

www-jmg.ch.cam.ac.uk/inchi

InChI canonical representifier for molecules **connecting** people

Cambridge virtual meeting April 2021



Educational Resources Bob Belford Large Molecules Evan Bolton Nanomaterials Iseult Lynch Markush/Variable Jonathan Goodman **Resolver** Markus Sitzmann Stereochemistry Andrey Erin **Organometallics** Colin Batchelor Mixtures Leah McEwen **Reactions** Gerd Blanke Tautomers Marc Nicklaus **QR codes** Richard Hartshorn

Extensions must play nicely

Be careful with layer labels: /b /c /f /h /I /m /o /p /q /s /t /z are already spoken for

Molecule separators: & or !

Component separators: <> <M> <V> <N> [] should allow nesting

Educational Resources Bob Belford Large Molecules Evan Bolton Nanomaterials Iseult Lynch Markush/Variable Jonathan Goodman **Resolver** Markus Sitzmann Stereochemistry Andrey Erin **Organometallics** Colin Batchelor Mixtures Leah McEwen **Reactions** Gerd Blanke Tautomers Marc Nicklaus **QR codes** Richard Hartshorn



Educational Resources Bob Belford Large Molecules Evan Bolton Nanomaterials Iseult Lynch Markush/Variable Jonathan Goodman **Resolver** Markus Sitzmann Stereochemistry Andrey Erin **Organometallics** Colin Batchelor Mixtures Leah McEwen **Reactions** Gerd Blanke Tautomers Marc Nicklaus **QR codes** Richard Hartshorn



Educational Resources By Belford

Large Molecules Evan Bolton

Nanomaterials Iseult Lynch

Markush/Variable Jonathan Goodman

Resolver Markus Sitzmann

Stereochemistry Andrey Erin

Organometallics Colin Batchelor

Mixtures Leah McEwen

Reactions Gerd Blanke

Tautomers Marc Nicklaus

QR codes Richard Hartshorn

Education/Training Resources



Open Meeting April 16, 5:00 PM UTC

13 people attended the open meeting

Two Major Outcomes

- Need to Redesign OER Landing Page
- Need for Training Materials



InChI OER



- tutorial on how to use OER
- a base introduction to InChI and use cases
 - Ehren Bucholtz @ UHSP will lead student project



InChI Trust Open Education Resource (InChI-OER)

The InChI OER is an Open Education Resource (OER) devoted to the use of InChI, the International Chemical Identifier. Chemical nomenclature underpins chemical communication and with its release in 2005, the InChI initiative of the International Union of Pure and Applied Chemistry (IUPAC) 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. This OER initiative, <u>IUPAC project 2018-012-3-024</u>, is being created to provide a resource on InChI related resources to assist practicing scientists and educators in learning about and benefiting from the use of InChI.

This resource contains a tag taxonomy-based filter that defaults to Open Access content, which can be further refined by clicking multiple tags. Each post has a short description and an information box that include various download options. OER posts include educational content that has been uploaded to the site along with open access off site content. All material uploaded to this site is open access, and by clicking "non OER" content you can extend the filter to include information on off-site non open access material.

Search	PUBLISHED	<u>TITLE/LINK</u>	CONTENT TYPE	
	11/29/20	Molecule2PubChemLCSS	OER	
	07/16/20	InChI: Measuring the Molecules	OER	
Content Types	02/12/20	Enhanced recognition and encoding of stereoconfiguration by InChI tools	Non OER	
	12/16/16	Identifying International Chemical Identifier (InChI) Enhancements—QR Codes and Industry Applications	OER	
OER Post	ost 01/01/14 Implementation of InChI for chemically modified large biomolecules			
Non OER		InChI Videos Launched	OER	
InChl Tags		IUPAC and the InChI Trust Agree Upon Conditions for Collaboration	OER	
		IUPAC InChI/InChIKey Project Joins Microsoft BioIT Alliance	OER	
InChI Algorithm and Description	01/01/18	InChI'ng forward: Community Engagement in IUPAC's Digital Chemical identifier	OER	
-Algorithm	03/22/13	International chemical identifier for chemical reactions	OER	
Toolkits Resolvers InChI Application -Data Extraction -Search	Current Pag	e=1 Previous 1 2 3 4 5 6 7 8 9 10 11 Next Last	Total Pages= 11	



