InChI for inorganic chemists: organometallics and molecular inorganics

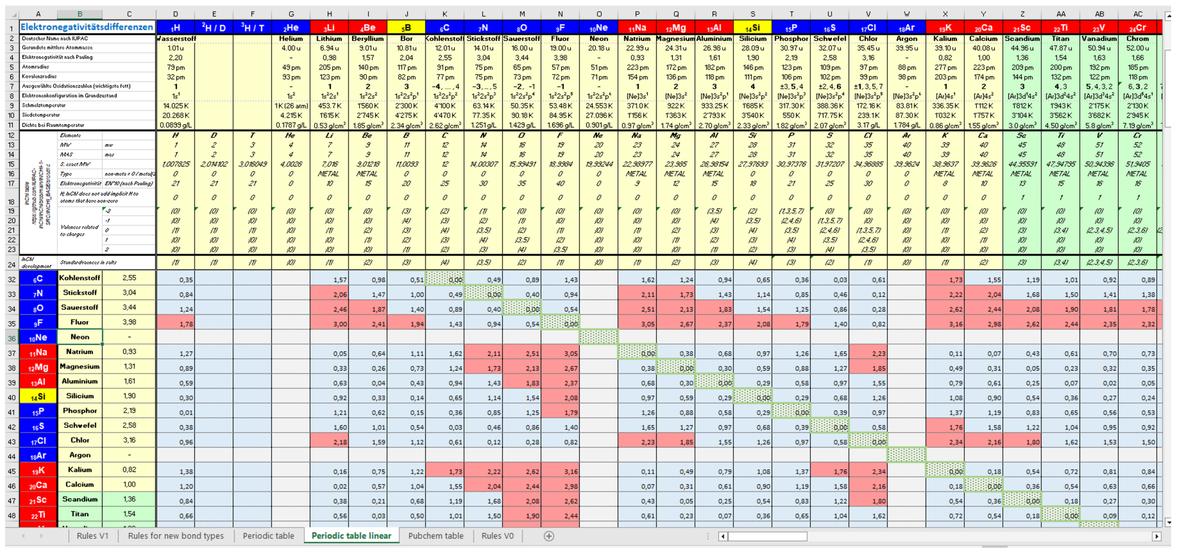
Gerd Blanke[1], Sonja H. Herres-Pawlis[2], Jonathan M. Goodman[3], Ulrich Schatzschneider[4], Andrey Yerin[5], Richard Hartshorn[6], Felix Bänsch[7], Nauman Khan[2], Djordje Baljozovic[2], Jan Brammer [2]

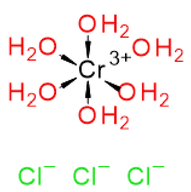
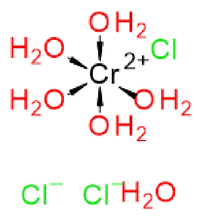
1. StructurePendium Technologies GmbH, Essen, Germany; 2. Institute of Inorganic Chemistry, RWTH Aachen University, Aachen, Germany; 3. Yusuf Hamied Department of Chemistry, University of Cambridge, CB2 1EW, UK; 4. Institute of Inorganic Chemistry, Julius Maximilians University Würzburg, Germany; 5. Advanced Chemistry Development Inc; 6. University of Canterbury, New Zeeland; 7. Beilstein Institut, Frankfurt(Main), Germany

The standard InChI, v1.07, was developed with organic molecules as the focus. The algorithm generates InChI for organometallics which can be helpful as identifiers in an organic chemistry context, but which neglect important aspects of these structures. Filling these gaps is a priority. For example, the NFDI4Chem initiative (www.nfdi4chem.de) will use InChI for molecular inorganic compounds as well as organic ones. A better description of organometallics is needed.

