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
- $\bullet \ \ remove fragment: method = keep largest \\$
- mesomerize
- tautomerize
- neutralize

- ullet transformations
- \bullet dearomatize

•	٨	•	١	

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

			NH	NH			
A	2	phthalimide	N- O	NH	[1.0]	[1.0]	[0]

		N N N N N N N N N N N N N N N N N N N	N N O N N N N N N N N N N N N N N N N N			
		N O O O O O O O O O O O O O O O O O O O	N N S N N N N N N N N N N N N N N N N N			
		N N N N N N N N N N N N N N N N N N N	N N N N N N N N N N N N N N N N N N N			
3	Viagra			[1.0, 1.0, 1.0]	[1.0, 0.91, 0.91]	[1, 21, 21]

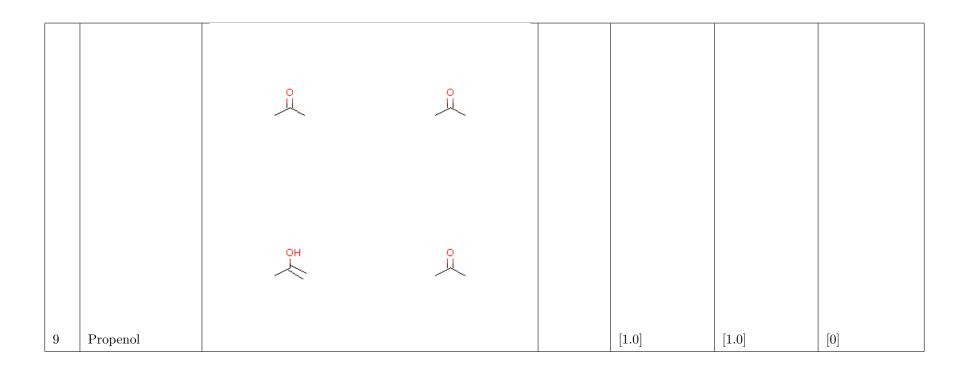
		OH L	ONH NH			
5	pyridinol	O I NH	O NH	[1.0]	[1.0]	[0]

		N N+N-	N N+TN-			
6	azide	N°N+N-	N-N+ ² N-	[1.0]	[1.0]	[0]

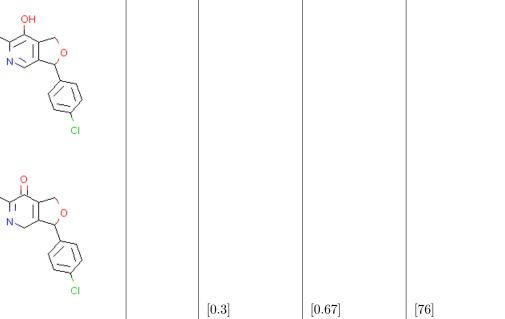
_		
_		
$\overline{}$	_	
	-	

		-0 P 0 -0 -0 -0 -0 -0 -0 -0 -0 -0 -0 -0 -0 -	но _Р , О			
8	phosphate	но _Р о но он	но _р о но он	[1.0]	[1.0]	[0]

_	-
_	



		→ PH3+	PH2			
10	phosphine	PH2	PH2	[1.0]	[1.0]	[0]



		OH O	О ОН				
12	propaneAcid	~~~°-	ОН	[1.0	D]	[1.0]	[0]

		N-\	N NH			
13	imidazole	N NH	N NH	[1.0]	[1.0]	[0]

۲	-			
	3	١	١	
•	_	٠	,	

		\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\	N —				
14	enamine	NH ————————————————————————————————————	N —	[1.0	0]	[1.0]	[0]

		O. S.=NH2 O	0. 5. NH2 0				
120	16 sulfonamide	9. -S-NH- 0	0. s-NH2 0	[1.0]	[1.0]	[0]	

		SHOH	S NH OH				
17	tiouracil	N H VO	SH NH NH O	[0.	35]	[0.5]	[54]

N	
r.	$\overline{}$
•	_

		N+ 1 N N N N N N N N N N N N N N N N N N	N+ N				
18	quatAmMesomer2	N+ N	N+ N	[:	1.0]	[1.0]	[0]

		-0 \ N+	-0 N+			
		HONNOH	HO NOH			
19	nitro	HO N+	HO N+	[0.67, 1.0, 0.67]	[0.56, 0.8, 0.6]	[58, 43, 56]

		O N N	NH NH			
		OH N	O NH			
20	107and108and109	O NH	O NH	[1.0, 1.0, 1.0]	[0.7, 0.7, 1.0]	[17, 17, 0]

