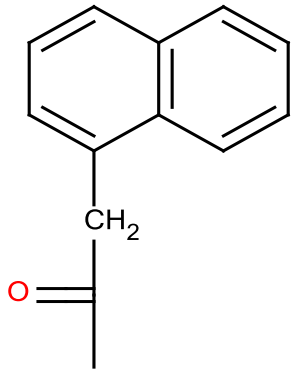
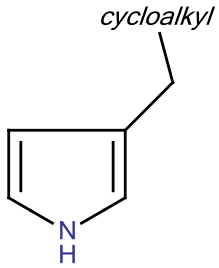
Structure Checker – *in silico* surgery for molecules

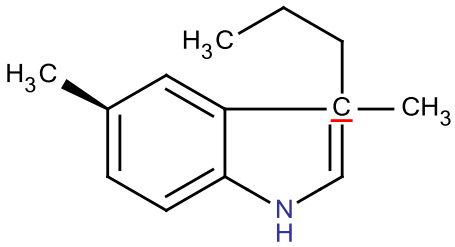
György Pirok, Zsolt Mohácsi, **Iván Solt**

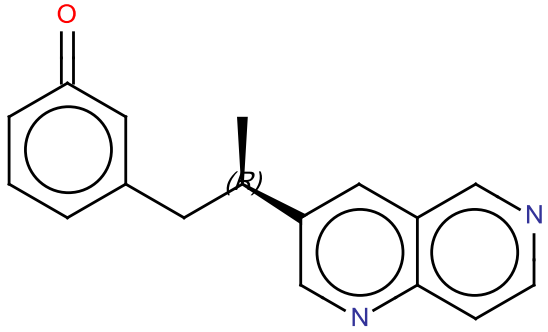
Structure issues

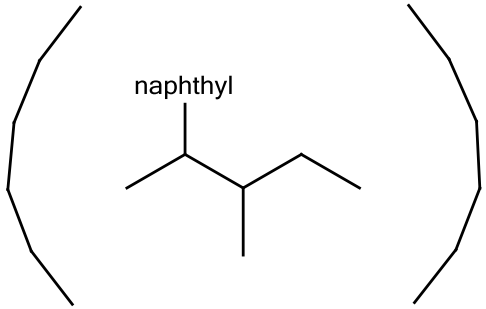
- Drawing errors
- Scanning errors
- Inconsistent, ambiguous representations
- Aliases
- etc.





