

Molecular standardization

June 14, 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.

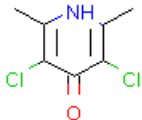
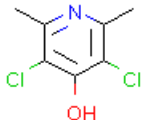
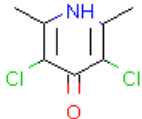
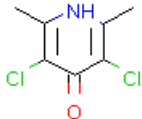
For each method a table is produced showing the input and output structures as drawn by RDKit for the different forms. The table also displays two fingerprint similarities between all pairs of output structures. If the standardization is successful the fingerprints should be identical. Furthermore, the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

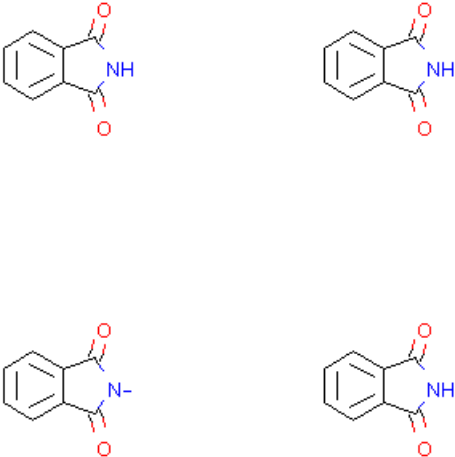
1 Standardization with ChemAxon Standardizer

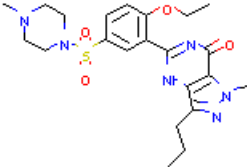
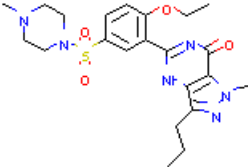
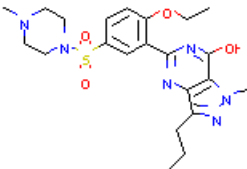
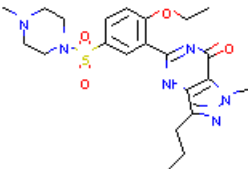
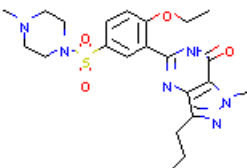
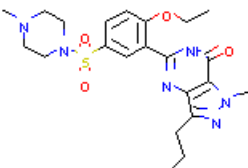
The ChemAxon standardizer was used with 7 actions, applied in the following order, together with the set of transformations defined in the GUI:

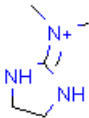
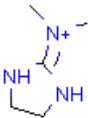
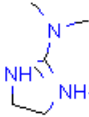
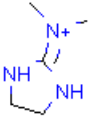
- removesolvents
- stripsalts
- removefragment:method=keeplargest
- mesomerize
- tautomerize
- neutralize

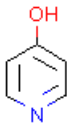
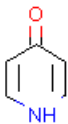
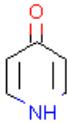
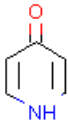
- transformations
- dearomatize

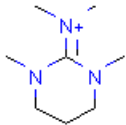
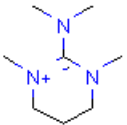
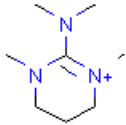
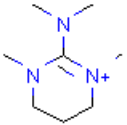
Idx	Name	Structure In and Out	Comment	TanTop	DiceMorgan	No. of DescDiff
1	clopidol	<div></div> <div></div>		[1.0]	[1.0]	[0]

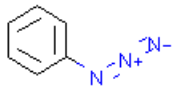
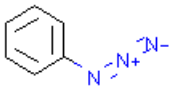
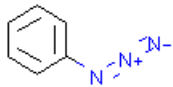
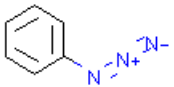
2	phthalimide		[1.0]	[1.0]	[0]
---	-------------	--	-------	-------	-----

