Metamolecular

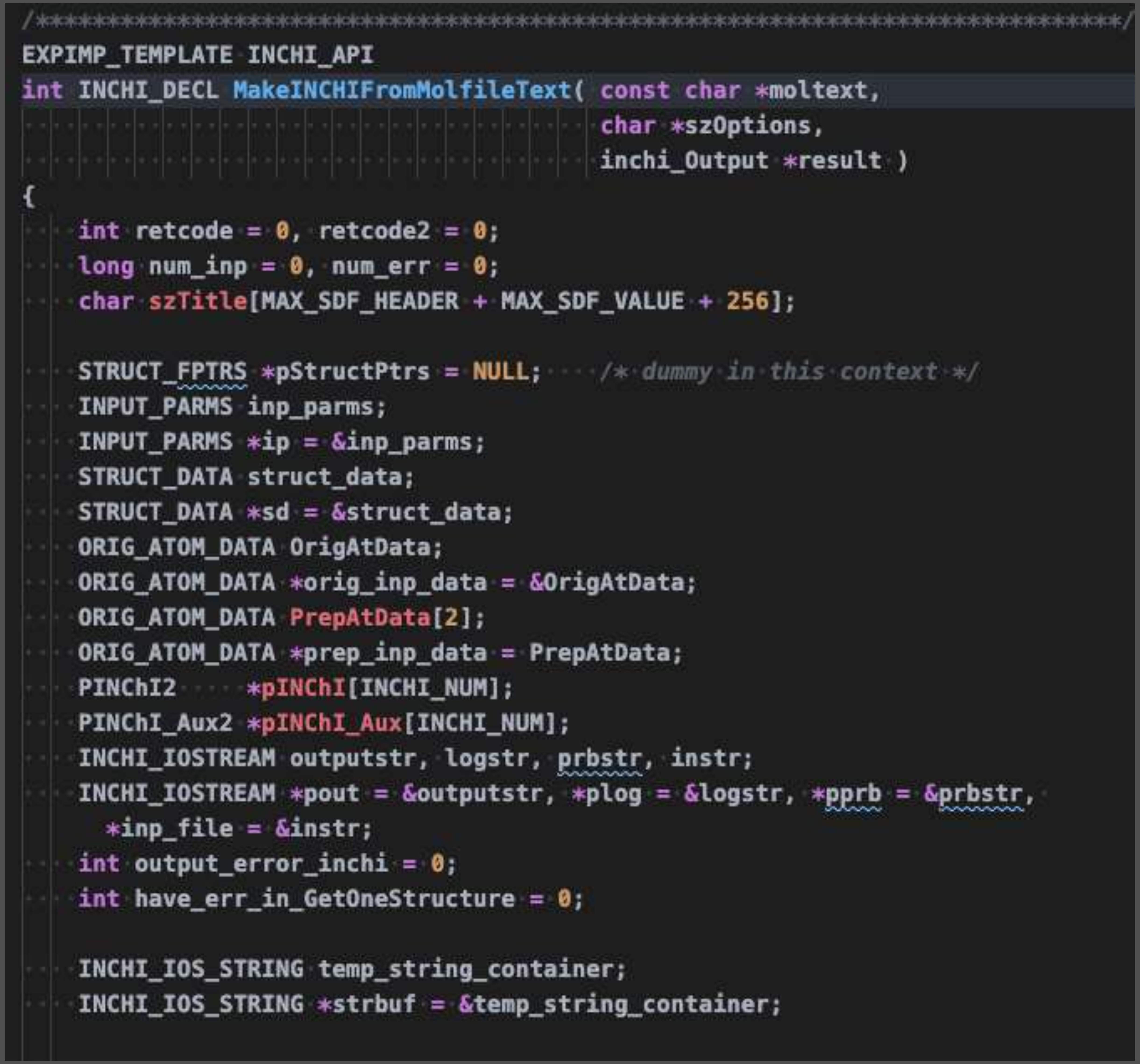
Running InChI Anywhere with WebAssembly

Rich Apodaca
(@rapodaca)

The InChI Value Proposition

- Unique Key: Does this database have this molecule?
- Foreign Key: What other databases have this molecule?
- Permissionless (cf. CAS)

InChI Source Code



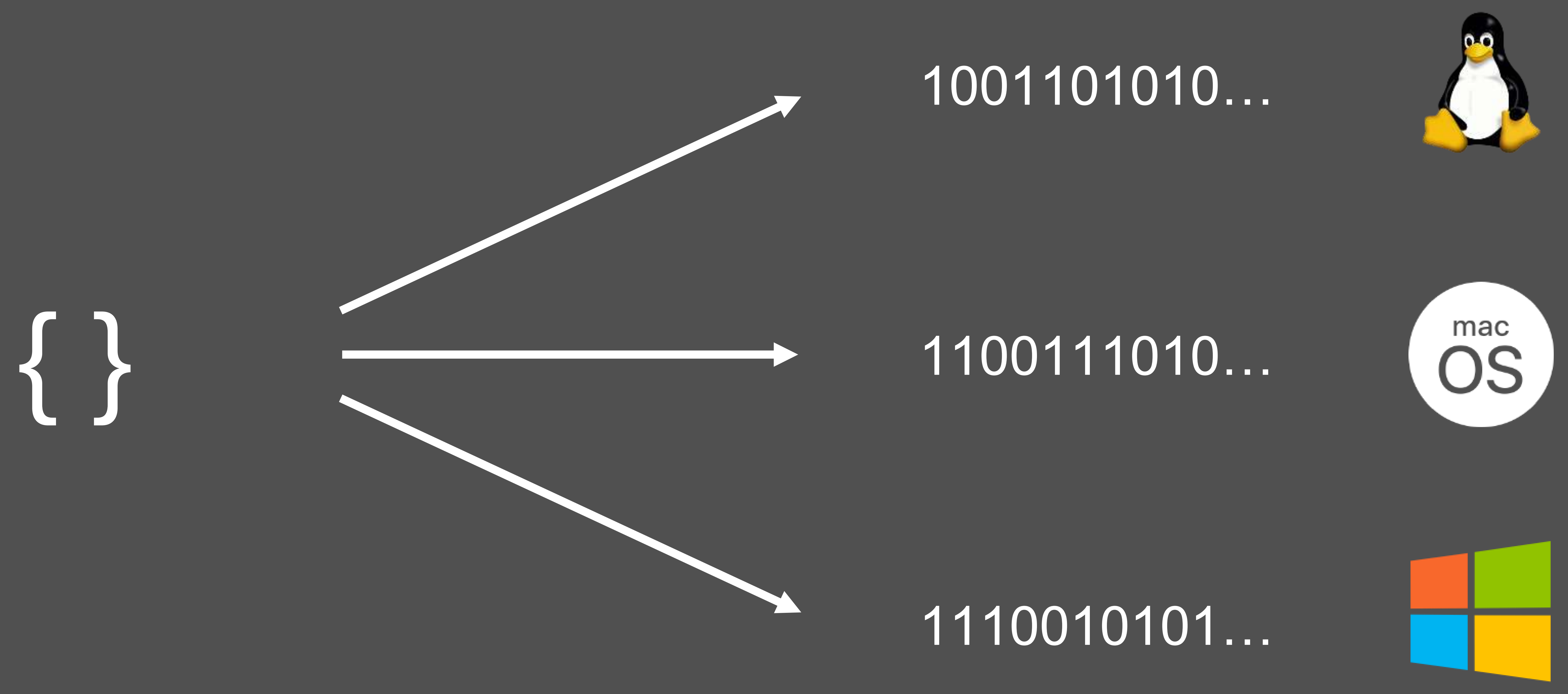
```c
/******************************************************************************/
EXPIMP_TEMPLATE INCHI_API
int INCHI_DECL MakeINCHIFromMolfileText( const char *moltext,
                                         char *szOptions,
                                         inchi_Output *result )
{
    int retcode = 0, retcode2 = 0;
    long num_inp = 0, num_err = 0;
    char szTitle[MAX_SDF_HEADER + MAX_SDF_VALUE + 256];

    STRUCT_FPTRS *pStructPtrs = NULL;    /* dummy in this context */
    INPUT_PARMS inp_parms;
    INPUT_PARMS *ip = &inp_parms;
    STRUCT_DATA struct_data;
    STRUCT_DATA *sd = &struct_data;
    ORIG_ATOM_DATA OrigAtData;
    ORIG_ATOM_DATA *orig_inp_data = &OrigAtData;
    ORIG_ATOM_DATA PrepAtData[2];
    ORIG_ATOM_DATA *prep_inp_data = PrepAtData;
    PINChI2     *pINChI[INCHI_NUM];
    PINChI_Aux2 *pINChI_Aux[INCHI_NUM];
    INCHI_IOSTREAM outputstr, logstr, prbstr, instr;
    INCHI_IOSTREAM *pout = &outputstr, *plog = &logstr, *pprb = &prbstr,
      *inp_file = &instr;
    int output_error_inchi = 0;
    int have_err_in_GetOneStructure = 0;

    INCHI_IOS_STRING temp_string_container;
    INCHI_IOS_STRING *strbuf = &temp_string_container;
```

