InChI encoding of polymers current results and further tasks

Andrey Yerin, ACD/Labs,
IUPAC Division VIII

Status and Future of the IUPAC InChI: context and use cases
August 16, 2017, Bethesda, US

InChI encoding for polymers – history

The work was started in 2009 via IUPAC project 2009-042-1-800
"InChI Requirements for Representation of Polymers".

Task group members

Andrey Yerin, ACD/Labs, Russia, project chairman;
Ted Wilks, DuPont contractor, USA;
Jaroslav Kahovec, Institute of Macromolecular Chemistry, Czech Republic;
Roger Schenck, Chemical Abstracts Service, USA;
Dmitrii Tchekhovskoi, NIST, USA.

Experts involved in discussions

Igor Pletnev, InChI developer, Moscow State University, Russia;
Keith Taylor, Accelrys, Inc., USA
Jonathan Brecher, PerkinElmer, Inc., USA;
Yulia Borodina, FDA, US.

InChI encoding for polymers – history

IUPAC project 2009-042-1-800 "InChI Requirements for Representation of Polymers".

Meetings organized and participated in:

Task group meeting, San Francisco, USA, 20 March 2010.
Task group meeting, London, UK, 29 November 2010
InChI Trust Board meeting, Berlin, Germany, 21 March 2011.
InChI Subcommittee meeting, InChI, Gaithersburg, USA, 21-22 March 2012.

The project aimed at

- **analysis of existing polymer representation standards – graphical and electronic;**
- **development of InChI encoding concepts for polymers;**
- **specification of necessary normalization procedures for polymer encoding;**

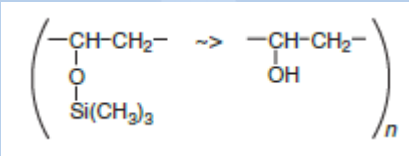
The project was completed in 2013 and the report was submitted to all involved parties.
(Note that the report is important and should include all involved IUPAC bodies!)

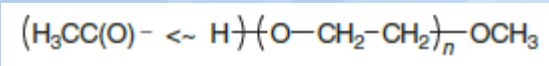
InChI version 1.05 released early this year includes encoding of polymers as experimental option for tests and collection of proposals for changes and improvements.

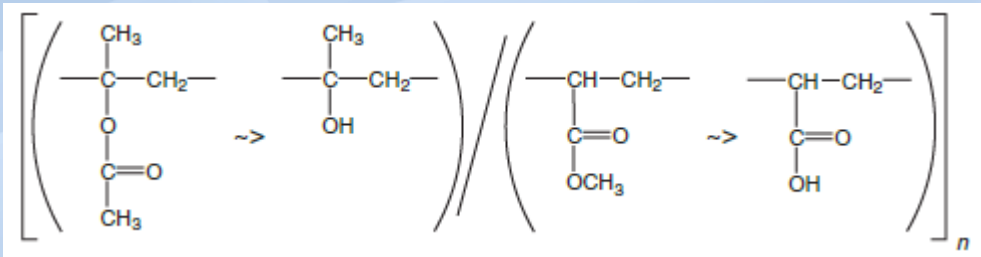
InChI encoding for polymers – lack of sources

While small molecules have good representation conventions, polymers really lack any general considerations of representation standards.

Old IUPAC recommendations on representation of polymers are dated about 1980s and now officially considered as obsolete.
Even some new polymer recommendations use structures created with MS Word or with some graphical objects that are beyond structure encoding.