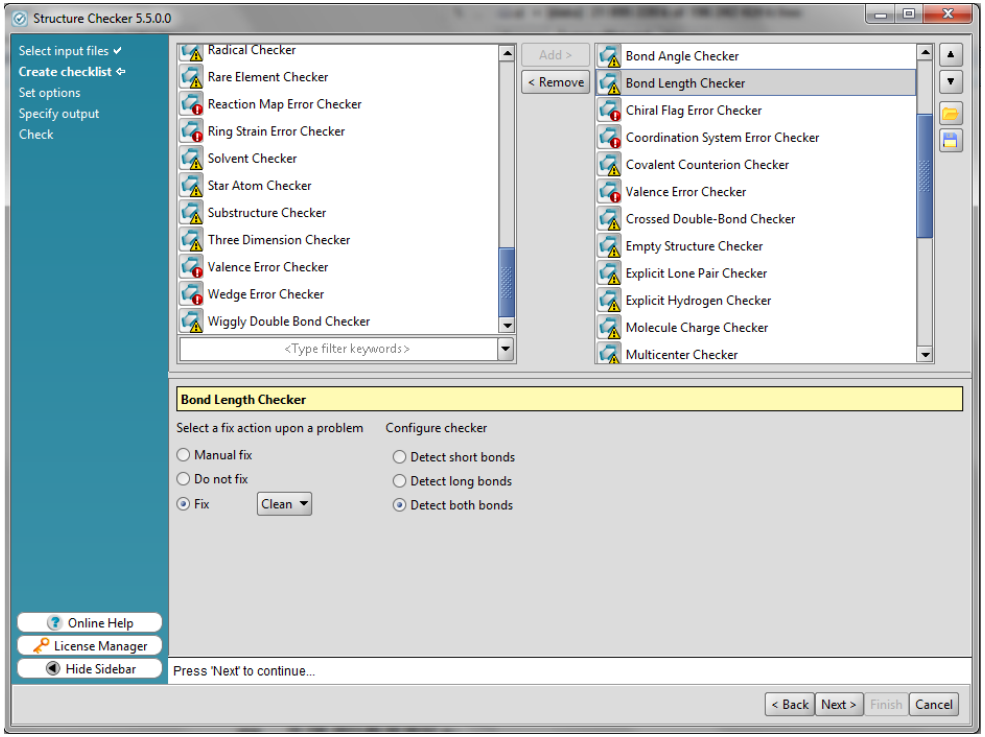






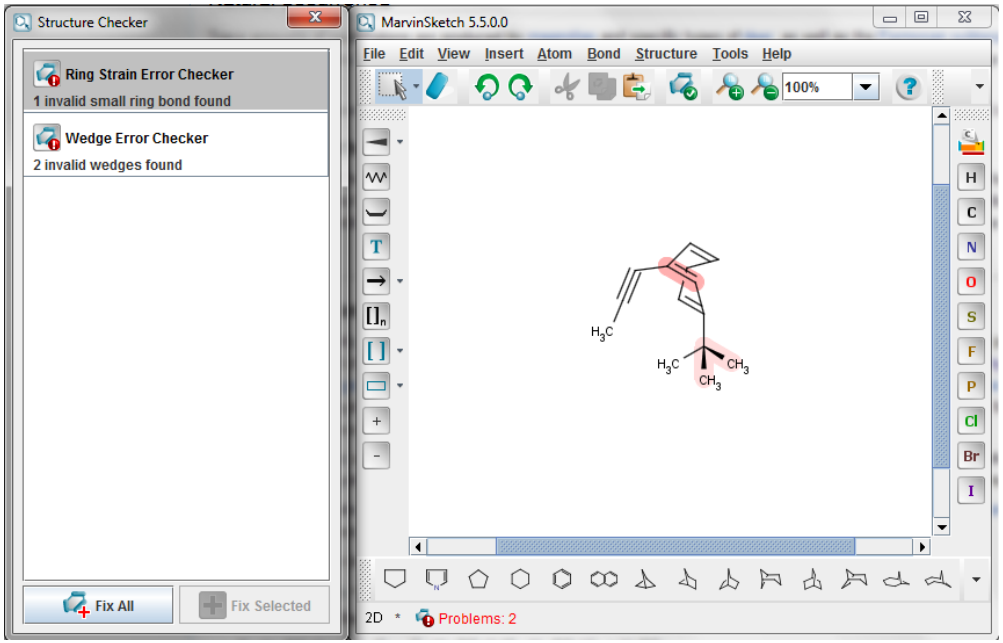
Structure Checker overview

- 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



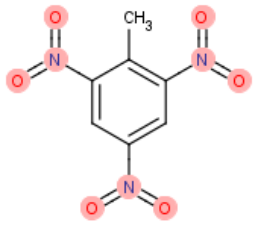
Checkers – error and issue reporting

- 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

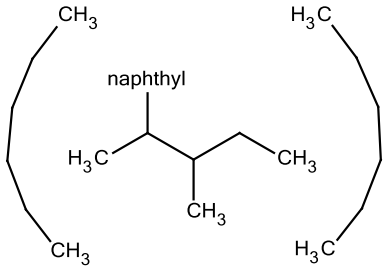
- Abbreviated group checker
- Alias checker
- Atom map checker
- Atom value checker
- Attached data checker
- Bond angle checker
- Bond length checker
- Covalent counterion checker
- Crossed double bond checker
- **Empty structure checker**
- Explicit hydrogen checker
- **Explicit lone pair checker**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

