

Methods for tautomer enumeration, -searching and -duplicate filtering

**József Szegezdi, Zsolt Mohácsi, Tamás Csizmazia,
Szilárd Dóránt, Ákos Papp, György Pirok,
Szabolcs Csepregi, Ferenc Csizmadia**

September 2010, Cxn US UGM, Boston



ChemAxon
Solutions for Cheminformatics

- Tautomerization plugin
 - Methods
 - Use 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 Tautomers plugin active. The main window displays a chemical structure of a pyridine derivative. The 'Tools' menu is open, highlighting 'Tautomers'. A 'Tautomers Options' dialog box is open, showing settings for calculation (All tautomers, Max. number of structures: 1000, Consider pH effect: 2.5, Single fragment mode checked). A 'Tautomers' window displays a table of six tautomers.

	2	3
	Select	

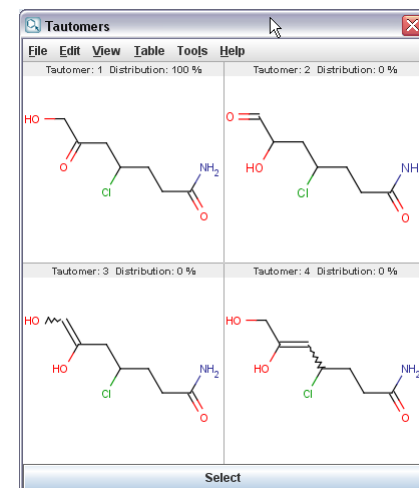
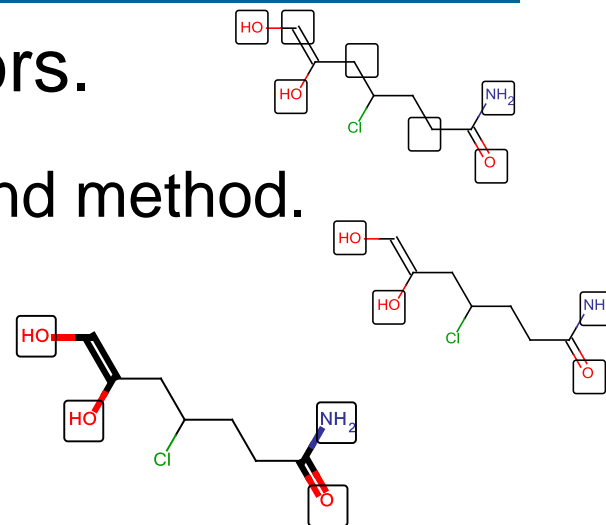
Tautomerization Plugin

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 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)



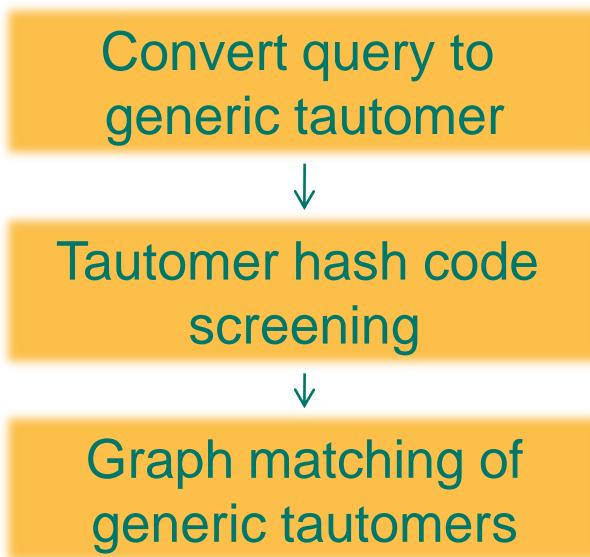
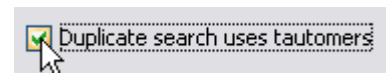
Enhancing property predictions

- 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.