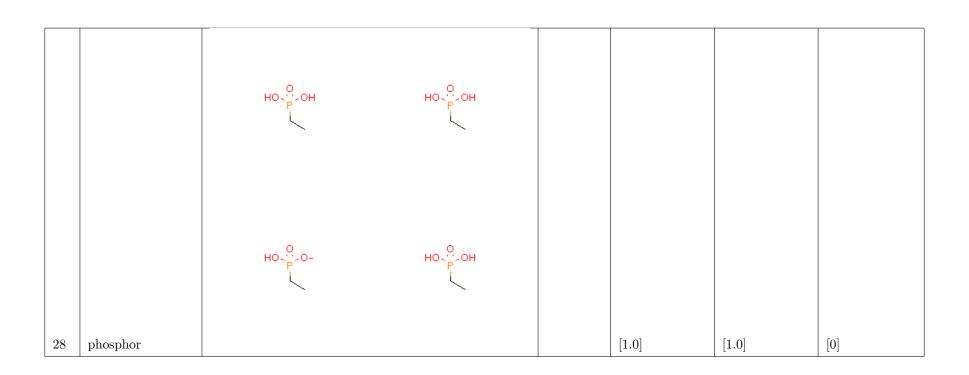
ı	ು
١	Α

	SH	S NH			
22 105and106	S I I NH	S NH	[1.0]	[1.0]	[0]

	0)s-20 H0	О.S. 7-О НО			
24 sulfon	0)s:0	0.35°,0 но	[1.0]	[1.0]	[0]

	•
f.	
c	

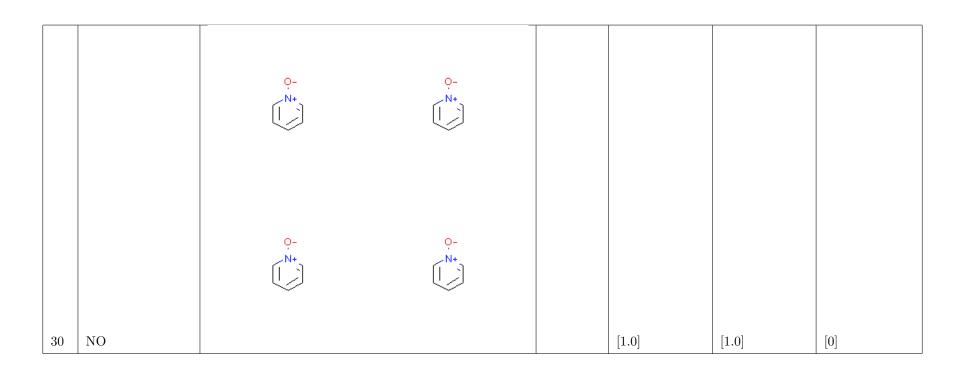
			(OH	OH 			
29	27	phenol	(<u> </u>	[0.4]	[0.25]	[68]



C	S
_	

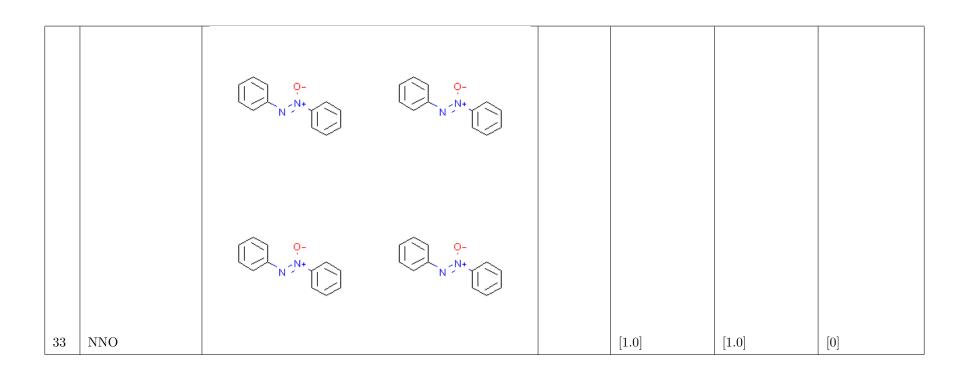
		SH SH	S S			
29 112	2 and113	S S S	S S S	[1.0]	[1.0]	[0]

٧.	·	
h	•	
ι		•



ಲ	
Δ	

		N OH	N OH			
32	benzaldoxime	N ^z -O	N ⁻⁰	[0.57]	[0.58]	[55]



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
		CI OH	NH CI O				
1	clopidol	NH CI CI	NH CI O		[1.0]	[1.0]	[0]

		NH	NH O			
2	phthalimide	N- O	NH O	[1.0]	[1.0]	[0]

