

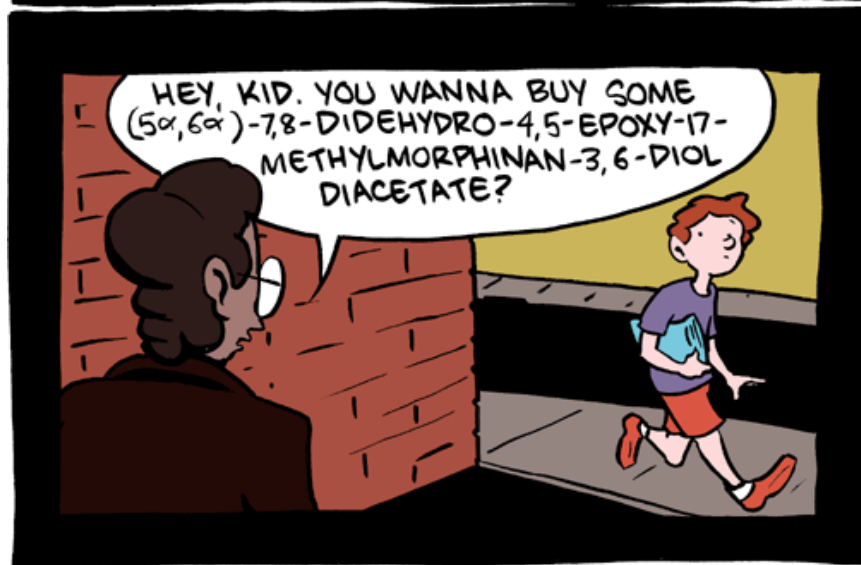
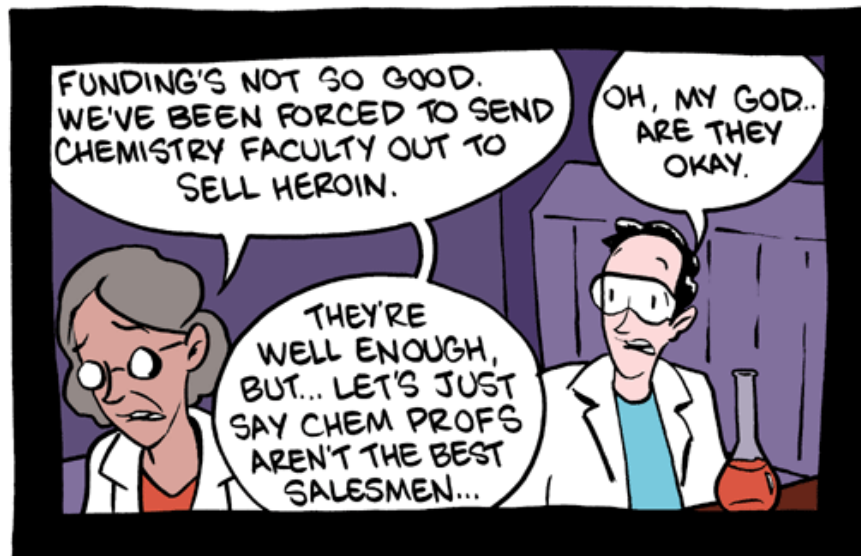
# Reliably converting names to structures with ChemAxon tools