- 165,000 lines of code
- Mostly C

Deployment



{ }

1001101010…

1100111010…

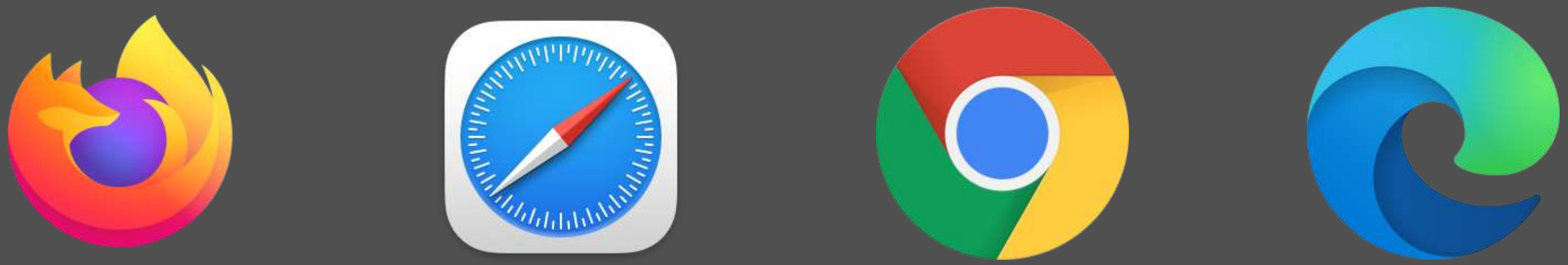
1110010101…

C Source

Build System

Native Executables

What About…

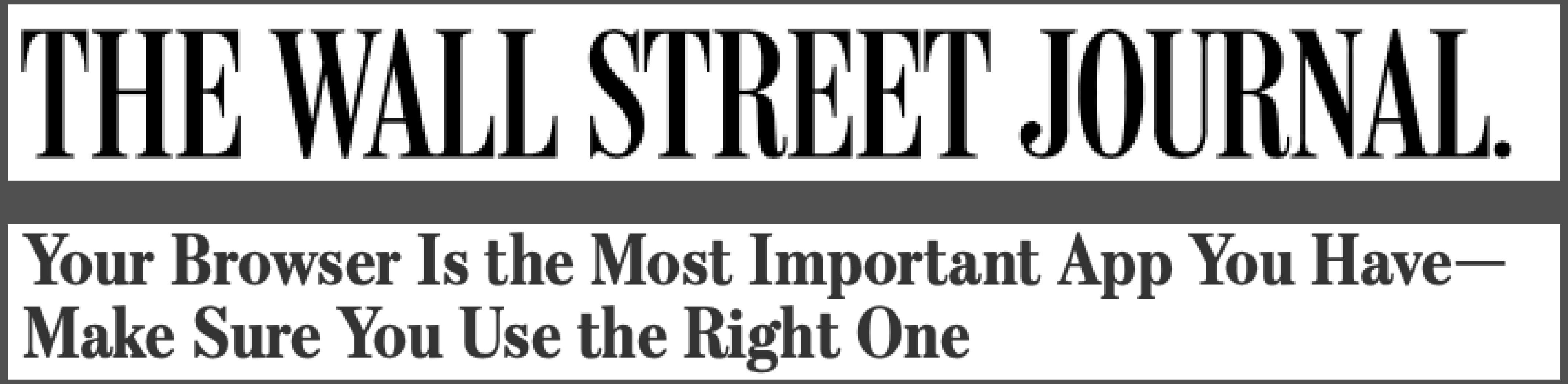


✓ Ubiquitous

X Native Code not Allowed!

Servers… Permission

The Most Important App



THE WALL STREET JOURNAL.

Your Browser Is the Most Important App You Have—
Make Sure You Use the Right One

https://www.wsj.com/articles/your-browser-is-the-most-important-app-you-havemake-sure-you-use-the-right-one-1537707600

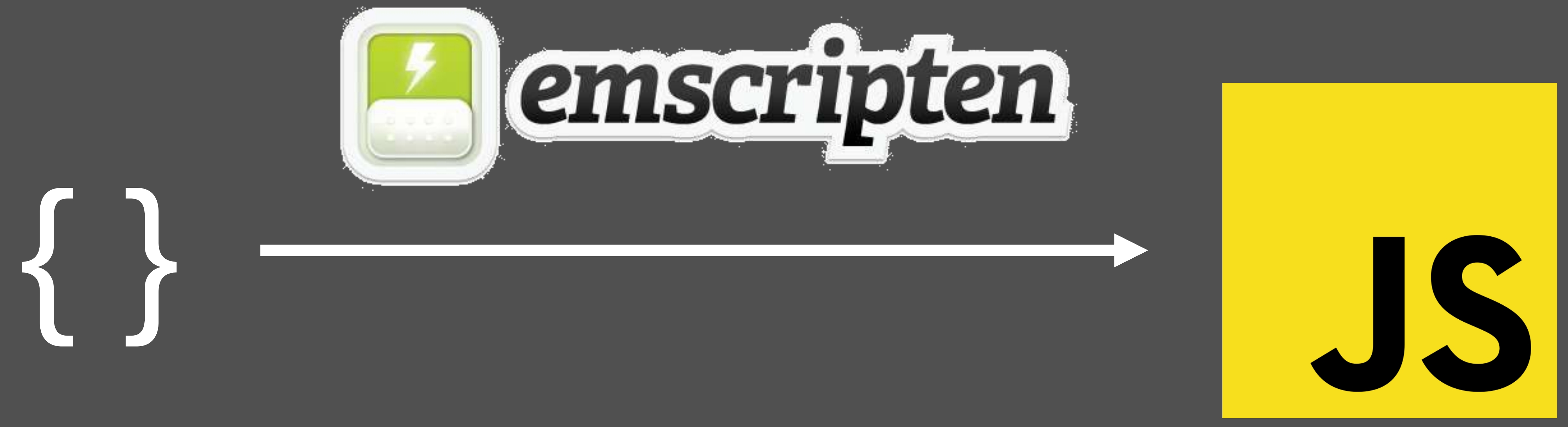
JavaScript

- Browser debut in 1994
- Universal browser support
- Pretty fast

InChI to JavaScript

Rewrite 160,000 lines of C to JS.

