ALGORITHMS FOR AUTOMATIC TAUTOMER GENERATION AND THEIR APPLICATIONS

Nikolay T. Kochev[1], Vesselina H. Paskaleva[1], Nina Jeliazkova[2]

[1]University of Plovdiv, Department of Analytical Chemistry and Computer Chemistry;

[2]Ideaconsult Ltd, 4 A. Kanchev str., Sofia 1000, Bulgaria

Ambit-Tautomer Basic Features

Software characteristics

- CDK.sf.net based structure representation, input, output and info processing
- Supports standard chemical formats: SMILES, InChI, MOL/SDF file, CML
- Exhaustive tautomer generation
- Customizable set of rules and post-generation filters
- Set of predefined rules
- Tautomer ranking based on simple empirical rules

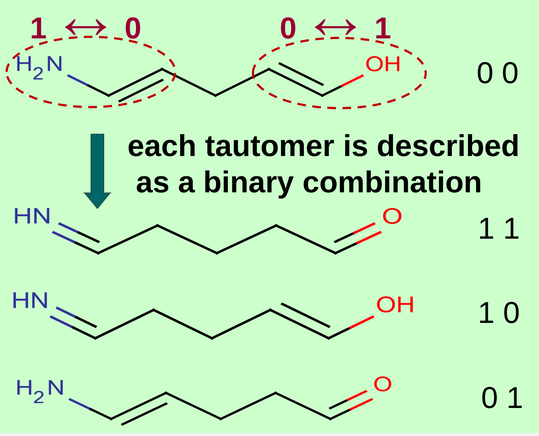
Customizable set of rules

- Basic set of 1-3 and 1-5 proton shift rules
- Additional rules: 1-7 proton shifts, chlorine atom shifts
- Rule description based on SMARTS

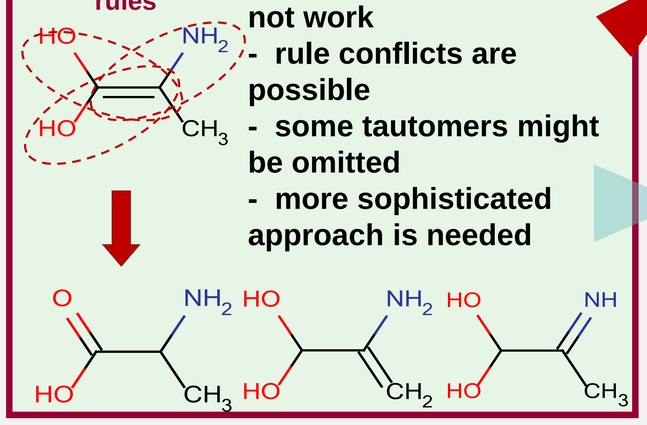
Tautomer generation algorithms

- Pure combinatorial algorithm
- Incremental approach (based on depth first search algorithm) for rule combination with local rule corrections and refinement on the way

