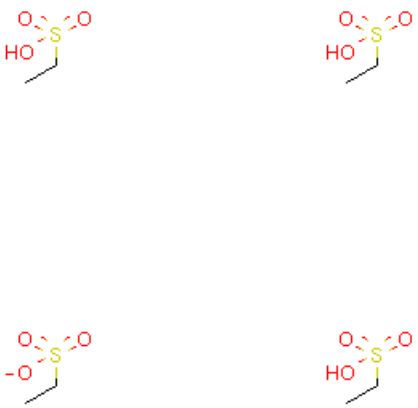
19	nitro			[0.67, 1.0, 0.67]	[0.56, 0.8, 0.6]	[58, 43, 56]

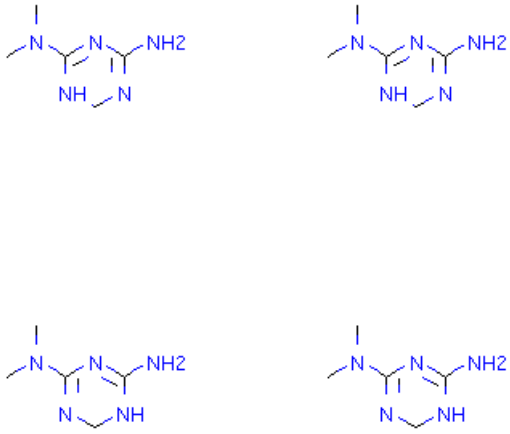
20	107and108and109	     		[1.0, 1.0, 1.0]	[0.7, 0.7, 1.0]	[17, 17, 0]

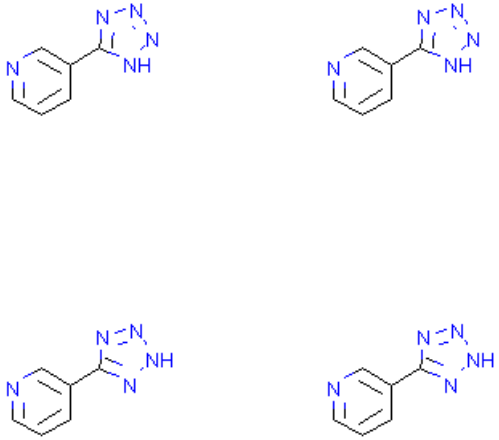
21	aminopyrimidine			[1.0]	[1.0]	[0]

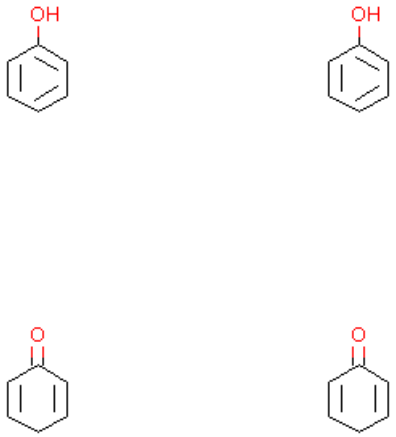
22	105and106			[1.0]	[1.0]	[0]

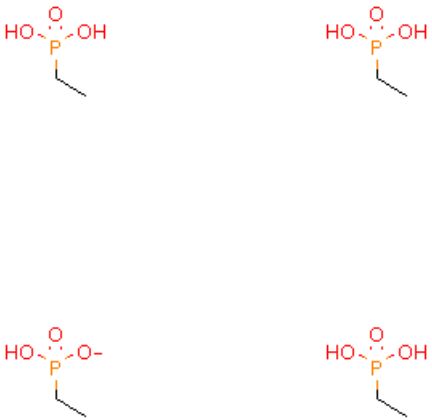
23	Rhodamine			[1.0]	[1.0]	[0]

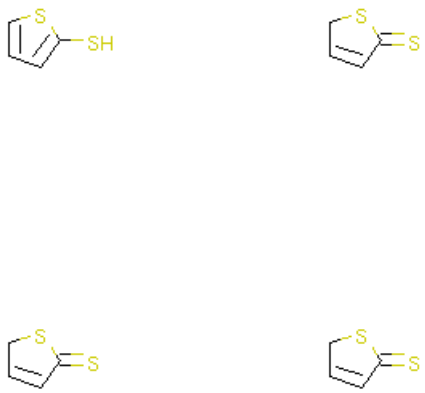
24	sulfon			[1.0]	[1.0]	[0]