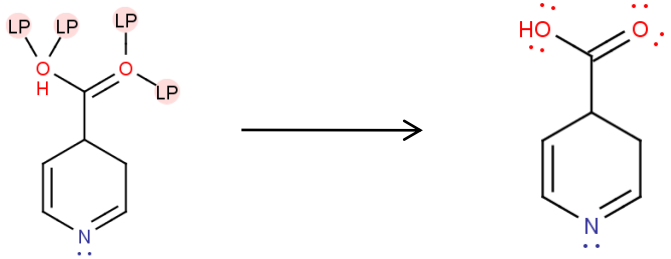


- OCR error checker
  - Errors stemming from misinterpretation of characters in OCR processes



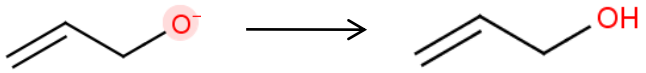
- Empty structure checker
  - Find and handle empty structure fields in multiple structure files

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

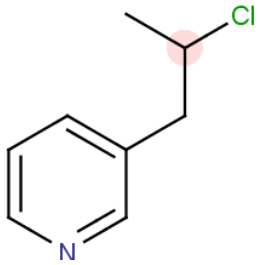


New Checkers 2

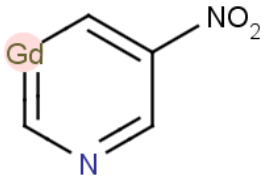
- Molecule charge checker
  - Finds molecules with non zero net charge
  - Fix: remove charge by addign/removing hydrogens



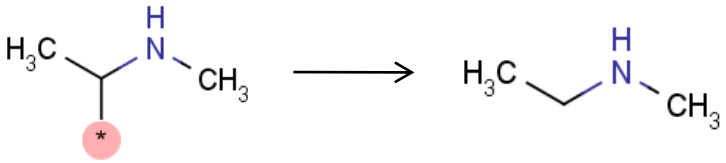
- Racemate checker
  - Finds molecules with chiral centers with no specific stereo configuration
  - Fix: only manual fixing available



