Nano Commons

Knowledge Infrastructure

NInChI working group:
Towards a first version of the NInChI standard

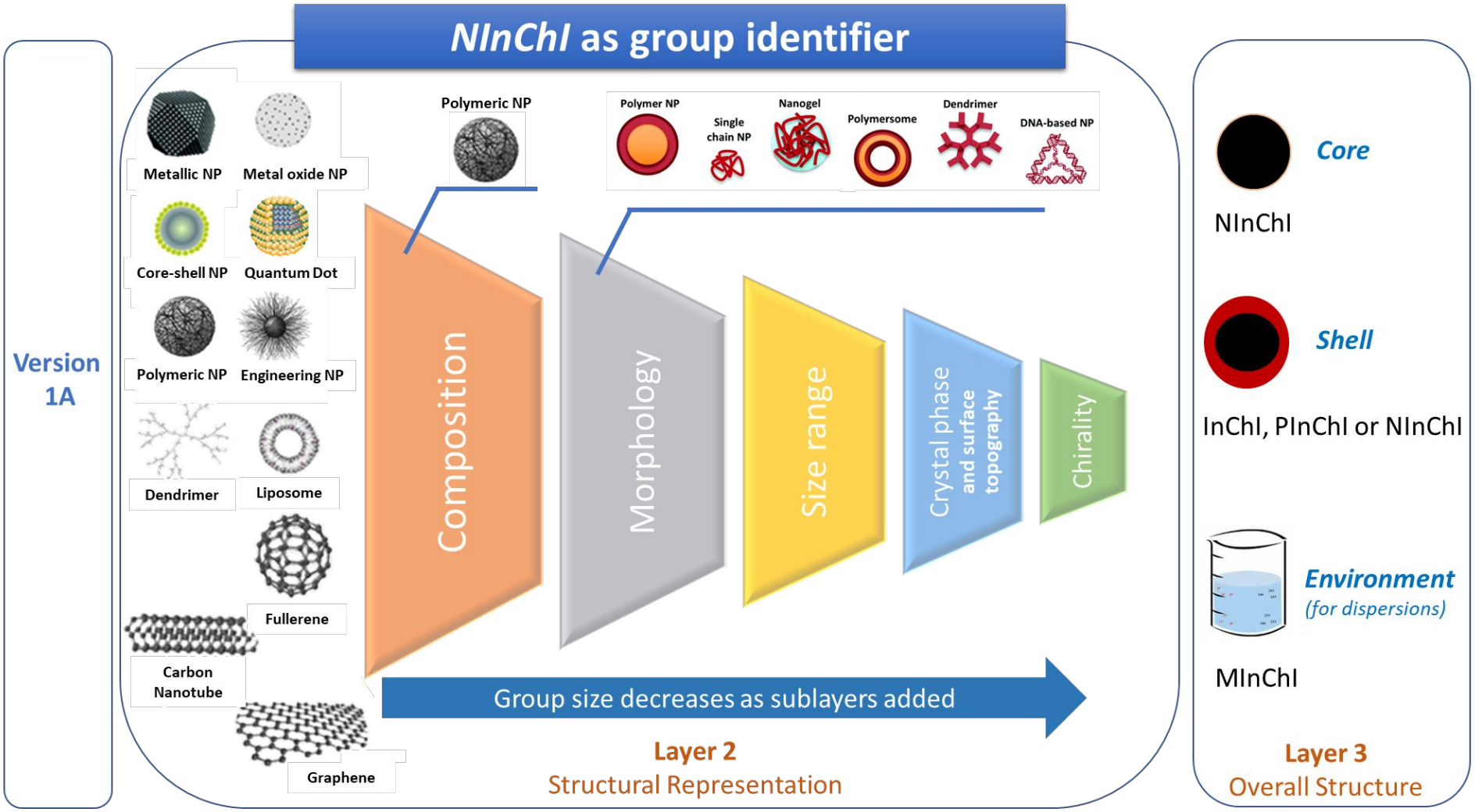
Thomas Exner (7P9)