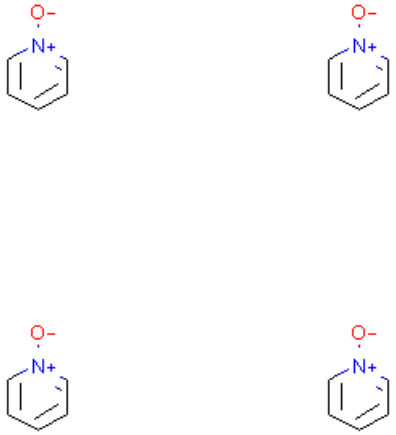
25	triazine			[0.7]	[0.7]	[21]

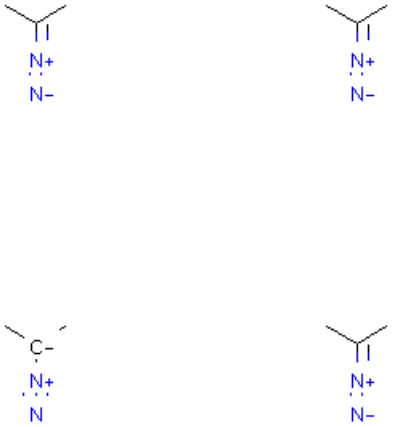
26	tetrazole			[1.0]	[0.73]	[22]

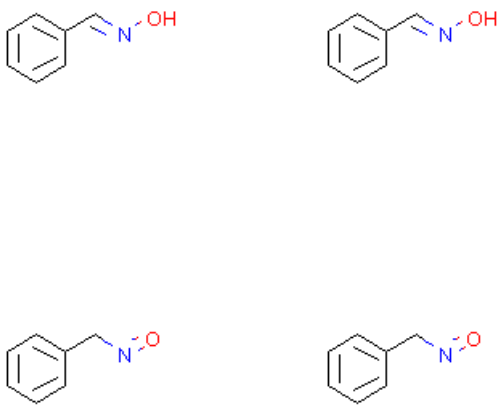
27	phenol			[0.4]	[0.25]	[68]

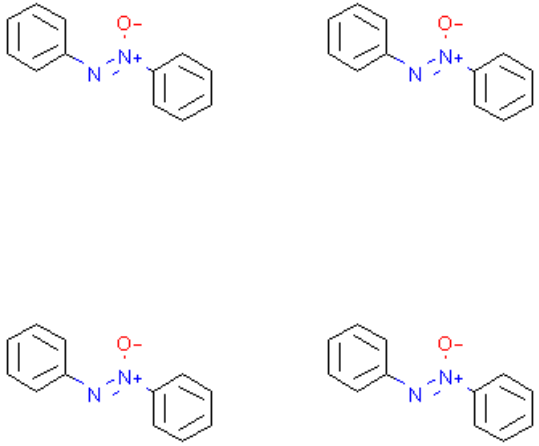
28	phosphor			[1.0]	[1.0]	[0]

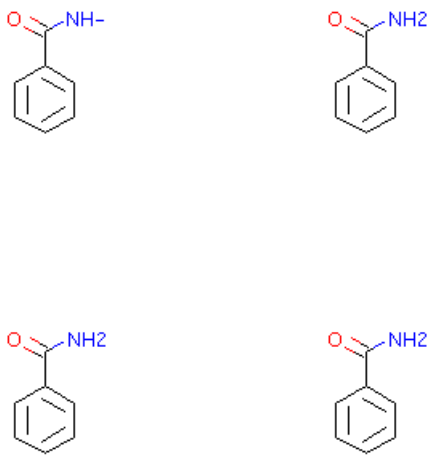
29	112and113			[1.0]	[1.0]	[0]

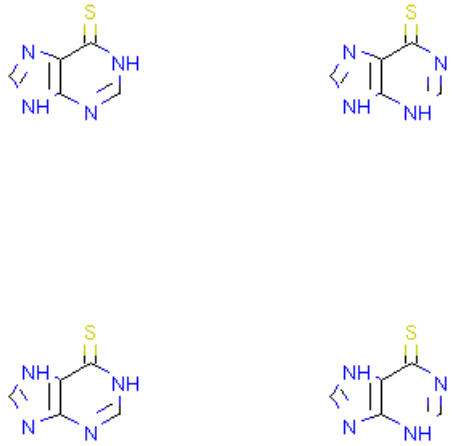
30	NO			[1.0]	[1.0]	[0]

31	diazo			[1.0]	[1.0]	[0]

32	benzaldoxime			[0.57]	[0.58]	[55]