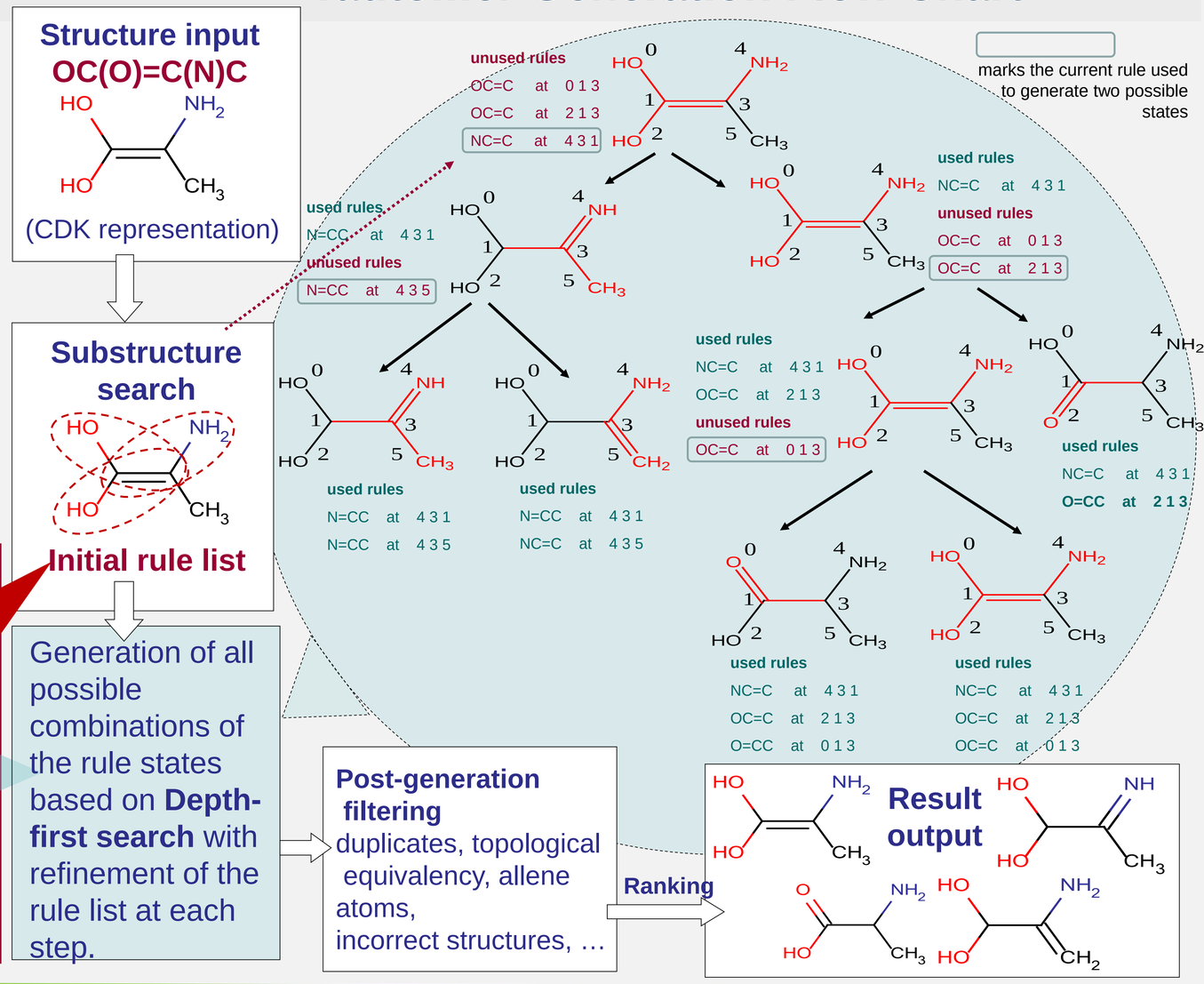
Combinations of non-overlapping rules



Overlapping rules



Tautomer Generation Flow Chart



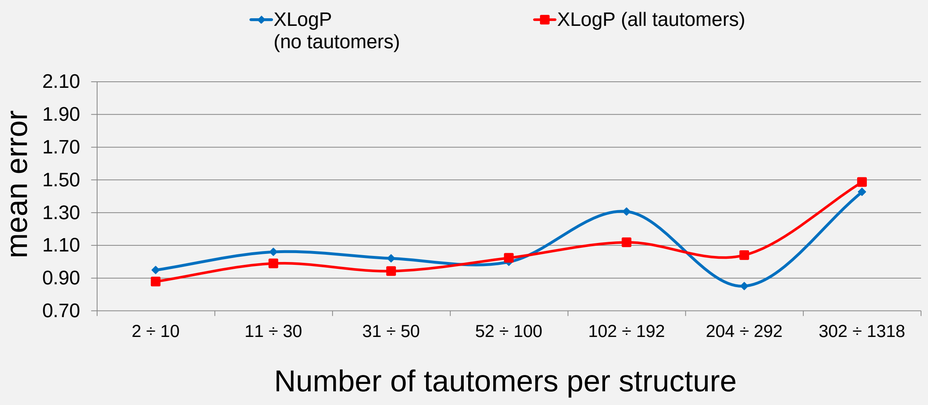


Figure 2. The mean absolute errors for XLogP model compared with the errors obtained from the averaged model values calculated for all tautomers for each testing structure. The statistics is calculated for 8327 test structures.