20 September 2022

NanoCommons
Nano-Knowledge Community

Prototype version

Recap of what we proposed



Recap of what we proposed

Component 1:
Composition: Au -> InChI=1S/Au
Morphology: shell -> sh
Size: thickness of 2 nm given as lower and upper radius -> 2t-9

Part of NInChI for component 1: Au/msh/s1t-9

Component 2:
Composition: Si -> InChI=1S/O2Si/c1-3-2
Morphology: sphere -> sp
Size: diameter of 20 nm -> 20d-9
Space group: amorphous -> 000

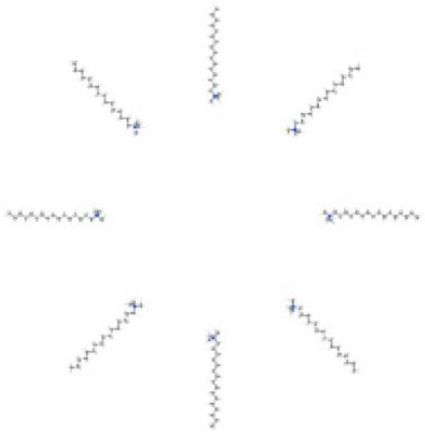
Part of NInChI for component 2: O2Si/c1-3-2/msp/s20d-9/k000

Layer 1: 1A
Layer 2: /Au/msh/s2t-9!O2Si/c1-3-2/msp/s20d-9/k000
Layer 3: /y2&1 (ordering from inside out)

Recap of what we proposed

Component 1:
Composition: Au -> InChI=1S/Au
Morphology: sphere -> sp
Size: diameter of 20 nm -> 20d-9
Space group: not specified

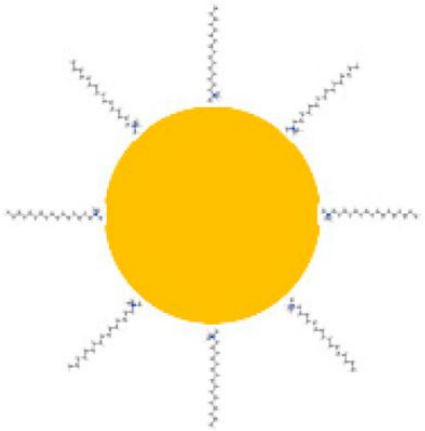
Part of NInChI for component 1: Au/msp/s20d-9



Component 2:
Composition: Si -> InChI=1S/C19H42N.BrH
/c1-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-20(2,3)4;
/h5-19H2,1-4H3;1H/q+1;/p-1
Morphology: none
Size: none
Space group: none

Part of NInChI for component 2: C19H42N.BrH
/c1-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-20(2,3)4;
/h5-19H2,1-4H3;1H/q+1;/p-1

Layer 1: 1A
Layer 2: /Au/msp/s20d-9
!C19H42N.BrH
/c1-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-20(2,3)4;
/h5-19H2,1-4H3;1H/q+1;/p-1
Layer 3: /y1&2 (ordering from inside out)



NanoCommons
Nano-Knowledge Community

Next version

Proposal for next version

NanoCommons
Nano-Knowledge Community

1) Keep as much as possible
2) Fix specific issues to integrate it into the InChI universe
3) Specific problem identified: Binding between different layers
4) Concentrate on general features

