

# Restoration of Molecular Artifacts

## Structure checker and fixer framework for registration systems

György Pirok



**ChemAxon**  
Solutions for Cheminformatics

# The Team

---



Attila Szabó



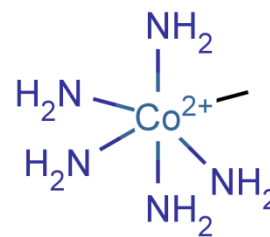
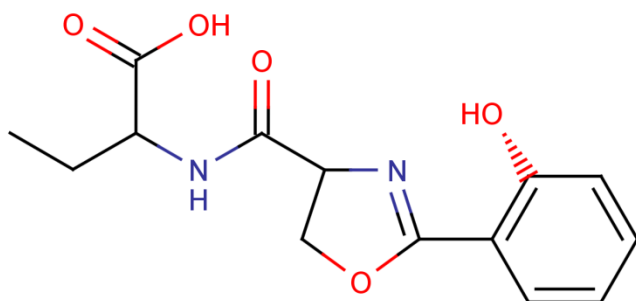
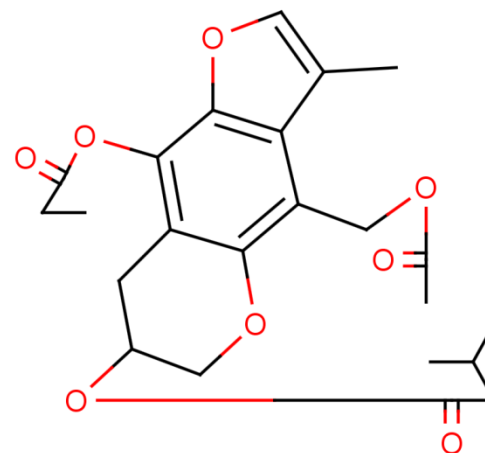
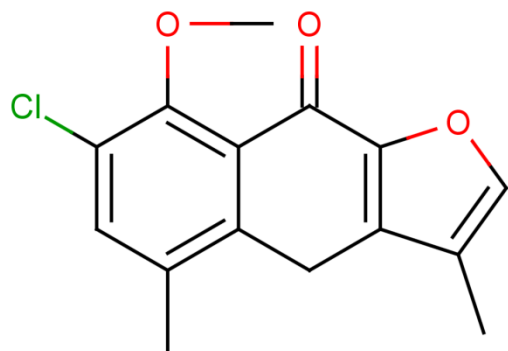
István Rábel



Zsolt Mohácsi

# The Problem

Molecules registered in databases often contains drawing errors or unpreferred representations. Some other issues are originated from custom workarounds due to the limitations of older chemistry applications.



$\text{No}^{3-}$

$\text{No}^{3-}$

# The Solution – Structure Checker

Many common errors can be identified by a new software tool called Structure Checker. Some of them can be fixed automatically, others can be displayed for manual correction. Structure Checker first appeared in the form of API and as a MarvinSketch feature to highlight specific structural problems.

The screenshot displays the MarvinSketch 5.3.2 interface. The main window shows a chemical structure of a benzene ring attached to a piperazine ring, which is further connected to a chloroethyl group. The Structure Checker tool is open on the left, listing two errors: "1 Invalid wedge problem found" and "2 bonds found with wrong angle". The Preferences dialog is open in the foreground, showing the "Checkers" tab with a list of error checkers: Alias Checker, Bond Length Checker, Chiral Flag Error Checker, and Coordination System Error Checker. A tooltip "Add new Checker..." is visible over the "+" button in the Preferences dialog. The status bar at the bottom of the main window indicates "2D \* Problems: 2".