Table 2. The values of Ames-mutagenecity model and XLogP model for all tautomers of viuoluric acid.

Violuric acid tautomers /SMILES notations/	Ames Mutagenicity (model)	XLogP
O=C1NC(=O)C(=NO)C(=O)N1	1	0.135
O=C1N=C(O)N=C(O)C1(=NO)	0	-0.086
O=C1N=C(O)C(=NO)C(O)=N1	0	0.267
O=C1N=C(O)C(=NO)C(=O)N1	1	0.041
O=C1N=C(O)NC(=O)C1(=NO)	1	0.361
O=NC1=C(O)N=C(O)N=C1(O)	1	-0.102
O=NC=1C(=O)NC(O)=NC=1(O)	1	-0.084
O=NC=1C(=O)N=C(O)NC=1(O)	1	1.230
O=NC=1C(O)=NC(=O)NC=1(O)	1	0.698
O=NC=1C(=O)NC(=O)NC=1(O)	1	0.363
O=NC1C(O)=NC(=O)N=C1(O)	1	-0.277
O=NC1C(=O)N=C(O)N=C1(O)	1	-1.056
O=NC1C(=O)NC(=O)N=C1(O)	1	-0.932
O=NC1C(=O)N=C(O)NC1(=O)	1	-1.038
O=NC1C(=O)NC(=O)NC1(=O)	1	-1.267

The structural information was processed according to the presented flow chart. We studied the influence of tautomers information on various processing stages: descriptor calculation (table 3), similarity searching (see table 1) and QSAR/QSPR modeling of Ames-Mutagenicity and LogP (see fig.2 and table 2).