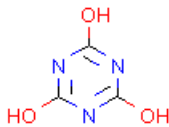
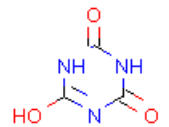
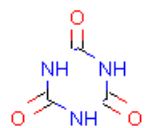
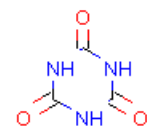
3	Viagra	     	[1.0, 1.0, 1.0]	[1.0, 0.91, 0.91]	[1, 21, 21]
---	--------	--	-----------------	-------------------	-------------

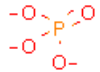

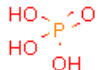

4	quatAmMesomer	 				
		 		[1.0]	[1.0]	[0]





5	pyridinol	 				
		 		[1.0]	[1.0]	[0]





6	quatAmMesomer4	 				
		 		[1.0]	[1.0]	[1]


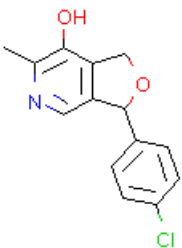
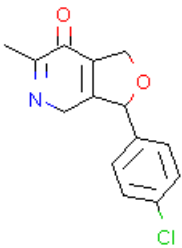
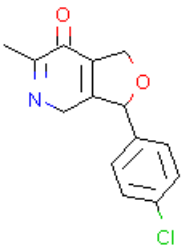
7	azide	   	[1.0]	[1.0]	[0]
---	-------	--	-------	-------	-----

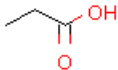
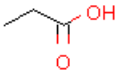
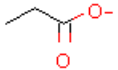
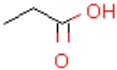
8	triazinol	   	[0.58]	[0.54]	[50]
---	-----------	--	--------	--------	------





9	phosphate	   		[1.0]	[1.0]	[0]
---	-----------	---	--	-------	-------	-----





10	Propenol	   		[1.0]	[1.0]	[0]
----	----------	---	--	-------	-------	-----





11	phosphine	<div><div></div><div></div></div> <div><div></div><div></div></div>		[1.0]	[1.0]	[0]
----	-----------	---	--	-------	-------	-----

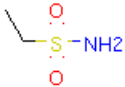
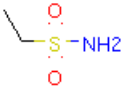
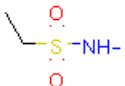
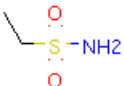
12	cicletanine	   	[0.3]	[0.67]	[76]
----	-------------	---	-------	--------	------

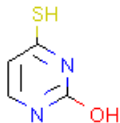
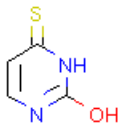
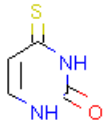
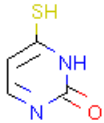
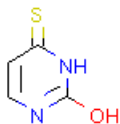
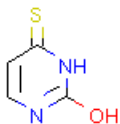

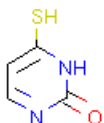
13	propaneAcid	   		[1.0]	[1.0]	[0]
----	-------------	--	--	-------	-------	-----

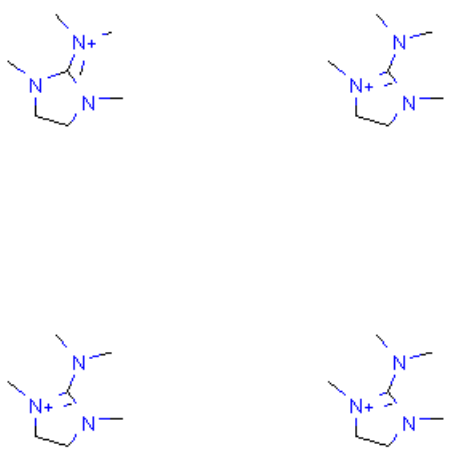
14	imidazole	   		[1.0]	[1.0]	[0]
----	-----------	--	--	-------	-------	-----

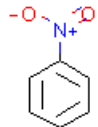
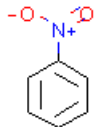
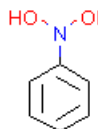
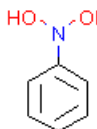
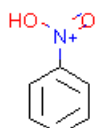
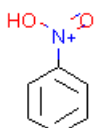
15	enamine	   		[1.0]	[1.0]	[0]
----	---------	--	--	-------	-------	-----

