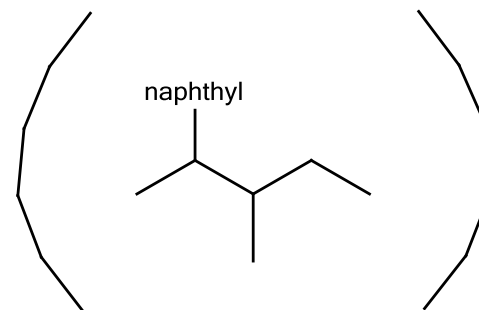
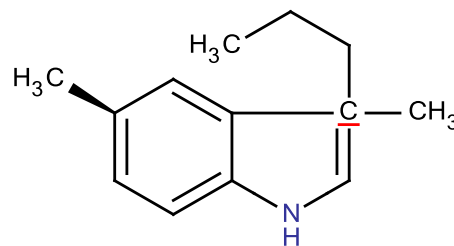
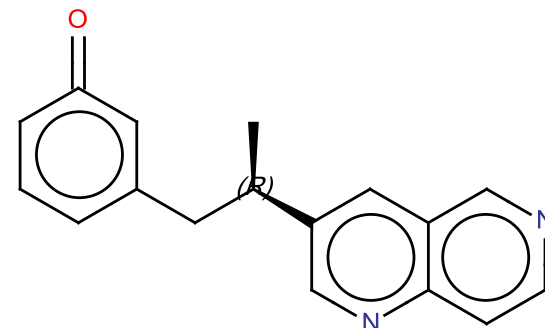
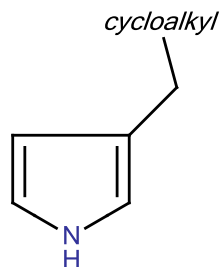
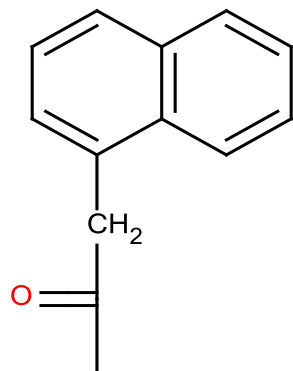

Structure Checker – *in silico* surgery for molecules

György Pirok, Zsolt Mohácsi, **Iván Solt**



Structure issues

- Drawing errors
- Scanning errors
- Inconsistent, ambiguous representations
- Aliases
- etc.