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

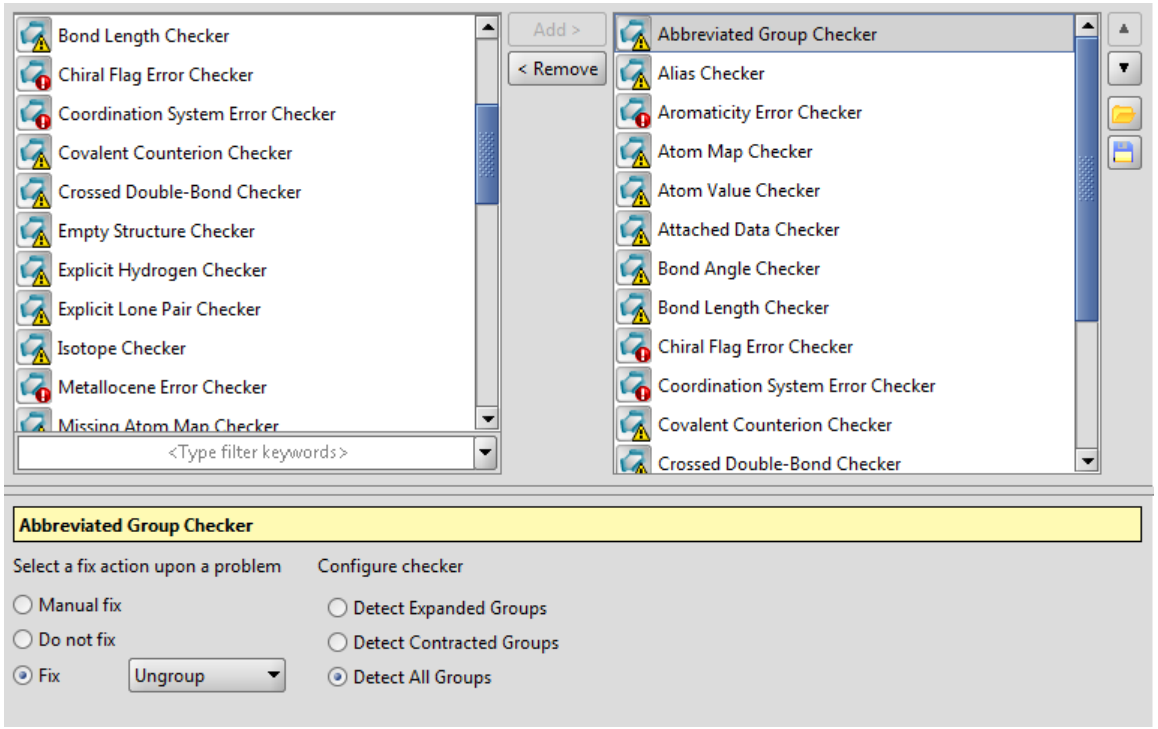


- Star atom checker
  - Finds star atoms
  - Fix: remove

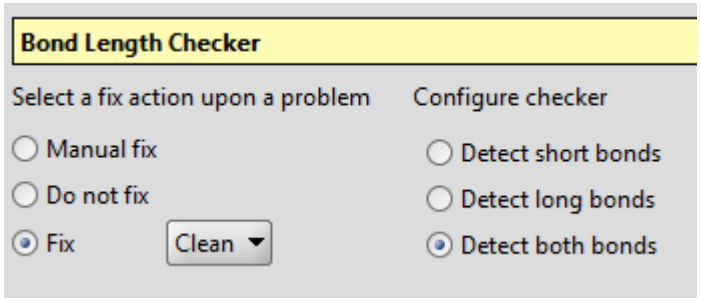


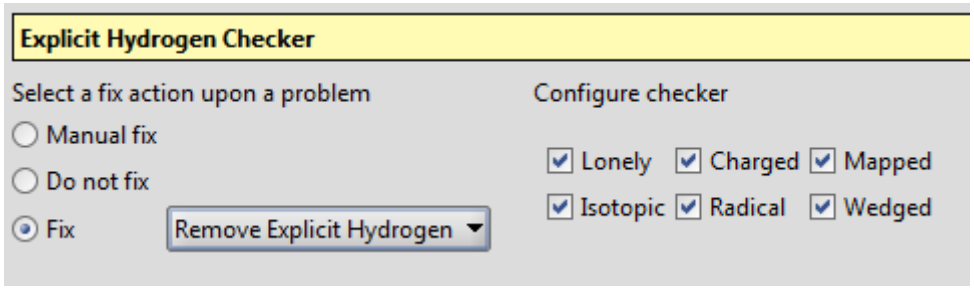
Configuration options

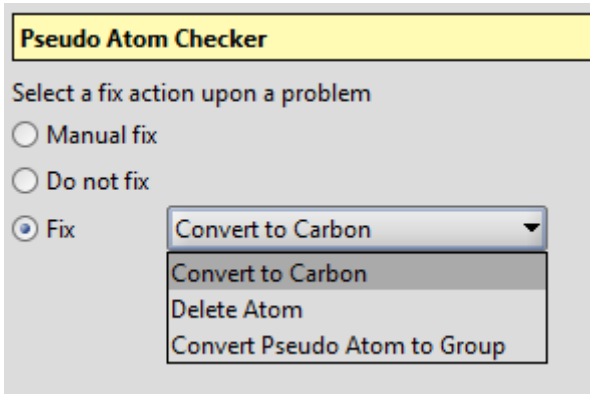
- Checker configuration
  - individual configuration for each checker
- Fixer actions
  - auto/manual fixing
  - fix/leave intact

