

Methods for tautomer enumeration, -searching and -duplicate filtering

**J Szegezdi, Z Mohácsi, T Csizmazia, S Dóránt,
Á Papp, G Pirok, S Csepregi, F Csizmadia**

May 2010, European UGM



ChemAxon
Solutions for Cheminformatics

Contents

- Tautomerization plugin
- Methods for generating tautomers
- How tautomers are used in property predictions
- Tautomer duplicate search
- Generic tautomer generation
- Handling of stereochemistry
- Tautomer substructure search
- Custom tautomerization: Standardizer

What is tautomerization?

Try Beta [Log in](#) / [create account](#)

[article](#) [discussion](#) [edit this page](#) [history](#)

Tautomer

From Wikipedia, the free encyclopedia

Tautomers are [isomers](#) of [organic compounds](#) that readily interconvert by a [chemical reaction](#) called **tautomerization**. It is common that this reaction results in the formal migration of a [hydrogen atom](#) or [proton](#), accompanied by a switch of a [single bond](#) and adjacent [double bond](#). The concept of tautomerizations is called **tautomerism**. Tautomerism is a special case of [structural isomerism](#) and can play an important role in non-canonical [base pairing](#) in [DNA](#) and especially [RNA](#) molecules.

Enol form **Keto form**

Lactam form **Lactim form**

Amide form **Imidic acid form**

Amine form **Imine form**

Tautomerization Plugin

The screenshot shows the MarvinSketch 5.3.0 interface with the Tautomerization Plugin. The main window displays a pyridine ring with an oxygen atom. The 'Tools' menu is open, showing 'Isomers' > 'Tautomers'. A 'Tautomers Options' dialog box is open, showing 'All tautomers' selected. A 'Tautomers' window displays a table of six tautomers.

Table		Tools		Help	
2		3			
5		6			
Select					

Tautomers Options

General Options | Advanced Options

Calculation

- All tautomers
- Canonical tautomer
- Generic tautomer
- Major tautomer
- Dominant tautomer distribution

Max. number of structures: 1000

Consider pH effect
at pH: 2.5

Single fragment mode

OK Cancel Restore Defaults

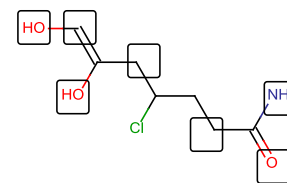
Tautomerization Plugin

Can generate:

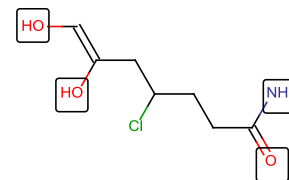
- **Dominant tautomer distribution with estimated ratio:** those likely to exist (e.g. at a given pH).
- **Major tautomer:** the most dominant one
- **All tautomers:** all theoretically possible
- **Generic tautomer:** used for duplicate tautomer searching
- **Canonical tautomer:** canonicalization based on empirical rules

Algorithm

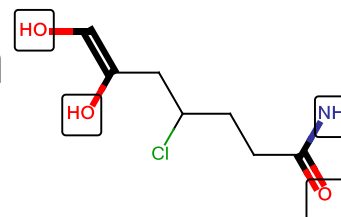
1. Identify H donors and acceptors.



2. Filter them depending on parameters and method.

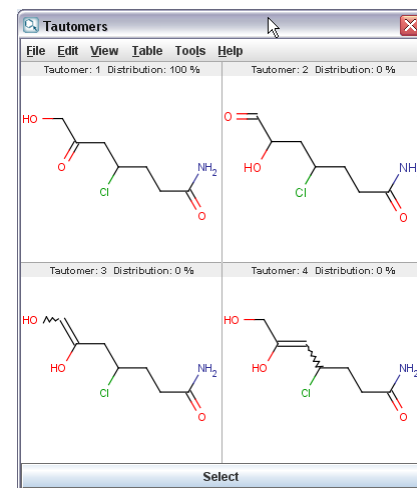


3. Consider bond paths between donors/acceptors.



4. Process results of 2 & 3:

- Combinatorially enumerate,
- Rank results (dominant & canonical),
- Calculate distribution (dominant),
- Filter & select (canonical)
- Combine paths into regions (generic)



Enhancing property predictions

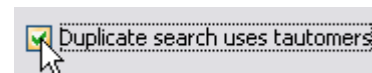
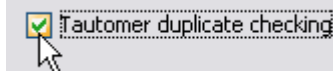
ChemAxon property predictions have option to consider tautomers:

- Log D: whole dominant tautomer distribution is used, weighted by the ratio of isomers.
- pKa, major microspecies, logP: the single major tautomer is used for the calculation.