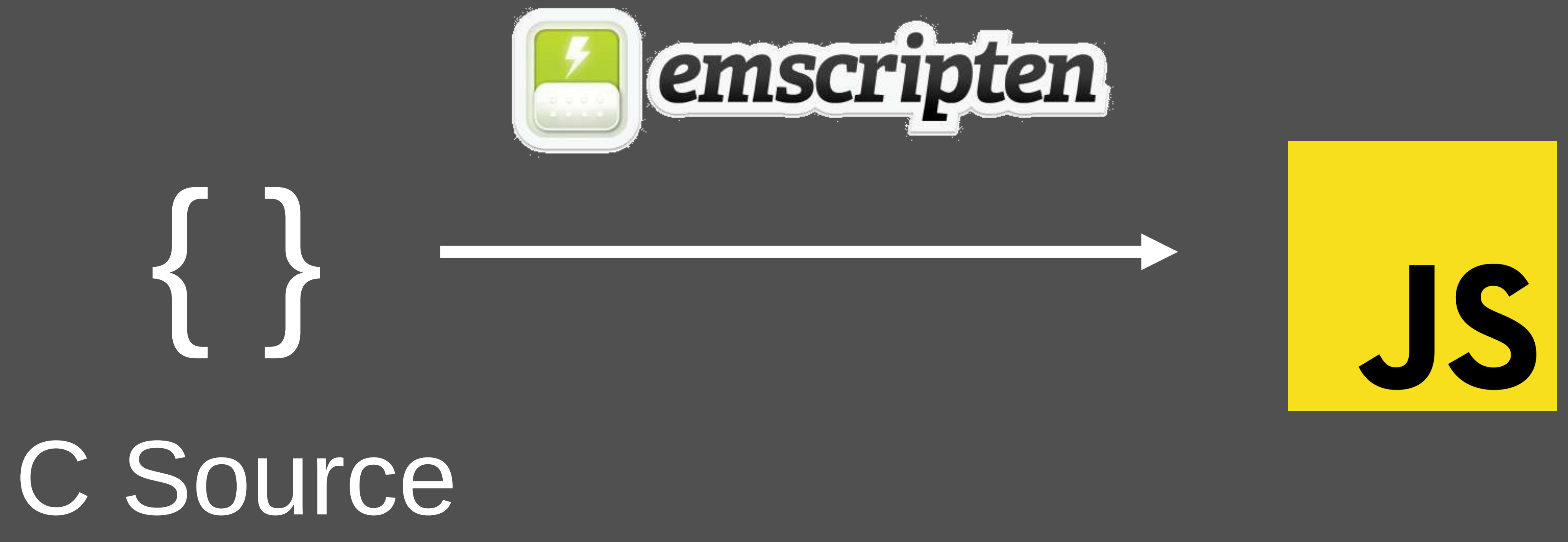
Transpile 160,000 lines of C to JS.



InChI
Source

https://github.com/metamolecular/inchi-js

JavaScript as Assembly Language



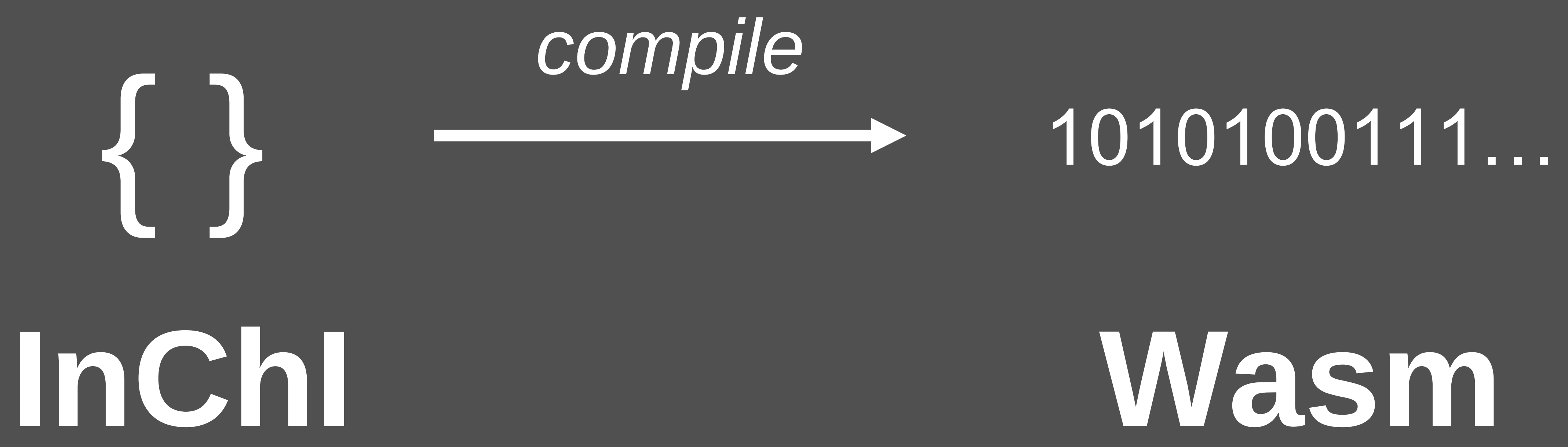
C Source

- Ad hoc solution
- Limited tooling
- Limited scope
- Gobs of glue code
- Not built for speed

WebAssembly (Wasm)

- Binary instruction format
- Fast, portable compile target
- Sandboxed for security
- Runs in all browsers, 2017
- W3C Standard, 2019
- **Runs InChI?**

Goals



{ }

compile

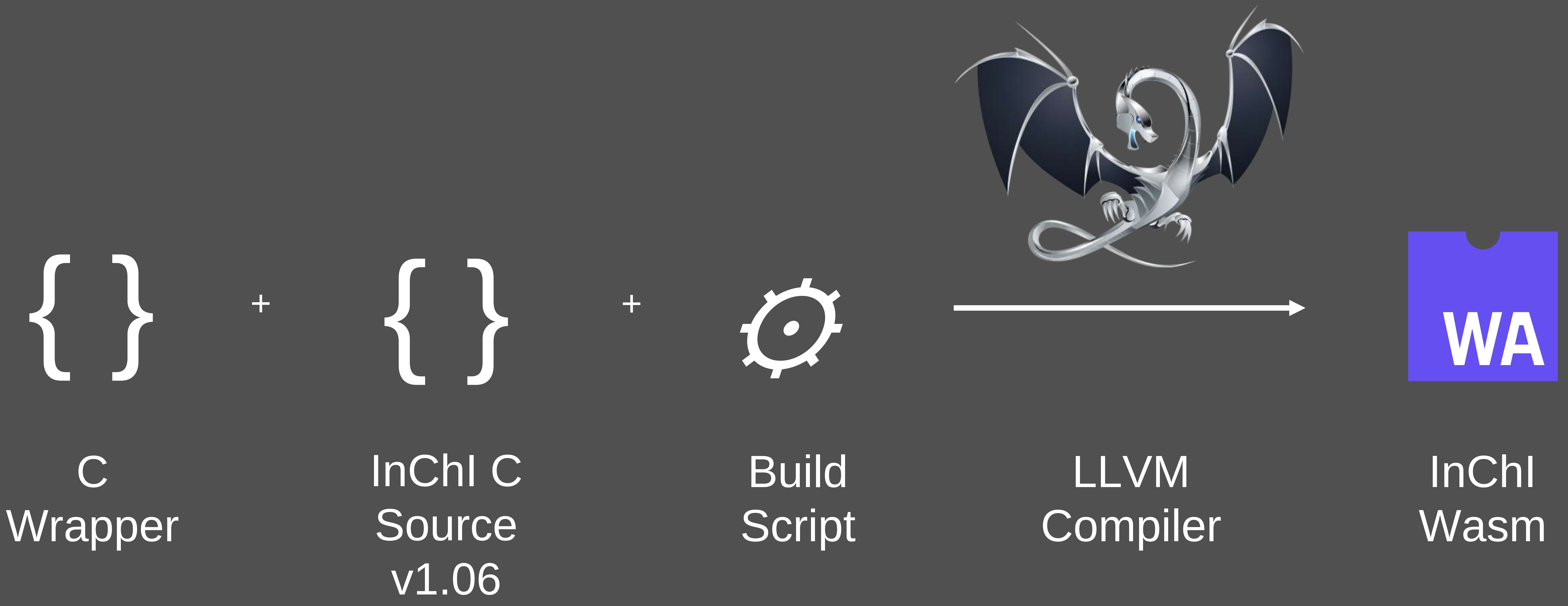
1010100111…

InChI

Wasm

- Minimal tooling
- No auto-generated glue code
- Use verbatim InChI source
- In other words, a build system