		N N O N N N N N	N N S N N N N N			
		N O O O O O O O O O O O O O O O O O O O	N N N N N N N N N N N N N N N N N N N			
		N N N N N N N N N N N N N N N N N N N	N N N N N N N N N N N N N N N N N N N			
3	Viagra			[1.0, 1.0, 1.0]	[1.0, 1.0, 1.0]	[0, 0, 0]

L		
,	•	
ι	•	•

		N+ NH NH	N NH			
4	${ m quatAmMesomer}$	NH+	N NH	[1.0]	[1.0]	[0]

	•	,

			OH N	O I I			
43	5	pyridinol	O III	O III	[1.0]	[1.0]	[0]

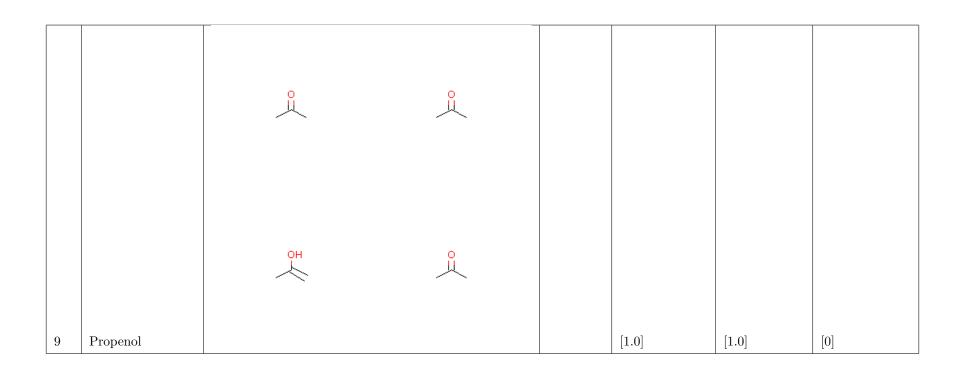
\sim
$\overline{}$
\rightarrow

		N-N+N-	N-N+TN-			
6	azide	N N+N-	N N+N-	[1.0]	[1.0]	[0]

		HO OH	NH NH			
7	triazinol	NH NH O	NH NH O	[1.0]	[1.0]	[0]

ı	ı	`		
١	٠	-	-	•
,	-	۹	۰	
١		,		J

		-0>P20 -0 0-	HO OH			
8	phosphate	НО _Р 20 НО ОН	HO OH	[1.0]	[1.0]	[0]



Ľ	ı		
		7	

		∕ РН3+	PH2			
10	phosphine	PH2	PH2	[1.0]	[1.0]	[0]

		O C C	OH C				
11	cicletanine	N CI	OH N	[1.0]	[1.0]	[0]	

		OH O	∕ H OH			
12	$\operatorname{propaneAcid}$	~ ↓ 0- 0	ОН ОН	[1.0]	[1.0]	[0]

			√NH NH	ſŢ <mark>NH</mark>			
51	13	imidazole	N- </th <th>N-<!--</th--><th>[1.0]</th><th>[1.0]</th><th>[0]</th></th>	N- </th <th>[1.0]</th> <th>[1.0]</th> <th>[0]</th>	[1.0]	[1.0]	[0]

	ر	•
١	Ċ	c
۰		•

		N N	×			
14	enamine	NH	×	[1.0]	[1.0]	[0]

	O: s.	o. /s.			
15 sulfoxide	•- ✓S+ 、	0; >S_	[1.0]	[1.0]	[0]

$\frac{7}{2}$

		O. -SNH2 O	O. S. NH2 O			
16	$\mathbf{sulfonamide}$	O. SNH- O	0. SNH2 0	[1.0]	[1.0]	[0]

•	5	τ	
•	5	τ	
-			

		SH Z OH	S NH NH O			
17	tiouracil	S I N O	S NH NH O	[1.0]	[1.0]	[0]

	3	τ
5		5

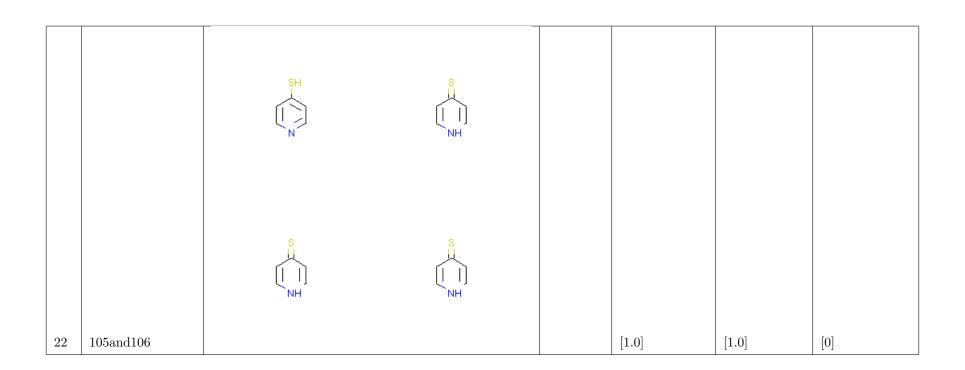
	N+ N N N N N N N N N N N N N N N N N N			
18	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	[1.0]	[1.0]	[0]