QSAR/QSPR Cheminfo Processing Flow Chart

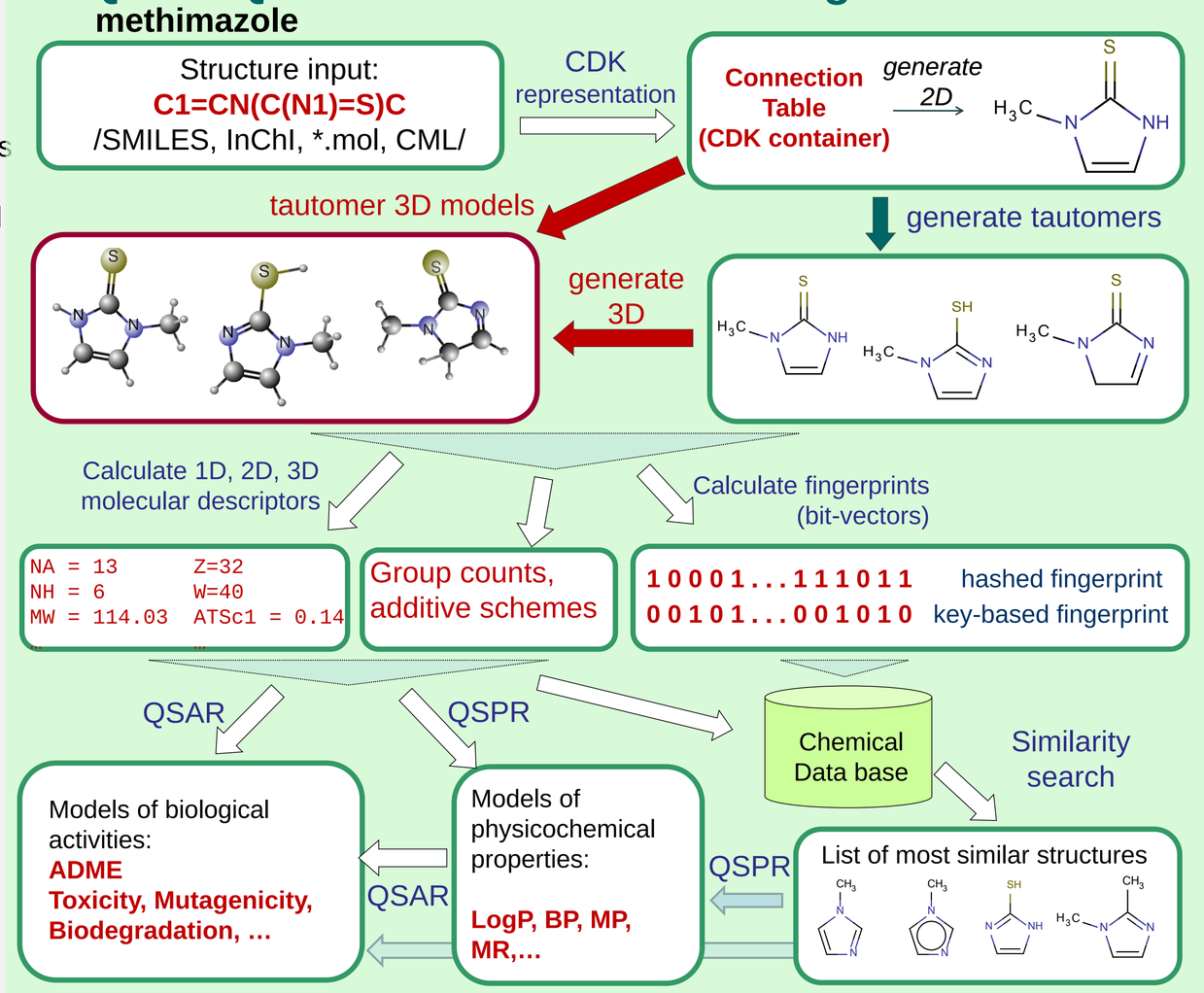


Table 3. The number of descriptors (out of total 863) which exhibit relative standard deviation (RSD due to the tautomerism) larger than particular thresholds: 0.1, 0.3, 0.5, 1.0

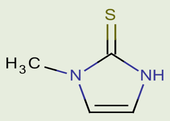
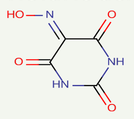
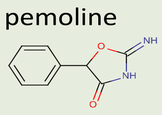
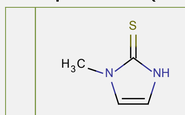
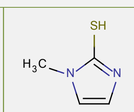
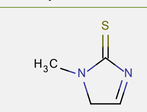
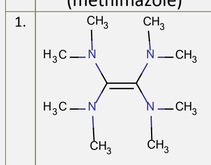
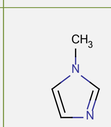
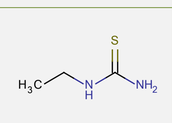
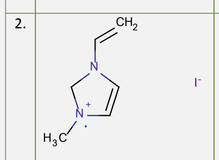
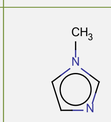
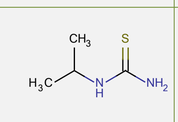
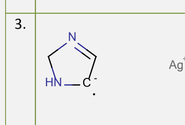
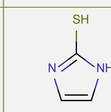
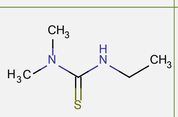
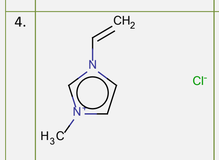
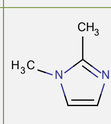
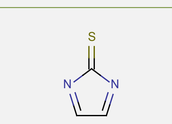
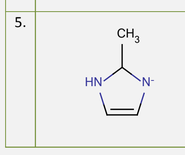
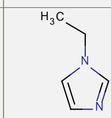
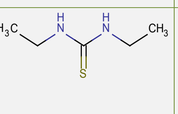
Structure	RSD threshold	Number of PaDEL descriptors that have RSD > $RSD_{threshold}$
methimazole 	0.1	180
	0.3	124
	0.5	99
	1.0	71
violuric acid 	0.1	217
	0.3	151
	0.5	108
	1.0	80
pemoline 	0.1	239
	0.3	168
	0.5	138
	1.0	113

Table 1. The similarity search results for the three tautomers of methimazole. Each column contains the five most similar structures to the tautomer. Similarity search is performed in a data base with 553477 compounds (subset of PubChem data base).