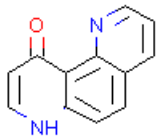
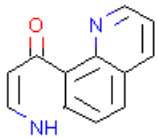
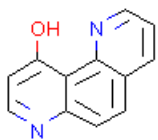
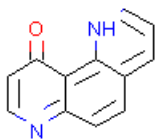
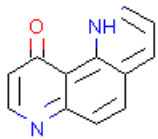
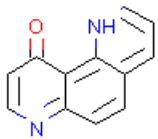
16	sulfoxide	 				
		 				
				[1.0]	[1.0]	[0]

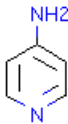
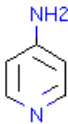
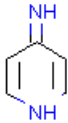
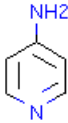
17	sulfonamide	   		[1.0]	[1.0]	[0]
----	-------------	--	--	-------	-------	-----


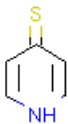


18	tiouracil	       	[0.35, 1.0, 0.35, 0.35, 1.0, 0.35]	[0.5, 1.0, 0.5, 0.5, 1.0, 0.5]	[54, 0, 54, 54, 0,
----	-----------	--	------------------------------------	--------------------------------	--------------------

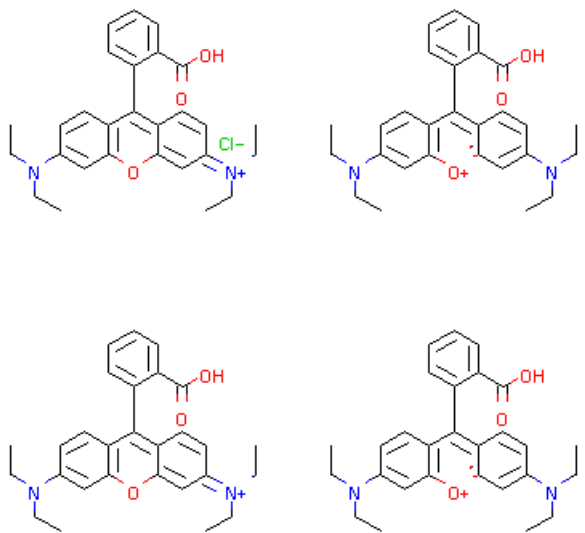
19	quatAmMesomer2		[1.0]	[1.0]	[0]
----	----------------	--	-------	-------	-----





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

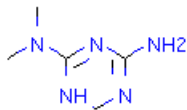
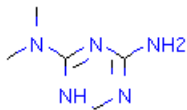
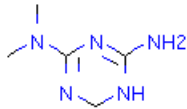
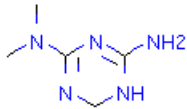
21	107and108and109	     	[1.0, 1.0, 1.0]	[0.7, 0.7, 1.0]	[17, 17, 0]
----	-----------------	--	-----------------	-----------------	-------------

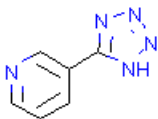
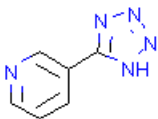
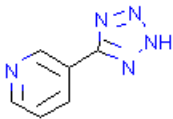
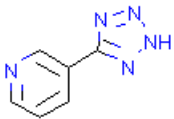
22	aminopyrimidine	   		[1.0]	[1.0]	[0]
----	-----------------	--	--	-------	-------	-----

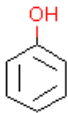
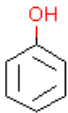

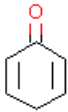
23	105and106	 				
		 		[1.0]	[1.0]	[0]