**Daniel Bonniot**

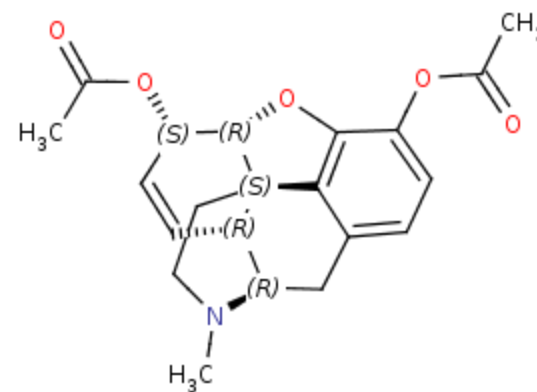
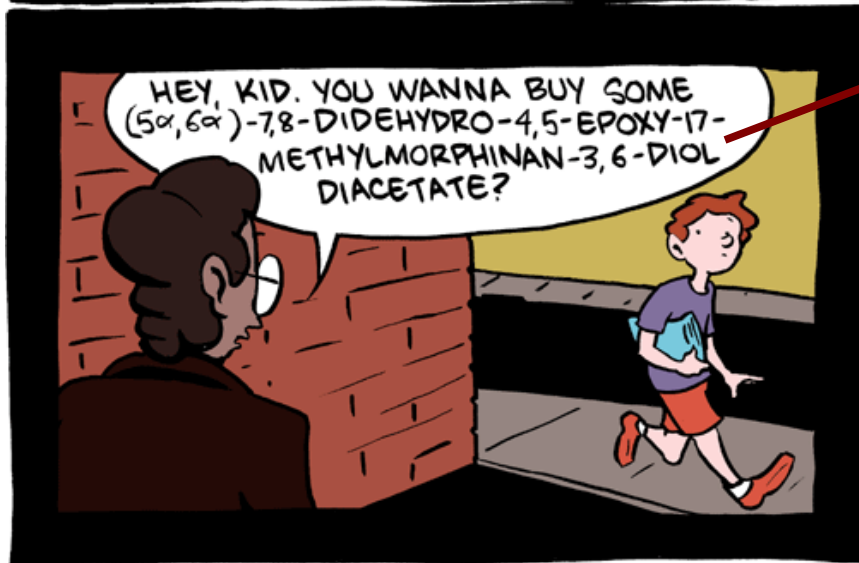
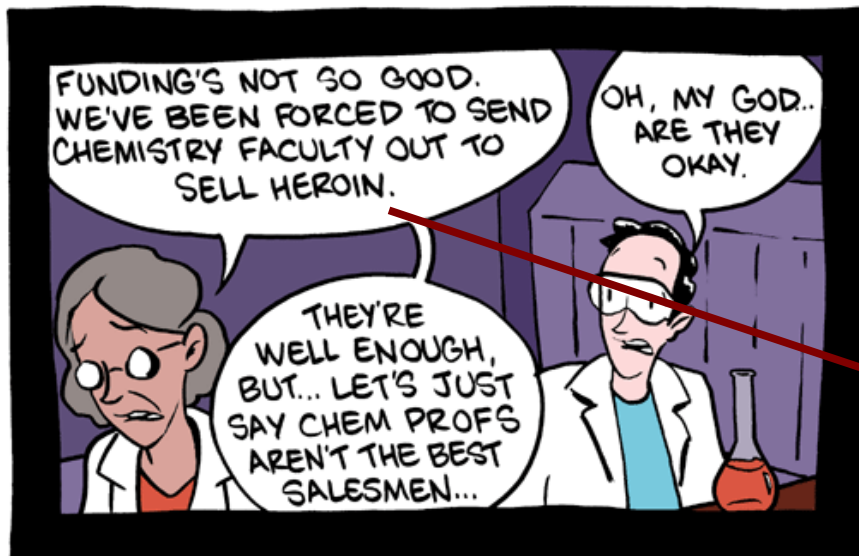


**ChemAxon**

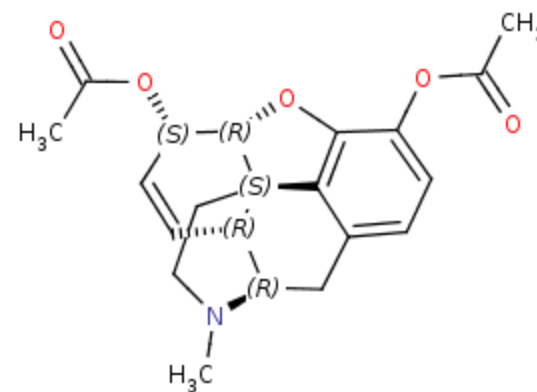
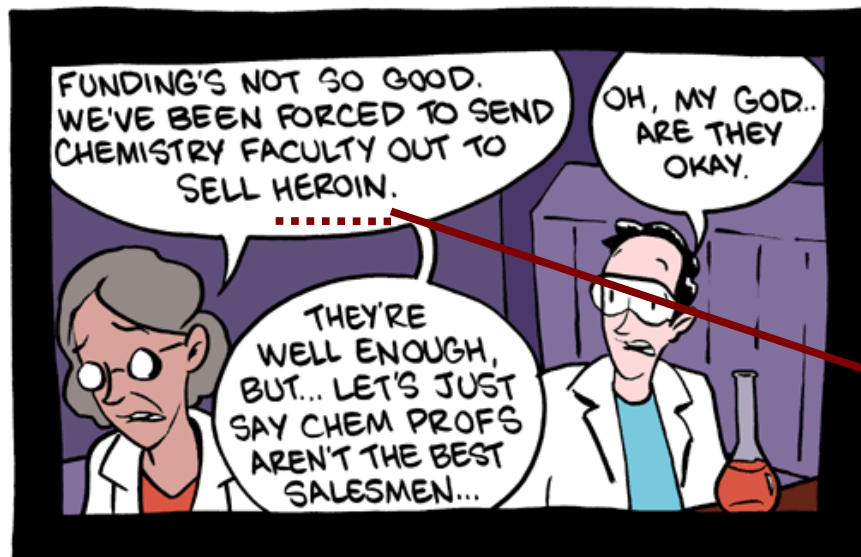
# A look at chemical literature