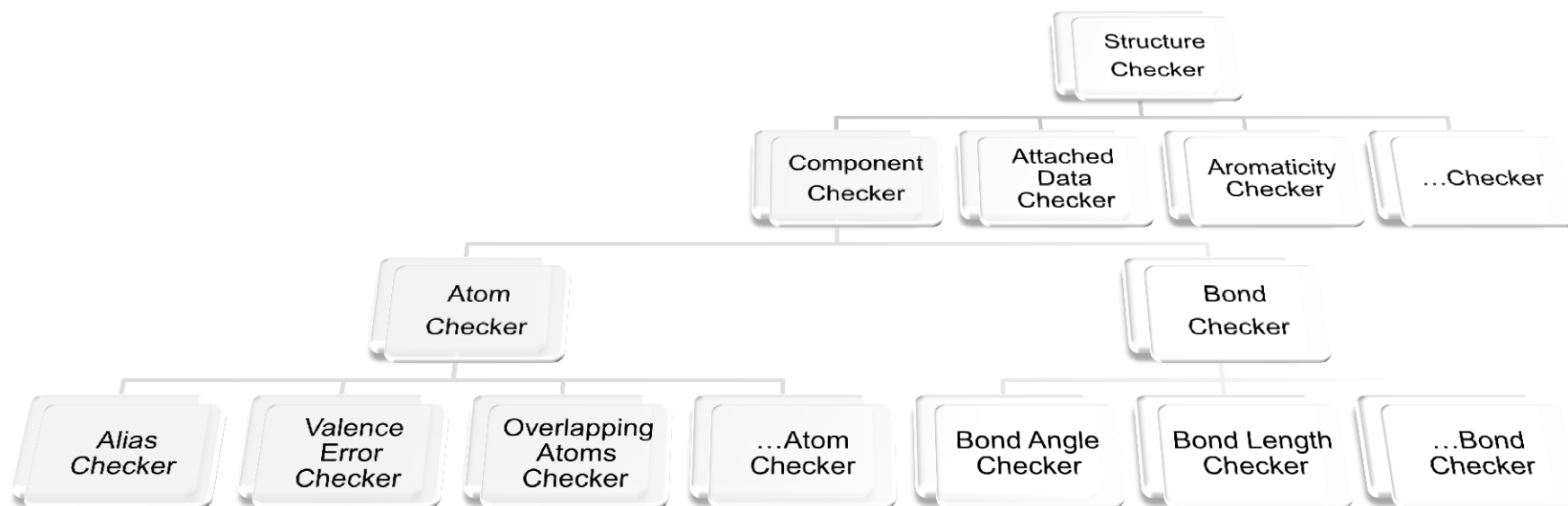
# About Checkers and Fixers

---

- **Checkers** are classes for the identification of specific issues in chemical structures (aromaticity, valence errors, etc). The `check()` method creates a result which contains all information of the problem.
- **Fixers** are classes responsible for correcting specific problems by using the result provided by checkers. The `StructureCheckerResult` object .

For the sake of flexibility, check and fix logics are well separated, checkers do not know about fixers, fixers do not know about checkers.

# Simplified Structure Checker Class Model



# Current Structure Checkers

---

## Error Checkers

- Aromaticity
- Chiral flag
- Coordination system
- Ring strain
- Reaction map
- Valence
- Wedge

## Issue Checkers

- Abbreviated group
- Alias
- Atom map
- Atom value
- Attached data
- Bond angle
- Bond length
- Covalent counterion
- Crossed double bond
- Explicit hydrogen
- Isotope
- Metallocene
- Missing atom map
- Multicenter
- Multicomponent
- Overlapping atoms
- Overlapping bonds
- Pseudo atom
- Query atom
- Query bond
- Pseudo atom
- Radical
- Solvent
- Three dimension
- Wiggly double bond