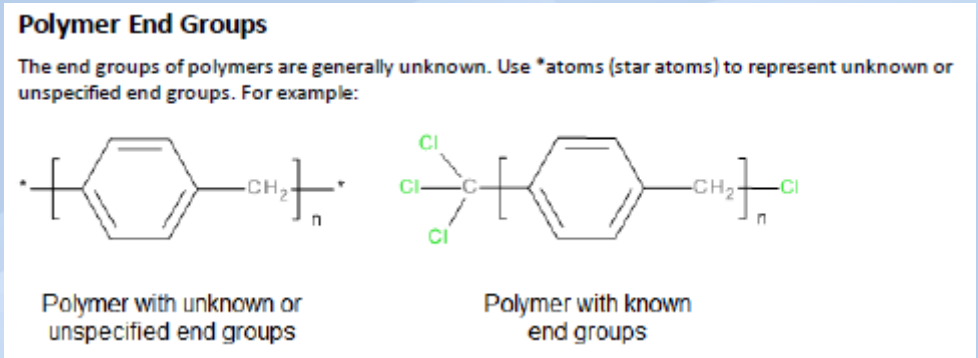
Anyway thus July during IUPAC General Assembly in Sao Paulo at the meeting of IUPAC Chemical Nomenclature and Structure representation Division VIII and Polymer division IV It was decided that new recommendations of representation of polymers are necessary.

We can hope to have new recommendations that take into account electronic representation.

InChI encoding for polymers – lack of sources

For implementation in InChI existing electronic conventions for polymers are taken into account with addition of some general principles of normalization to detect equivalence of alternative representations.

Mostly Accelrys Draw (originating from ISIS/Draw and resulting in BIOVIA Draw)
Like the following example from BIOVIA Structure representation Guide.



All basic principles are taken into account and already implemented in InChI 1.05

Thanks to Igor Pletnev – one the main InChI developers.

InChI encoding for polymers – InChI version 1.05

Polymer encoding is introduced via new 'polymer' layer.
This layer is treated as modification and designated with '/z'

InChI=1B/.../c.../h.../q.../p.../z.../b.../t.../m.../s...
(between protonation and double bond stereo layers)

The status of polymer support in InChI 1.05:

The support of polymers is currently an experimental option.
A special command line option should be used to allow treatment of polymers.

InChI/InChIKey for a polymer uses the 'B' flag character (for "Beta"), instead of 'S' or 'N' for standard/non-standard InChI.
This flag will be replaced by common standard/non-standard conventions if and when InChI for polymers is officially adopted.

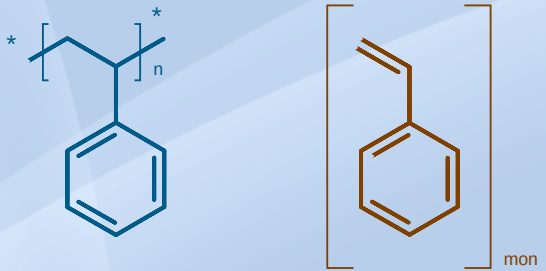
InChI encoding for polymers – InChI version 1.05

Source-based and structure-based representations.

The terms “source-based” and “structure-based” are normally used in association with chemical names:

Structure-based	**poly(1-phenylethylene) or poly(1-phenylethane-1,2-diyl)**
Source-based	**polystyrene or poly(ethenylbenzene)**

While representation of polymer with monomer structure is a part of electronic representation standard for about three decades, it is not so accepted term.



MDL structure- and source-based

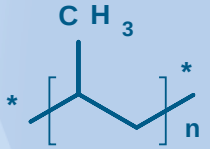
Reasonable alternative

So a source-based structure representation is further validated with InChI tools.

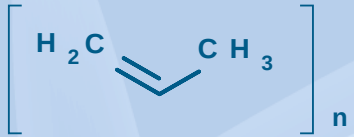
InChI encoding for polymers – InChI version 1.05

Source-based and structure-based representations.

InChI for polymers is designed so that, in many practically important cases, InChIs and InChIKeys for source- and structure-based representations have the same first part