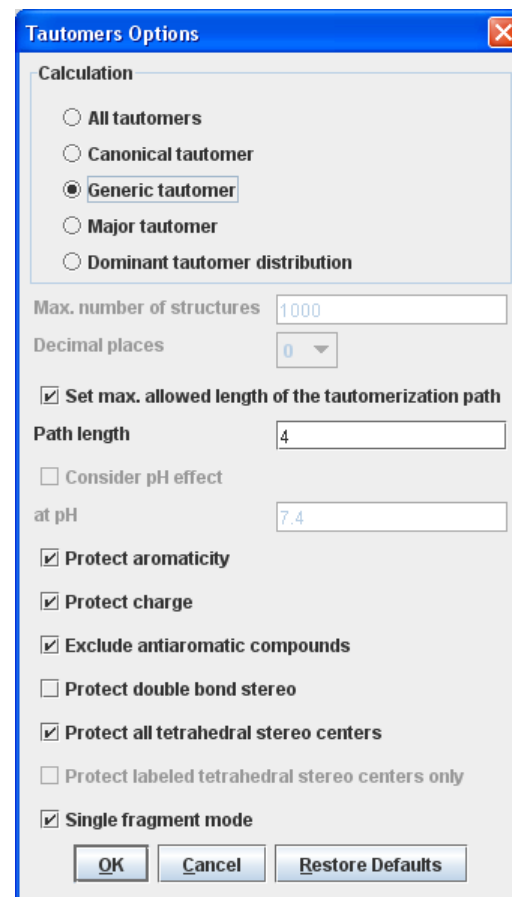
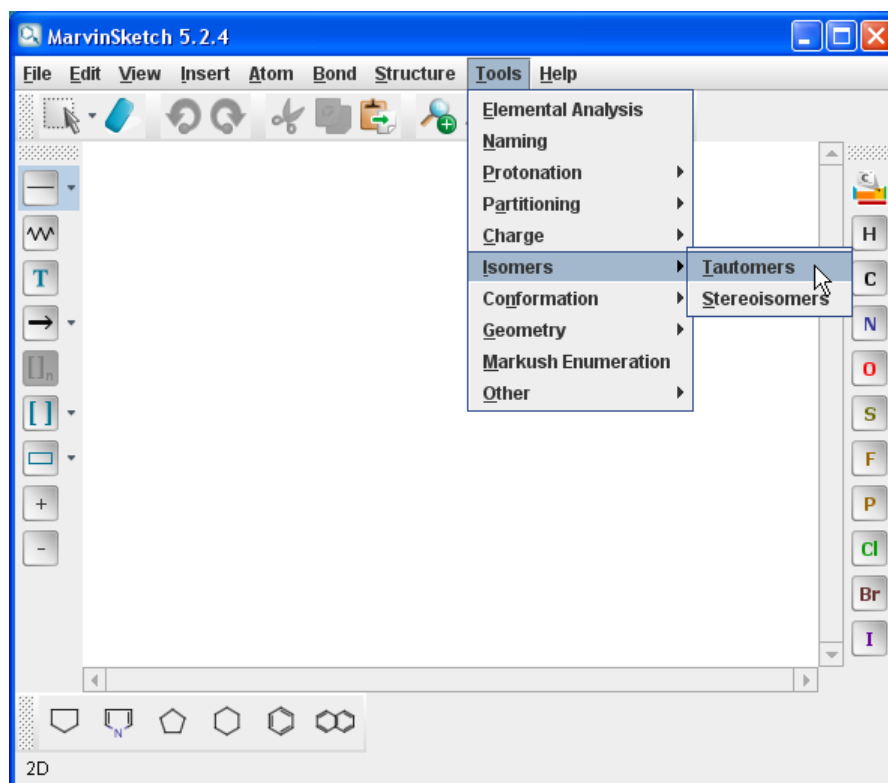
Tautomer searching methods in JChem

Table/index option

- „Duplicate search uses tautomers” –
- It makes all duplicate searches „tautomer search”
- Algorithm:

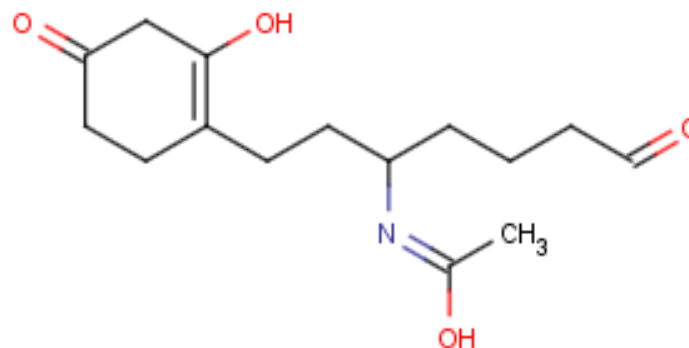


Generic tautomer

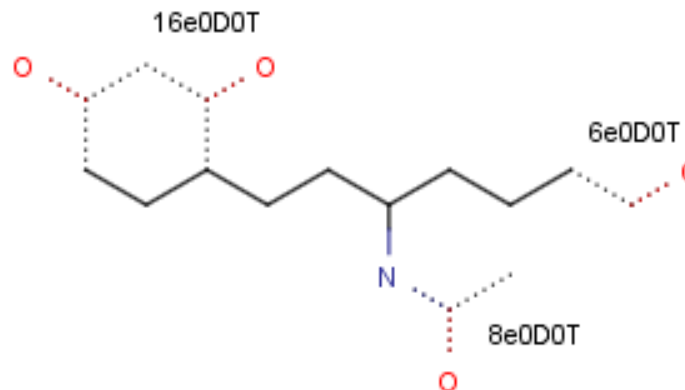


Generic tautomer

Example:



Generic
tautomer



Generic tautomer – algorithm

H donor and acceptor atoms are identified



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

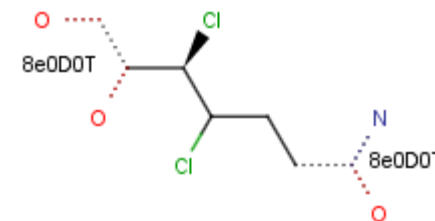
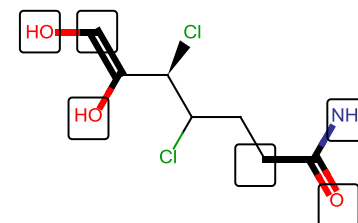
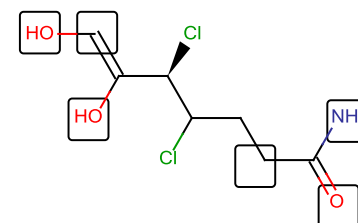
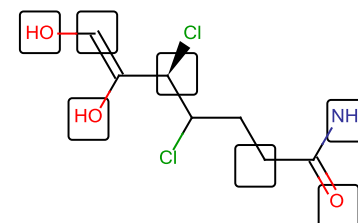


Tautomer regions are identified based on alternating single/double bonds, etc. between H donor and acceptor atoms.



Assignment of tautomer regions

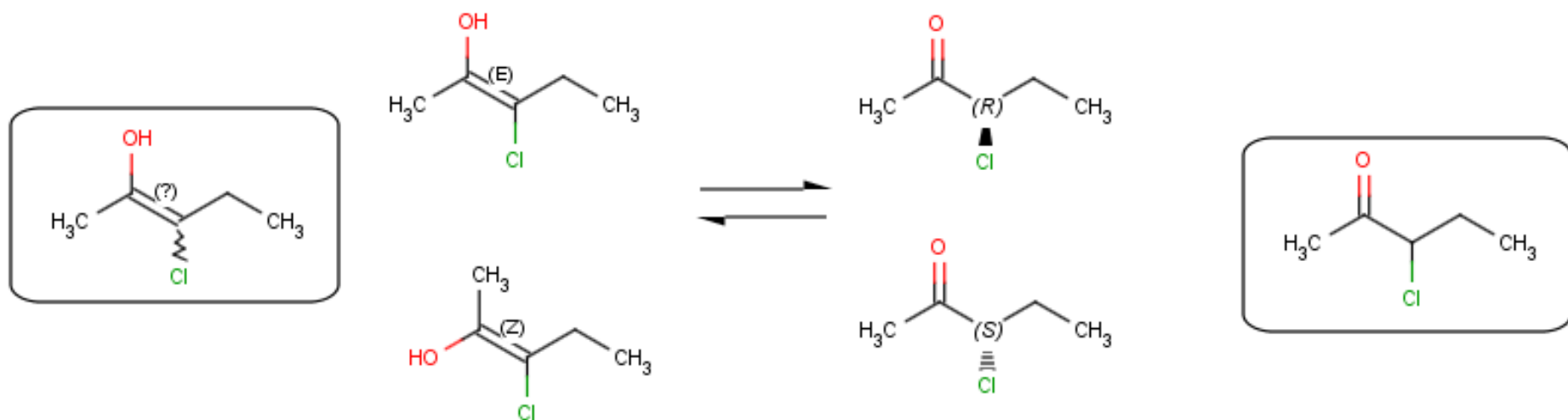
1. Variable bonds are replaced by any bonds.
2. Bonding electron count, number of D and T in the region is attached to the region as data. (Needed for graph matching.)