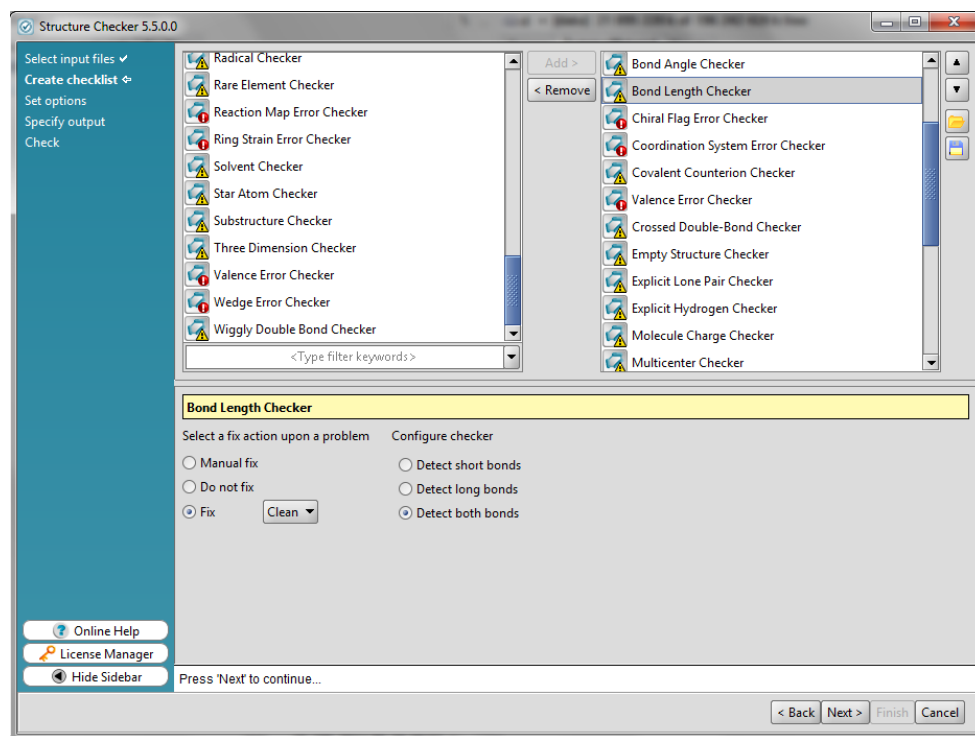
Structure Checker overview

- Reporting and/or fixing structure issues
- Manual or automatic
- Configurable checker modules for different

issues

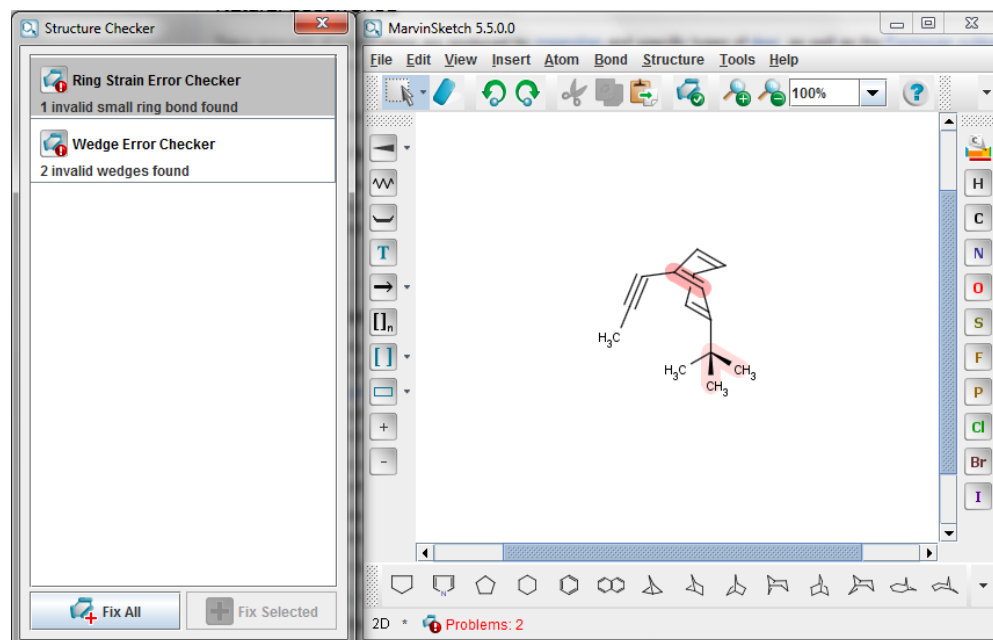
- aromaticity, attached data...
- atom issues (chirality, aliases, overlap, valence...)
- bond issues (length, angle, crossing...)
- substructure issues
- reaction map errors

- Configurable fixing protocols



























Integration
















- Accessibility:
 - MarvinSketch
 - standalone GUI and batch mode
 - via *Chemical Terms* in
 - Instant JChem
 - JChem for Excel
 - JChem Cartridge
 - JChem Web Services



Checkers – error and issue reporting

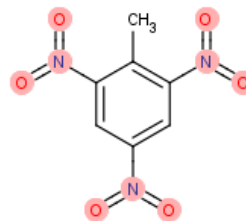
-  Aromaticity error checker
-  Chiral flag error checker
-  Coordination system error checker
-  **Metallocene error checker**
-  **OCR error checker**
-  Ring strain error checker
-  Reaction map error checker
-  Valence error checker
-  Wedge error checker

-  Abbreviated group checker
-  Alias checker
-  Atom map checker
-  Atom value checker
-  Attached data checker
-  Bond angle checker
-  Bond length checker
-  Covalent counterion checker
-  Crossed double bond checker
-  **Empty structure checker**
-  Explicit hydrogen checker
-  **Explicit lone pair checker**
-  Isotope checker
-  Missing atom map checker
-  **Molecule charge checker**