Keep as much as possible

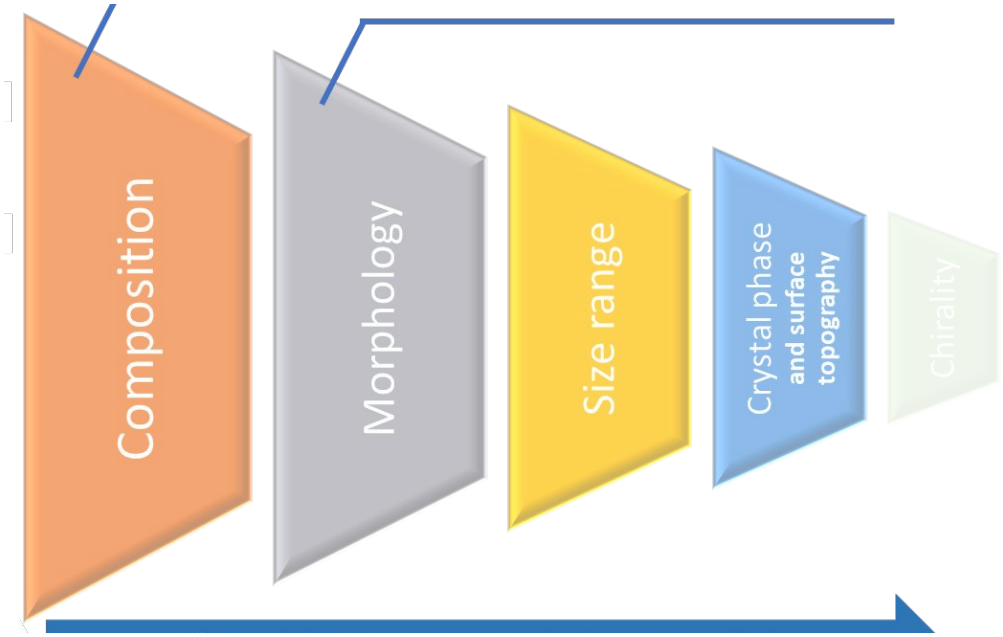
We know that the covered information is not enough for a full characterisation and some use cases would profit from more.

However, it will always be possible to add more using the layer approach of InChI later

But can we even reduce the number of properties?

Gerd: The critical issue we may discuss on Tuesday are the sizes. After reviewing our discussion in Limassol I think we were not that far apart. Fred's contra argument was only related to the general definition of Nanomaterials that are defined being up to 100 Nm as far as I understood. But beside this general definition we need the size as additional parameter because it has major influence to the properties of the materials. If we distinguish between general definition (Fred's argument) and the actual size of the material we may do a step forward.

Keep as much as possible



Alignment with the InChI universe

InChI and NInChI use letters to specify the specific layers and sublayers.

Selection of NInChI letters partly conflicted with the InChI and InChI extension specifications.

Easy solution:

Add “N” to each identifier

→ **“/s” for size becomes “/Ns”**

Representation issues identified

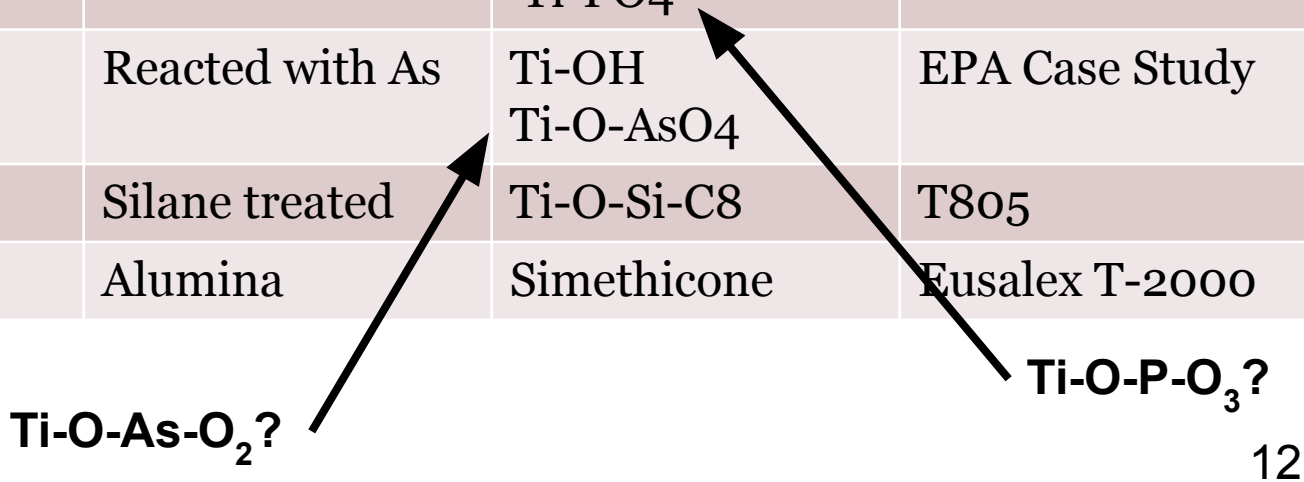
The original proposal specified the layers in an inside-out approach. However, how these layers are bound to each other was not described but was seen as one of the major characteristics of complex particles.