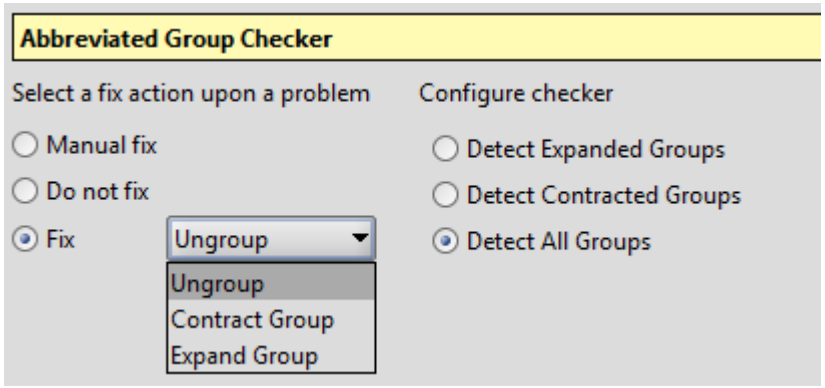


Configuration options – examples

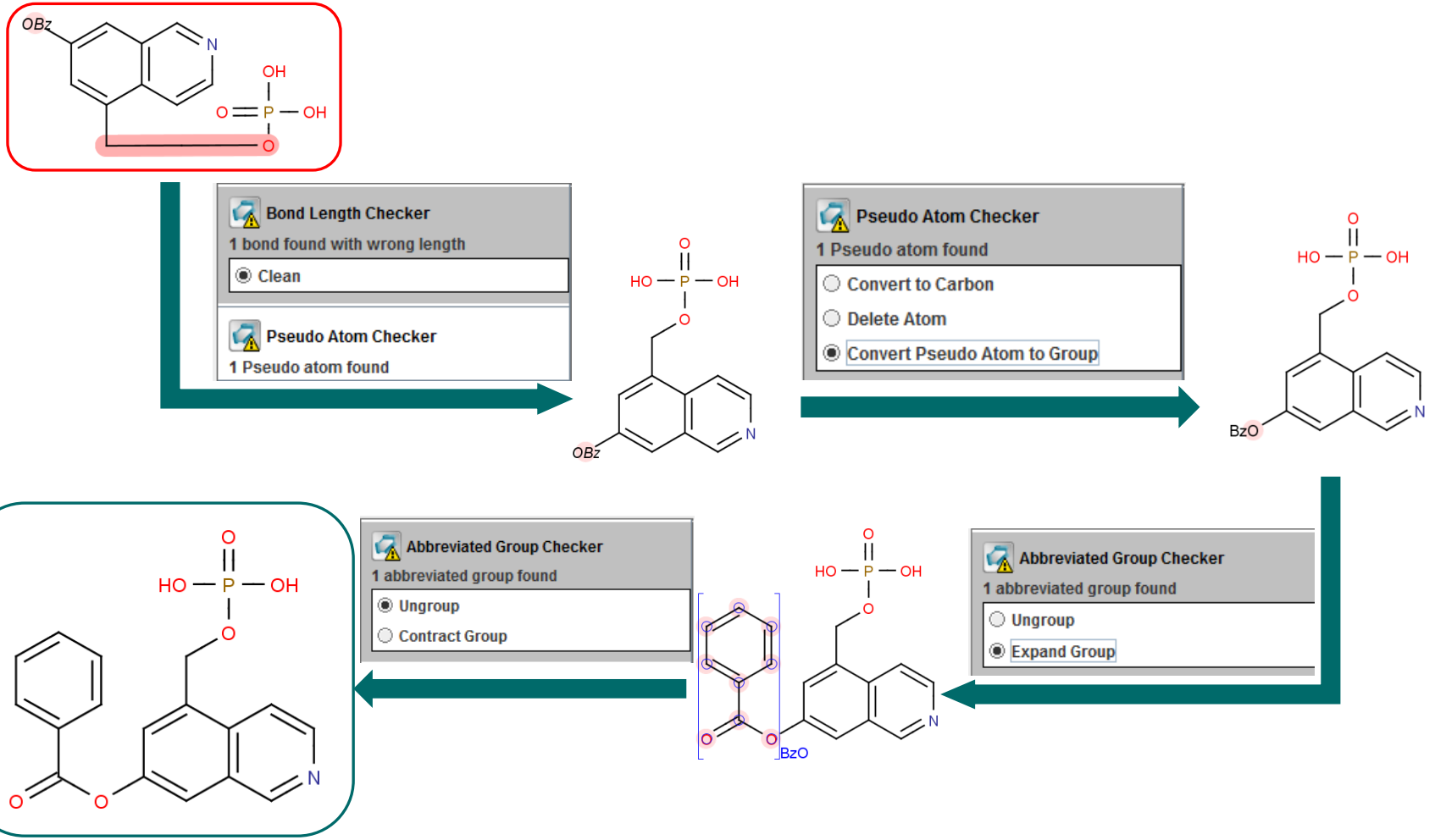




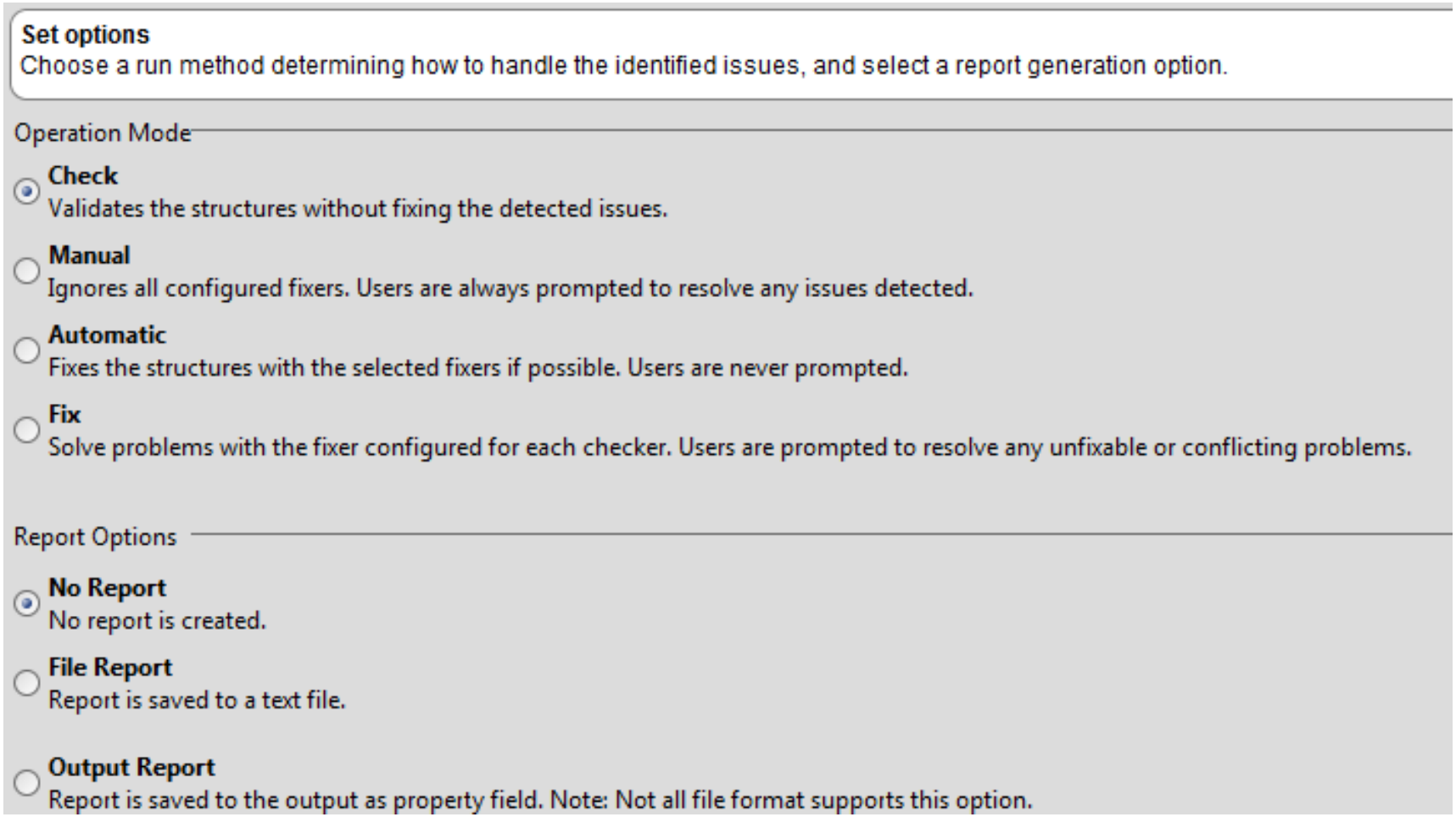




Series of checkers and