InChI-Wasm Project



C Wrapper + InChI C Source v1.06 + Build Script → LLVM Compiler → InChI Wasm

https://github.com/rapodaca/inchi-wasm

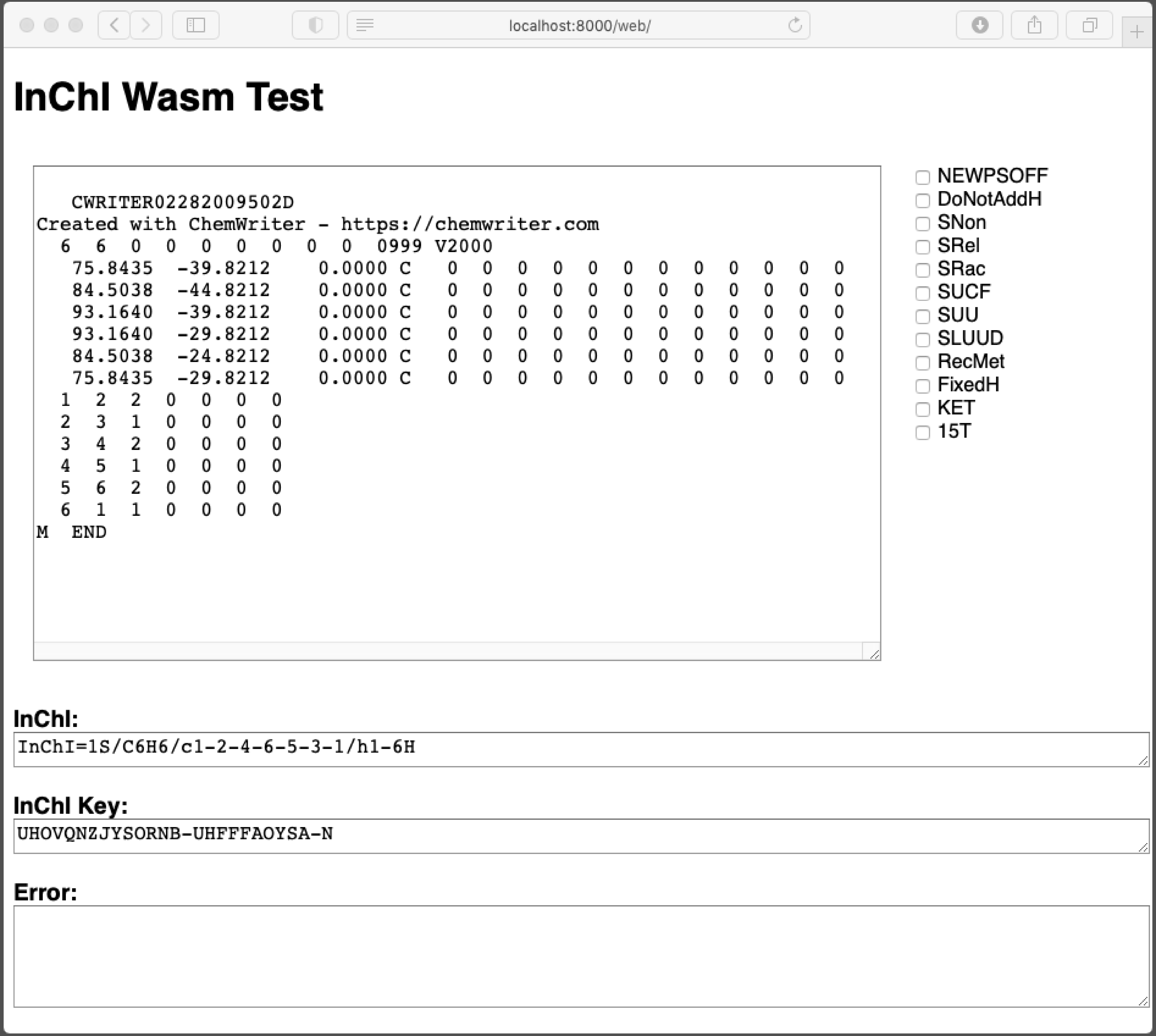
Application Stack

User Interface

JS-Wasm Interface (2 KB)

InChI Wasm (601 KB)

Standard Library (Static, 69 KB)



InChI Wasm Test

```

  CWRITER02282009502D
Created with ChemWriter - https://chemwriter.com
  6  6  0  0  0  0  0  0  0  0999 V2000
   75.8435  -39.8212    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   84.5038  -44.8212    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   93.1640  -39.8212    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   93.1640  -29.8212    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   84.5038  -24.8212    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   75.8435  -29.8212    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
  1  2  2  0  0  0  0
  2  3  1  0  0  0  0
  3  4  2  0  0  0  0
  4  5  1  0  0  0  0
  5  6  2  0  0  0  0
  6  1  1  0  0  0  0
M  END
```

☐ NEWPSOFF
☐ DoNotAddH
☐ SNon
☐ SRel
☐ SRac
☐ SUCF
☐ SUU
☐ SLUUD
☐ RecMet
☐ FixedH
☐ KET
☐ 15T

InChI:

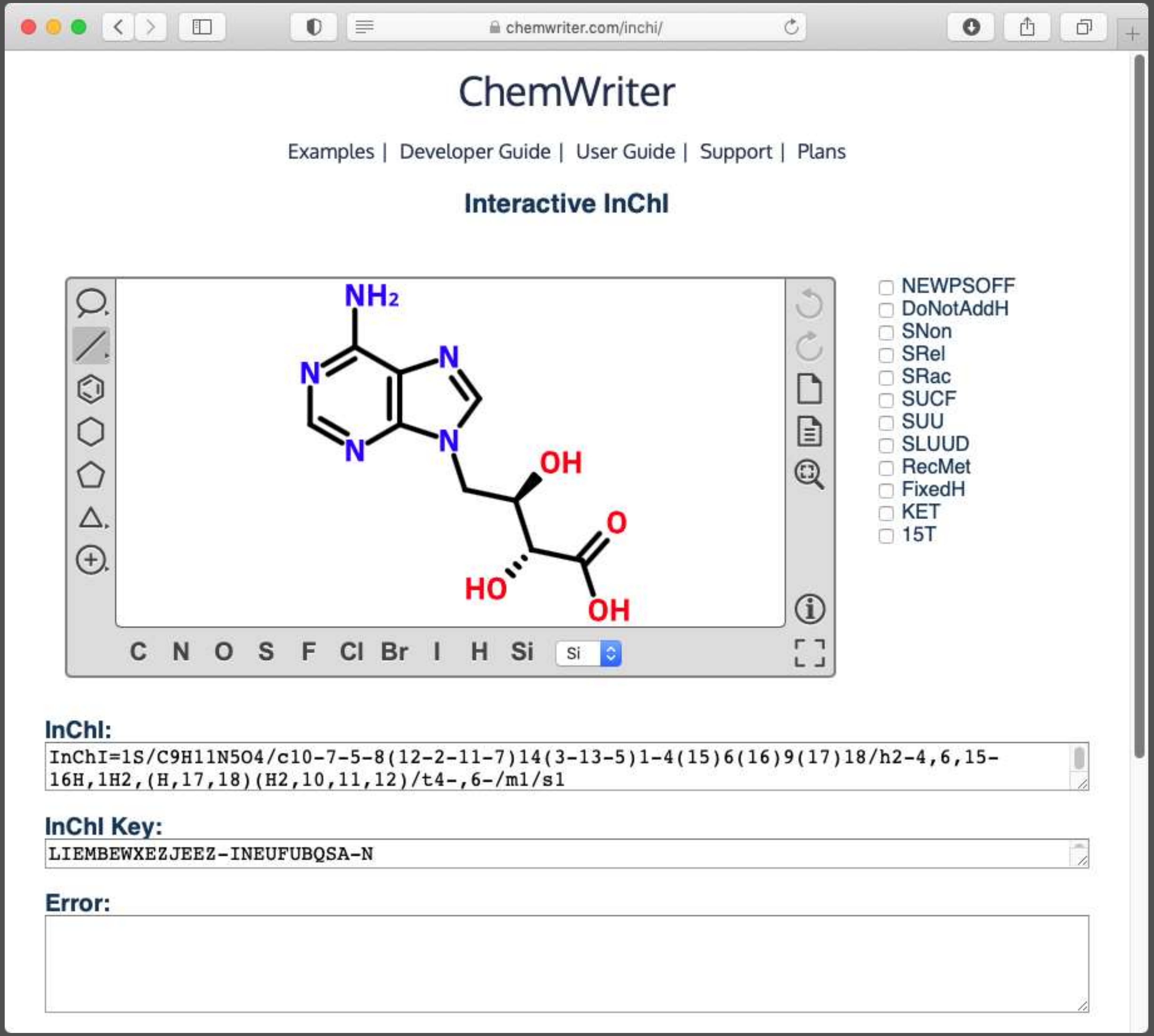
```
InChI=1S/C6H6/c1-2-4-6-5-3-1/h1-6H
```