Proposed solution:

- Grouping of components with **()** if this is not better done by the use of **MInChI** to represent the composition
- Components in one shell/core could be combined with **“&”** and **“|”**, however this overlaps with the MInChI specification
- Covalently bound shells are linked to inner shells/core by **“>>”**
- Non-covalently bound shells get **“>”**
- **Rebound atoms** in definitions of composition

Examples

Substrate	Surface Coating	Projecting Species	Product or Example
$TiO_2 \cdot 0.7\ H_2O$	None	Ti-OH; H20	Metatinic Acid
TiO2	None	Ti-OH	P-25
Al: TiO2	Alumina	Al-OH	R-100 & Uf-1
Al:TiO2	Alumina/ alumina silicate	Al-OH Si-OH	Uf-2
TiO2	PBS-present	Ti-OH Ti-PO4	Uf-3 (P25)
TiO2	Reacted with As	Ti-OH Ti-O-AsO4	EPA Case Study
TiO2	Silane treated	Ti-O-Si-C8	T805
TiO2 (??)	Alumina	Simethicone	Eusalex T-2000



Examples

Substrate	Surface Coating	Projecting Species	Product or Example
TiO2	Silane treated	Ti-O-Si-C8	T805

NinChI=0.00.1/

C10H23O3Si.**Ti**/c1-4-5-6-7-8-9-10-14(11,12-2)13-3;/h4-10H2,1-3H3;/q-1;+1
!
O2Ti/c1-3-2/**Nm**sh/**Ns**2d-8

/**Ny**2**>>**1

Examples

Substrate	Surface Coating	Projecting Species	Product or Example
TiO2	PBS-present	Ti-OH Ti-PO4	Uf-3 (P25)

NinChI=0.00.1/

H2O.**Ti**/h1H2;/q;+1/p-1
!
H3O4P.**Ti**/….
!
/O2Ti/c1-3-2/**Nm**sh/**N**sXXd-9

/**Ny**3**>>**(1|2)

Concentrate on general features

NanoCommons
Nano-Knowledge Community

Carbon nanotubes were specifically considered in the prototype.

However, this was more to show the flexibility of the InChI approach but it is probably better to leave out such specific materials in the first version of the NInChI standard.

Next steps

1. Discussion of the proposed changes
2. Start to draft the NInChI specifications
3. Implement the changes in the reference implementations (NovaM and UM)
4. **Input format definition** (needs to be compatible with InChI libraries)
5. Proposal to InChI Trust
6. Request for better support of binding/dummy atoms

NanoCommons
Nano-Knowledge Community

Future topics

(after proposal of next version)

AuxInfo in InChI

NanoCommons
Nano-Knowledge Community

11.1. What is the 'Auxiliary Information' (AuxInfo) in the InChI output?

The InChI Software output complements the Identifier itself with a range of additional information. This includes warnings and errors messages, as well as a specific 'Auxiliary Information' (AuxInfo) string.