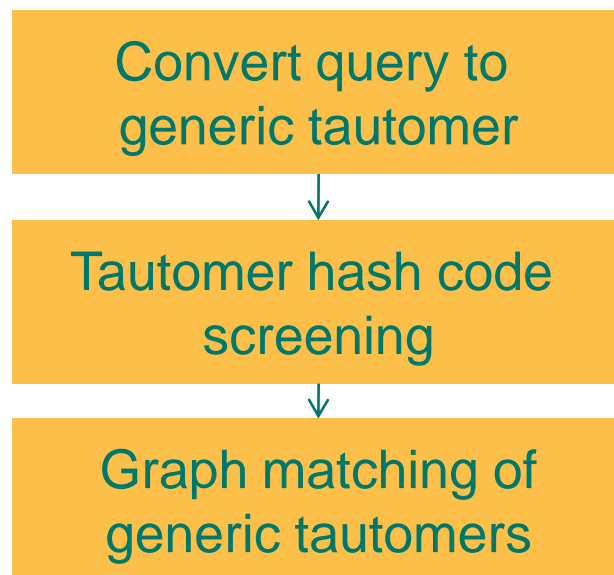
Tautomer searching methods in JChem

Table/index option

- „Duplicate search uses tautomers” – check box at table / index creation.

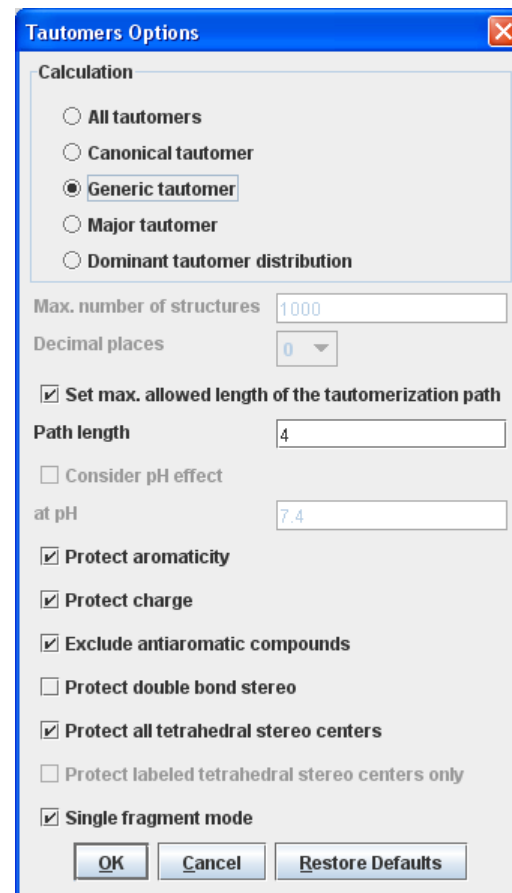
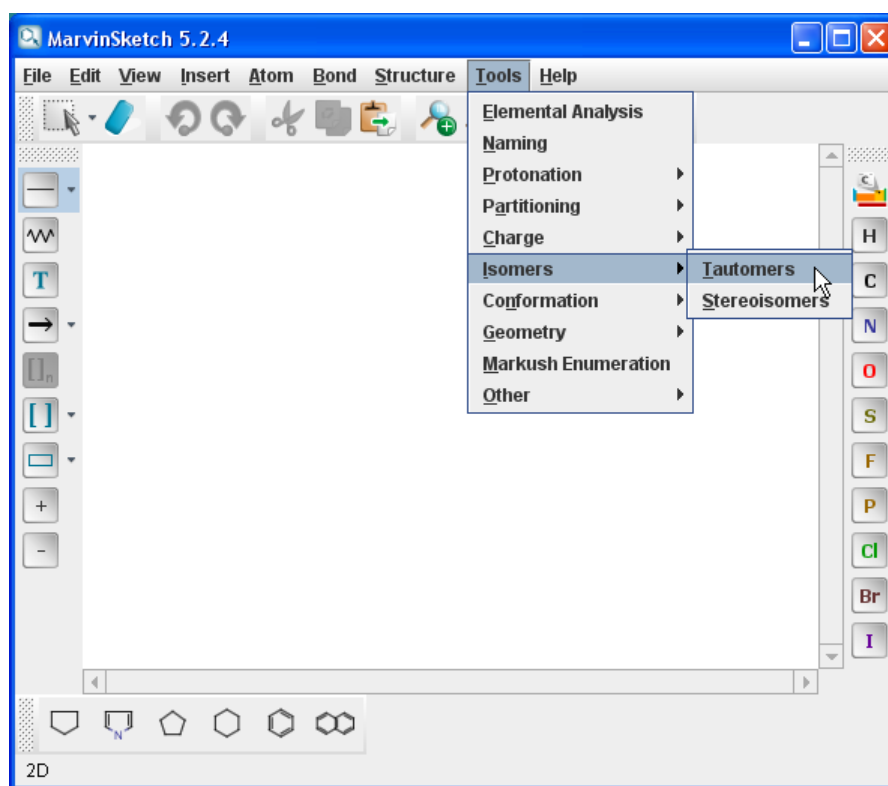