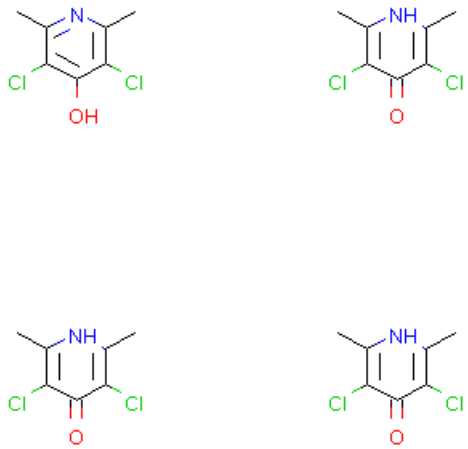
33	NNO			[1.0]	[1.0]	[0]

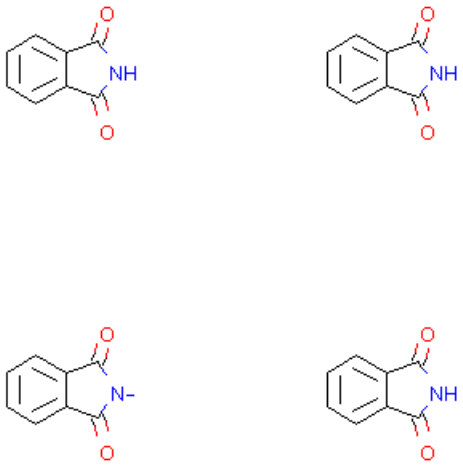
34	amide			[1.0]	[1.0]	[0]

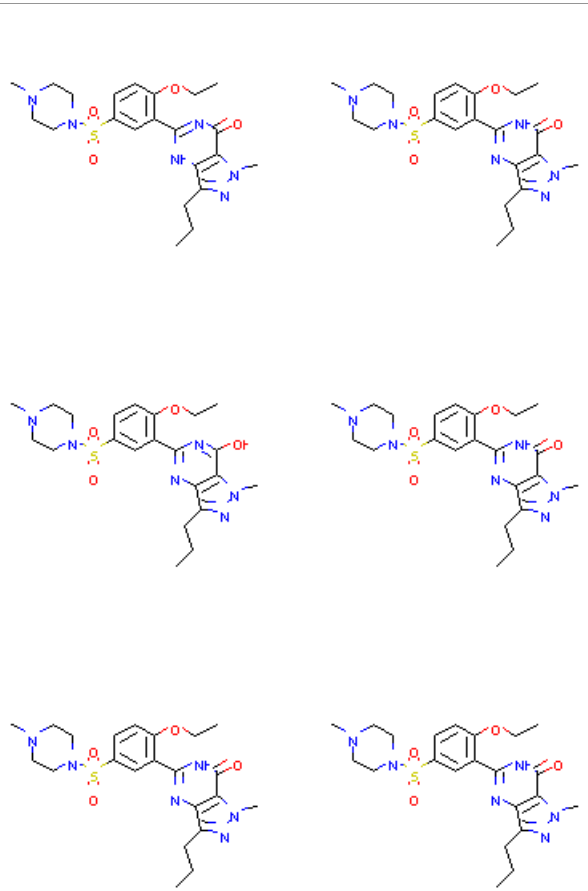
35	thiopurine			[1.0]	[0.72]	[21]

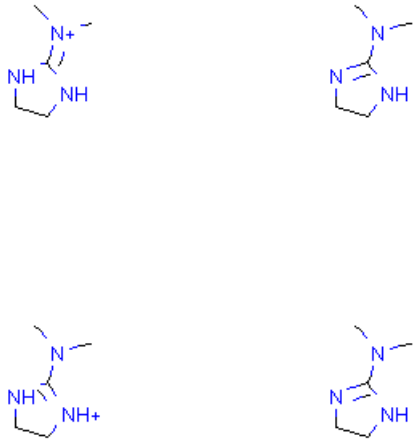
2 Additional standardization with MolVS

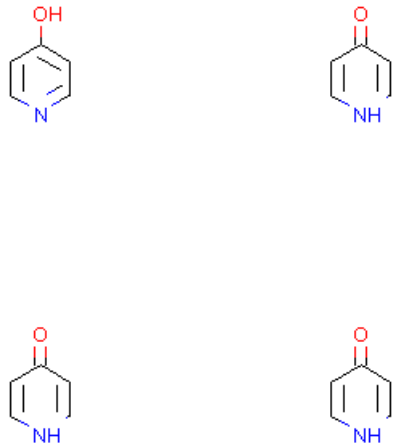
In addition to running Standardizer by ChemAxon, the canonicalize and uncharge options of MolVS were used.