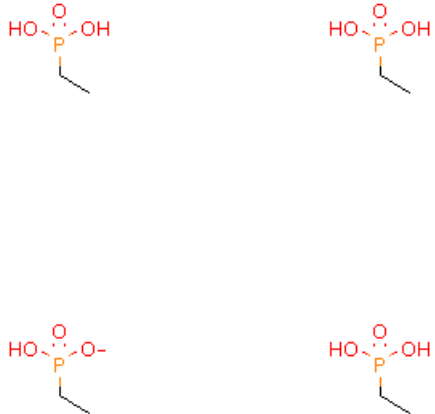
24	Rhodamine	 <p>The image displays four chemical structures of Rhodamine, arranged in a 2x2 grid. Each structure consists of a central xanthene core with two diethylamino groups (N(CH₂CH₃)₂) at the 3 and 6 positions and a phenyl ring at the 10 position. The phenyl ring is substituted with a carboxylic acid group (-COOH) at the para position. The four structures represent different ionization states: 1. Top-left: Neutral form with a carbonyl oxygen (O) at position 2 and a chloride counterion (Cl⁻) at position 4. 2. Top-right: Zwitterionic form with a carbonyl oxygen (O) at position 2 and a positive charge on the carbonyl carbon (C=O⁺) at position 4. 3. Bottom-left: Radical cation form with a carbonyl oxygen (O) at position 2 and a positive charge on the carbonyl carbon (C=O⁺) at position 4, with a radical dot on the carbonyl oxygen. 4. Bottom-right: Radical cation form with a carbonyl oxygen (O) at position 2 and a positive charge on the carbonyl carbon (C=O⁺) at position 4, with a radical dot on the carbonyl carbon.</p>	[1.0]	[1.0]	[0]
----	-----------	--	-------	-------	-----

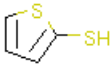



25	sulfon	   	[1.0]	[1.0]	[0]
----	--------	---	-------	-------	-----

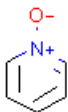
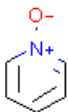
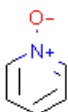
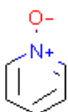
26	triazine	   		[0.7]	[0.7]	[21]
----	----------	--	--	-------	-------	------

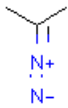
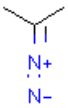
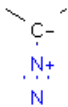
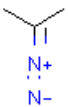
27	tetrazole	   	[1.0]	[0.73]	[22]
----	-----------	--	-------	--------	------

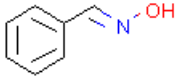
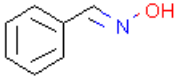
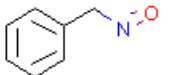
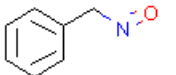
28	phenol	 				
		 		[0.4]	[0.25]	[68]

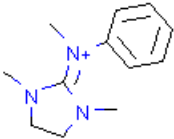
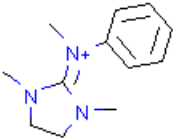
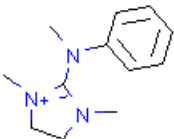
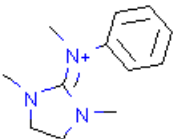
29	phosphor			[1.0]	[1.0]	[0]
----	----------	--	--	-------	-------	-----

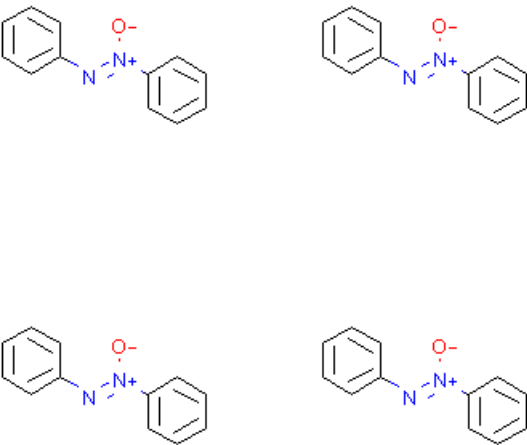
30	112and113	   	[1.0]	[1.0]	[0]
----	-----------	---	-------	-------	-----

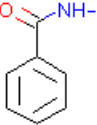
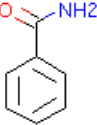
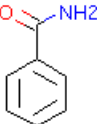
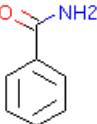
31	NO	   		[1.0]	[1.0]	[0]
----	----	--	--	-------	-------	-----