InChI=1B/C3H6/c1-3-2/h3H,1H2,2H3/z200-1-3
InChIKey=QQONPFPTGQHPMA-DJNVOPQRBA-N



InChI=1B/C3H6/c1-3-2/h3H,1H2,2H3/z101-1-3(1.3)
InChIKey=QQONPFPTGQHPMA-KDIWSDSHBA-N

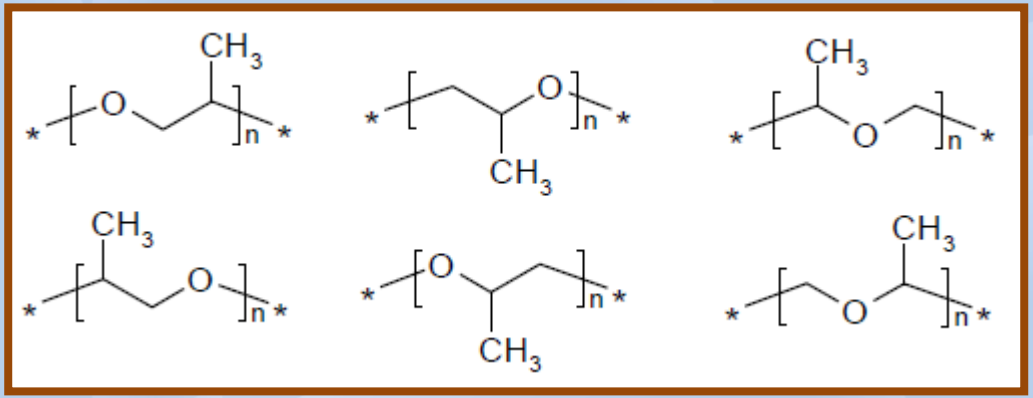
Obviously not in every case the main layers and the first block of InChIKey are the same. First of all as soon as the same polymer can be prepared from different monomers.

Anyway this option is useful and may cover most important cases.

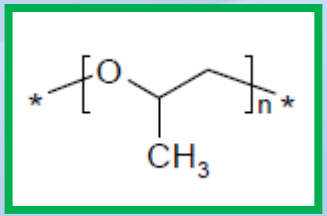
InChI encoding for polymers – InChI version 1.05

Structure normalization in polymers

The same polymer can be represented in several ways using different CRUs (Constitution Repeating Unit). So called ‘phase shift”.



Six equivalent presentations



Canonical form

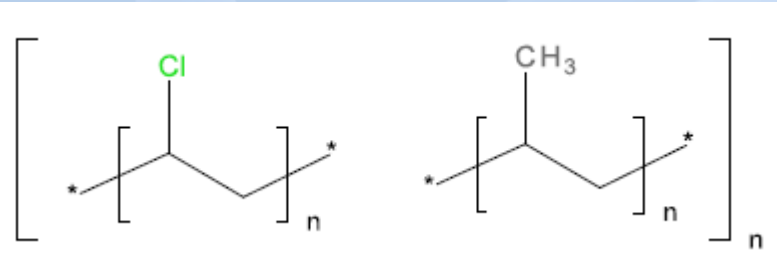
Canonicalization for polymers is implemented in InChI tools

Note that they are intentionally close as possible to IUPAC recommendations on nomenclature of regular single strand polymers in choice of preferred CRU.

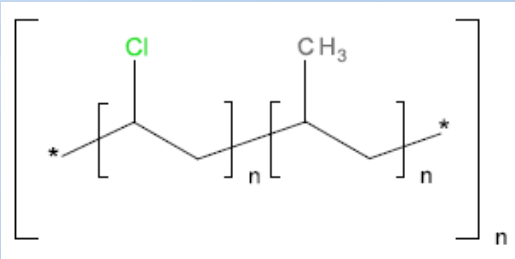
InChI encoding for polymers – InChI version 1.05

Restrictions and reported issues:

1. Only single-strand and source-based are supported;
2. A combination of structure- and source-based representation is not allowed;
3. Hydrogen atoms for end-groups are not supported;
4. Normalization with specified end-groups is not supported;
5. Folding of CRU is not supported ([-CH_2-CH_2-] > [-CH_2-]);
6. Some restrictions for copolymers



supported



not supported and not recommended

Several additional issues are mentioned previously and will be mentioned during this meeting.

Improvements in InChI treatment of polymers

Mostly those that are currently mentioned as restrictions:

1. **Normalization with terminal groups;**
2. **Hydrogen for terminal groups;**
3. **CRU folding/simplification;**
4. **Consider other types of polymers – examples needed;**
5. **Collect examples and requests for currently problematic cases.**

Please send examples of polymer representations

Thank you!