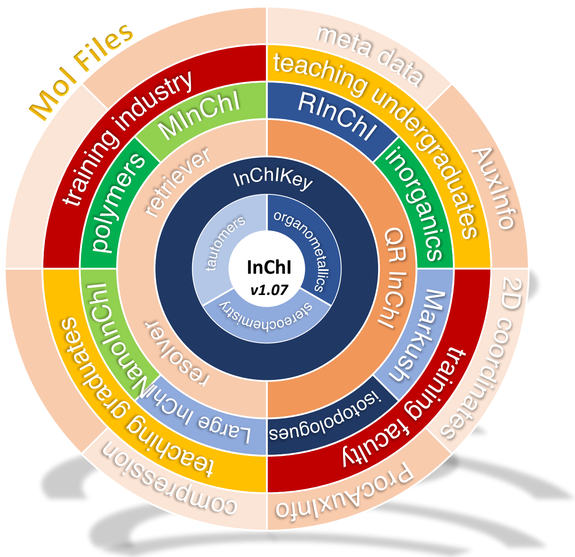
A revision of the InChI is being developed which includes more information about the organometallics and molecular inorganics. This requires an improved bond recognition and a more sophisticated stereochemical model.

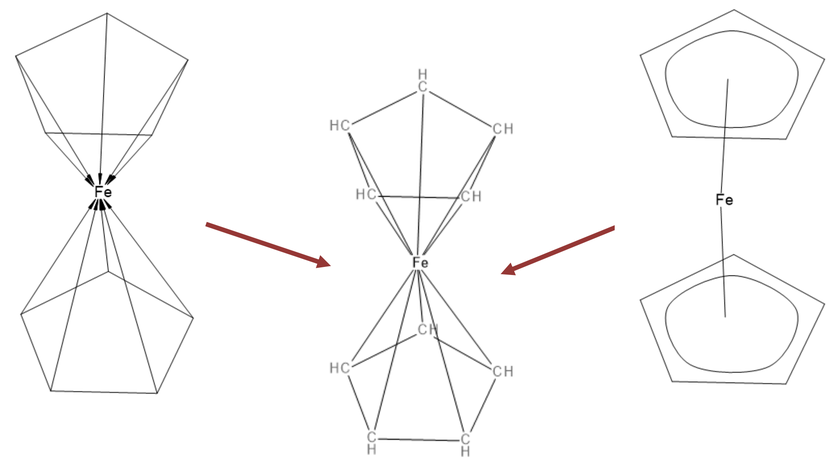
Disconnection Check Table



Depiction	Current standard InChI (1.07)	Developing standard InChI
	InChI=1S/3ClH.Cr.6H2O/h3*1H;;6*1H2/q;;;+3;;;;;;/p-3	InChI=1S/ClCrH10O5.2ClH.H2O/c1-2(3,4,5,6)7;;;/h3-7H2;2*1H;1H2/q+2;;;/p-2
		InChI=1S/3ClH.Cr.6H2O/h3*1H;;6*1H2/q;;;+3;;;;;;/p-3

A compound can only be identified by one InChI but may be represented by different structure diagrams. The new InChI program will interpret diverse diagrams of inorganics and organometallics. Structure depictions of the same compound may especially differ in the bond types used for the chemical representation. Beside single, double, and triple bonds, other bond types like coordinative or haptic bonds of the input structure depictions must be processed accordingly to ensure the creation of unique InChIs. Hydrogen atoms directly bound to metal atoms must always be drawn explicitly. Guidance on constructing accurate diagrams will be available.