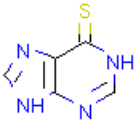



32	diazo	 				
		 		[1.0]	[1.0]	[0]

33	benzaldoxime	 				
		 		[0.57]	[0.58]	[55]

34	quatAmMesomer3	   		[1.0]	[1.0]	[0]
----	----------------	---	--	-------	-------	-----

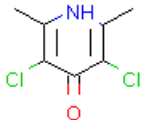
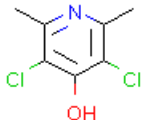
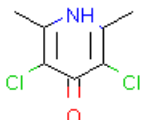
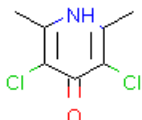
35	NNO			[1.0]	[1.0]	[0]
----	-----	--	--	-------	-------	-----

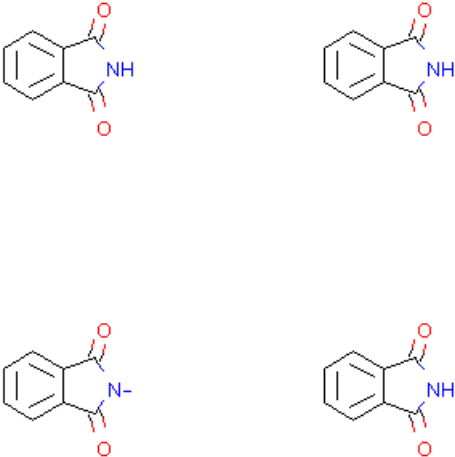
36	amide	   		[1.0]	[1.0]	[0]
----	-------	---	--	-------	-------	-----

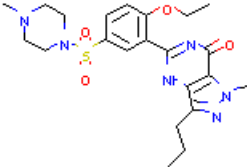
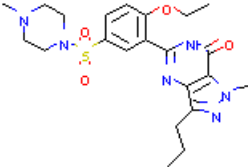
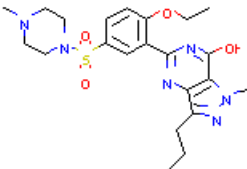
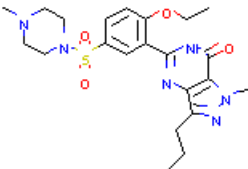
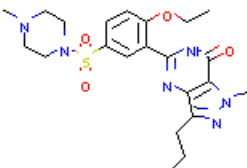
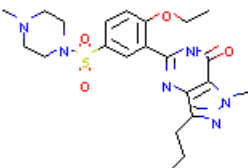
37	thiopurine	   		[1.0]	[0.72]	[21]
----	------------	--	--	-------	--------	------

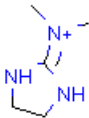
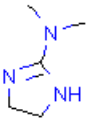
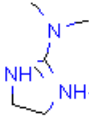
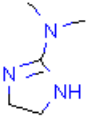
2 Additional standardization with MolVS

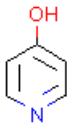
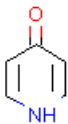
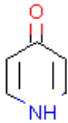
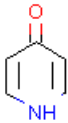
In addition to running Standardizer by ChemAxon, the canonicalize and uncharge options of MolVS were used.

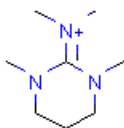
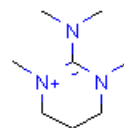
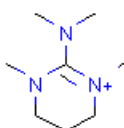
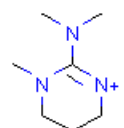
Idx	Name	Structure In and Out	Comment	TanTop	DiceMorgan	No. of DescDiff
1	clopidol	<div></div> <div></div>		[1.0]	[1.0]	[0]