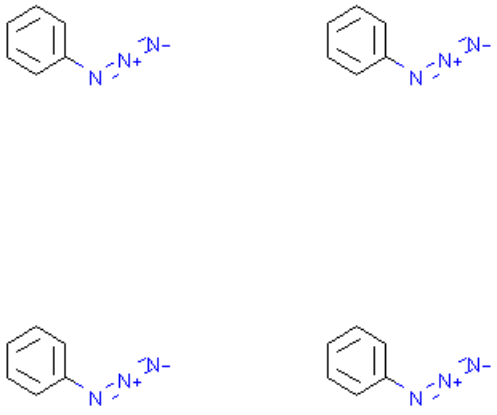
Idx	Name	Structure In and Out	Comment	TanTop	DiceMorgan	No. of DescDiff
1	clopidol			[1.0]	[1.0]	[0]

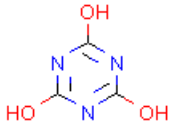
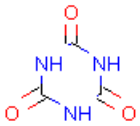
2	phthalimide			[1.0]	[1.0]	[0]

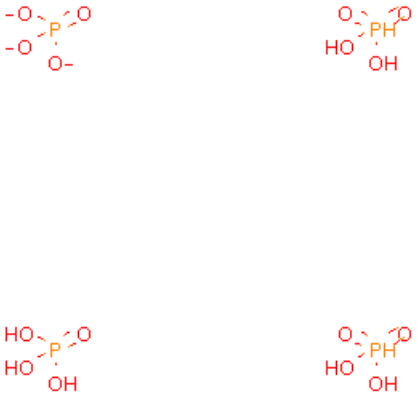
3	Viagra			[1.0, 1.0, 1.0]	[1.0, 1.0, 1.0]	[0, 0, 0]

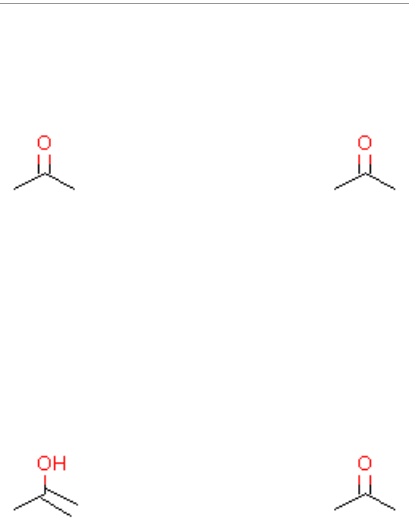
4	quatAmMesomer			[1.0]	[1.0]	[0]

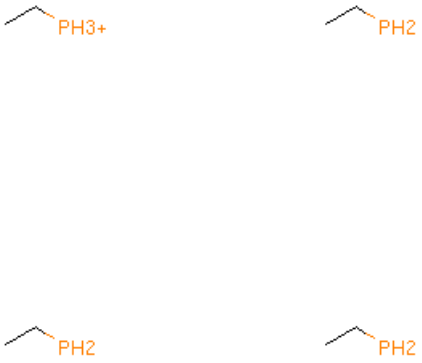
5	pyridinol			[1.0]	[1.0]	[0]

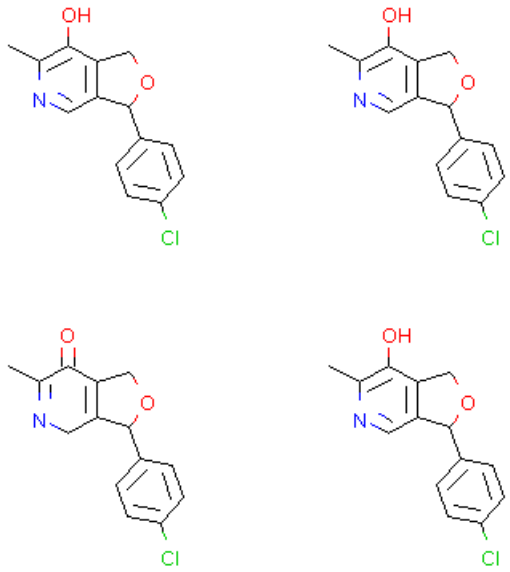
6	azide			[1.0]	[1.0]	[0]