InChI OER

- Need for Training Material
- Possible project for developing training material for use of InChI in Patent Searches
- Possible project for developing training material for use of InChI in Environmental Health and Safety



- New undergraduate course in Cheminformatics for Fall of 2021
- Will expand on material regarding use and applications of InChI



Educational Material Development within

LibreText HyperLibrary



Educational Resources Bob Belford

Large Molecules Evan Bo on

Nanomaterials Iseult Lynch

Markush/Variable Jonathan Goodman

Resolver Markus Sitzmann

Stereochemistry Andrey Erin

Organometallics Colin Batchelor

Mixtures Leah McEwen

Reactions Gerd Blanke

Tautomers Marc Nicklaus

QR codes Richard Hartshorn



Educational Pesources Bob Belford

Large Molecules Evan By on

Nanomanenais Iseult Lynch

How big is a large molecule?

The transition from not-large to large needs to be watched

The structure of large molecules may not be precisely defined

Canonicalisation is important

Need a specification



Educational Resources Bob Belford

Large Molecules Evan Bolton

Nanomaterials Iseult Lv h

Warkush variable Jonathan Goodman

Resolver Markus Sitzmann

Stereochemistry Andrey Erin

Organometallics Colin Batchelor

Mixtures Leah McEwen

Reactions Gerd Blanke

Tautomers Marc Nicklaus

QR codes Richard Hartshorn





InChl for Nanomaterials (NInChl)

Prof. Iseult Lynch, University of BirminghamDr. Antreas Afantitis, Novamechanics Ltd.Dr. Thomas Exner, Seven Past Nine

Nano & advanced materials VAST space - need pragmatic approach



Next steps for NanoInChI working group

- 1. Select some datasets of nanomaterials families with different specifications and challenges to work through what is essential to the NanoInChl itself versus what can/should be in AuxInfo file; e.g.,
 - Various complexities of ligands on Au NMs reactions to form the ligands (thiol functionalisation), grafting reactions to attach the ligands, ligand density (low or high), etc. Graphene oxides - edges, ends, oxidation states of the Oxygens, distribution and destiny of
 - oxygens etc.
- 1. Define what options exist from other InChIWGs
 - a. e.g. polymer people's use of Anchor points easier for grafted polymers then physisorbed
 b. How to specify amounts of core / shell etc. relative to one another (do we care if < 1% of the material for
 - example)....
 - c. Covalent versus non-covalent bonds between layers / components
- 1. Establish strong collaboration with MixInChI and RxnInChI in terms of SUVC and JSON for example

Joint meeting to be organised ASAP

1. Next NanolnChl meeting: Wednesday 12th May 2021, 4-6 pm UK time



Educational Resources Bob Belford

Large Molecules Evan Bolton

Nanomatorials Iseult Lynch

Markush/Variable Jone In Goodman

Resolver Iviarkus Sitzmann

Stereochemistry Andrey Erin

Organometallics Colin Batchelor

Mixtures Leah McEwen

Reactions Gerd Blanke

Tautomers Marc Nicklaus

QR codes Richard Hartshorn

Canonical Identifier for Variability: Markush and other variations



Canonical Identifier for Variability: Markush and other variations





Educational Resources Bob Belford Large Molecules Evan Bolton Nanomaterials Iseult Lynch Markuch (Variable Jonathan Goodman Resolver Markus Sit lann Stereochemistry Andrey Erin **Organometallics** Colin Batchelor Mixtures Leah McEwen **Reactions** Gerd Blanke Tautomers Marc Nicklaus **QR codes** Richard Hartshorn

InChI Resolver Project – Current Status

InChI Resolver Protocol specification (JSON:API)

Search or jump to	7 Pull requests Issues Marketplace Exp	lore	+• 🜮		
Inchiresolver / inchires	olver	⊕ Watch 👻	0 🖧 Star 1 😵 Fork 0		
↔ Code ① Issues 5	1 Pull requests 🕑 Actions 🛄 Projects 1 🕮	Wiki 🕛 Security 🗠 Insig	nts 🐵 Settings		
🐉 master 👻 🐉 5 branches	© 4 tags	Add file ▼ 💆 Code ▼	About 🕸		
🔅 markussitzmann project file	🗴 markussitzmann project file updates, adding logo files 🗸 cd7da17 8 days ago 🕥 112 commits				
📄 .circleci	fixing circleci config	17 months ago			
appsite	project file updates, adding logo files	8 days ago			
📄 client	first release preparations	13 months ago	A RED D Classe L'annua		
docs	project file updates, adding logo files	8 days ago	Me BSD-3-Clause License		
nginx nginx	minor bug fixes	8 months ago	D. L		
.gitignore	testing API behaviour	15 months ago	Releases 4		
Dockerfile	project file updates, adding logo files	8 days ago	on 22 Oct 2020		
LICENSE	license and readme	2 years ago	+ 3 releases		
README.rst	documentation improvements	5 months ago			
🗅 build	recreated project on Chembience 0.2.12	2 years ago	Packages		
Chombionco diango onv	project file updates, adding logo files	one such 8	No packages published		