- It makes all duplicate searches „tautomer search” on that table
- Algorithm:



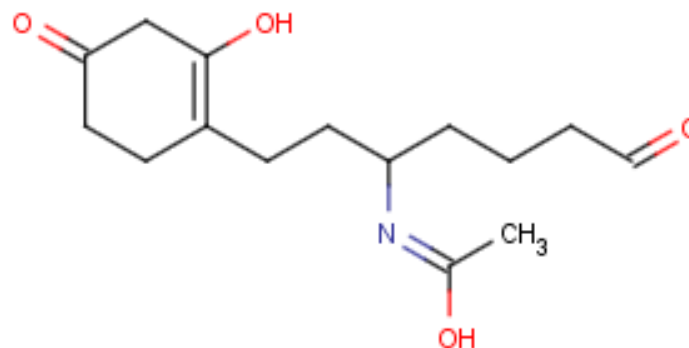
Generic tautomer

Generic tautomer calculation is accessible in Marvin for understanding/checking behavior:

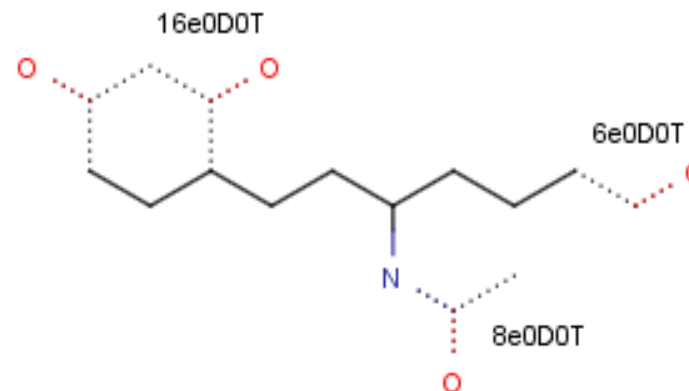


Generic tautomer

Example:



Generic
tautomer



Generic tautomer – algorithm

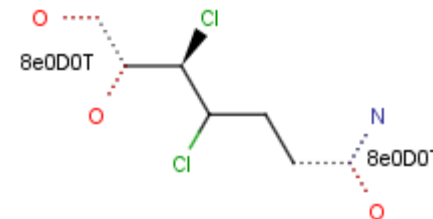
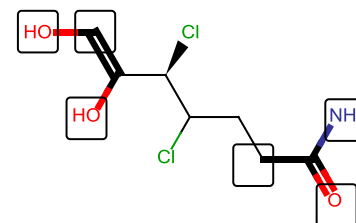
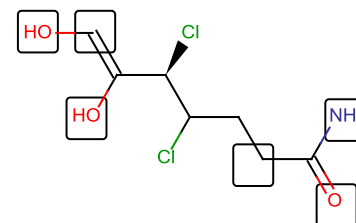
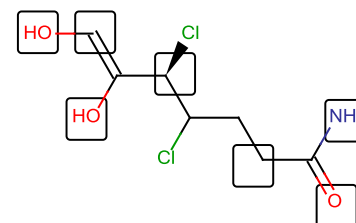
H donor and acceptor atoms are identified

Excluded parts are located.
(e.g. Protected stereo, aromaticity, etc.)

Tautomer regions are identified based on alternating single/double bonds, etc. between H donor and acceptor atoms.