fixers: example

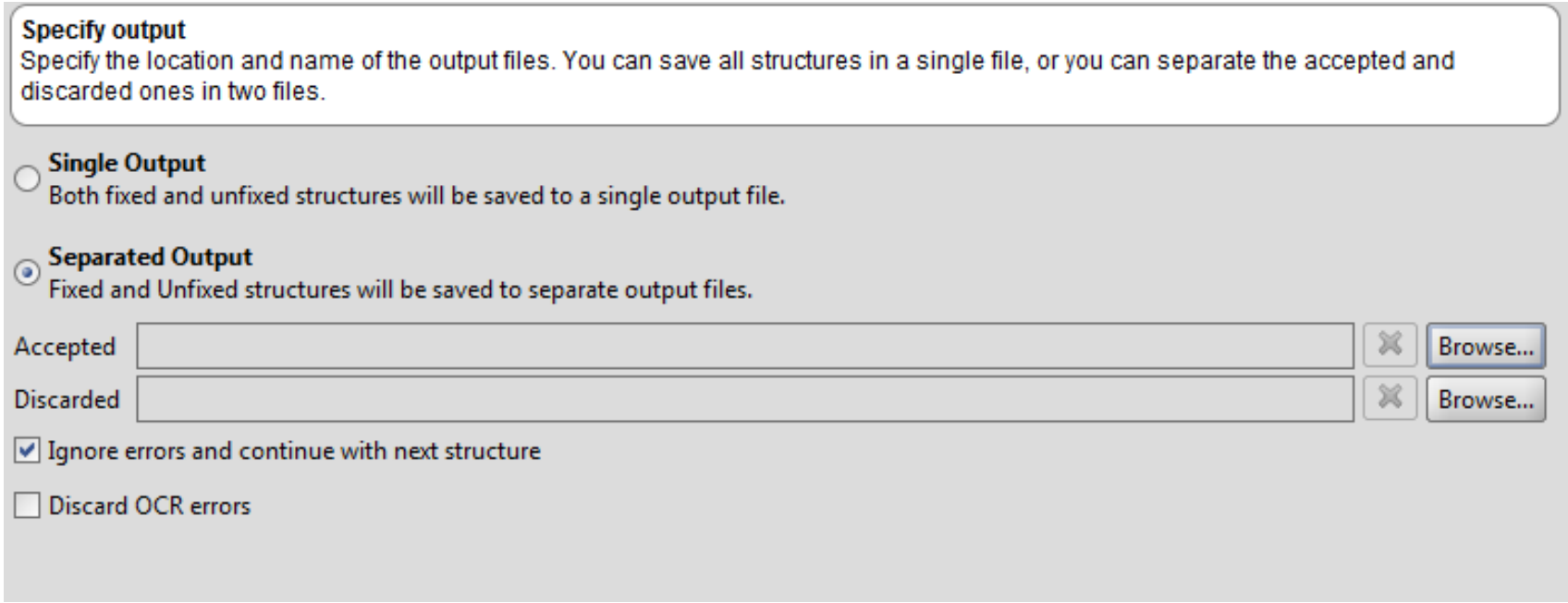


Reporting or fixing



