InChI OUTREACH

Issue #2 – March 2022

INCHI OUTREACH newsletter distributes 4 issues per year. The goal is to share the latest information about the InChI Project, the InChI Trust and their usage. Please send any comments, questions, or suggestions to Outreach Director, Rudy Potenzone (rudy@inchi-trust.org).

What's New

Call for Papers -ACS Fall Meeting in Chicago Aug 21-25

We have a session: "InChI in the Wild: Celebrating Over 20 years of InChI Development in Memory of InChI Developer Igor Pletnev" at the ACS meeting.

We invite presentations that demonstrate practical application of InChI to enable the implementation of chemical information systems, the advancement of cheminformatics, interoperability of data across resources, and reliable communication of the outputs of chemistry.

There are two options to present your project:

1. Formal Presentation

You will need to submit your title and abstract to the ACS. Final abstracts are due March 14. https://acsnm264.abstractcentral.com/submission

2. "Interesting InChI Projects" Presentation

If your InChI project might not need a full presentation we are planning an "Interesting InChI Projects" session. If you could provide a short description, maybe a slide or three, we will include it in this session.

April/June InChI Meetings: Steve Heller, InChI Project Director, has been trying to organize the next face to face gathering of at least a few of the project teams for late April in Cambridge UK. Covid pandemic challenges for international travel as well as logistics and funding have made this most difficult to schedule. The plan now is to hold **two meetings**:

1. **Open InChI Days 2022** April 5th and 6th 2022, from 4pm to 7:30pm (GMT). Gerd Blanke (gerd@inchi-trust.org) and Jonathan Goodman are organizing this virtual meeting. Sign up HERE for more information.

2. **Face to face and virtual meeting on June 17-19**, just after the ICCS meeting in Noordwijkerhout (June 12-16). The session will be in Cambridge UK. Steve Heller is organizing this meeting and will have the first information packet for distribution shortly. Sign up HERE to get on the mailing list for that meeting.

Technology: Gerd has started assembling a working group to focus on v2 of the InChI core software. He is also coordinating with the other Working Groups.

Outreach: Rudy Potenzone (rudy@inchi-trust.org) has released a new version of the website homepage to bring more information and organization for visitors. This includes links to papers about InChI technology and projects using InChI for scientific studies. Updated information on the working groups is also included.

Join our *InChI OUTREACH* Mailing List here

Latest Publications & Presentations

The following papers were published recently

About the InChI code, InChI Key, Extensions, Technology

InChIs and Registry Numbers, Leigh, Jeffrey, J Cheminform 13 40 (2021). **DOI:** https://doi.org/10.1515/ci.2012.34.6.2

About Using the InChI code in Projects or Research

De Novo Molecular Design with Chemical Language Models, Grisoni, F., Schneider, G., Artificial Intelligence in Drug Design (2022).

A computer-aided drug design approach to discover tumour suppressor p53 protein activators for colorectal cancer therapy

Rui P.S. Patrício, Paula A. Videirab, Florbela Pereira

Bioorganic & Medicinal Chemistry 53, 116530 (2022)

Data Centric Molecular Analysis and Evaluation of Hepatocellular Carcinoma Therapeutics Using Machine Intelligence-Based Tools

Rengul Cetin-Atalay; Deniz Cansen Kahraman; Esra Nalbat; Ahmet Sureyya Rifaioglu; Ahmet Atakan; Ataberk Donmez; Heval Atas; M. Volkan Atalay; Aybar C. Acar; Tunca Do?an

J Gastrointestinal Cancer (2021)

Using deep neural networks to explore chemical space, Martin Vogt, Expert Opinion on Drug Disc (2021).

A chemical kinetic mechanism for combustion and flame propagation of CH2F2/O2/N2 mixtures, Donald R. Burgess Jr., Valeri I. Babushok, Jeffrey A. Manion, Int. J. Chem. Kinet. 61 (2021) DOI: https://doi.org/10.1002/kin.21549.

Working Groups

Nanomolecules

Group Leader: Iseult Lynch

This project proposes to identify a specific nanomaterial/nanoform through an extension of the IUPAC InChI called NInChI. Their approach is to encode information (composition, size, shape and surface chemistry) required to unambiguously identify these materials. As nanomaterials are particulates, there is an intimate relationship between the core and surface compositions that challenges traditional naming conventions, thereby limiting information exchange among scientists, computational modelers, regulatory agencies and industry. Leveraging best practices with other InChI working groups e.g. Mixture InChI, Reaction InChI & Polymer InChI is planned.

