The IUPAC International Chemical Identifier (InChI): Promotion and Extension (project 2004-039-1-800)

Extension to polymeric systems

Notes of a meeting at the Institute of Macromolecular Chemistry, Heyrovského námestí 2 CZ-162 06 Praha 6, Czech Republic, on June 3rd 2005 at 9.00 am

Present: Dr Stephen Heller, Dr Jaroslav Kahovec, Dr Alan McNaught, Dr Andrey Yerin

1.0 **Objective.** The meeting was convened to consider what methods for two-dimensional depiction of macromolecular structures could be advantageously incorporated into the InChI specification. In discussion, reference was made to the 1994 IUPAC Recommendations on Graphic Representations of Macromolecules :http://www.iupac.org/publications/pac/1994/pdf/6612x2469.pdf

2.0 **Questions.** Dr Stephen Stein (absent Task Group member) had suggested the following questions for discussion:

2.1 *What, if any, features of polymer structure can be usefully 'canonicalized'?*

This was discussed in relation to representations recommended in the 1994 IUPAC Recommendations (see minute 1.0); conclusions are given in minute 3.0.

2.2 *Is the Chemical Abstracts representation of polymers helpful/adequate?*

Task Group members considered that the largely text-based representations used by CAS were not directly relevant to the extension of InChI.

2.*3 Is there any need to reflect source-based in addition to structure-based depiction?*

For a variety of reasons it was considered that InChI should continue to reflect actual structure without regard for origin. In particular, it was noted that there was no generally accepted source-based 2-D representation, and that the same macromolecular structure could arise from several sources. However, source information might usefully be included in the Auxiliary Information field.

2.4 *To what extent should we deal with the variability inherent in polymeric structure (crosslinking, variable end-groups, etc). Is there or should there be a polymer markup language (PolymerML?) or equivalent to separate the variables*

Task Group members were not aware of any moves to establish a markup language for categorising the parameters defining a macromolecular structure.

3.0 **Recommendations.** The following recommendations arose from consideration of the 1994 IUPAC Recommendations referred to above (minute 1.0) in the light of the questions listed in minute 2.0. Reference numbers (e.g. 2-E8) quoted for structures are those used in the 1994 paper.

3.1 Work to include polymers should be categorised as follows:

(a) Regular single-strand organic polymers (essential)

(b) Copolymers (desirable)

(c) Irregular polymers (low priority; not for the next version)

3.2 Acceptable 2-D representations should show the constitutional repeating unit (CRU) with 'free valences' as bonds, enclosed (or crossed) by parentheses (), brackets [ ] or braces { }, followed by a letter subscript, e.g. $[-CH_2-CH=CH-CH_2-]_n$. Steric configuration should be included if known, but there is no requirement to recognise Fischer representations (e.g. 2-E2).

3.3 The InChI should be the same regardless of the direction of reading of the backbone chain, and regardless of where the CRU is considered to begin. A CRU beginning within a ring should not be allowed, i.e. the number of CRU free valences should be minimised. There is no requirement to handle 'ladder' polymers, e.g. 2-E-14 (where the CRU can only start within a ring).

For example $[-O-CH_2-CH_2-CH_2-C_6H_4-CH_2-CH_2-]_n$ can have 14 different representations, all of which should have the same InChI.

3.4 The molecular formula field could contain the formula of the CRU plus a polymer flag.

3.5 End groups should be included as a separate 'layer' if known; however the algorithm will need to deal with the fact that the end groups vary according to where the CRU starts. It would be necessary to relate end group citation to a specified CRU starting point.

3.6 There is no requirement to deal with inorganic polymers, e.g. 2-E17, 2-E18.

3.7 Alternating and periodic copolymers (4.1-E1 to E3) can be represented and treated like regular single-strand polymers.

3.8 Statistical (4.2-E1 to 7) and block (4.3-E1 to E6) copolymers should both be included: $[-A-]_x-[-B-]_y$ (statistical and random where $x + y = 1$; block where x and y are both > 1); perhaps in the former case they can be treated like mixtures?

3.9 Graft and star copolymers are not to be covered at this stage.

Alan McNaught
13 June 2005