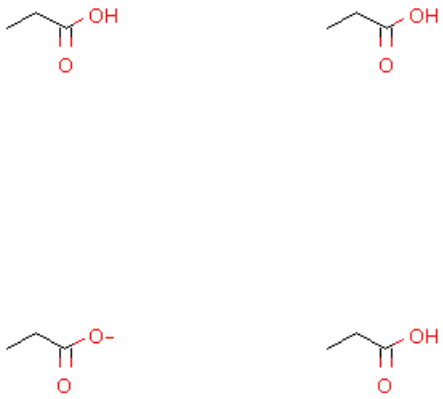
7	triazinol	   		[1.0]	[1.0]	[0]

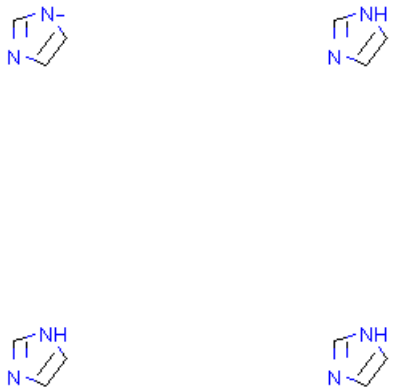
8	phosphate			[1.0]	[1.0]	[0]

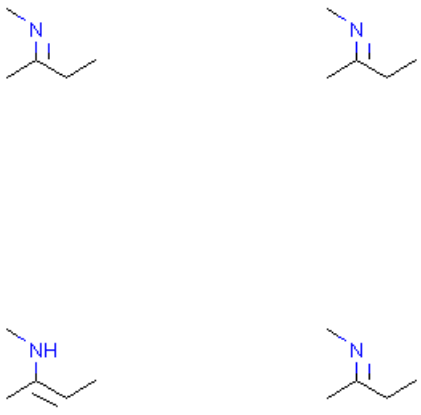
9	Propenol			[1.0]	[1.0]	[0]

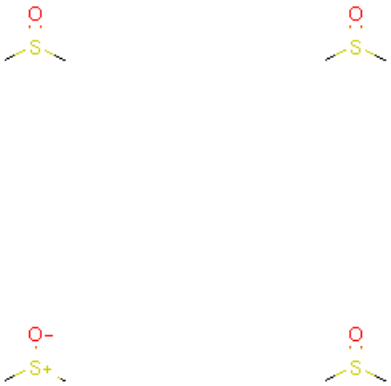
10	phosphine			[1.0]	[1.0]	[0]

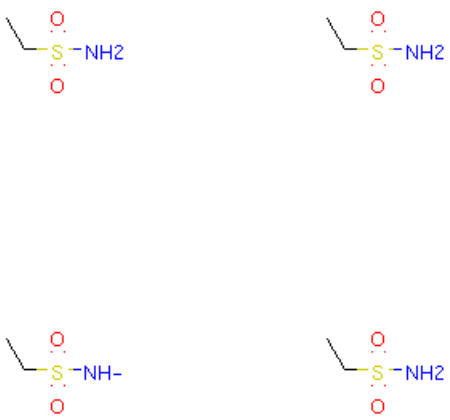
11	cicletanine			[1.0]	[1.0]	[0]

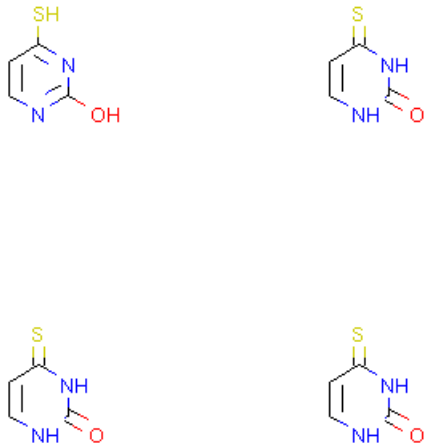
12	propaneAcid			[1.0]	[1.0]	[0]

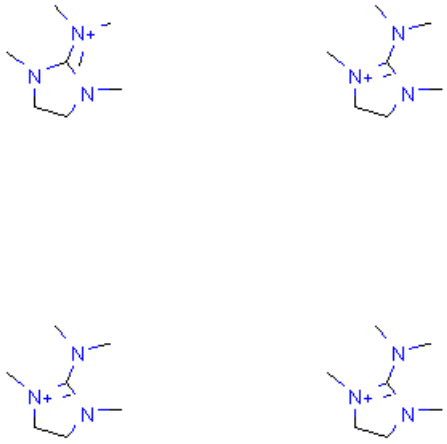
13	imidazole			[1.0]	[1.0]	[0]