AuxInfo contains, in particular, atom non-stereo equivalence information, mapping input atom positions to output positions, and 'reversibility' information for re-drawing the structure. To see how AuxInfo is decrypted, look at Section "Auxiliary Information Output" of the InChI Software User Guide, (it may also be necessary to consult the file InChI_UserGuide.pdf).

AuxInfo is generated by the inchi-1 executable by default (this behavior may be turned off by using the switch AuxNone).

Note that the AuxInfo string itself is a valid input for the inchi-1 generator and may be used to regenerate the source structure.

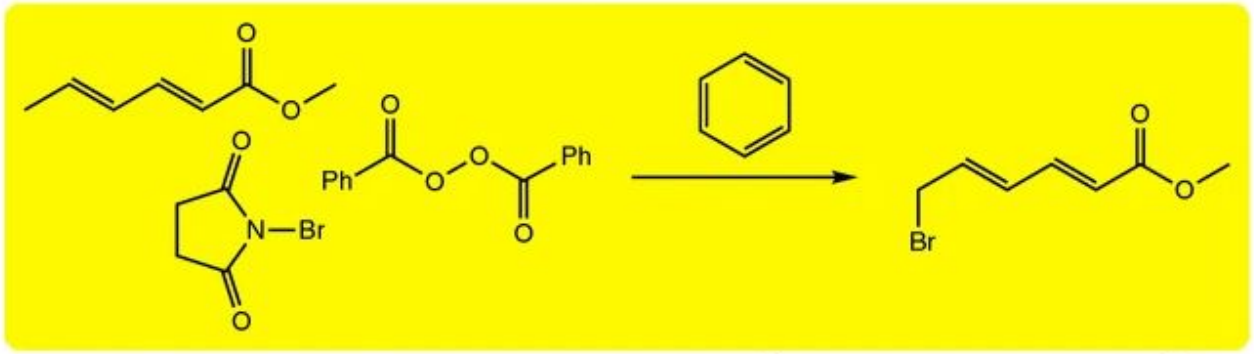
AuxInfo in InChI

```
AuxInfo=
{version}1
/{normalization_type}
```

```
* Input_File: "C:\inchi-samples\benzoicacid.mol"
Structure: 1
InChI=1S/C7H6O2/c8-7(9)6-4-2-1-3-5-6/h1-5H,(H,8,9)
AuxInfo=1/1/


InChIKey=WPYMKLBDIGXBTP-UHFFFAOYSA-N
XHash1=58b69502210f14434087af02eac658408b0e4577bf8fafa8
XHash2=80fc1c149afbf4c8996fb92427ae41e4649b934ca495991b7852b855
```

AuxInfo in reaction InChI



convert to RInChI
key reaction data only

RInChI=1.00.1S/C14H10O4/c15-13(11-7-3-1-4-8-11)17-18-14(16)12-9-5-2-6-10-12/h1-10H!C4H4BrNO2/c5-6-3(7)1-2-4(6)8/h1-2H2!C7H10O2/c1-3-4-5-6-7(8)9-2/h3-6H,1-2H3/b4-3+,6-5+<>C7H9BrO2/c1-10-7(9)5-3-2-4-6-8/h2-5H,6H2,1H3/b4-2+,5-3+<>C6H6/c1-2-4-6-5-3-1/h1-6H/d+

Auxiliary Information
(RAuxInfo)
also generated:

RAuxInfo=1.00.1/0/N:10,16,9,11,15,17,8,12,14,18,7,13,1,4,5,6,2,3/E:(1,2)(3,4,5,6)(7,8,9,10)(11,12)(13,14)(15,16)(17,18)/rA:18nCOOOOOOCCCCCCCCCC/rB:s1;s2;s3;d1;d4;s1;d7;s8;d9;s10;s7d11;s4;d13;s14;d15;s16;s13d17;/rC:-.5452,.1111,0,-.166,-.0952,0,.1693,.1111,0,.5265,-.0952,0,-.5452,.5256,0,.5265,-.5077,0,-.9024,-.0952,0,-1.2597,.1111,0,-1.6169,-.0952,0,-1.6169,-.5077,0,-1.2597,-.7139,0,-.9024,-.5077,0,.8925,.1042,0;1.2445,-.1108,0;1.6067,.0865,0;1.6169,.4989,0;1.2649,.7139,0,.9927,.5166,0;0/N:1,2,5,3,8,4,6,7/E:(1,2)(3,4)(7,8)/rA:8nCOCNCOOBr/rB:s1;s2;s3;s184;d5;d3;s4;/rC:-.5256,.2062,0,-.5256,.2063,0,-.1313,-.3337,0,.1111,0,0,-.1313,.3337,0,-.5039,.726,0,-.5039,-.726,0,.5256,0,0;0/N:1,8,2,3,4,5,6,9,7/rA:9nCOCCCCOCO/rB:s1;d2;s3;d4;s5;s6;d7;d8;/rC:-1.2503,-.3094,0,-.8931,-.1031,0,-.5359,-.3094,0,-.1786,-.1031,0,.1786,-.3094,0,.5359,-.1031,0,.8931,-.3094;0;1.2503,-.1031,0,.5359,.3094,0;<>0/N:8,3,4,2,5,1,6,10,9,7/rA:10nCCCCCCOOBr/rB:s1;d2;s3;d4;s5;s6;s7;d8;s1;/rC:-1.2503,-.1031,0,-.8931,.1031,0,-.5359,-.1031,0,-.1786,.1031,0,.1786,-.1031,0,.5359,.1031,0,.8931,-.1031,0;1.2503,.1031,0,.5359,.5156,0;-1.2503,-.5156,0;<>0/N:1,2,6,3,5,4/E:(1,2,3,4,5,6)/rA:6nCCCCCCrB:d1;s2;d3;s4;s1d5;/rC:-.3572,.2062,0,-.3572,-.2062,0,0,-.4125,0;.3572,-.2062,0;.3572,.2062,0;0,.4125,0;

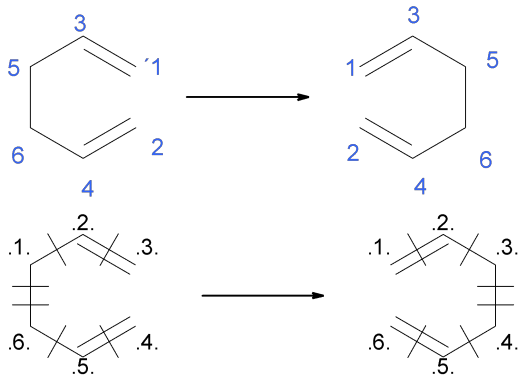
Grethe, G., Blanke, G., Kraut, H. *et al.* International chemical identifier for reactions (RInChI).
J Cheminform **10,** 22 (2018). https://doi.org/10.1186/s13321-018-0277-8

AuxInfo in reaction InChI

- The atom mapping is used to describe the transfer of the atoms involved in a reaction from the starting materials to the products
  - Example: Esterification

MapAuxInfo=1.00.1/1-1<>1-2;1-2<>1-4;1-3<>1-8;2-1<>1-1;2-2<>1-3;2-3<>1-5;2-4<>1-6;2-(5,6)<>1-7;2-(5,6)<>2-1

  - Example: Cope elimination



RInChI=1.00.1S/C6H10/c1-3-5-6-4-2/h3-4H,1-2,5-6H2<>C6H10/c1-3-5-6-4-2/h3-4H,1-2,5-6H2/d+

MapAuxInfo=1.00.1/1-1<>1-5;1-2<>1-6;1-3<>1-3;1-4<>1-4;1-5<>1-1;1-6<>1-2

Grethe, G., Blanke, G., Kraut, H. *et al.* International chemical identifier for reactions (RInChI). *J Cheminform* **10,** 22 (2018). https://doi.org/10.1186/s13321-018-0277-8