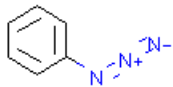
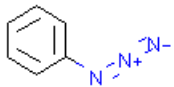
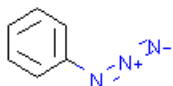
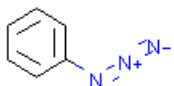
2	phthalimide	 <p>The image displays four chemical structures of phthalimide and its zwitterionic form, arranged in a 2x2 grid. The top-left and top-right structures are neutral phthalimide molecules, consisting of a benzene ring fused to a five-membered imide ring. The bottom-left and bottom-right structures are zwitterionic forms of phthalimide, where the nitrogen atom carries a negative charge (N-) and the carbonyl groups are represented as C=O+.</p>	[1.0]	[1.0]	[0]
---	-------------	---	-------	-------	-----

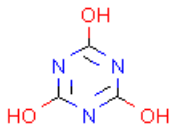
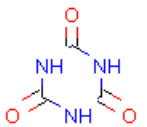
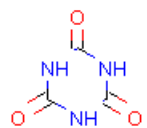
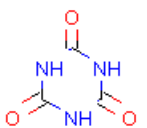
3	Viagra	     	[1.0, 1.0, 1.0]	[1.0, 1.0, 1.0]	[0, 0, 0]
---	--------	--	-----------------	-----------------	-----------





4	quatAmMesomer	   	[1.0]	[1.0]	[0]
---	---------------	--	-------	-------	-----



5	pyridinol	 				
		 		[1.0]	[1.0]	[0]





6	quatAmMesomer4	 				
		 		[1.0]	[1.0]	[1]


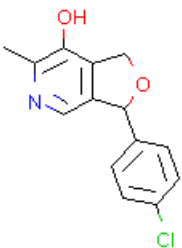
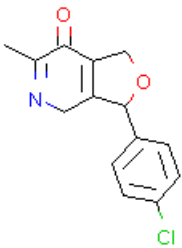
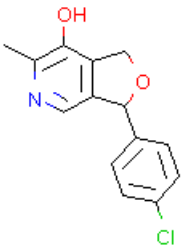
7	azide	   	[1.0]	[1.0]	[0]
---	-------	--	-------	-------	-----

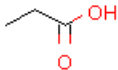
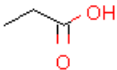
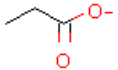
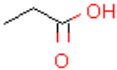
8	triazinol	   	[1.0]	[1.0]	[0]
---	-----------	--	-------	-------	-----





9	phosphate	   		[1.0]	[1.0]	[0]
---	-----------	---	--	-------	-------	-----





10	Propenol	<div></div> <div></div>		[1.0]	[1.0]	[0]
----	----------	---	--	-------	-------	-----





11	phosphine	<div><div></div><div></div></div> <div><div></div><div></div></div>	[1.0]	[1.0]	[0]
----	-----------	---	-------	-------	-----

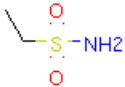
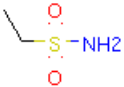
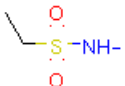
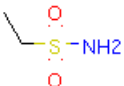
12	cicletanine	   	[1.0]	[1.0]	[0]
----	-------------	---	-------	-------	-----

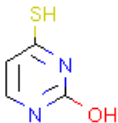
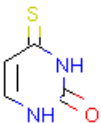

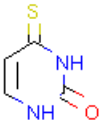
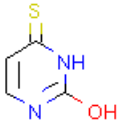
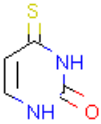
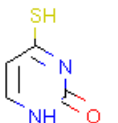
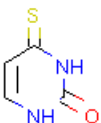
13	propaneAcid	   	[1.0]	[1.0]	[0]
----	-------------	---	-------	-------	-----

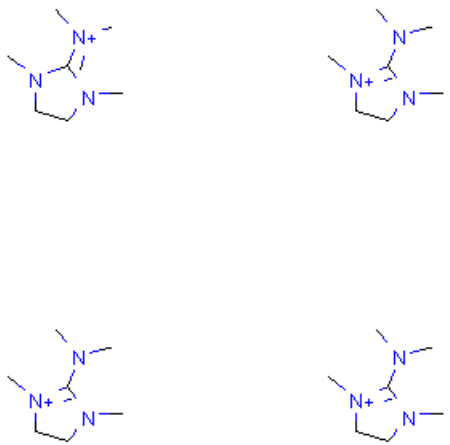
14	imidazole	   		[1.0]	[1.0]	[0]
----	-----------	--	--	-------	-------	-----