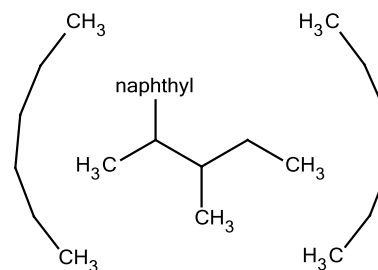
-  Multicenter checker
-  Multicomponent checker
-  Overlapping atoms checker
-  Overlapping bonds checker
-  Pseudo atom checker
-  Query atom checker
-  Query bond checker
-  **Racemate checker**
-  Radical checker
-  **Rare element checker**
-  Solvent checker
-  **Star atom checker**
-  **Substructure checker**
-  Three dimension checker
-  Wiggly double bond checker

New Checkers 1

- **Substructure checker**
 - Find SMARTS-defined structural elements
 - Fix: only manual fix available

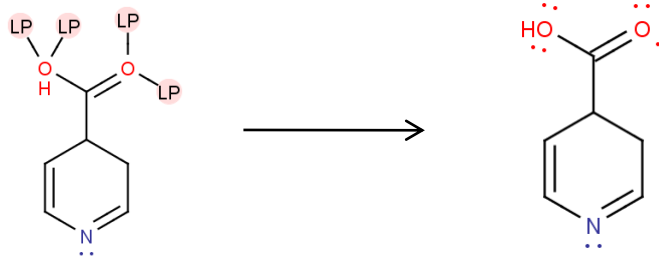


- **OCR error checker**
 - Errors stemming from misinterpretation of characters in OCR processes



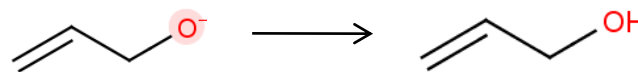
- **Empty structure checker**
 - Find and handle empty structure fields in multiple structure files

- **Explicit lone pair checker**
 - Find explicitly drawn lone pairs
 - Fix: remove

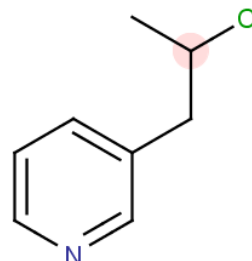


New Checkers 2

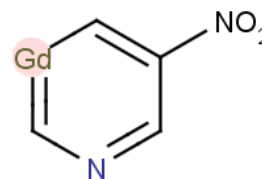
- Molecule charge checker
 - Finds molecules with non zero net charge
 - Fix: remove charge by addign/removing hydrogens



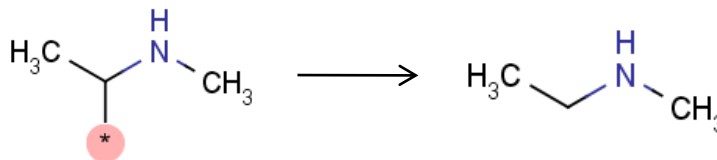
- Racemate checker
 - Finds molecules with chiral centers with no specific stereo configuration
 - Fix: only manual fixing available



- Rare element checker
 - Finds rare elements in structures
 - Fix: only manual fixing available



- Star atom checker
 - Finds star atoms
 - Fix: remove



Configuration options

- Checker configuration
 - individual configuration for each checker
- Fixer actions
 - auto/manual fixing
 - fix/leave intact

The screenshot shows a configuration window with two columns of checkers. The left column contains: Bond Length Checker, Chiral Flag Error Checker, Coordination System Error Checker, Covalent Counterion Checker, Crossed Double-Bond Checker, Empty Structure Checker, Explicit Hydrogen Checker, Explicit Lone Pair Checker, Isotope Checker, Metallocene Error Checker, and Missing Atom Man Checker. The right column contains: Abbreviated Group Checker, Alias Checker, Aromaticity Error Checker, Atom Map Checker, Atom Value Checker, Attached Data Checker, Bond Angle Checker, Bond Length Checker, Chiral Flag Error Checker, Coordination System Error Checker, Covalent Counterion Checker, and Crossed Double-Bond Checker. Below the columns is a search bar with the text '<Type filter keywords>'. The 'Abbreviated Group Checker' is selected and highlighted in yellow. Below this, there are two sections: 'Select a fix action upon a problem' with radio buttons for 'Manual fix', 'Do not fix', and 'Fix' (selected), and a dropdown menu set to 'Ungroup'; and 'Configure checker' with radio buttons for 'Detect Expanded Groups', 'Detect Contracted Groups', and 'Detect All Groups' (selected).