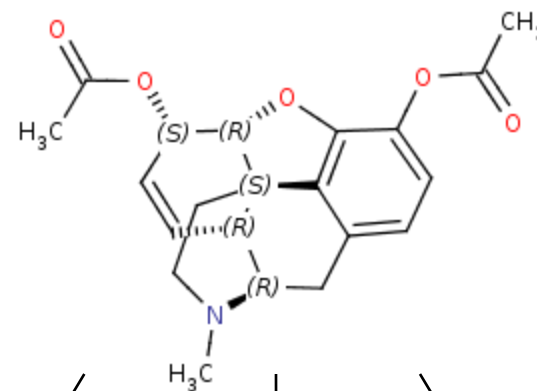
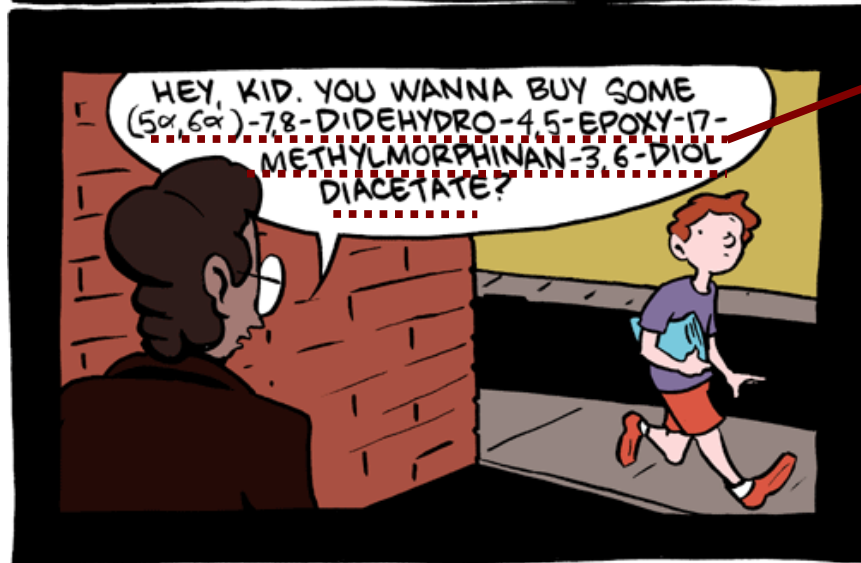
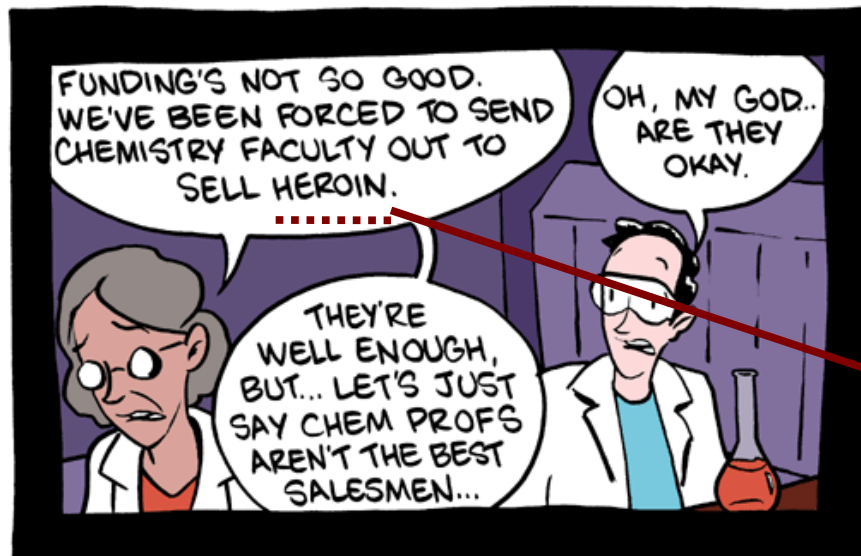
# Name to structure