	 (methimazole)	Similarity		Similarity		Similarity
1.		0.62		0.71		0.47
2.		0.6		0.71		0.45
3.		0.59		0.64		0.44
4.		0.58		0.57		0.44
5.		0.54		0.57		0.43

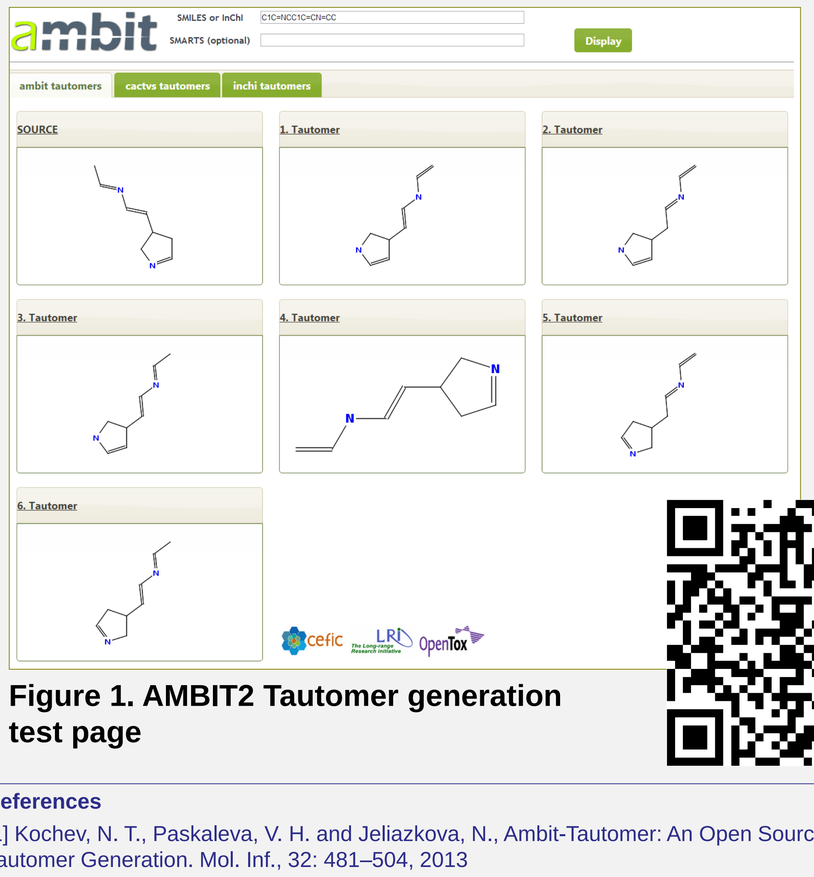


Figure 1. AMBIT2 Tautomer generation test page

Ambit-Tautomer [1] is part of the Ambit2 software package [2], distributed under LGPL license and using the Chemistry Development Kit (CDK) library [3] for basic chemoinformatics functionality. Ambit-Tautomer utilizes a depth-first search algorithm, combined with a set of rules for tautomeric transformations.The Ambit implementation of OpenTox Web [4] services for predictive toxicology, are being extended to include the tautomer generation algorithm. A web page, providing online tautomer generation by several different algorithms, including Ambit-Tautomer, is available at: **http://apps.ideaconsult.net:8080/ambit2/depict/tautomer**.

References

[1] Kochev, N. T., Paskaleva, V. H. and Jeliazkova, N., Ambit-Tautomer: An Open Source Tool for Tautomer Generation. Mol. Inf., 32: 481–504, 2013

[2] AMBIT project, http://ambit.sourceforge.net

[3] Steinbeck C., Hoppe C., Kuhn S., Guha R., Willighagen E.L., “Recent Developments of the Chemistry Development Kit (CDK) – An Open-Source Java Library for Chemo- and Bioinformatics”. Curr. Pharm. Des. 2006; 12(17):2111-2120 (DOI: 10.2174/138161206777585274)

[4] Jeliazkova N., Jeliazkov V., AMBIT RESTful web services: an implementation of the Open Tox application programming interface, Journal of Chemoinformatics 2011, 3:18, doi: 10.1186/1758-2946-3-18.;