Stereochemistry

Theory:

- Spontaneous tautomerization: racemization (lose stereo)
-> Ignore stereo in tautomer region
- Option to protect



Stereochemistry

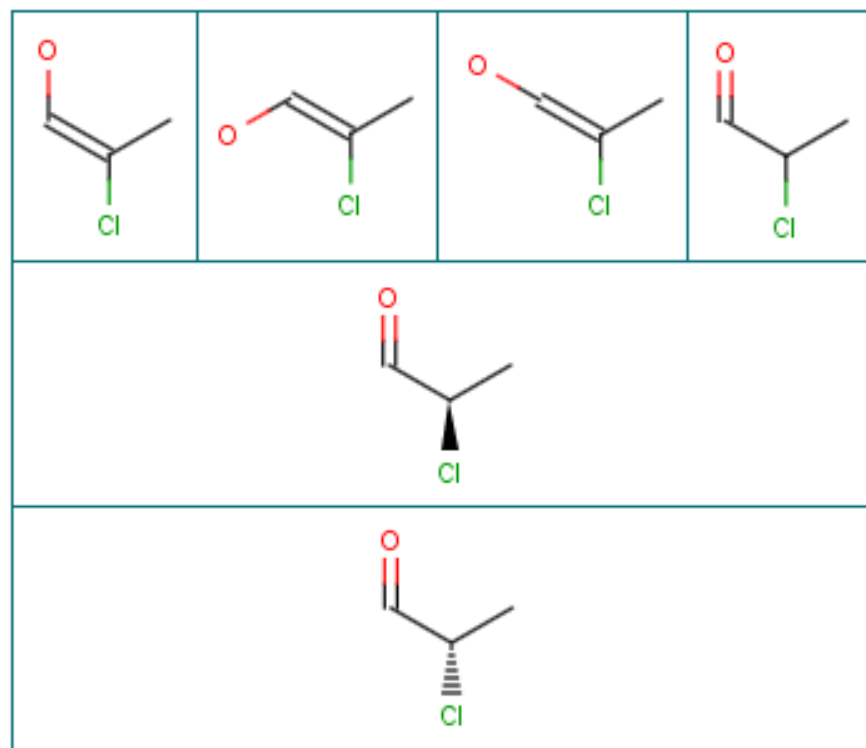
Current practices (considering structure registration):

- Tetrahedral stereo centers are protected

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

Stereochemistry

Example: Tautomer duplicate categories with stereochemistry



("protect double bond stereo"
option is off)

("protect all tetrahedral stereo
centers" option is on)