# Name extraction from document



# Name extraction from document

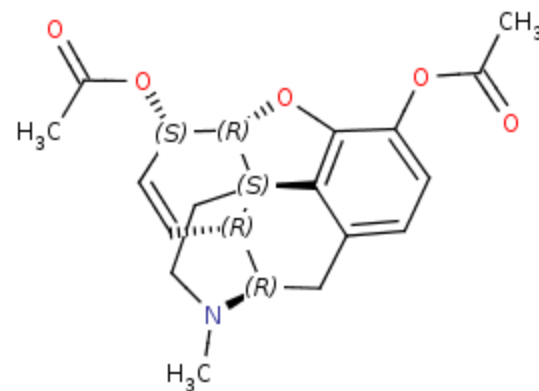
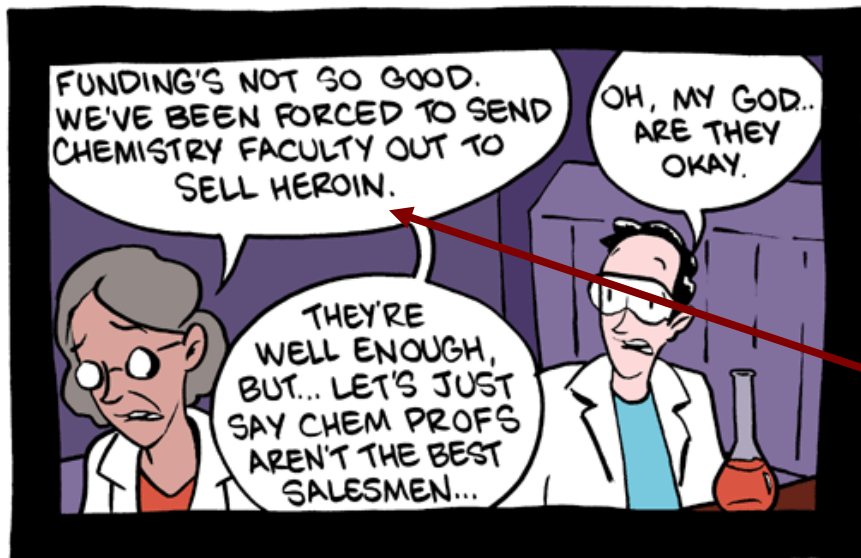


Export

Database

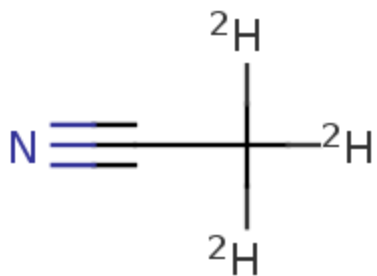
...

# Structure to Name

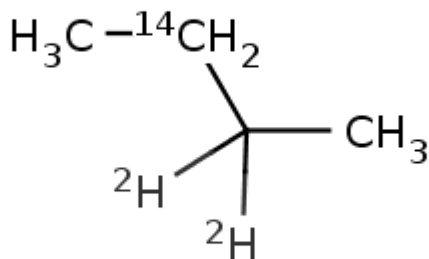


# Structure to Name

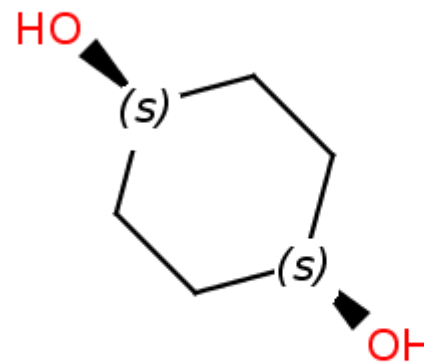
- Support for Isotopes, pseudo-asymmetric stereocenters, ...:



( $^2\text{H}_3$ )acetonitrile



(3- $^{14}\text{C}$ ,2,2- $^2\text{H}_2$ )butane



(1s,4s)-cyclohexane-1,4-diol

- Generates more traditional names (optional)
- 2x speedup in 5.4.0 (november 2010)
- Various fixes and smaller improvements



# Name to Structure

---

Continuous improvements since last year:

- Coverage (how many names we understand correctly)
- Correction (few errors)