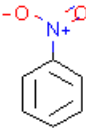
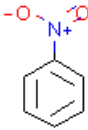
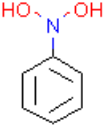
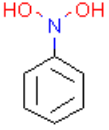
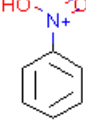
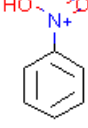
15	enamine	   		[1.0]	[1.0]	[0]
----	---------	--	--	-------	-------	-----

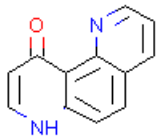
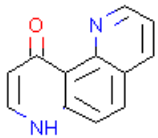
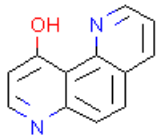
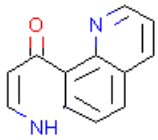
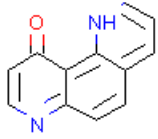
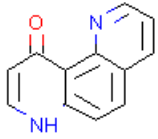
16	sulfoxide	 				
		 		[1.0]	[1.0]	[0]

17	sulfonamide	   	[1.0]	[1.0]	[0]
----	-------------	--	-------	-------	-----


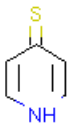


18	tiouracil	       	[1.0, 1.0, 1.0, 1.0, 1.0, 1.0]	[1.0, 1.0, 1.0, 1.0, 1.0, 1.0]	[0, 0, 0, 0, 0, 0]
----	-----------	---	--------------------------------	--------------------------------	--------------------

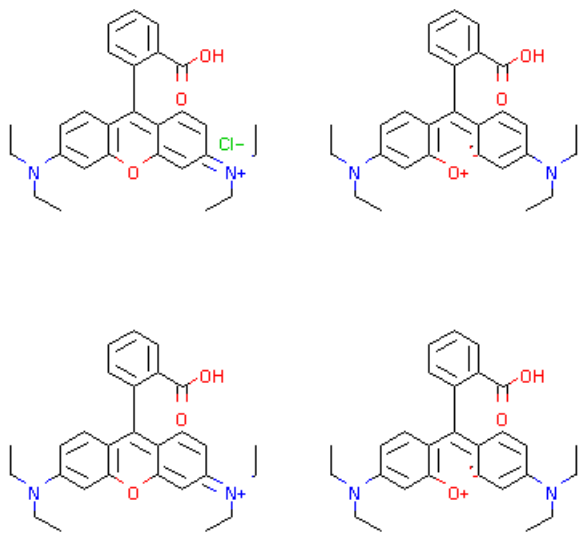
19	quatAmMesomer2		[1.0]	[1.0]	[0]
----	----------------	--	-------	-------	-----





20	nitro	 				
		 				
		 				

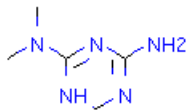
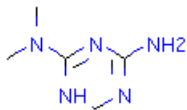
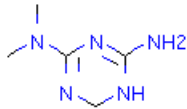
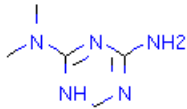
21	107and108and109	     	[1.0, 1.0, 1.0]	[1.0, 1.0, 1.0]	[0, 0, 0]
----	-----------------	--	-----------------	-----------------	-----------

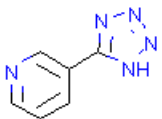
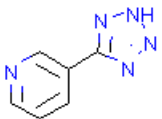
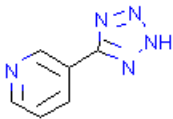
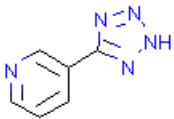
[illegible]

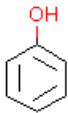
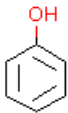
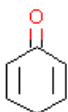
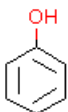
23	105and106	 				
		 		[1.0]	[1.0]	[0]