InChI Key:

```
UHOVQNZJYSORNB-UHFFFAOYSA-N
```

Error:

chemwriter.com/inchi



ChemWriter

Examples | Developer Guide | User Guide | Support | Plans

Interactive InChI

- □ NEWPSOFF
- □ DoNotAddH
- □ SNon
- □ SRel
- □ SRac
- □ SUCF
- □ SUU
- □ SLUUD
- □ RecMet
- □ FixedH
- □ KET
- □ 15T

InChI:

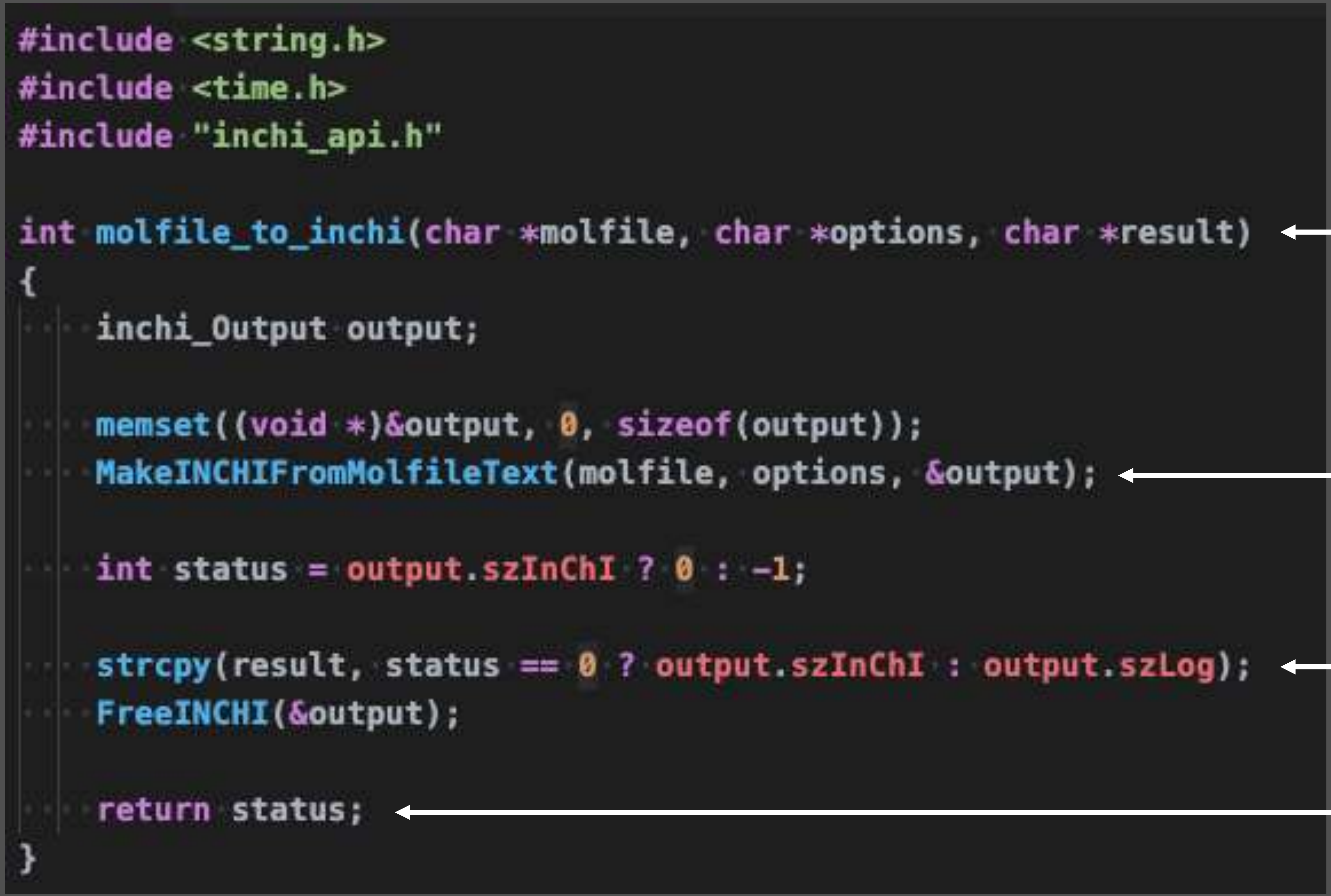
```
InChI=1S/C9H11N5O4/c10-7-5-8(12-2-11-7)14(3-13-5)1-4(15)6(16)9(17)18/h2-4,6,15-
16H,1H2,(H,17,18)(H2,10,11,12)/t4-,6-/m1/s1
```

InChI Key:

```
LIEMBEWXEZJEEZ-INEUFUBQSA-N
```

Error:

C Wrapper



```
#include <string.h>
#include <time.h>
#include "inchi_api.h"

int molfile_to_inchi(char *molfile, char *options, char *result)
{
    inchi_Output output;

    memset((void *)&output, 0, sizeof(output));
    MakeINCHIFromMolfileText(molfile, options, &output);

    int status = output.szInChI ? 0 : -1;

    strcpy(result, status == 0 ? output.szInChI : output.szLog);
    FreeINCHI(&output);

    return status;
}
```

Wasm interface

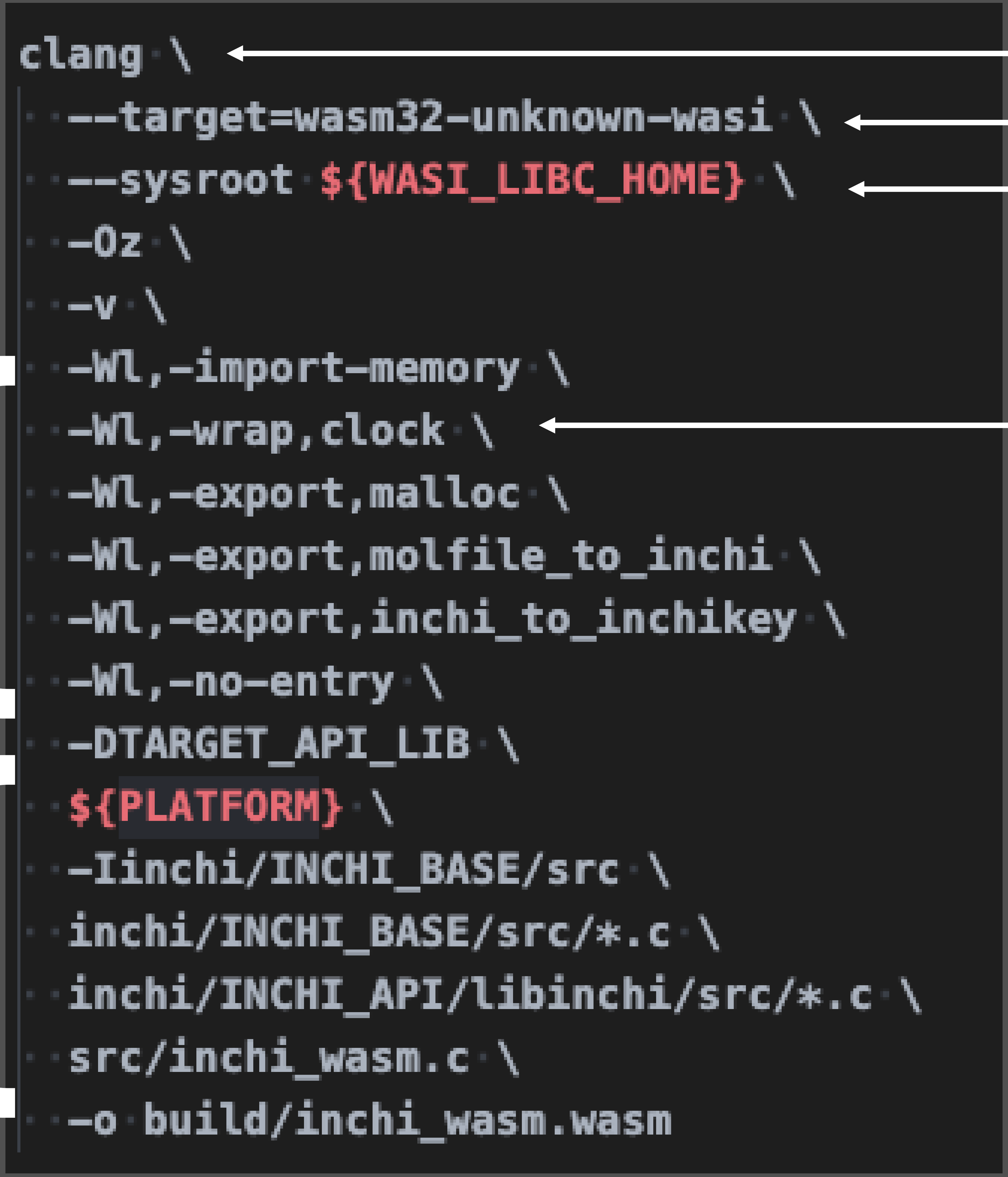
heavy lifting 💪

copy result (InChI or Error)