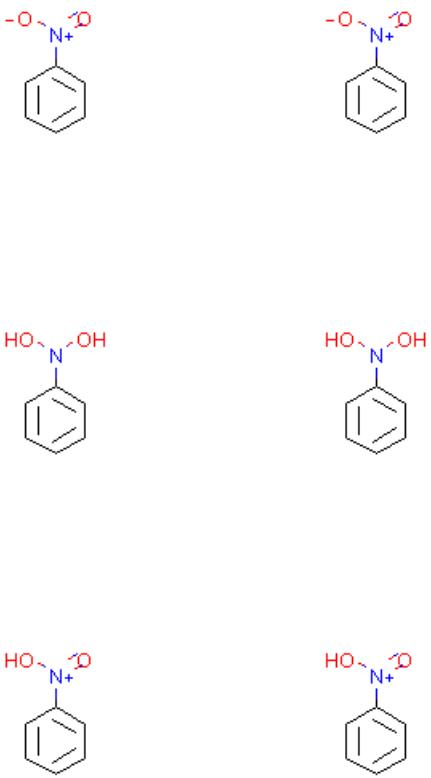
14	enamine			[1.0]	[1.0]	[0]

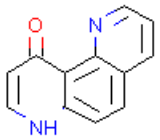
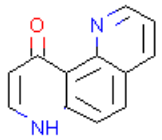
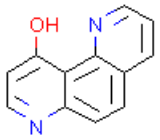
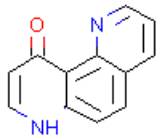
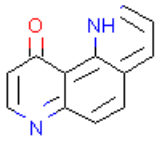
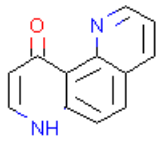
15	sulfoxide			[1.0]	[1.0]	[0]

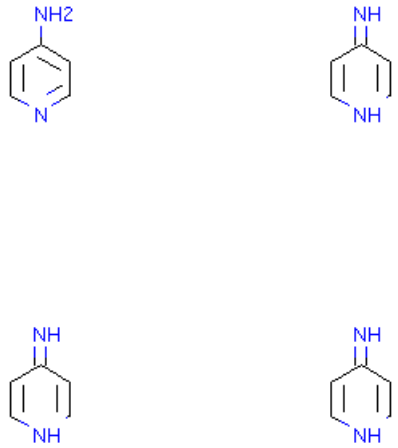
16	sulfonamide			[1.0]	[1.0]	[0]

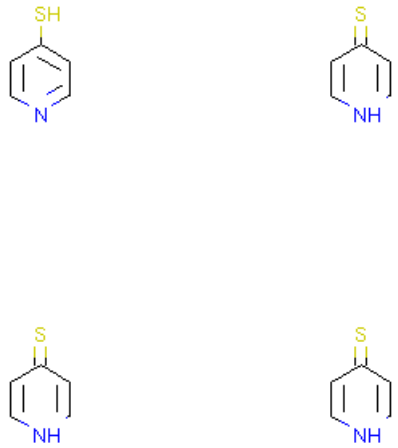
17	tiouracil			[1.0]	[1.0]	[0]

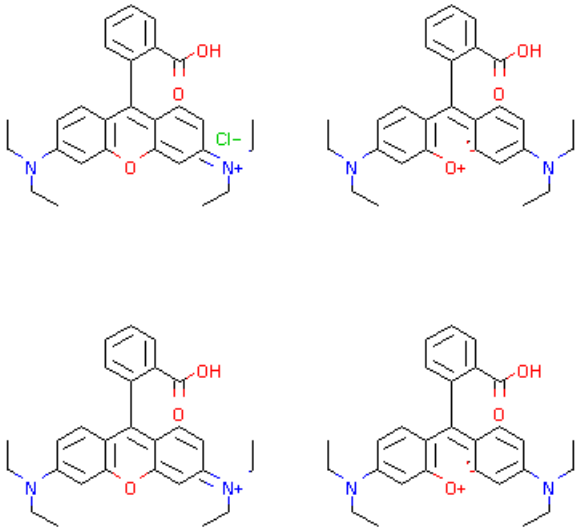
18	quatAmMesomer2			[1.0]	[1.0]	[0]