InChI Resolver Reference Implementation

downloadable software & online



https://github.com/inchiresolver/inchiresolver



InChI Resolver Project – Outlook

- more documentation (including use cases)
- addition of more data & resources (to the online reference service, e.g. linking Chemical Identifier Resolver, PubChem, ChEMBL, and CAS Common Chemistry)
- definition of additional InChI Resolver resources (MInChI, RInChI, ...)
- representation of the InChI Resolver protocol as a single schema file
- implementation of a InChI Resolver reference client
- how to run the system (in the future)?

Search or jump to	7 Pull requests	ssues Marketplace Explore	+• 😍		
🖵 inchiresolver / inchir	esolver	۲	Watch + 0 1 1 Fork 0		
↔ Code ① Issues 5	11 Pull requests ③ Actions	III Projects 1 III Wiki 🛈 Security	└─ Insights 영 Settings		
양 master - 양 5 brand	hes 🛭 🟷 4 tags	Go to file Add file - Code	e - About 🕸		
🔅 markussitzmann project	Repository of the InChI Resolver project - linking InChI related data and resources on the Internet				
.circleci	fixing circleci config	17 months a	ago		
appsite appsite	project file updates, adding	logo files 8 days a	ago		
🖿 client	first release preparations	13 months a	ago		
docs	project file updates, adding	logo files 8 days a	ago		
nginx	minor bug fixes	8 months a	ago Balances d		
🗅 .gitignore	testing API behaviour	15 months a	ago		
Dockerfile	project file updates, adding	logo files 8 days a	ago on 22 Oct 2020		
LICENSE	license and readme	2 years a	ago + 3 releases		
README.rst	documentation improveme	nts 5 months a			
🗅 build	recreated project on Chem	pience 0.2.12 2 years a	2 years ago Packages		
🕒 chombionco diango onv	project file undator, adding	logo filor 9 days :	No packages published		

https://github.com/inchiresolver/inchiresolver





Educational Resources Bob Belford

Large Molecules Evan Bolton

Nanomaterials Iseult Lynch

Markush/Variable Jonathan Goodman

Resolver Markus Sitzmann

Stereochemistry Andre Erin

Organometanics Colin Batchelor

Mixtures Leah McEwen

Reactions Gerd Blanke

Tautomers Marc Nicklaus

QR codes Richard Hartshorn



InChI stereo – good, but still not enough

InChI works well with traditional stereogenic units of organic structures: tetrahedral centers, double bonds, chirality axis in allenes.

Still there are some areas where further work is necessary. Two situations are possible for problems with recognition of stereoconfigurations:

- 1. Not recognized configurations different substances are treated as the same;
- 2. Not recognized equivalence same substances are treated as different.

IUPAC project 2019-017-2-800 approved in September 2019 Enhanced Recognition and Encoding of Stereoconfiguration by InChI Tools

Task group members:

Andrey Yerin, Gerd Blanke, Igor Pletnev, Burt Leland, Jürgen Kammere, Clare Tovee, Yulia Borodina, Jane Frommer



InChI stereo project tasks

Specific tasks in the scope of the task group

- 1. Not recognized configurations:
 - atropisomers;
 - long cumulenes only cases with two and three double bonds are recognized;
 - non-tetrahedral configuration and tetrahedral in coordination structures;
- 2. Not recognized equivalence same substances are treated as different:
 - Haworth and chair representations of carbohydrates;
 - Erroneous configurations for wrong or ambiguous representations.
- **3.** Borderline case with 'random' outcome:

Not supported MOL V3000 enhanced definition of configuration. This representation is not so frequent yet, but already accepted by industry and is requested by RInChI group.