JS error reporting

Also, `inchi_to_inchikey` with help from Bob Hanson

Build Script



```
clang \
  --target=wasm32-unknown-wasi \
  --sysroot ${WASI_LIBC_HOME} \
  -Oz \
  -v \
  -Wl,-import-memory \
  -Wl,-wrap,clock \
  -Wl,-export,malloc \
  -Wl,-export,molfile_to_inchi \
  -Wl,-export,inchi_to_inchikey \
  -Wl,-no-entry \
  -DTARGET_API_LIB \
  ${PLATFORM} \
  -Iinchi/INCHI_BASE/src \
  inchi/INCHI_BASE/src/*.c \
  inchi/INCHI_API/libinchi/src/*.c \
  src/inchi_wasm.c \
  -o build/inchi_wasm.wasm
```

LLVM compiler

Wasm Target

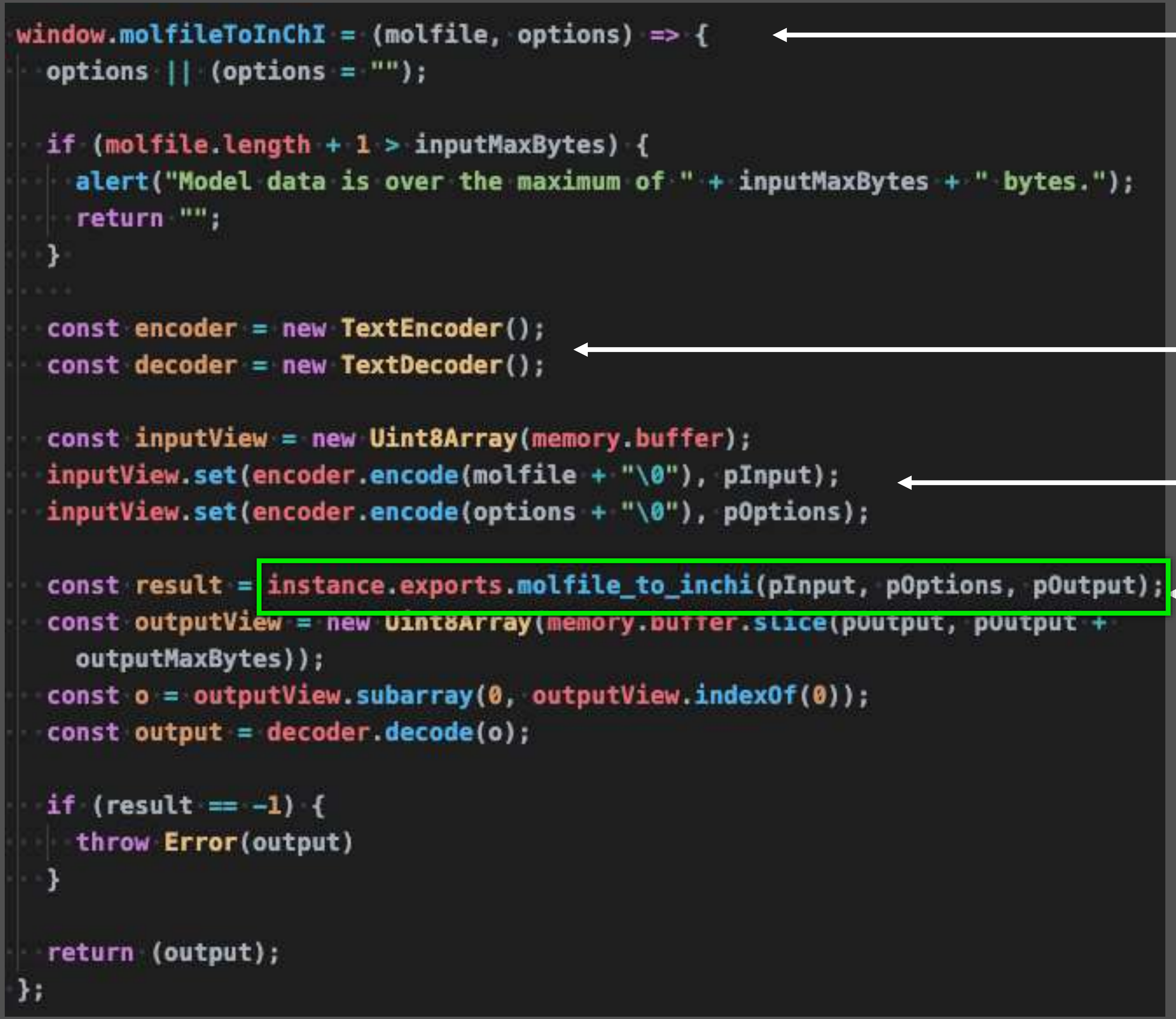
Standard Library

Obscure, hard-won hack

Wasm options

Standard build stuff

Data Schlepping in JS



```
window.molfileToInChI = (molfile, options) => {
  options || (options = "");

  if (molfile.length + 1 > inputMaxBytes) {
    alert("Model data is over the maximum of " + inputMaxBytes + " bytes.");
    return "";
  }

  const encoder = new TextEncoder();
  const decoder = new TextDecoder();

  const inputView = new Uint8Array(memory.buffer);
  inputView.set(encoder.encode(molfile + "\0"), pInput);
  inputView.set(encoder.encode(options + "\0"), pOptions);

  const result = instance.exports.molfile_to_inchi(pInput, pOptions, pOutput);
  const outputView = new Uint8Array(memory.buffer.slice(pOutput, pOutput +
    outputMaxBytes));
  const o = outputView.subarray(0, outputView.indexOf(0));
  const output = decoder.decode(o);

  if (result == -1) {
    throw Error(output)
  }

  return (output);
};
```