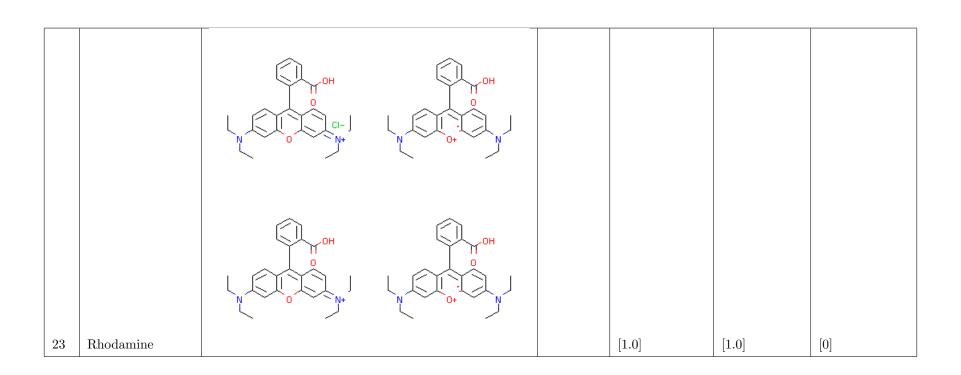
		-0 N+	-0 N+			
		HONNOH	HONNOH			
19	nitro	HO N+	HO N+	[0.67, 1.0, 0.67]	[0.56, 0.8, 0.6]	[58, 43, 56]

		NH NH	ON NOTE OF THE PROPERTY OF THE			
		OH N	ON N			
20	107and108and109	NH NH	O N N	[1.0, 1.0, 1.0]	[1.0, 1.0, 1.0]	[0, 0, 0]

	_			
ĺ	Ī		3	٦
٦	۰	-	1	
		C		7

			NH2	NH NH			
59	21	aminopyrimidine	NH L	H H H H H H H H H H H H H H H H H H H	[1.0]	[1.0]	[0]





$\overline{}$	5
	2
	_

		0) \$ 10 H0	HOSSO			
24	$_{ m sulfon}$	0)s ² 0	0.5.°О НО	[1.0]	[1.0]	[0]

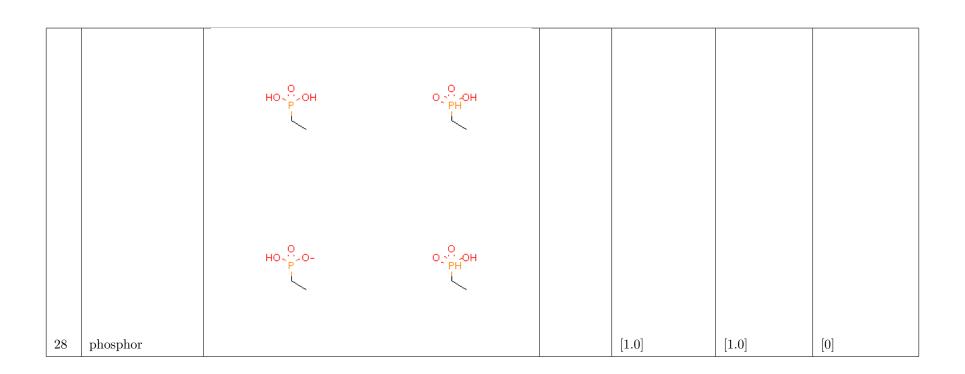
		NH NH2	NH _ N			
25	triazine	N NH2 N NH	NH NH2	[1.0]	[1.0]	[0]

	1	

		N N N N N N N N N N N N N N N N N N N	N N N N N N N N N N N N N N N N N N N			
26	tetrazole	N NH NH	N S NH	[1.0]	[1.0]	[1]

\sim	•

		OH ↓	OH			
65	27 phenol		OH 	[1.0]	[1.0]	[0]



`	

			SH SH	SH SH			
67	29	112and113	S ►s	SH SH	[1.0]	[1.0]	[0]

	1	3	٦
1	Ē	v	ō

		O N+	о- N+			
30	NO	O- N+	О- N+	[1.0]	[1.0]	[0]

)11 N+ N-) T N+ N-			
31	diazo	,) N+ N-	[1.0]	[1.0]	[0]

		N OH	N OH			
32	benzaldoxime	N ⁻ O	N OH	[1.0]	[1.0]	[0]