The Working Group has submitted a proposal to the InChI subcommittee for review.

Organometallics/Inorganics

Group Leader Organometallics: Hinnerk Rey, with Sonja Herres-Pawlis

The Working Group is currently working slowly towards a proposal to investigate how the disconnection and reconnection code within InChI affects a set of organometallic compounds.

IUPAC Project No.: 2009-040-2-800

https://iupac.org/project/2009-040-2-800

Overview:

To establish requirements for extending the applicability of the IUPAC International Chemical Identifier to organometallic and coordination compound structures, so that developers contracted to the InChI Trust can enhance the InChI software accordingly.

Inorganic compounds are being handled similarly to organometallics in the InChI framework.

Extended Stereochemistry

Group Leader: Andrey Yerin

The project is making gradual progress.

The following tasks have been already investigated with possible solutions for further considerations by the InChI Subcommittee and InChI developers:

1. Enhanced stereo marks;

2. Atropisomers;

3. Support of longer allenes

4. Additional tetrahedral cases, including special case of stereo at spiro atoms (*last part is already implemented in InChI 1.06*).

The WG is expected to produce recommendations to resolve problems with recognition of configurations of carbohydrates represented in Haworth and chair forms.

UPAC Project No.:2019-017-2-800

https://iupac.org/project/2019-017-2-800

Extended Tautomers

Group Leader: <u>Marc C. Nicklaus</u>

IUPAC Project No.: 2012-023-2-800

https://iupac.org/project/2012-023-2-800

90 tautomeric transforms were identified and six could be incorporated in an experimental version of the InChI code (based on InChI version 1.06). Initial tests of these six new rules plus the existing non-standard tautomeric InChI options KET and 15T have been performed. See the report "Tautomers in InChI" <link broken for now> presented at the March 2021 NIH InChI Workshop <workshop link>.

One step that still needs to be performed before a version of the InChI tool with the six new rules incorporated can be offered as an experimental release to the public is a round-tripping test of the nature struct -> InChI -> struct' -> InChI', with comparison of InChI' vs. InChI.

The papers forming the scientific background of this project have been published (https://doi.org/10.1021/acs.jcim.9b01080 and https://doi.org/10.1021/acs.jcim.9b01156).

Mixtures (MInChI)

Group Leader: Leah McEwen

IUPAC Project No.: 2015-025-4-800

https://iupac.org/project/2015-025-4-800

The MInChI notation is tailored for a primary use case to describe what is definitively reported for source components at the time of mixing. The initial phase is focused on

components with molecular structures that produce a well defined standard InChI identifier. Separate, discrete components allows for easy association of arbitrary concentrations. Other types of components may be considered for future phases of the project. While the core InChI identifier lends some canonical characteristics to the notation, for most use cases it is not appropriate to assume that two equivalent mixtures will have an identical MInChI description. The definition of a mixture is not a discrete concept as concentrations always have uncertainty associated with them (whether specified or not) and there are frequently unknown components (which may or may not be anticipated).

Detailed project information, specifications, presentations, links, etc. are available: https://github.com/IUPAC/MInChI

Reactions

Group Leader: Gerd Blanke

IUPAC Project No.: 2009-043-2-800

https://iupac.org/project/2009-043-2-800

Overview:

The group is now working on v2 to extend the machine-readable, indexable and searchable representation of chemical reactions already built into the IUPAC International Chemical Identifier (InChI).

InChI Reactions already has implemented a) Atom mapping for reactions; b) Stereochemistry representation; c) Representation of failing reactions +

Currently under discussion are:

● Handling of reaction conditions (ProcAuxInfo)

● New format has been developed based on JSON to make the procauxinfo easier to handle within data workflows e.g., in the context of machine learning.

● First draft is undergoing discussion with academic and industry members working on reaction optimization, predictions, and automation.