# Chemical Terms Integration

---

- **isValid(...)** returns a boolean value of **false** if the given checkers detected any errors or issues, **true** otherwise
- **check(...)** returns a string of “passed” or “failed” and provides additional **description** text on the identified problems
- **fix(...)** returns a **structure**, corrected if possible

where ... is an action string defining the checkers and optionally their settings like “queryAtom..bondAngle.. aromaticity:basic”

```
$ evaluate -e 'check("valence..overlappingAtoms..wedge..metallocene")' molecules.sdf
failed, Overlapping Atoms Checker: 2 overlapping atoms found
failed, Valence Error Checker: 1 atom with valence problem found
passed
failed, Wedge Error Checker: 1 Invalid wedge problem found
failed, Overlapping Atoms Checker: 4 overlapping atoms found
passed
failed, Valence Error Checker: 1 atom with valence problem found,
failed, Metallocene Checker: 1 metallocene found
passed
failed, Valence Error Checker: 1 atom with valence problem found
```

# Integration

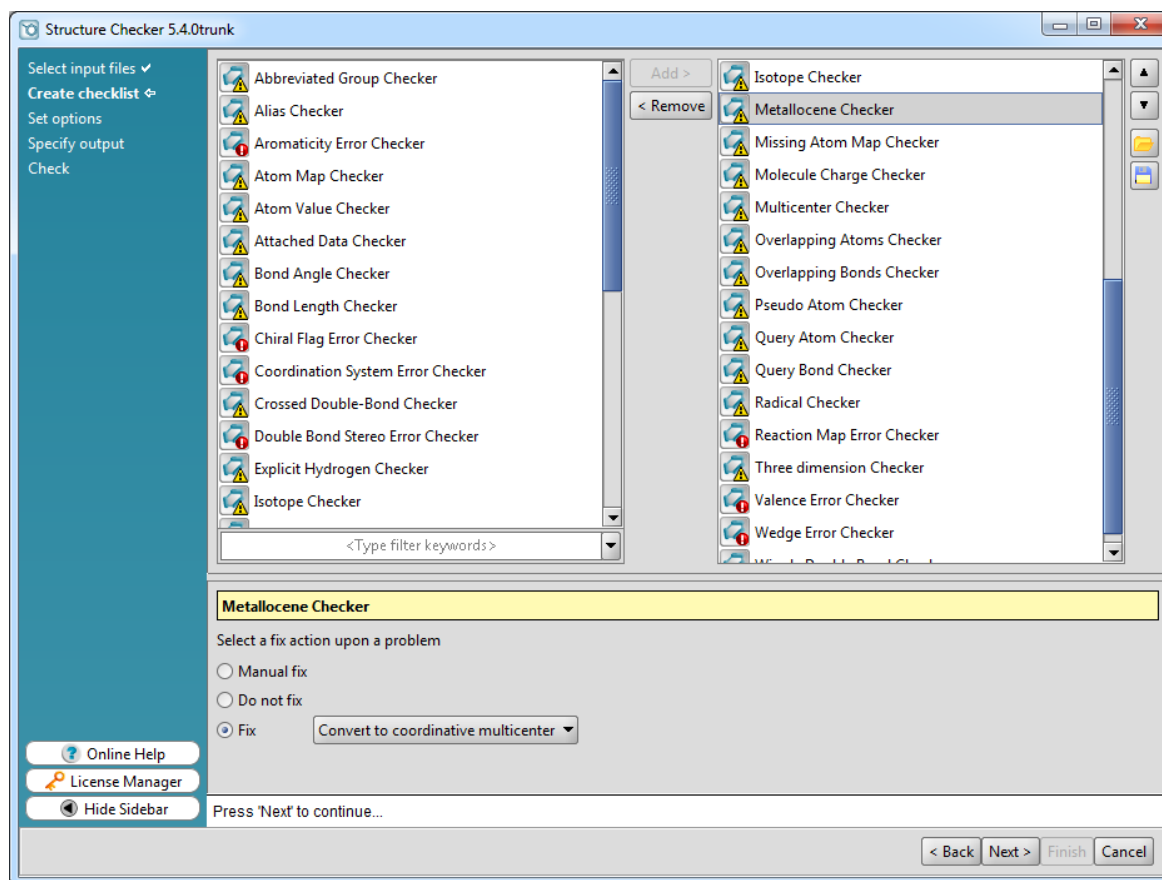
---

Structure Checker is currently available in the form of the `structurecheck` command line tool, it is smoothly integrated in MarvinSketch as a “chemical spell checker” and available in many other applications via Chemical Terms functions, for example:

- Instant JChem
- JChem for Excel
- JChem Cartridge
- JChem Web Service

# Structure Checker Wizard

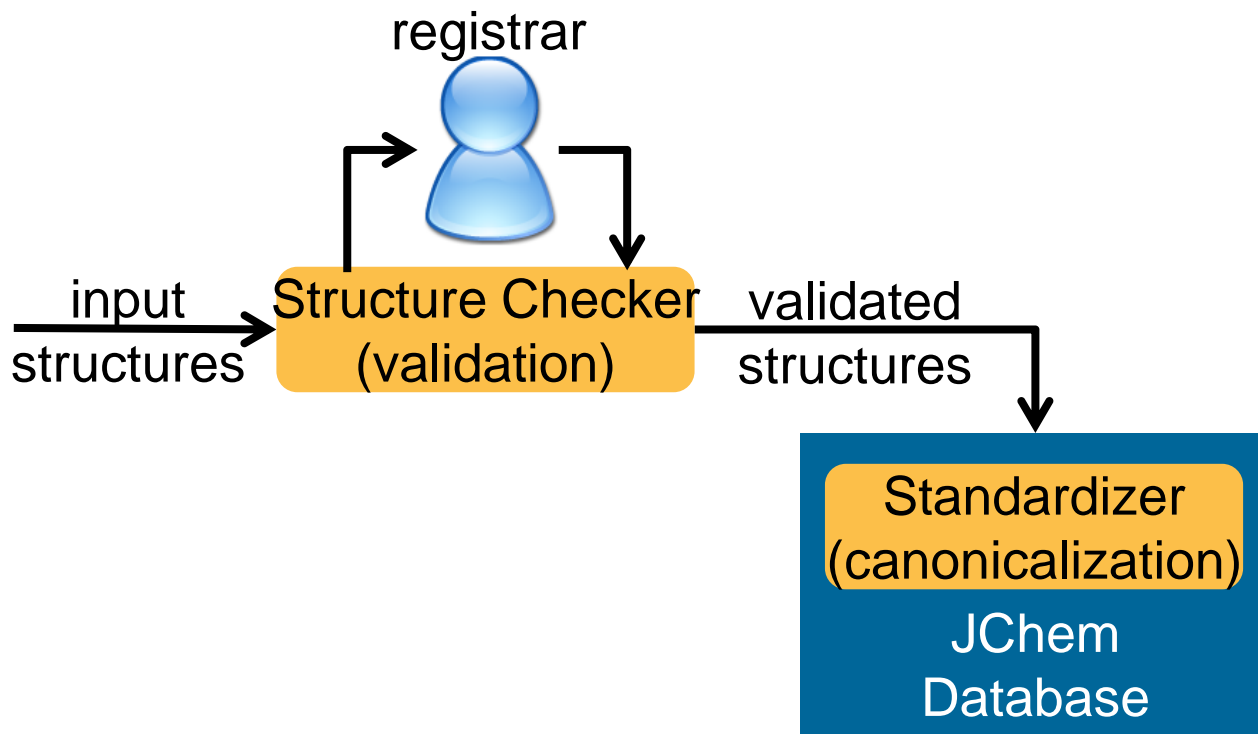
See live demo of the Structure Checker application providing wizard-like interface for the batch validation and correction of molecule files...



# Registration Systems

Structure Checker can be integrated in registration systems for the manual or automatic validation of structures before saving them in a database.

Standardizer provides automatic canonicalization for correct chemical structure search.



---

# Questions