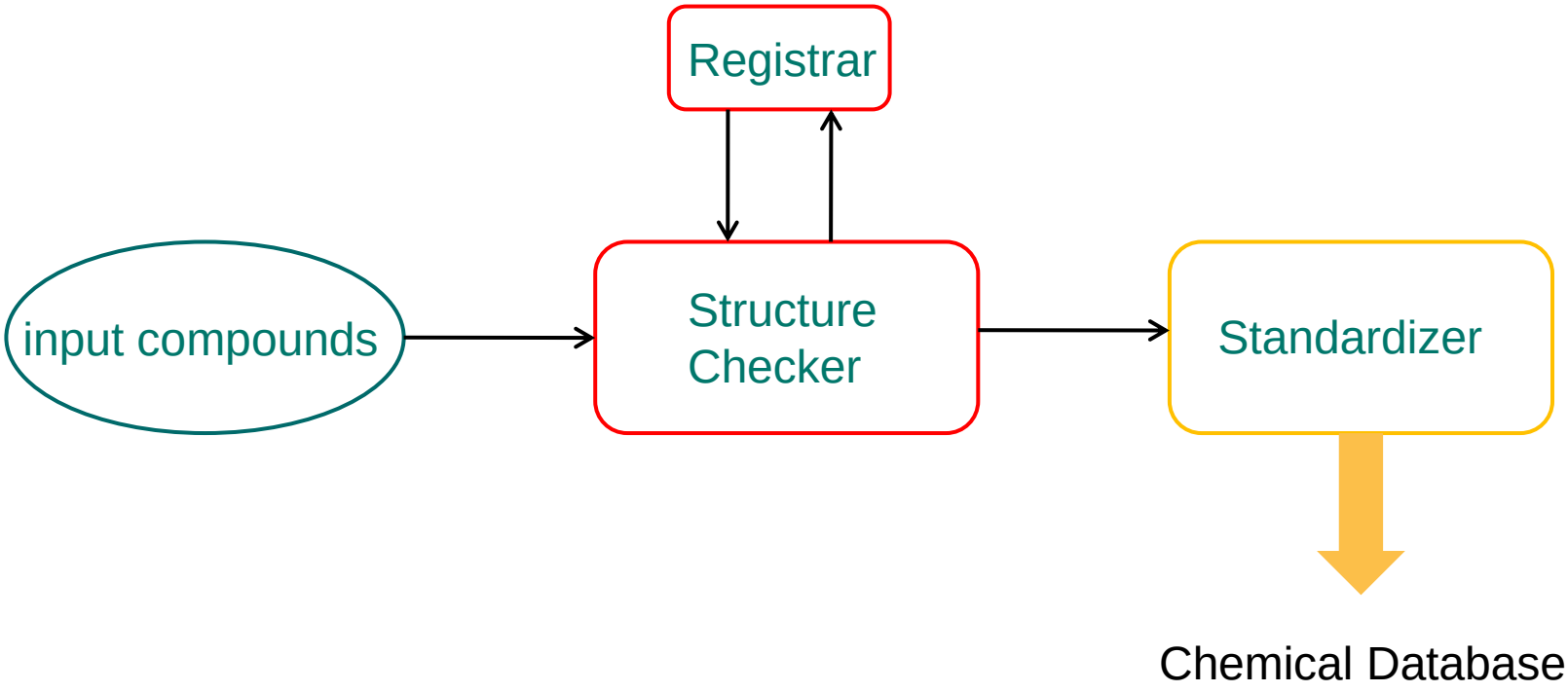
Output options

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