Inputs: molfile and options

Byte wranglers

Translate input

Call InChI Wrapper

Write output

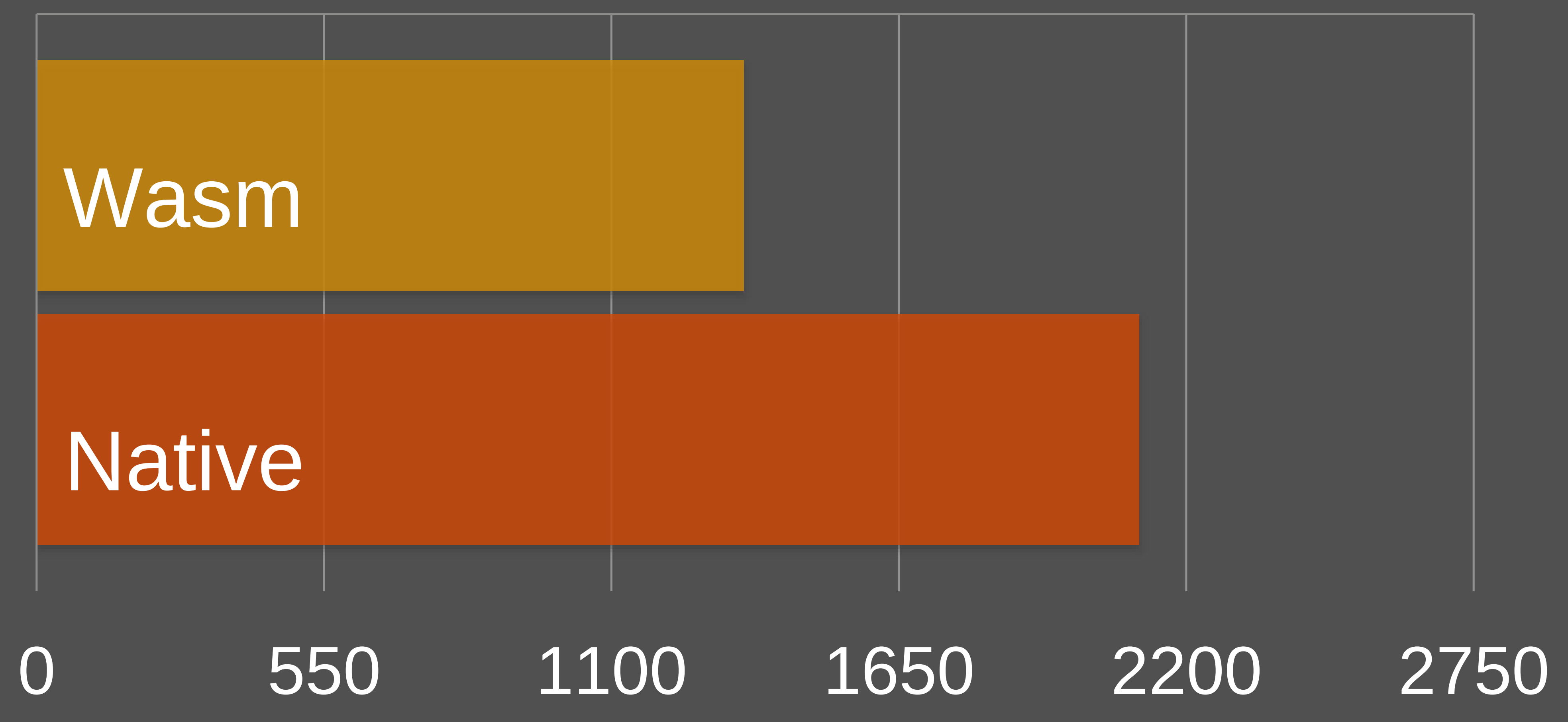
Return result

Performance

“WebAssembly **aims to execute at native speed** by taking advantage of common hardware capabilities available on a wide range of platforms.”

- webassembly.org

Benchmark

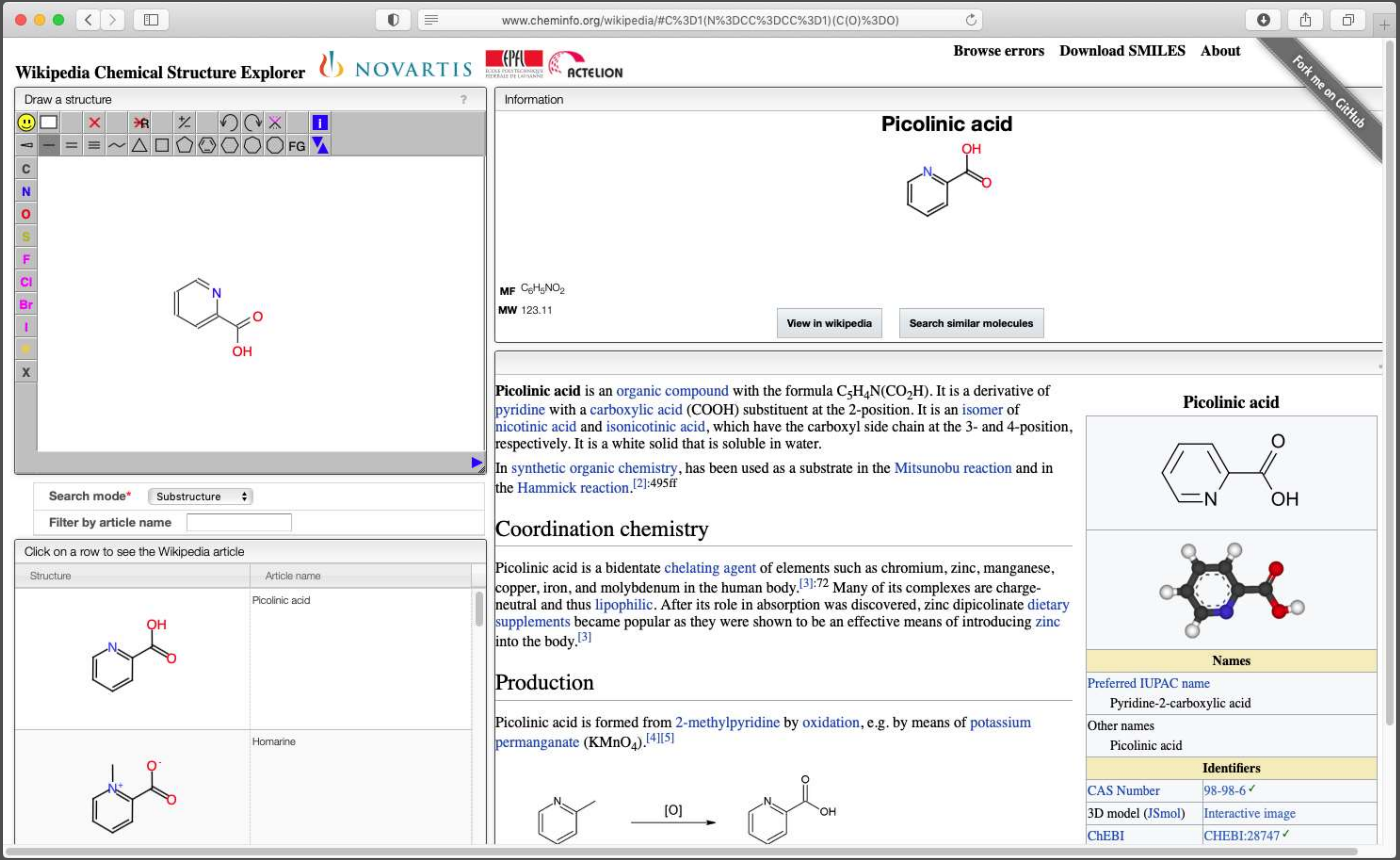


InChIs/sec

- 114 K SureChEMBL Update
- Test Wasm & native
- Same InChI output
- 155 ChEMBL discrepancies
- 3 Errors

https://github.com/rapodaca/inchi-benchmark

Rethinking the Web Browser for Chemistry



www.cheminfo.org/wikipedia/#C%3D1(N%3DCC%3DCC%3D1)(C(O)%3DO)

Wikipedia Chemical Structure Explorer

Browse errors Download SMILES About

Fork me on GitHub

Draw a structure

Search mode* Substructure

Filter by article name

Click on a row to see the Wikipedia article

Structure	Article name
	Picolinic acid
	Homarine

Information

Picolinic acid

MF $C_6H_5NO_2$

MW 123.11

View in wikipedia Search similar molecules

Picolinic acid is an organic compound with the formula $C_5H_4N(CO_2H)$. It is a derivative of pyridine with a carboxylic acid (COOH) substituent at the 2-position. It is an isomer of nicotinic acid and isonicotinic acid, which have the carboxyl side chain at the 3- and 4-position, respectively. It is a white solid that is soluble in water.

In synthetic organic chemistry, has been used as a substrate in the Mitsunobu reaction and in the Hammick reaction.[2]:495ff

Coordination chemistry

Picolinic acid is a bidentate chelating agent of elements such as chromium, zinc, manganese, copper, iron, and molybdenum in the human body.[3]:72 Many of its complexes are charge-neutral and thus lipophilic. After its role in absorption was discovered, zinc dipicolinate dietary supplements became popular as they were shown to be an effective means of introducing zinc into the body.[3]