InChI stereo project – current results and plans

What is ready for implementation

Atropisomers; Longer allenes – extension to 10 double bonds; Enhanced stereo marks;

Also, some additional tetrahedral cases are already fixed in 1.06

What to do next

Correct recognition of carbohydrate stereo in Haworth and chair representation; Principles of recognition and encoding of nontetrahedral configurations (in collaboration with InChI organometallics project);

Improvements for specific arrangements and designations.

We also need to work with IUPAC Technical report on enhanced stereo representation to make it officially recognized by IUPAC.

For the completed tasks, the corresponding concepts and principles are defined in the documents assuming further consideration by InChI subcommittee and developers.



Educational Resources Bob Belford Large Molecules Evan Bolton

Nanomaterials Iseult Lynch

Markush/Variable Jonathan Goodman

Resolver Markus Sitzmann

Stereochemistry Andrey Erin

Organometallics Colin tchelor

Mixtures Lean McEwen

Reactions Gerd Blanke

Tautomers Marc Nicklaus

QR codes Richard Hartshorn

Organometallics (status as of 2021-04-26)

- Alex Clark's work, both code and processed data, has clarified matters considerably, but there are still unresolved questions.
- Specifically, we are yet to explore the full effects of the normalization and disconnection stages in the InChI algorithm on the compounds we are interested in.
- To explore this will require a developer with detailed knowledge of the workings of the InChI code.
- The developer would be asked to produce experimental code which has the option to normalise the structure prior to and/or after metal disconnection, and an option to provide a coordination layer.
- Igor Pletnev has had a preliminary look at this suggestion and describes it as "attractive and [...], in principle, feasible". He also has a lot of questions.
- The working group will then examine different InChIs produced by this code and Alex Clark's work for structures from CCDC, Elsevier and the FDA.
- The open meeting last week revealed a wide variety of views on the problems that need to be solved. We need to explain more clearly what we are doing.
- Next meeting TBD. Agenda items: Igor's questions and review of equivalent structures.



Mixtures InChI (MInChI) – Resources

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

github.com/IUPAC/MInChI

- Spec doc (DRAFT)
- Validation data (JSON, preliminary set, 150 diverse examples, mixfile format)
- DEMO (181 example mixtures, interactive, open source)
- Mixture Editor (CDD-PoC, 30k mixture files extracted from text sources, mixfile format, open source)





MInChI=0.00.1S/C4H9.Li/c1-4(2)3;/h1-3H3;/q-1;+1&C5H12 /c1-3-5-4-2/h3-5H2,1-2H3/n{1&2}/g{17mr-1&}

Capturing mixture composition: an open machine-readable format for representing mixed substances, A.M. Clark, L.R. McEwen, P. Gedeck, B.A. Bunin, Journal of Cheminformatics 2019. DOI: 10.1186/s13321-019-0357-4

Mixtures InChI (MInChI) – Next Steps

Use cases

- Fine chemicals
- Formulations
- Eutectic solvents
- UVCBs

To Do:

- Finalize Phase 1
- Incorporate other InChI flavors
- Explore non-structure-based substance identifiers
- Metadata for process, form, other immutable properties
- Explore MInChI as query

<u>Unknown or Variable composition,</u> <u>Complex reaction products or Biological materials</u>

- Fatty acids, C16-18 and C18-unsatd., hexaesters with dipentaerythritol
- Phenol, 4-methyl-, reaction products with dicyclopentadiene and isobutylene
- Residual oils (petroleum), catalytic dewaxed
- Soybean oil, epoxidized, acrylate
- Naphtha, petroleum, light hydrocracked
- Distillates, petroleum, steam-cracked, C5-12 fraction

Joint meeting with RInChI / MInChI / NInChI

May/June (TBD) | Topic: metadata & AuxInfo



The <u>In</u>ternational <u>Ch</u>emical <u>Identifier for Reactions</u> (Reaction InChI or RInChI)

InChI Working Groups Meeting – April 2021

RInChI section - 19-April-2021

<u>Gerd Blanke¹</u>, Günter Grethe², Hans Kraut³, István Öri⁴, Jan Holst Jensen⁵, Jonathan Goodman⁶

¹-StructurePendium Technologies GmbH, Essen, Germany; ² San Diego, CA, US; ³ InfoChem GmbH, Munich, Germany; ⁴ ChemAxon, Budapest Hungary; ⁵ Biochemfusion AsP, Copenhagen, Denmark; ⁶ University of Cambridge, Department of Chemistry, Cambridge, UK