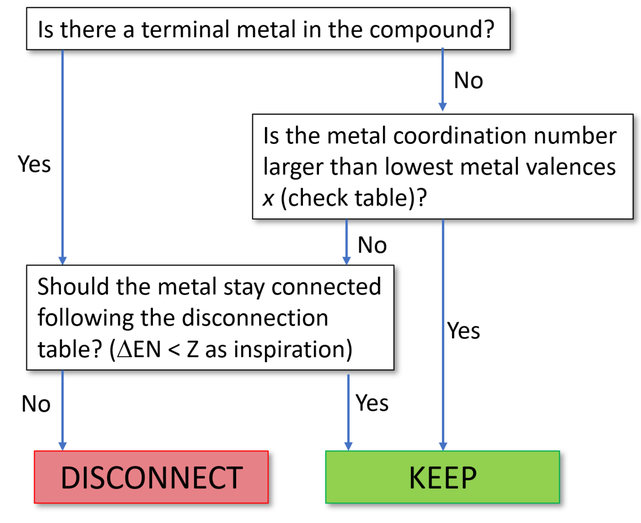


InChI=1/2C5H5.Fe/c2*1-2-4-5-3-1;/h2*1-5H;/**rC10H10Fe /c1-2-4-5-3(1)11(1,2,4,5)6-7(11)9(11)10(11)8(6)11/h1-10H**

The standard InChI, v1.07, breaks all bonds to metal atoms; the new standard InChI will be more selective. A data-driven approach will be used to decide whether bonds to metals should be retained.

Join the discussion on: https://github.com/IUPA-InChI/InChI/issues

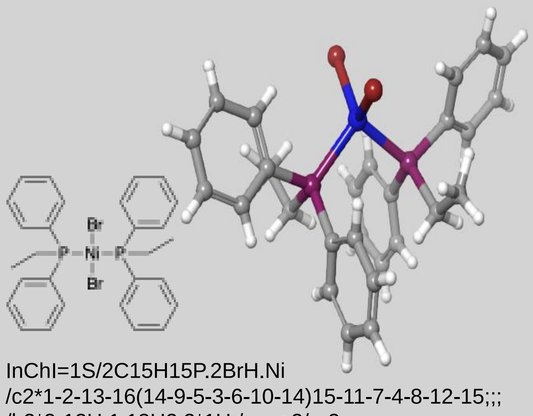
Preprocessing flow chart for inorganic compounds



To be noted:
- This flow chart has to be run for every metal.
- The „Disconnection Table“ step has to be run for every bond to a metal.
- If atoms are not connected in the input, they stay disconnected. No new bonds are formed that are not already in the input.

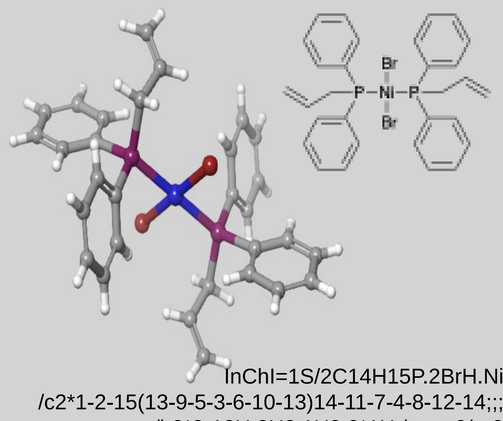
The stereochemistry recognition of InChI must be extended to the more complex relationships in inorganic compounds. While tetrahedral stereocentres may lead up to 2n stereoisomers (with n = number of stereocentres), just one octahedral centre leads up to 30 isomers.

Tetrahedral
No stereoisomers



InChI=1S/2C15H15P.2BrH.Ni/c2*1-2-13-16(14-9-5-3-6-10-14)15-11-7-4-8-12-15;;;/h2*2-12H,1,13H2;2*1H;/q;;;;+2/p-2

Square planar
Possibility of *cis* / *trans* stereoisomers



InChI=1S/2C14H15P.2BrH.Ni/c2*1-2-15(13-9-5-3-6-10-13)14-11-7-4-8-12-14;;;/h2*3-12H,2H2,1H3;2*1H;/q;;;;+2/p-2

CSD refcode: XIKWEX, M.L.Clapson, D.J.Nelson, M.W.Drover (2023) *ACS Org. Inorg. Au*,**3**,217

Test the InChI developments

References:

1.InChI Source Code: https://github.com/IUPAC-InChI/InChI
2.Making the InChI FAIR and sustainable by moving to open-source on GitHub: https://chemrxiv.org/engage/chemrxiv/article-details/66a9584301103d79c547b086