AuxInfo in reaction InChI

- Reaction data / ProcAuxInfo, failing reaction
  - Reaction related data
    - Data related to the entire reaction like protecting atmosphere, total yield, reaction vessel
    - Timepoint depending data like temperature, pressure, etc.
- Example: Esterification with 50% yield in a flask
  - Component data

	Time	Component	2-1	2-2	3-1	3-2	4-1
Summary	1800		0.5	1	.6	0,6	60
Unit			l	mol	mol	mol	ml
Timepoint	0		0.5	1	0	0	0
	300		0.5	.9	.1	.1	10
	600		0.5	.8	.2	.2	20
	900		0.5	.7	.3	.3	30
	1200		0.5	.6	.4	.4	40
	1500		0.5	.5	.5	.5	40
	1800		0.5	.4	.6	.6	40

  - Reaction data

	Time	Temperature		pH		Stirring	
Summary	1800.0	20.0	100.0	7.0	5.0	1000.0	1500.0
Timepoint		°C				rpm	
0.0		20.0		7.0		1000.0	
300.0		40.0		6.5		1000.0	
600.0		60.0		6.0		1000.0	
900.0		90.0		5.5		1000.0	
1200.0		100.0		5.0		1000.0	
1500.0		100.0		5.0		1500.0	
1800.0		100.0		5.0		1500.0	

InChIKey

Atom connection sub-layer

Charge layer

InChI=1S/C2H4ClNO2/c3-1(4)2(5)6/h1H,4H2,(H,5,6)/p+1/t1-/m0/s1/i3+0

InChI version

Chemical formula sub-layer

Hydrogen atoms sub-layer

Stereochemical layer

InChIKey=UWPWWENWLZPQGU-WRFRXMDISA-O

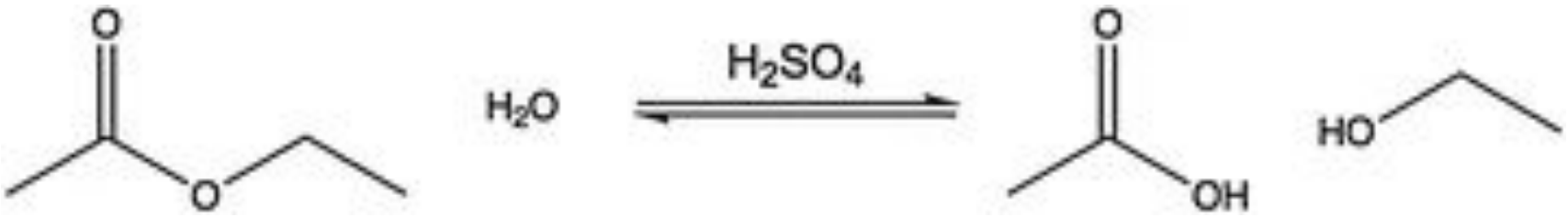
Main layer

Stereochemistry & isotopes

version

protons

InChIKey for reactions



Long-RInChIKey=SA-EUHFF-QTBSBXVTEAMEQO-UHFFFAOYSA-N-
LFQSCWFLJHTTHZ-UHFFFAOYSA-N--XEKOWRVHYACXOJ-UHFFFAOYSA-N-
XLYOFNOQVPJJNP-UHFFFAOYSA-N--QAOWNCQODCNURD-UHFFFAOYSA-N

Short-RInChIKey=SA-EUHFF-JJFIATRHOH-UDXZTNISGZ-QAOWNCQODC-
NUHFF-NUHFF-NUHFF-ZZZ

Web-RInChIKey=SMUHAWIQPXIVCEVKG-NUHFFFADPSCTJSA

Decisions:

- **For what would a NInChIKey be used for?**
- **Can we have a fixed-length version?**

https://jcheminf.biomedcentral.com/articles/10.1186/s13321-018-0277-8