

The Status of the IUPAC InChI Chemical Structure Standard

Today and the Future

by Ray Boucher, Stephen Heller, and Alan McNaught

For almost 100 years, IUPAC has been well known around the world for its efforts in standardizing nomenclature in chemistry. At the start of the present century, it became clear to all involved in chemical structure representation work that, with the extensive use of computers and electronic information in all aspects of chemistry and related sciences, an IUPAC standard was necessary. From this critical need, the IUPAC International Chemical Identifier—InChI—project was launched in cooperation with the US standards agency NIST. The result of this effort has been the development, maintenance, and expansion of capabilities of the open source nonproprietary International Chemical Identifier (InChI), first by NIST and now by the InChI Trust, a not-for-profit UK charity. Over 100 chemical information specialists and computational chemists volunteer to examine and test the software before a public release; this optimal quality control by a world-wide user community has led to improvements to and releases of the software with very few problems. The reliance on input from many volunteers enables the project staff to be restricted to two part-time contractors, a project director and a programmer, thus minimizing the running costs of the Trust.

This brief discussion of InChI will highlight ongoing efforts to strengthen and extend this standard for chemical structures and its hashed form, the InChIKey. Information standards are critical to enable effective and efficient communication of scientific content. Validation and reproducibility of research results are critical to advances in science. Without a chemical structure standard, it was becoming impossible to find and share all the reported results needed for a particular purpose. The costs of experiments are ever increasing, hence the need for increased efficiency in labs around the world. Open Access, Open Data, and Open Standards are areas that are expanding rapidly and are facilitating faster and more effective research discovery. However, before you can share data about a chemical, you need to find where the information has been made available on the Internet. Collaborative, interoperable, and global dissemination standards are essential in a more networked world.

The InChI is an open-source, widely adopted standard found in most chemical information-containing databases, including those from Chemical Abstracts,

Reaxys, ChemSpider, ChEMBL, Open PHACTS, PubChem, DrugBank, PDB, SigmaAldrich, and many others, including internal Chemical and Pharma corporate databases. The InChI distills diverse chemical representations into a single form, and is a string that enables easier linking and integration of scientific content, especially with printed and electronic data sources. For example, one can easily look up chemical structures in internet search engines such as Google, Bing, and Yahoo using an InChIKey. InChI is continually and actively being extended to increase its applicability and usability. The initial version, released in 2009, was able to handle almost 99 % of the chemicals which scientists are concerned with every day. Additional work is underway to improve the treatment of inorganics and organometallics and to handle biopolymers and their positional isomers and chemical mixtures. The latest release, in 2017 (version 1.05), added polymer support and multithreading capabilities. InChI has also been incorporated in a chemical reaction identifier (RInChI) and its use in labeling via QR codes is being explored (see *Chem Int* Nov 2016, p. 22; <https://doi.org/10.1515/ci-2016-0616>). Funding to maintain and enhance InChI comes from most major chemistry publishers (CAS, Elsevier, Wiley, RSC, Springer-Nature, Taylor & Francis) and database and chemical suppliers and providers (Sigma-Aldrich, ChemAxon, BioRad, ACD/Labs, OpenEye, RELX group), as well as from US governmental agencies (NIH, FDA, and NIST). This funding helps to ensure that the future development of InChI meets the needs of the scientific community and it also helps to support other potential avenues for its use.

InChI is a valuable addition to other compound identifiers (e.g., systematic and trivial names, registry numbers, and various versions of SMILES) in a database; it is not intended to be a replacement. With the implementation of the ISO identification of medicinal products (IDMP) and the related ISO 11238 standards, adding and having an InChI will allow for an easier, effective, and more complete search for information on a particular chemical, be it a drug, a pollutant, or a chemical for other commercial and/or noncommercial use.

Details on the project can be found at <https://www.iupac.org/inchi> and <http://www.inchitrust.org>

For a recent view paper about InChI, documenting its design, layout and algorithms, see Stephen R. Heller, et al, InChI the IUPAC International Chemical Identifier, *Journal of Cheminformatics* 2015, 7:23 (30 May 2015), <https://doi.org/10.1186/s13321-015-0068-4>

The authors are all members of the InChI Trust, Cambridge, UK.