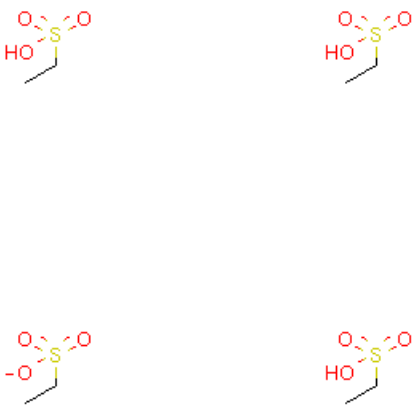
19	nitro			[0.67, 1.0, 0.67]	[0.56, 0.8, 0.6]	[58, 43, 56]

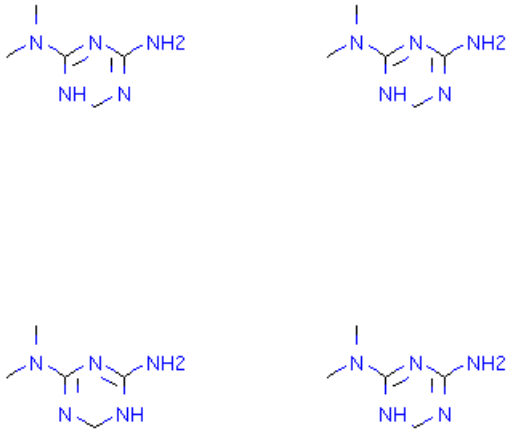
20	107and108and109	     		[1.0, 1.0, 1.0]	[1.0, 1.0, 1.0]	[0, 0, 0]

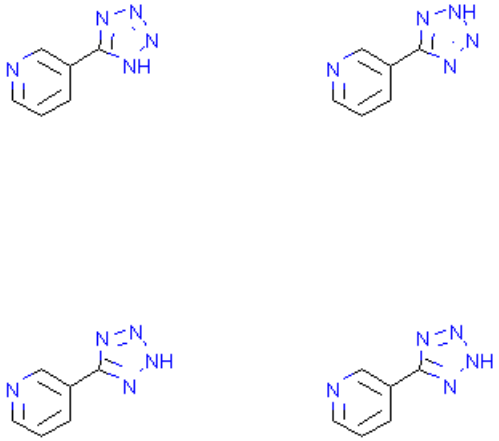
21	aminopyrimidine			[1.0]	[1.0]	[0]

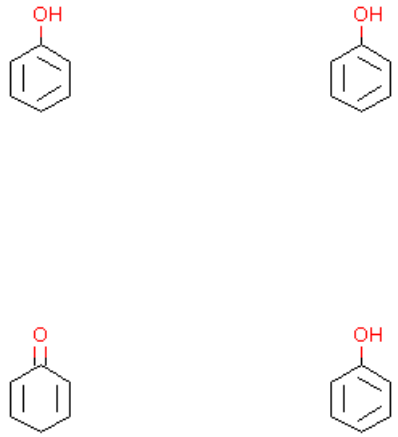
22	105and106			[1.0]	[1.0]	[0]

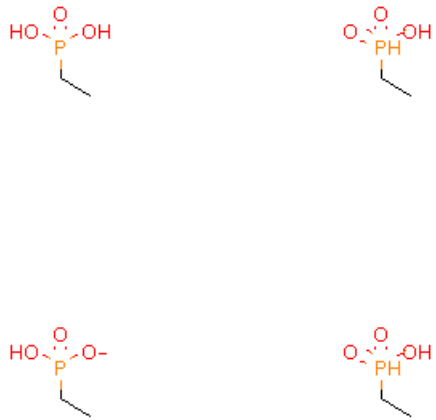
23	Rhodamine			[1.0]	[1.0]	[0]

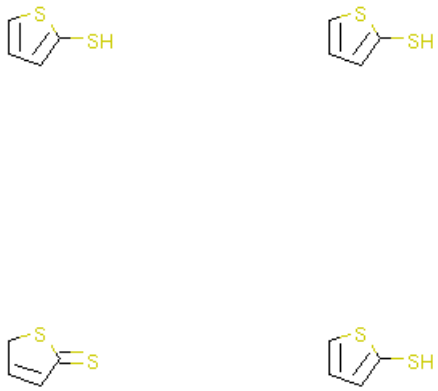
24	sulfon			[1.0]	[1.0]	[0]

