Some Applications of InChI

Peter J. Linstrom

Office of Data and Informatics, NIST

A long introduction to a short talk

- InChI is a tool

- Ways we have used it
  - Analyzing computational results
  - Matching database entries
  - Helping people navigate databases

- Ways we might use it in the future
  - Now you can run InChI code in the web browser (WebAssembly)!

InChI is a tool

Basic functionality

- Uniquely identify a chemical species
  - Problems with some classes of compounds
- Identify related species
  - Stereo and geometric isomers
  - Isoanalogues
  - Protonated species

Applications

- Searching
  - Data collections
  - Internet (InChIKey)
- Merging database entries
- Miscellaneous computational activities
  - “diff” for structures
  - Counting duplicates

Example – Computational results

- Testing of computational method
- Generate results based on different models / input conditions
- Result is 3d coordinates
  - Can’t compare results directly
  - Coordinates → molecular graph → InChI

Do these data refer to the same species?

Example NIST Gas Chromatographic Retention Index Database.

- Gas Chromatography data from literature, NIST, and affiliated labs
- Inconsistent identifiers (nomenclature) → have to draw structure, often multiple times
- Python script + InChI code to identify species which may be the same → **manual** review of data

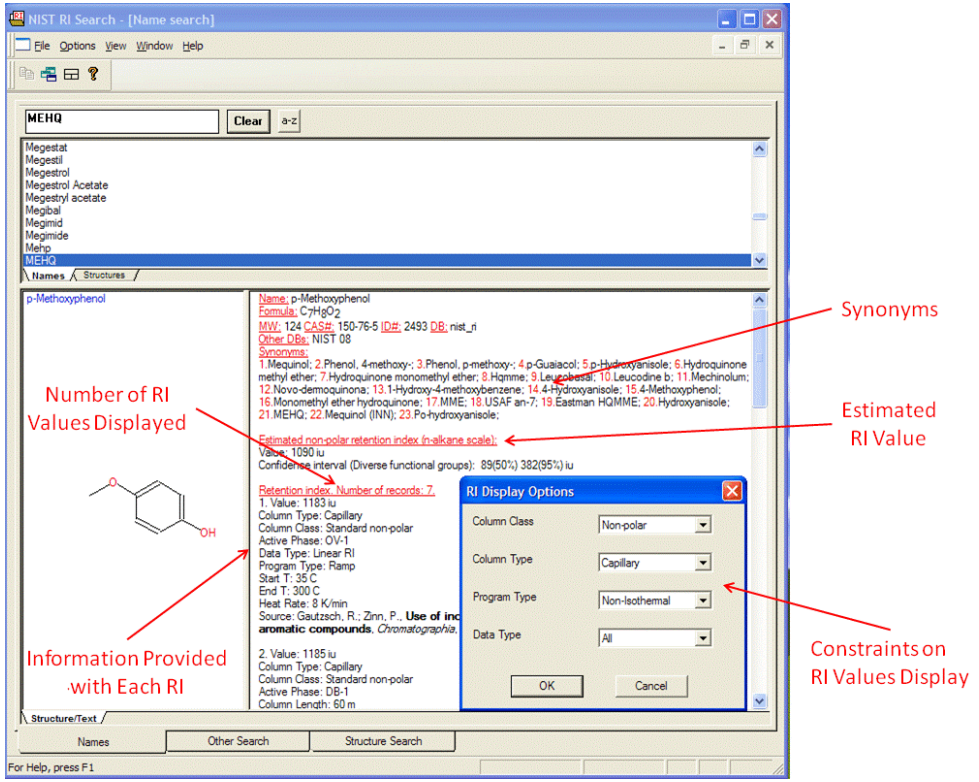


Image credit: chemdata.nist.gov

Navigating databases

The layered nature of InChI makes it easy to identify “duplicate” entries, non-structural isomers, and isoanalogues

We have “historical” data: some “duplicates” are really structures with incomplete information.

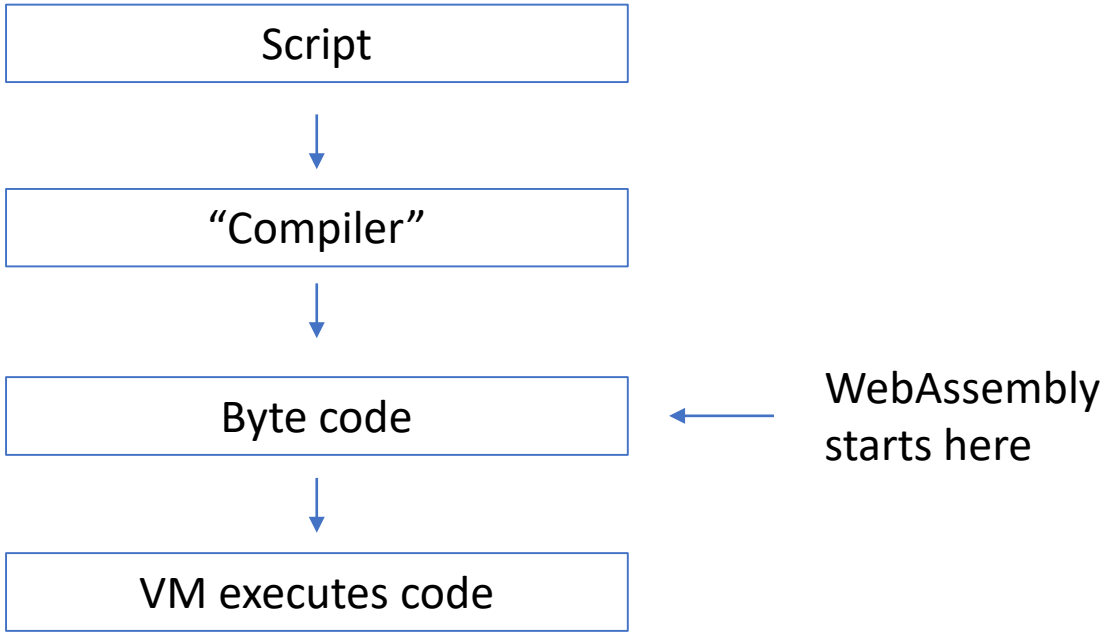
- **Chemical structure:**

This structure is also available as a 2d Mol file or as a computed 3d SD file
The 3d structure may be viewed using Java or Javascript.