Configuration options – examples

Bond Length Checker

Select a fix action upon a problem

Manual fix

Do not fix

Fix

Clean ▾

Configure checker

Detect short bonds

Detect long bonds

Detect both bonds

Explicit Hydrogen Checker

Select a fix action upon a problem

Manual fix

Do not fix

Fix

Remove Explicit Hydrogen ▾

Configure checker

Lonely Charged Mapped

Isotopic Radical Wedged

Pseudo Atom Checker

Select a fix action upon a problem

Manual fix

Do not fix

Fix

Convert to Carbon ▾

Convert to Carbon

Delete Atom

Convert Pseudo Atom to Group

Abbreviated Group Checker

Select a fix action upon a problem

Manual fix

Do not fix

Fix

Ungroup ▾

Ungroup

Contract Group

Expand Group

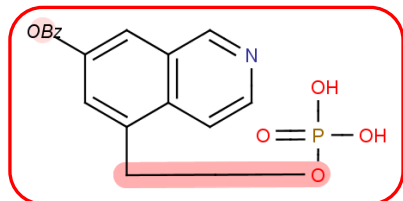
Configure checker

Detect Expanded Groups

Detect Contracted Groups

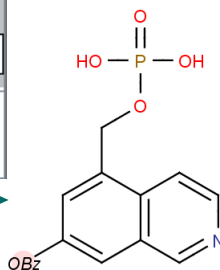
Detect All Groups

Series of checkers and fixers: example

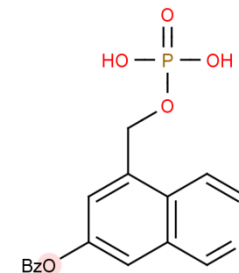


Bond Length Checker
1 bond found with wrong length
 Clean

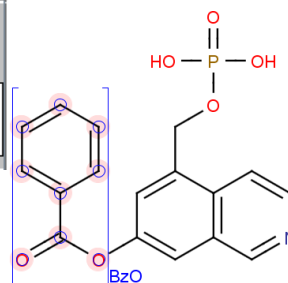
Pseudo Atom Checker
1 Pseudo atom found



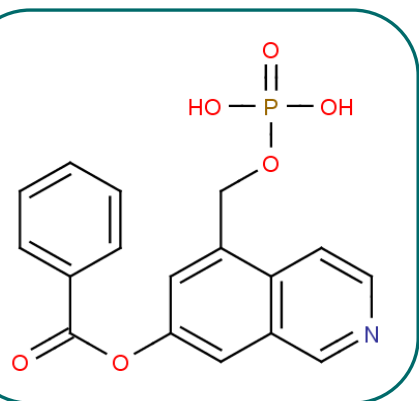
Pseudo Atom Checker
1 Pseudo atom found
 Convert to Carbon
 Delete Atom
 Convert Pseudo Atom to Group



Abbreviated Group Checker
1 abbreviated group found
 Ungroup
 Contract Group



Abbreviated Group Checker
1 abbreviated group found
 Ungroup
 Expand Group



Reporting or fixing

Set options

Choose a run method determining how to handle the identified issues, and select a report generation option.

Operation Mode

- Check**
Validates the structures without fixing the detected issues.
- Manual**
Ignores all configured fixers. Users are always prompted to resolve any issues detected.
- Automatic**
Fixes the structures with the selected fixers if possible. Users are never prompted.
- Fix**
Solve problems with the fixer configured for each checker. Users are prompted to resolve any unfixable or conflicting problems.

Report Options

- No Report**
No report is created.
- File Report**
Report is saved to a text file.
- Output Report**
Report is saved to the output as property field. Note: Not all file format supports this option.

Output options

- Optionally separate output files for accepted and discarded structures
- „Failsafe” mode
- Automatic discarding of OCR errors

Specify output

Specify the location and name of the output files. You can save all structures in a single file, or you can separate the accepted and discarded ones in two files.

Single Output

Both fixed and unfixed structures will be saved to a single output file.