Large Molecules

Group Leader: Evan Bolton

IUPAC Project No.: 2013-010-1-800

https://iupac.org/project/2013-010-1-800

The objective for this project is to establish requirements and guidelines for the generation of a unique name for biological sequences including chemically modified. The intended outcome is the world-wide adoption of a standard with the ensuing more comprehensive and more accurate handling of biological structures that have been chemically modified in, e.g., databases, appropriate books, and journal articles etc.

A presentation titled 'Progress towards "Large Molecule" support within InChI' was given at the March 2021 NIH Workshop on InChI by Evan Bolton.

QR Codes

Group Leader: Richard M. Hartshorn

This project is complementary to another user-focused project that is developing InChI for states and mixtures. That project is about the specificity and usefulness of the information being encoded in the InChI, while this one will focus on deployment and use at scale with key stakeholders globally.

A proof-of-principle implementation was made available: www-rinchi.ch.cam.ac.uk/qrinchi. No feedback has been received.

A revised draft of the InChI QR code specification document, formulated as an IUPAC Recommendation, was prepared and sent out for approval. The recommendation was then revised and submitted to PAC where it is currently under review.

IUPAC Project No.: 2015-019-2-800

https://iupac.org/project/2015-019-2-800

Education/Academic Training

Group Leader: Robert E. Belford

The InChI OER is an Open Education Resource (OER) devoted to the use of InChI. Chemical nomenclature underpins chemical communication and with its release in 2005, the InChI initiative supports the advancement of chemical nomenclature into the digital age. InChI is evolving to handle reactions, mixtures and other needs of 21st century scientific communication, and yet there is little educational material available on the use of InChI.

The OER initiative provides a resource on InChI related resources to assist practicing scientists and educators in learning about and benefiting from the use of InChI. Each post has a short description and an information box that includes various download options. OER posts include educational material uploaded to this site. https://inchi-trust.org/oer/

IUPAC Project No.: 2018-012-3-024 https://iupac.org/project/2018-012-3-024

Monomer Atoms

Group Leader: Evan Bolton

Overview:

Efforts towards polymer handling have involved working with the Pistoia Alliance HELM project on a core set of monomers for amino acids, nucleic acids, and glycans. This harmonization of core monomers will help to enable use cases for biopolymer handling (e.g., input) using a pseudo-atom approach that will sit on top of InChI using the Zz atom.

Variability/Positional Isomers

Group Leader: Jonathan Goodman

Overview:

How can an InChI provide a canonical encoding of variable structures? The working group is assembling test cases, working out what would be useful, and beginning to focus on the issues that need to be solved to get an effective variable InChI. An ordered list of InChI addresses many of the criteria but requires processes to generate it from information sources and will be unwieldy for the very long lists of molecules which are important use cases. Various ways of summarizing and compressing are being considered.

Isotopologues

Group Leader: Hunter Moseley

COVID 19 has delayed most of the efforts. The working group will work on the isotopologue InChI manuscript sometime in the spring semester. It proposes enhanced specifications within the regular InChI standard for representing isotopologues and isotopomers. More specifically, augment the isotopic layer specifications of the regular InChI standard so that specific isotopologues, isotopomers, partial isotopomers, and isotopologue fragments can be represented by a single InChI string and used to identify isotope-informative analytical features.

Resolver

Group Leader: Markus Sitzmann

The InChI Resolver Protocol is a common API protocol which allows any InChI Resolver instance to be findable and browsable for an (automated) client system in a systematic, predefined manner. It will also allow for the implementation of a network of federated InChI Resolvers.

Benefits:

- web resources providing data & information based on InChI become findable & interlinked on the web
- their content is accessible in a generic way
- (long-term) an autonomous client can find and collect information based in InC

Open-Source Development Environment

Group Leader: Gerd Blanke

Having just started, Gerd is in the early stages of assembling this team to review the requirements for an open-source code repository. This must not only support the complex and diverse development community but also protects the InChI assets through the InChI Trust. With a better understanding of the requirements, the group will consider the various options available in terms of languages, development tools and repositories, etc. Given these options, they can then review financial aspects of the various options and work with the InChI supporters and InChI Trust to make recommendations.

Anyone interested in getting involved with this effort should reach out to Gerd directly (gerd@inchi-trust.org)

InChI OUTREACH

Links to InChI Trust social media:

Twitter   YouTube   LinkedIn

Subscribe to InChI OUTREACH