Assignment of tautomer regions

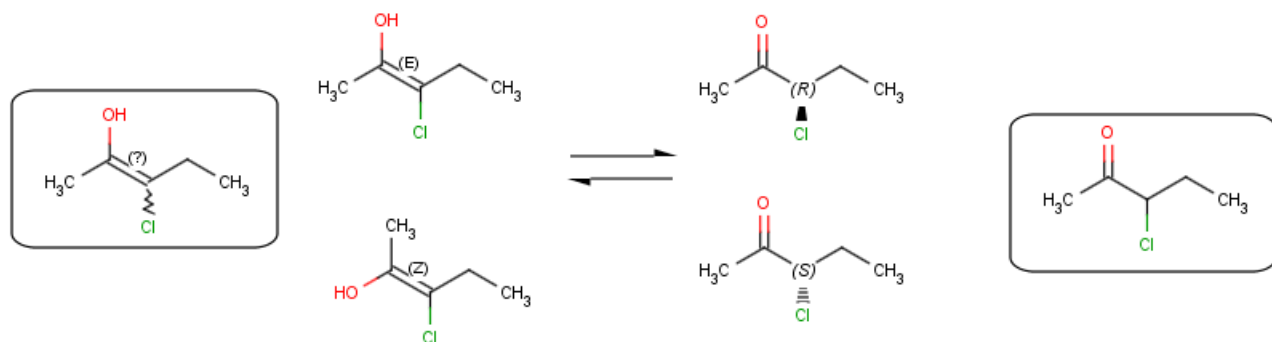
1. Variable bonds are replaced by any bonds.
2. Bonding electron count, number of D and T in the region is attached to the region as data. (Needed for graph matching.)



Stereochemistry

Theory:

- When spontaneous tautomerization happens in nature, the specific stereo configuration is lost easily in two steps of back and forth tautomerization. (E.g. racemization.)
- For this reason, in general we ignore stereochemistry in tautomer regions, but there are options to protect different types of stereochemistry.



Stereochemistry

Current practices in JChem (considering structure registration):

- Tetrahedral stereo centers (which are marked by wedge bonds) are excluded from tautomerization (= they are never included in tautomer regions).

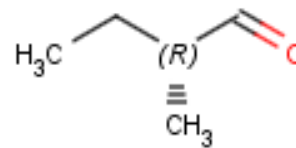
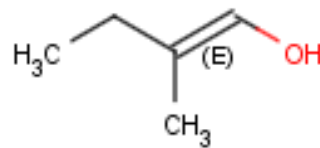
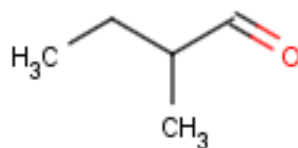
Explanation: wedge stereo drawing is usually not accidental. This way the chemists express that a separate stereoisomer can be isolated.

- Double bond(E/Z) stereochemistry is allowed to tautomerize.

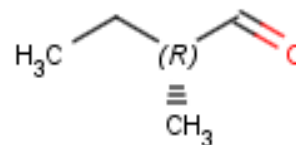
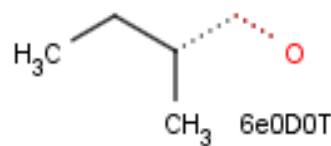
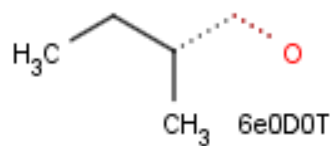
Explanation: It is very easy to „accidentally” define E/Z stereochemistry. – Just by the atom coordinates. This way it is not easy to decide what the original intention was.

Stereochemistry

Examples:

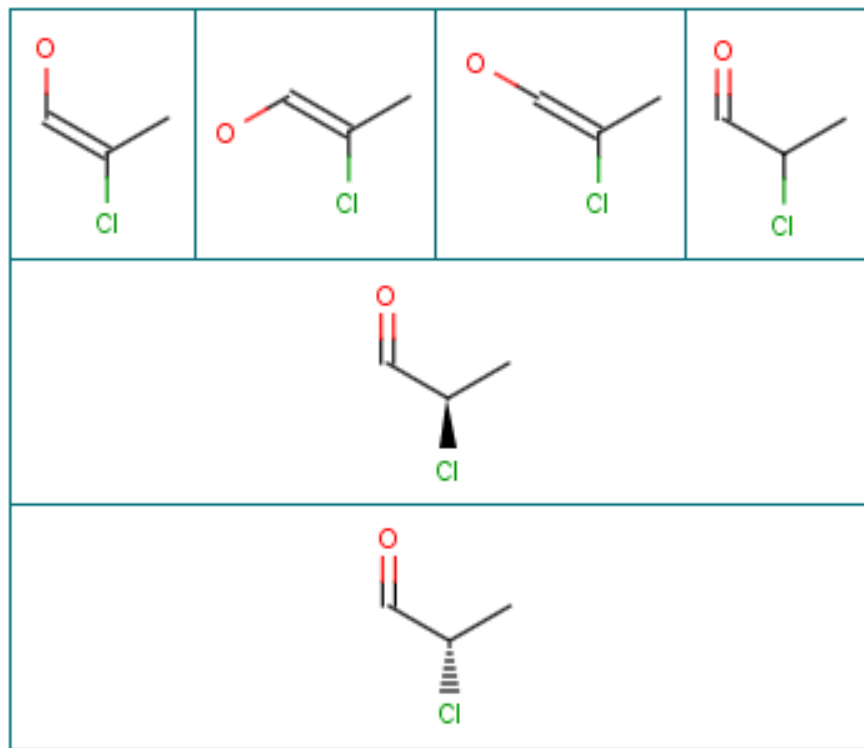


Generic
tautomer



Stereochemistry

Example 2: As a result, each row below will be recognized as different compounds. Molecules in the same row are recognized as tautomers of each other:



(“protect double bond stereo” option is off)

(“protect all tetrahedral stereo centers” option is on)

(“protect all tetrahedral stereo centers” option is on)

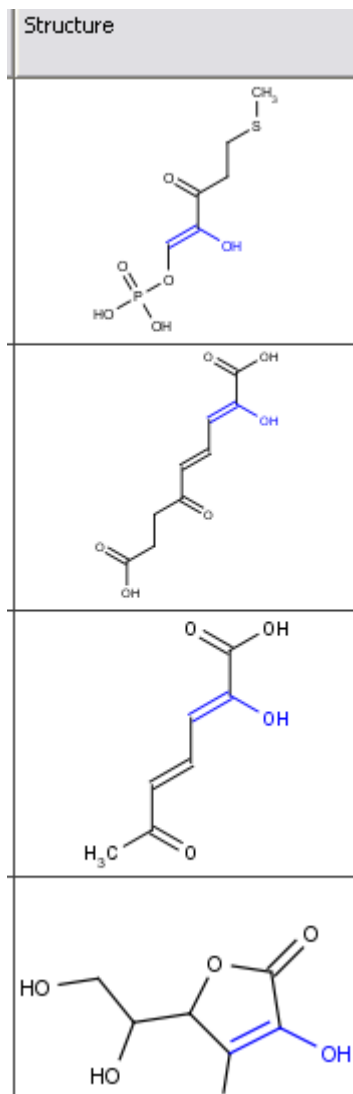
Tautomer search option

Query molecule is enumerated

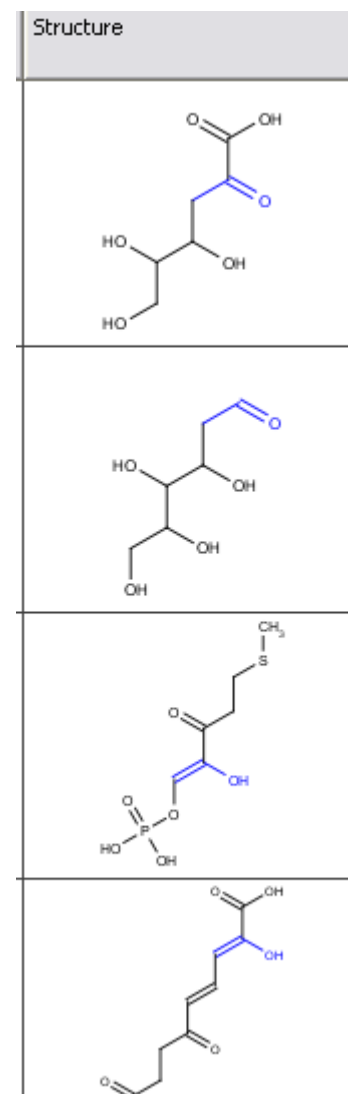
Query:



non-tautomer search:

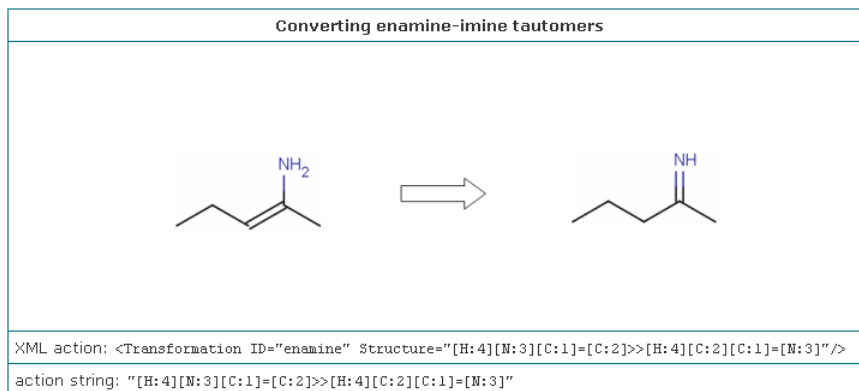
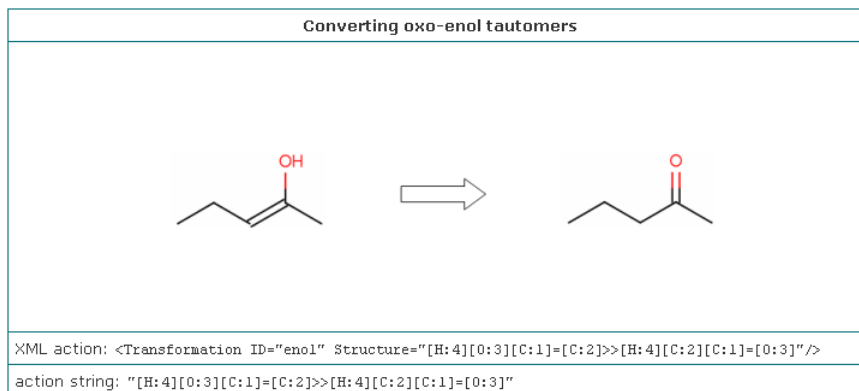


tautomer search:



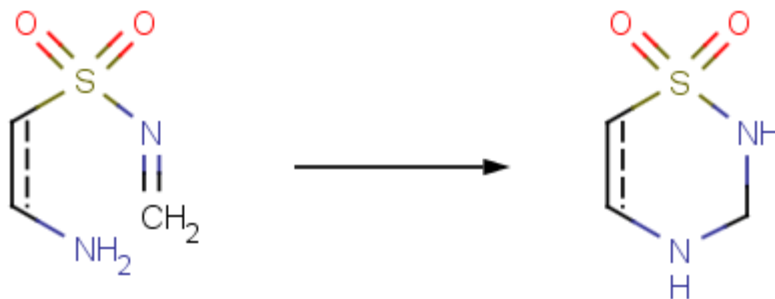
Standardization methods

- Canonical tautomer  Tautomerize
- Custom transforms – examples:



Standardization method

- Custom transform example for ring-chain tautomerism:



Mollica et al, *J.Pharmaceutical Sciences*, Vol 60, No 9, p1380,
1971

Comparison of the methods

Method	Available search types				Registration speed	Tautomer search speed
	tautomer		non-tautomer			
	Duplicate	Substructure	Duplicate	Substructure		
Duplicate tautomer table option	OK	-	-	OK	Small overhead	Fast
Tautomer search option	-	OK	OK	OK	As normal	May be slow
Canonical tautomer in standardization	OK	-	-	-	Small overhead	Fast
Custom standardization transforms	OK	OK	-	-	Small overhead	Fast

Summary

- Tautomer calculation plugin offers various methods
- Tautomers can improve property predictions
- JChem offers 4 ways for handling tautomers

Acknowledgements

- József Szegezdi – development of the tautomerization algorithms
- Zsolt Mohácsi – plugin framework, standardizer
- Szilárd Doránt, Tamás Csizmazia, Ali Baharev – tautomer search
- Ákos Papp, György Pirok, Ferenc Csizmadia, Szabolcs Csepregi – concepts

References

- [1] Martin, Y.C.: Let's not forget tautomers J Comput Aided Mol Des (**2009**) 23:693–704, DOI 10.1007/s10822-009-9303-2
- [2] Szegezdi, J.; Csizmadia, F: Tautomer generation. pKa based dominance conditions for generating dominant tautomers. American Chemical Society meeting, Aug 19-23rd, **2007**
http://www.chemaxon.com/conf/Tautomer_generation_A4.pdf
- [3] Pirok, G. et al: Standardizer - Molecular Cosmetics for Chemoinformatics. Drug Discovery Technology, August 7-10th, **2006**
<http://www.chemaxon.com/conf/standardizer.pdf>