Separated Output

Fixed and Unfixed structures will be saved to separate output files.

Accepted



Browse...

Discarded



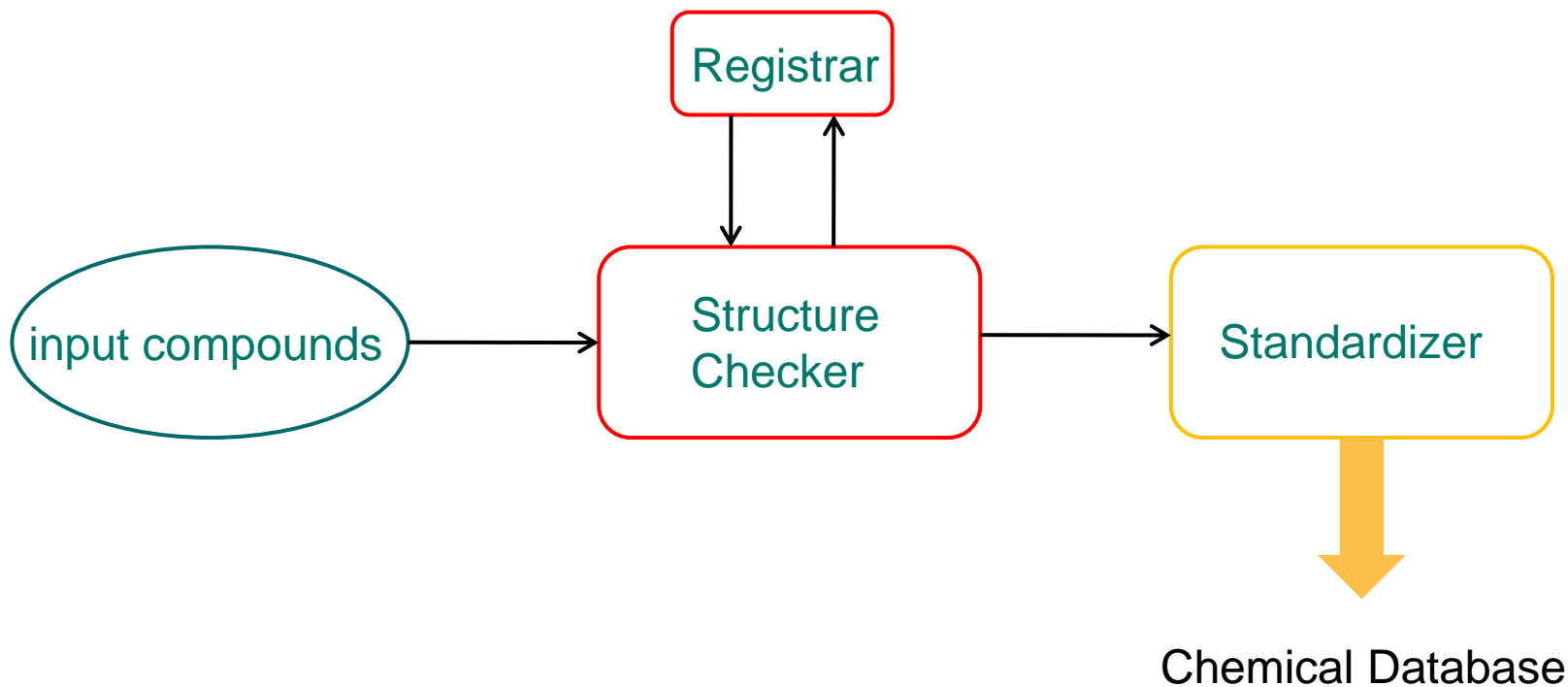
Browse...

Ignore errors and continue with next structure

Discard OCR errors

Registration system

- Key component for registration systems
- Combined with Standardizer



Future plans

- New checkers:
 - Reaction Checker
 - Unbalanced Reaction Checker
 - Valence Property Checker
 - R-group Checker
 - R-group Error Checker
 - Invalid R-group Checker
 - Polymer Checker
- Improve Structure Checker command line application
- Customizable fixers for Substructure Checker

Thank you



György Pirok



Attila Szabó



István Rábel



Zsolt Mohácsi