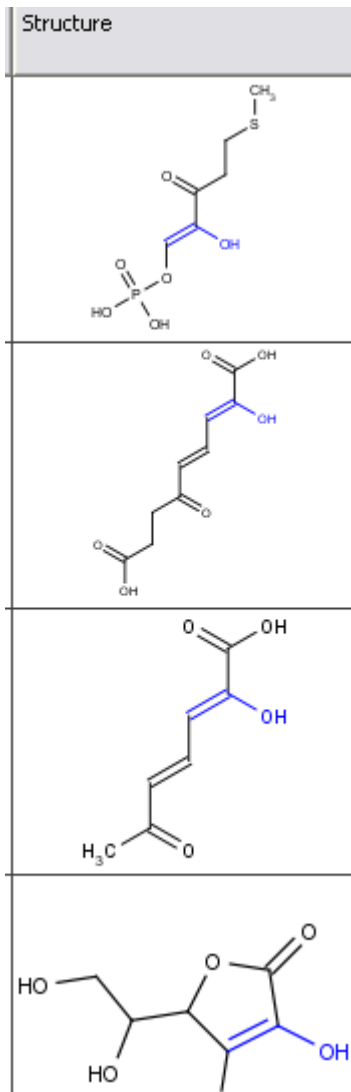
Tautomer search option

Query molecule is
enumerated

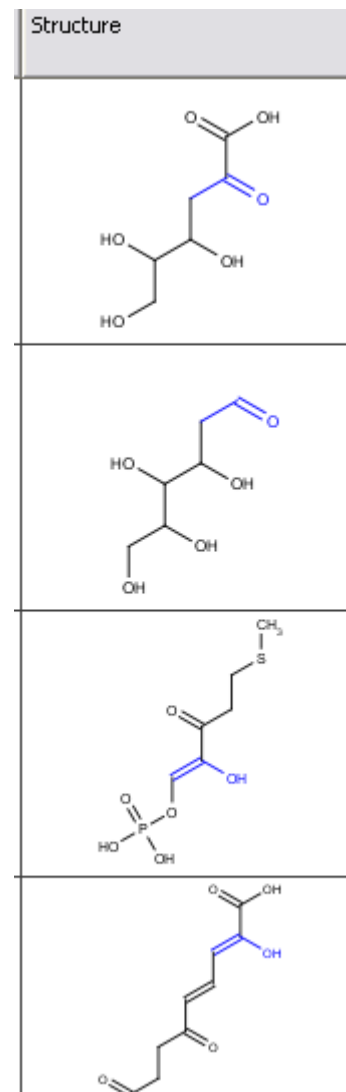
Query:



non-tautomer
search:

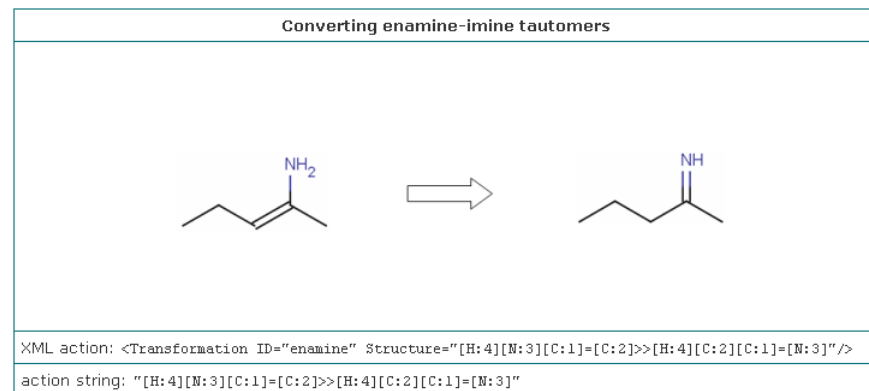
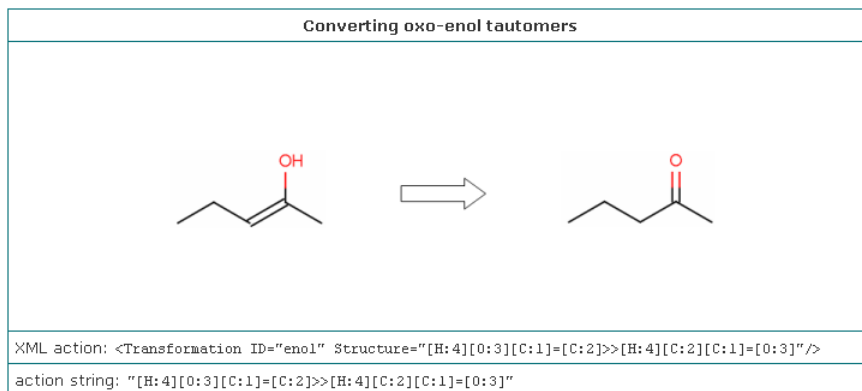


tautomer
search:

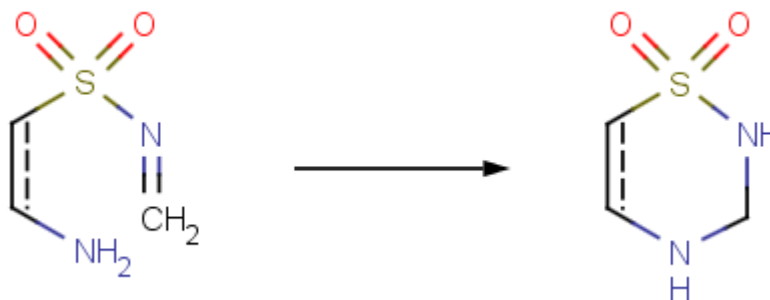


Standardization methods

- Canonical tautomer  Tautomerize
- Custom transforms – examples:



- Flexible, e.g. ring-chain tautomers



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	- *	- (under dev.)	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

* Available via tautomer search option (covered in row 2).

Under development

- Further filters (e.g. customizable stable groups)
- Built-in ring-chain tautomerization in tautomer plugin
- Switchable tautomer search for tautomer table option in JChem

Summary

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