The <u>In</u>ternational <u>Ch</u>emical <u>I</u>dentifier for <u>R</u>eactions (RInChI) in the InChI Ecosystem



Interest and Feedback

- Agent concept: for ELN purposes, you should be able to distinguish between catalysts, solvents and so on in an additional auxinfo layer.
- Make sure that Mixture InChIs are compatible with the RInChI formats
- Experiences with JSON format: how many data make sense to be transported
 - There are similar discussions going on in the Nano InChI group
- Necessary for the further development: extended stereochemistry representation and organo metallics
- More financial support by companies using RInChI to insure the further technical development of RInChI



Educational Resources Bob Belford Large Molecules Evan Bolton Nanomaterials Iseult Lynch Markush/Variable Jonathan Goodman **Resolver** Markus Sitzmann Stereochemistry Andrey Erin **Organometallics** Colin Batchelor Mixtures Leah McEwen **Poactions** Gerd Blanke Tautomers Marc Nich us

UK COULES KIChard Hartshorn

Tautomerism

Compared with a comprehensive set of tautomeric rules:

- Current Standard InChI recapitulates ~30% of amenable compounds
- Current Non-Standard InChI (KET, 15T) recapitulates ~37% of compounds
- Relative to Standard InChI, Non-Standard InChI (KET, 15T) equates 3.5% more compounds as tautomers of other compounds in a typical large database (e.g. PubChem)

Working group achievements:

- Six new prototropic rules were added to InChI code
- Relative to Standard InChI, "Tauto InChI" (KET, 15T, 6 new rules) equates 7% more compounds as tautomers of other compounds, i.e. yet 3.5% more than Non-Standard InChI
- Tauto InChI is different InChI: many InChIKeys are different. Do not mix with non-standard InChI!
- Should have kind indicator "T" instead of "N": WCHQBIYPPGCACF-UHFFFAOYTA-N
- When to release Tauto InChI? In version 1.06x? Or 1.07? Or wait for InChI V.2? It is ready to go...
- Question: Which rules are realistic, which ones may be too strict?

Future outlook:

- Prototropic transforms: doubtful whether more can be added
- Ring-chain, valence tautomerism: likely incompatible with current InChI chemical structure model
- To be able to add more rules, InChI code likely needs to be re-written



Large Molecules Evan Bolton Nanomaterials Iseult Lynch

Markush/Variable Jonathan Goodman

Resolver Markus Sitzmann

Stereochemistry Andrey Erin

Organometallics Colin Batchelor

Mixtures Leah McEwen

Reactions Gerd Blanke

Marc Nicklaus

QR codes Richard P .shorn

QR InChl

Richard Hartshorn, Jeremy Frey, and Leah McEwan

Draft Specification Ready for PAC



- Specified for InChI and InChIKey
 - InChI = 'InChI text'
 - InChIKey = 'InChIKey text'
- Could be included in a search string, or a URL resolving to a supplier/organization site.

Next Steps

- Implement website service to provide QR InChl codes
- QR InChI App that would read and link to InChI enabled services
- Discussion with GS1
- Workshop to select/examine use cases (June-July?)
- QR InChI business opportunities?
- Links to Resolver
- MInChI and RInChI could be used similarly (probably in their Key form)

meta da		Change core code?	Technical issues	Develop specification	Ready for testing
teaching dia	Education :			(yes)	yes
industry Chi Rincy ng	Large Molecules :			yes	
Nu con con T	Markush :		yes	yes	yes
InChIKey	Mixtures :			yes	yes
	CNanomaterials :			yes	
OG TO THE OF CS O	Organometallics :	possibly	yes	yes	
	QR codes :			yes	(yes)
S S S S S S S S S S S S S S S S S S S	Reactions :			yes	yes
	Resolver :				yes
	Stereochemistry :	yes		yes	
Confolodoros (Confolodoros)	Tautomers :	barely			yes
10,050 compression (MINDO) 10 compression	Useful to	ools: InC Wo Co	ChI2InChI (ba orkflows (da llection of u	ad InChI to good ta to InChI) se-cases	InChI)

www-jmg.ch.cam.ac.uk/inchi





InChI canonical representifier for molecules **connecting** people

> Thank you to all working groups and their chairs and to all participants in the last fortnight's meetings