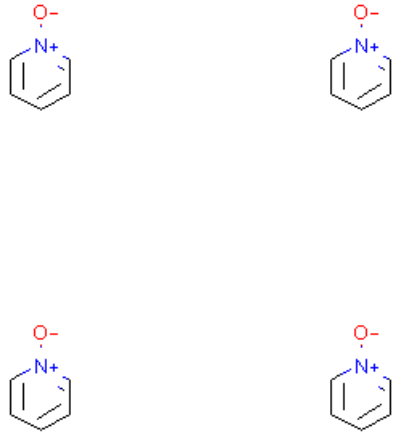
25	triazine			[1.0]	[1.0]	[0]

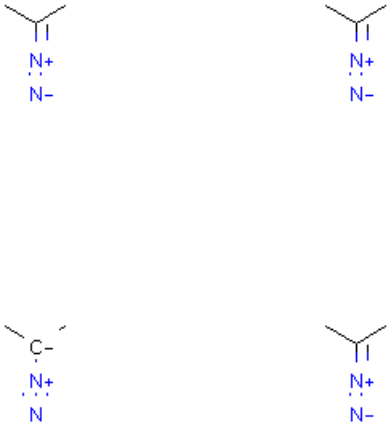
26	tetrazole			[1.0]	[1.0]	[1]

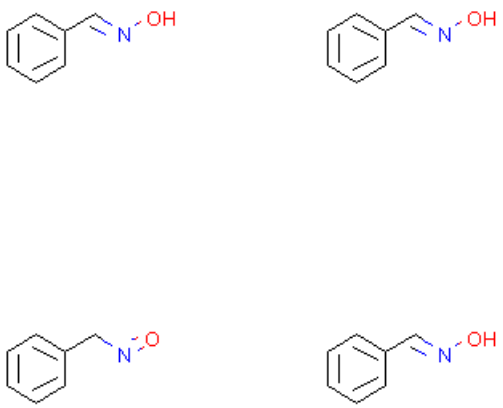
27	phenol			[1.0]	[1.0]	[0]

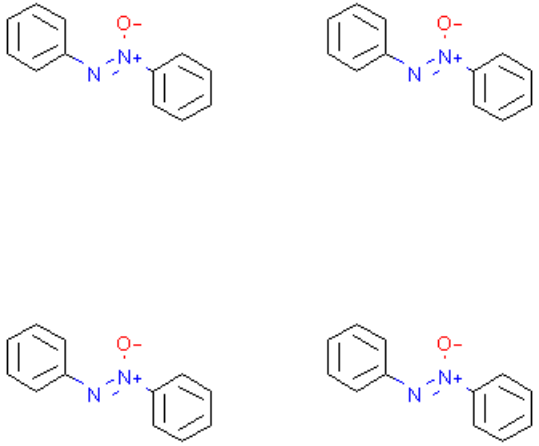
28	phosphor			[1.0]	[1.0]	[0]

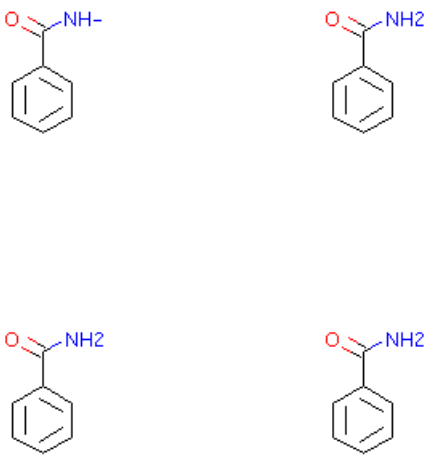
29	112and113			[1.0]	[1.0]	[0]

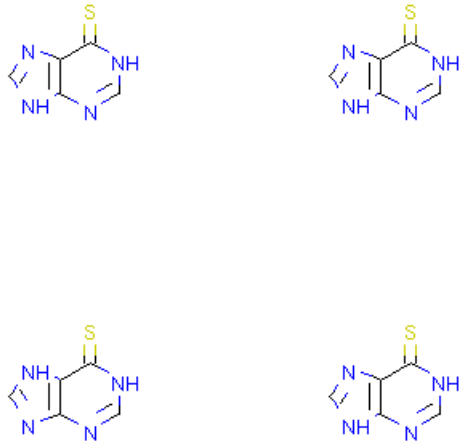
30	NO			[1.0]	[1.0]	[0]

31	diazo			[1.0]	[1.0]	[0]

32	benzaldoxime			[1.0]	[1.0]	[0]

33	NNO			[1.0]	[1.0]	[0]

34	amide			[1.0]	[1.0]	[0]

35	thiopurine			[1.0]	[1.0]	[0]