

# 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

The screenshot shows the NIST RI Search interface. The search term 'MEHQ' is entered in the search bar. The results list includes 'MEHQ', 'Megestat', 'Megestil', 'Megestrol', 'Megestrol Acetate', 'Megestryl acetate', 'Megibal', 'Megimid', 'Megimide', and 'Mehp'. The detailed view for 'p-Methoxyphenol' shows its chemical structure, name, formula (C<sub>7</sub>H<sub>8</sub>O<sub>2</sub>), MW (124), CAS# (150-75-5), ID# (2493), DB (nist\_ri), and other DBs (NIST 08). It lists 23 synonyms, including Mequinol, Phenol, 4-methoxy-, 3-Phenol, p-methoxy-, 4-p-Guaiacol, 5-p-Hydroxyanisole, 6-Hydroquinone methyl ether, 7-Hydroquinone monomethyl ether, 8-Hamme, 9-Leucobess, 10-Leucodine b, 11-Mechinolum, 12-Novo-demoquinona, 13.1-Hydroxy-4-methoxybenzene, 14.4-Hydroxyanisole, 15.4-Methoxyphenol, 16-Monomethyl ether hydroquinone, 17.MME, 18 USAF an-7, 19-Eastman HQMME, 20-Hydroxyanisole, 21.MEHQ, 22. Mequinol (INN), 23. Po-hydroxyanisole. It also shows the estimated non-polar retention index (n-alkane scale) as 1090 iu with a confidence interval of 89(50%) 382(95%) iu. The retention index is listed as 1183 iu with 7 records. The 'RI Display Options' dialog box is open, showing settings for Column Class (Non-polar), Column Type (Capillary), Program Type (Non-isothermal), and Data Type (All). Red arrows point to various parts of the interface: 'Synonyms' points to the 'Other DBs' field; 'Estimated RI Value' points to the 'Estimated non-polar retention index (n-alkane scale)'; 'Constraints on RI Values Display' points to the 'RI Display Options' dialog box; 'Information Provided with Each RI' points to the 'Retention index, Number of records, 7.' field; and 'Number of RI Values Displayed' points to the 'Retention index, Number of records, 7.' field.

Image credit: [chemdata.nist.gov](http://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](#).

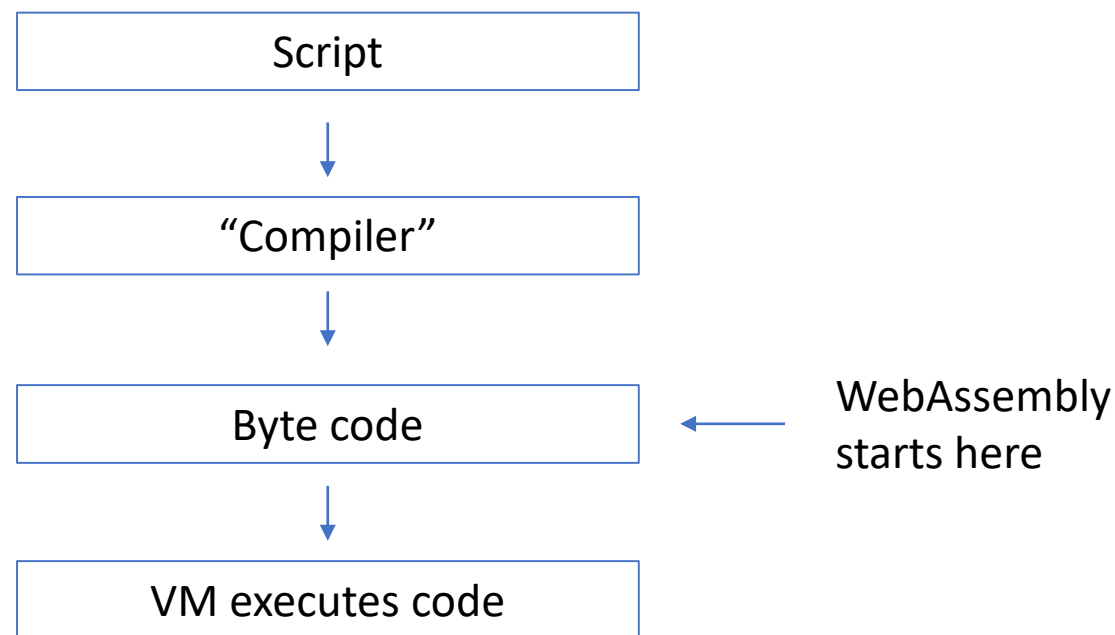
- **Isotopologues:**

- [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__`)

## Mol file

```
C4H8O
APtclcactv08151709183D 0 0.00000 0.00000

13 12 0 0 0 0 0 0 0 0 0999 v2000
 0.7124 -0.8445 -0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-1.8789 -0.5308 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 1.9864 0.0028 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-0.3475 1.2593 -0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-0.4927 0.0604 -0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0.6944 -1.4736 -0.8900 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0.6944 -1.4736 0.8900 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

## Run

Press to create InChI!

## Results

InChI: InChI=1S/C4H8O/c1-3-4(2)5/h3H2,1-2H3

InChIkey: ZWEHNKRNPVOVGH-UHFFFAOYSA-N



# 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>