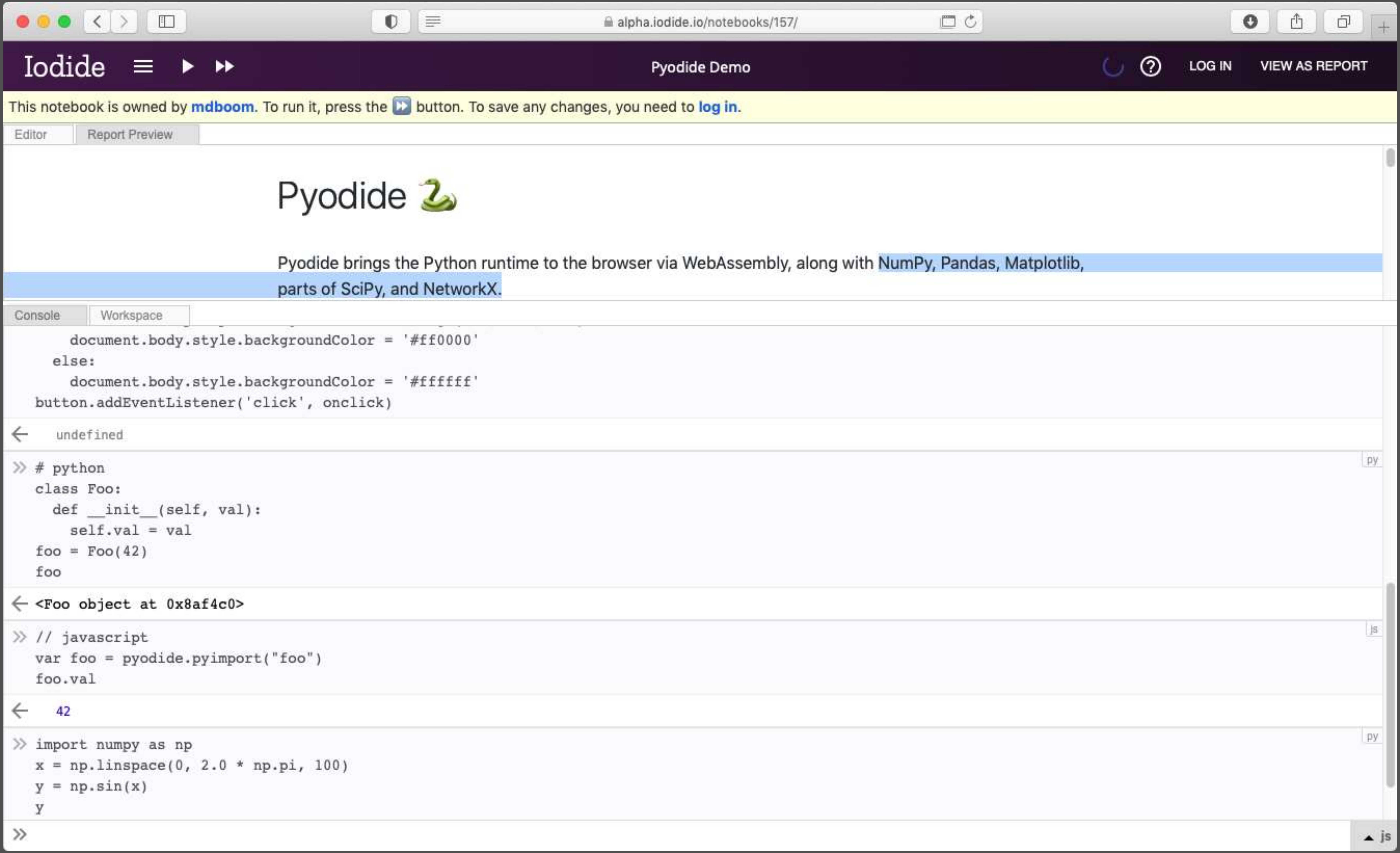
Production

Picolinic acid is formed from 2-methylpyridine by oxidation, e.g. by means of potassium permanganate ($KMnO_4$).[4][5]

Picolinic acid

Names	
Preferred IUPAC name	
Pyridine-2-carboxylic acid	
Other names	
Picolinic acid	
Identifiers	
CAS Number	98-98-6 ✓
3D model (JSmol)	Interactive image
ChEBI	CHEBI:28747 ✓

Ertl et. al. https://doi.org/10.1186/s13321-015-0061-y



This notebook is owned by mdboom. To run it, press the ⏩ button. To save any changes, you need to log in.

Pyodide 🐍

Pyodide brings the Python runtime to the browser via WebAssembly, along with NumPy, Pandas, Matplotlib, parts of SciPy, and NetworkX.

```
    document.body.style.backgroundColor = '#ff0000'
  else:
    document.body.style.backgroundColor = '#ffffff'
button.addEventListener('click', onclick)
```

```
←   undefined
```

```
» # python
  class Foo:
    def __init__(self, val):
      self.val = val
  foo = Foo(42)
  foo
```

```
← <Foo object at 0x8af4c0>
```

```
» // javascript
  var foo = pyodide.pyimport("foo")
  foo.val
```

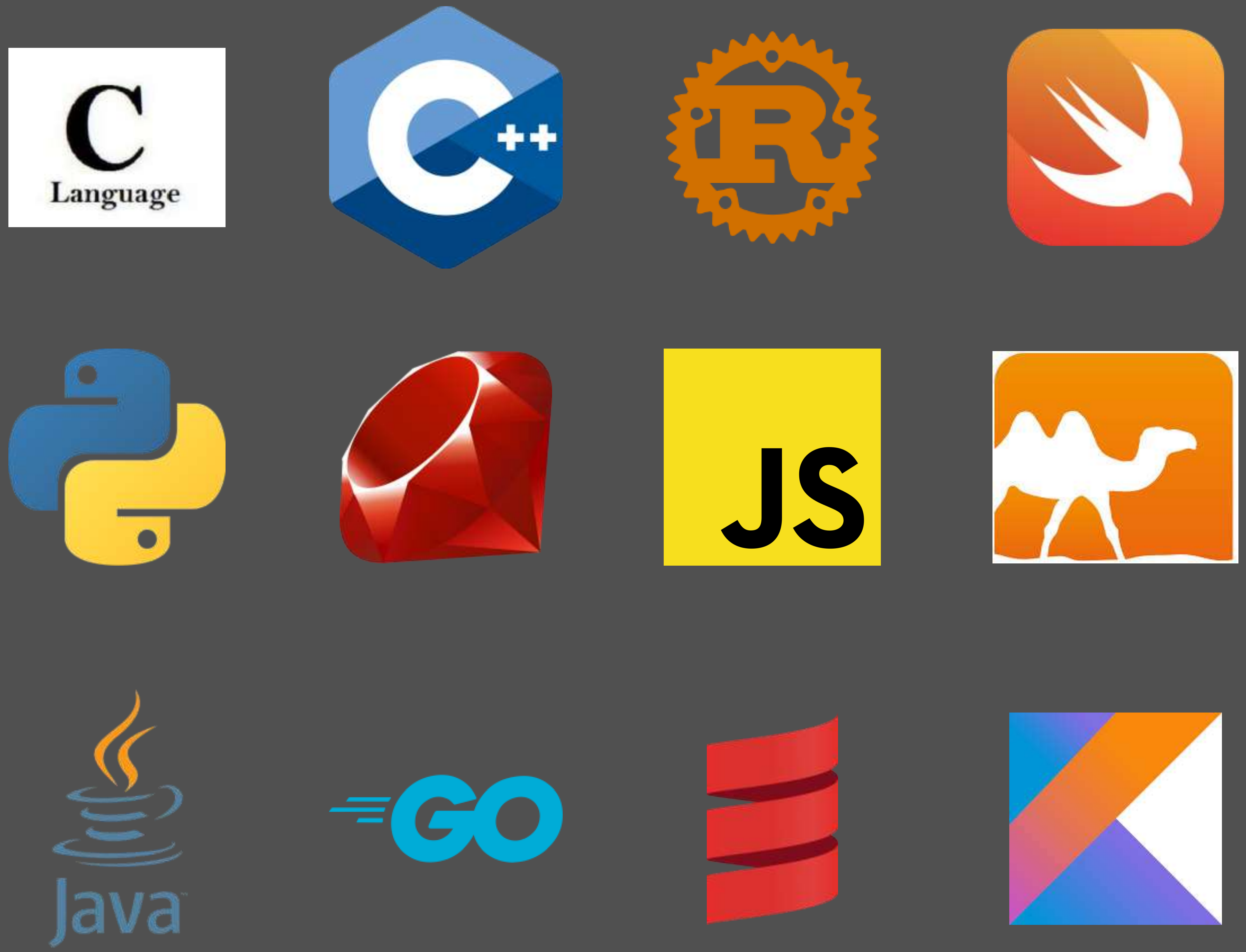
```
←   42
```

```
» import numpy as np
  x = np.linspace(0, 2.0 * np.pi, 100)
  y = np.sin(x)
  y
```

```
»
```

Pyodide Project https://github.com/iodide-project/pyodide

Beyond the Browser



A Template for Future Work

- Legacy code written in C, C++, or FORTRAN
- Incentive for in-browser and/or serverless deployment
- Compile to Wasm with LLVM
- User Interface in HTML5

https://metamolecular.com