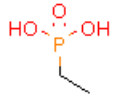
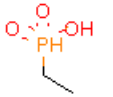
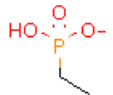
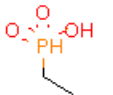
24	Rhodamine	 <p>The image displays four chemical structures of Rhodamine, arranged in a 2x2 grid. The top row shows the ground state (left) and the excited singlet state (right). The bottom row shows the ground state (left) and the excited triplet state (right). Each structure consists of a xanthene core with two diethylamino groups at the 3 and 6 positions and a phenylcarboxylic acid group at the 10 position. In the ground state, the central oxygen atom is double-bonded to the ring and carries a negative charge, while the nitrogen atom at the 4 position is double-bonded to the ring and carries a positive charge. In the excited singlet state, the central oxygen atom is single-bonded to the ring and carries a positive charge, while the nitrogen atom is double-bonded to the ring and carries a negative charge. In the excited triplet state, the central oxygen atom is single-bonded to the ring and carries a positive charge, while the nitrogen atom is double-bonded to the ring and carries a negative charge. The structures are color-coded: the xanthene core is black, the diethylamino groups are blue, the phenylcarboxylic acid group is red, and the central oxygen and nitrogen atoms are green.</p>	[1.0]	[1.0]	[0]
----	-----------	---	-------	-------	-----

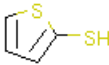
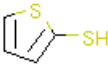

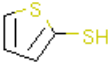
25	sulfon	   				
				[1.0]	[1.0]	[0]

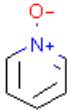
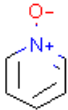
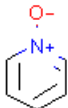
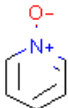
26	triazine	   	[1.0]	[1.0]	[0]
----	----------	--	-------	-------	-----

27	tetrazole	   	[1.0]	[1.0]	[1]
----	-----------	--	-------	-------	-----

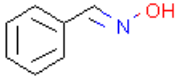
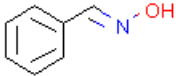
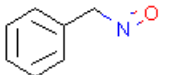
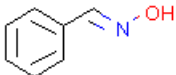
28	phenol	 				
		 		[1.0]	[1.0]	[0]

29	phosphor	   	[1.0]	[1.0]	[0]
----	----------	---	-------	-------	-----

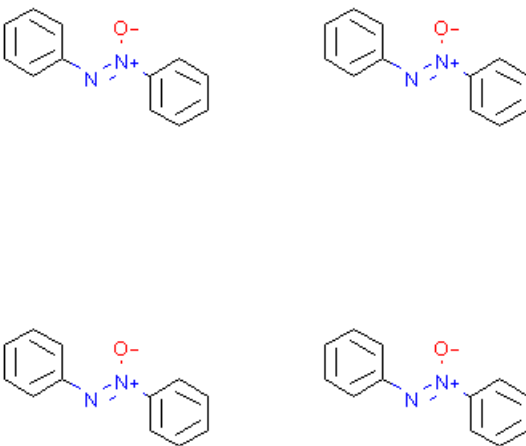
30	112and113	   	[1.0]	[1.0]	[0]
----	-----------	--	-------	-------	-----

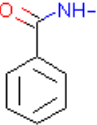
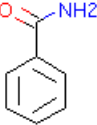
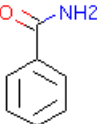
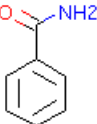
31	NO	   		[1.0]	[1.0]	[0]
----	----	--	--	-------	-------	-----

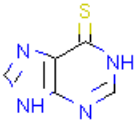
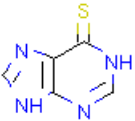


32	diazo			[1.0]	[1.0]	[0]

33	benzaldoxime	 				
		 				
				[1.0]	[1.0]	[0]

34	quatAmMesomer3	 				
				[1.0]	[1.0]	[0]

35	NNO					
			[1.0]	[1.0]	[0]	

36	amide	 				
		 		[1.0]	[1.0]	[0]

37	thiopurine	   		[1.0]	[1.0]	[0]
----	------------	---	--	-------	-------	-----