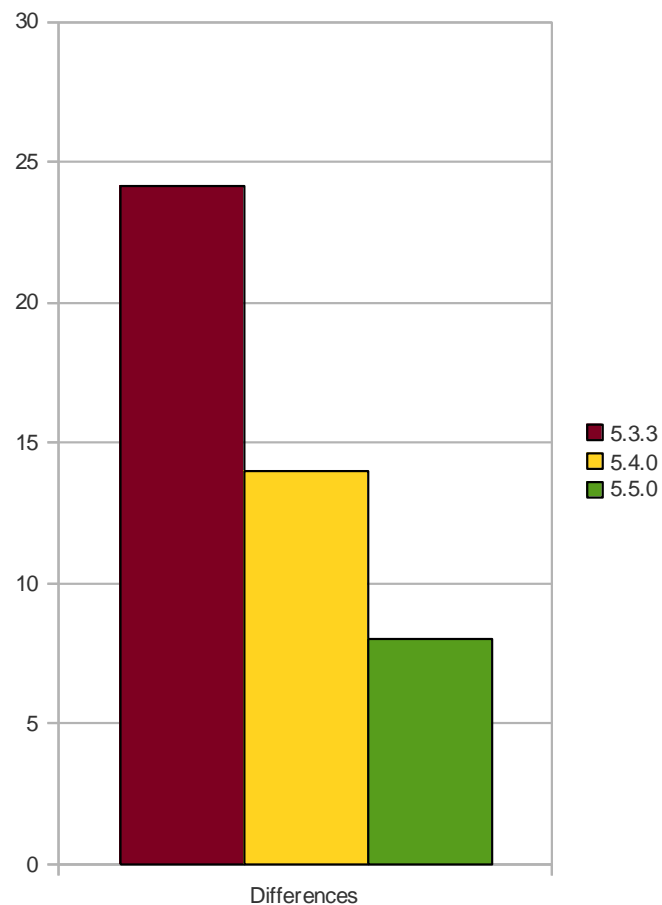
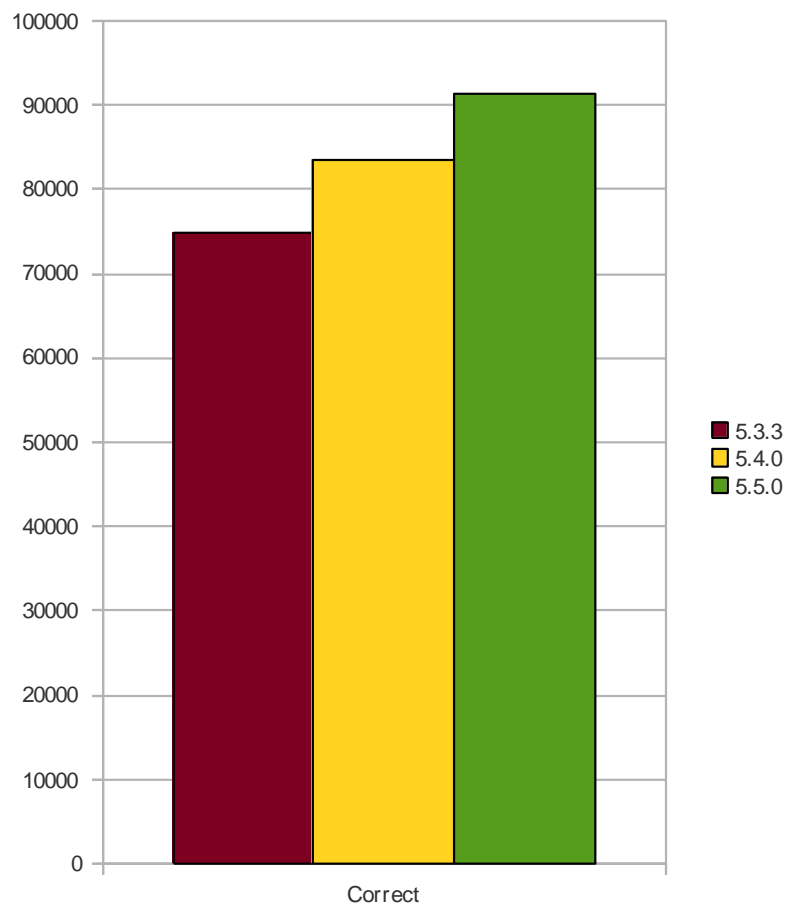
5.4.0 and 5.5.0 both include major upgrades:

- Coverage+
- Correction++



# Name to Structure improvements

Various names, mixed nomenclatures (hard)



# Document to Structure

---

- Extracts names from text documents
- Returns the discovered structures and their location
- OCR and syntax correction constantly developed
  - 3-methyl-1-methoxynaphthalene
  - 3-methyl-1-methoxynaphthalene
- Since 5.4: extraction from **PDF** documents



# Summary

---

Structure to Name: incremental improvements, mature

Name to Structure: improving steadily, major upgrades in 5.4 and 5.5

Document to Structure: improves with n2s, PDF support, more OCR and syntax correction

*Try them, get support, ask questions!*