- **Isotopologues:**
  - Benzene-d6 <-- Link found via InChI
- **Other names:** Benzol; [6]Annulene; Benzole; Coal naphtha; Cyclohexatriene; Phenyl hydride; Pyrobenzol; Pyrobenzole; Benzolene; Bicarburet of hydrogen; Carbon oil; Mineral naphtha; Motor benzol; Benzeen; Benzen; Benzin; Benzine; Benzolo; Fenzen; NCI-C55276; Phene; Rcra waste number U019; UN 1114; NSC 67315; 1,3,5-Cyclohexatriene

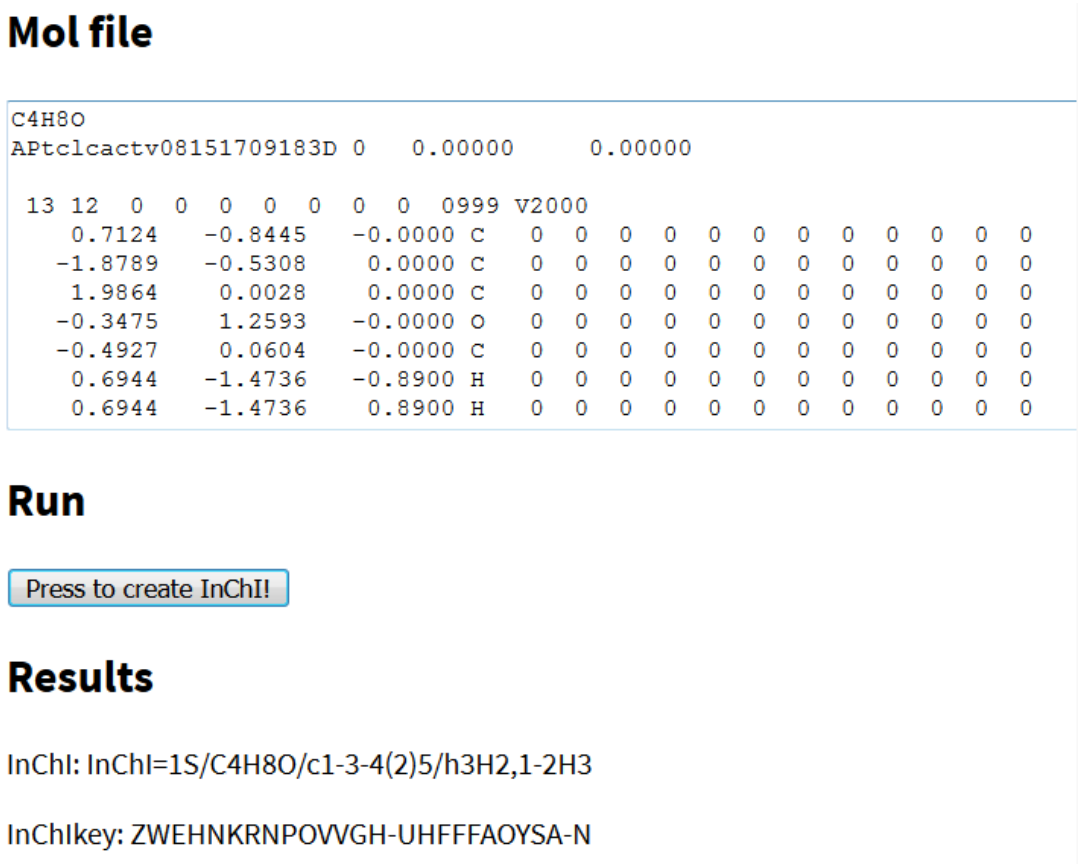
InChI code can run in a web browser!

- WebAssembly allows non-JavaScript code to run in web browsers
- Supported by:
  - Current versions of Firefox and Chrome
  - Current version of Edge (but needs a flag turned on for activation)
  - Safari 11 (currently in beta)



Demo of InChI in WebAssembly

- We can now compile C/C++ code to WebAssembly with Emscripten
  - Other languages / tools coming
  - Produces WASM, JS, and HTML files
- InChI v. 1.05 code compiles with Emscripten
  - LLVM based (__APPLE__)



Files in demo

- Index.html
  - HTML web page
  - Based on template from the compiler
  - Small JavaScript section to communicate with WebAssembly
- Inchi-demo.wasm
  - WebAssembly code (compiled from C/C++)
  - ~1.4 MBytes
- Inchi-demo.js
  - Loads WebAssembly
  - Generated by compiler

Links

- Demo (on this machine)
- Demo from Epic games: https://s3.amazonaws.com/mozilla-games/ZenGarden/EpicZenGarden.html
- List of asm.js, WebAssembly ports: https://github.com/kripken/emscripten/wiki/Porting-Examples-and-Demos