IUPAC DIVISION (VIII) OF CHEMICAL NOMENCLATURE AND STRUCTURE REPRESENTATION

Report to IUPAC Bureau and Council, August 2005

I. Highlights

I.1 **IUPAC International Chemical Identifier (InChI).** Version 1 released; enthusiastic response and widespread implementation in chemical databases (item II.3.1)

I.2 **Organic Preferred Names.** Final revision of the new Blue Book, including recommendations for selecting Preferred IUPAC Names (PINs), following ICTNS and public review (item II.3.2)

I.3 **Revised Red Book.** Scheduled for publication in August 2005 (item II.3.3).

I.4 **Fullerenes.** "Numbering of Fullerenes" published (item II.3.5).

I.5 **Graphical representation standards for chemical structures.** Guidelines for graphical representation of configuration undergoing final revision, as the first component of this project (item II.3.8).

II. Report of activities 2004-2005

II.1 ***Context.*** The work of the Division of Chemical Nomenclature and Structure Representation is concerned entirely with standards for the transmission of chemical information, and as such addresses the following long-range IUPAC goals:

b) IUPAC will facilitate the advancement of research in the chemical sciences through the tools that it provides for international standardization and scientific discussion.

d) IUPAC will foster communication among individual chemists and scientific organizations, with special emphasis on the needs of chemists in developing countries.

II.2 ***Constitution of the Division.*** Division VIII consists of a Division Committee (sixteen members from eight nations and nine National Representatives) and an Advisory Subcommittee (44 people from 18 nations). The Joint IUPAC-IUBMB Commission on Biochemical Nomenclature (nine members from seven nations) reports to the Division Committee. Division members have access to a Web Discussion Board, on which minutes of meetings and current draft recommendations are posted for comment.

II.3 ***Current projects***

II.3.1 IUPAC International Chemical Identifier (InChI).

A further test version of the software was distributed in July 2004, and comments were accommodated in a final test version issued in February 2005. To allow trademark copyright and licensing issues to be resolved, the name of the Identifier was changed to International Chemical Identifier (InChI) [formerly IUPAC-NIST Chemical Identifier (INChI)]. Version 1 was

finally released in April 2005: see www.iupac.org/inchi and an announcement in *Chemistry World*:

"International chemical identifier goes online", *Chem. World*, 2005, **6**, 7: http://www.rsc.org/chemistryworld/Issues/2005/June/this_month/International_chemical_identifier.asp

The Identifier has been incorporated into Chemical Markup Language and its potential is being explored by various groups; see for example:

P. Murray-Rust, H. S. Rzepa and Y. Zhang, "Googling for INChIs; A remarkable method of chemical searching": http://lists.w3.org/Archives/Public/public-swls-ws/2004Oct/att-0019/index.html

S.J. Coles, N.E. Day, P. Murray-Rust, H.S. Rzepa and Y. Zhang, "Enhancement of the chemical semantic web through the use of InChI identifiers," *Org. Biomol. Chem*., 2005, **3**(10), 1832-1834: http://pubs.rsc.org/ej/OB/2005/b502828k.pdf

M.D. Prasanna, J. Vondrasek, A. Wlodawer and T.N. Bhat, "Application of InChI to Curate, Index, and Query 3-D Structures," *Proteins: Structure, Function, and Bioinformatics*, 2005, **60**, 1-4

One of the graduate students in Murray-Rust's group has prepared a very useful InChI information source:

http://wwmm.ch.cam.ac.uk/inchifaq/

InChI is currently being incorporated into a variety of public and commercial chemistry databases:

- NIST - 150,000 structures
- NIH/NCBI/PubChem project - 800,000+ structures
- ISI - 2+ million structures
- NCI Database - 23 million+ structures
- EPA-DSSToX Database - 1450 structures
- KEGG database - 9584 structures
- UCSF ZINC - 3.3 million structures
- Chemical Entities of Biological Interest (ChEBI) database of the European Bioinformatics Institute – 5000 structures
- BRENDA database (University of Cologne) – 36,000 structures

Other InChI developments include:

- Integration by ACD/Labs into their chemical structure drawing program, ChemSketch.
- Discussions on adoption by the new *Nature Chemical Biology*
- Discussions on use in the new *Beilstein Journal of Organic Chemistry*.

A Supplementary Project is now in progress with the following objectives:

- to promote the use of the Identifier throughout the chemical information community
- to extend its applicability to include polymeric structures
- to explore the need for other extensions, including the ability to handle Markush structures, and to include information on other attributes such as phases and excited states

A meeting in Prague in June 2005 has established requirements for extension to polymers.

To enable development of InChI facilities and applications in an Open Source context, a project to encompass this work has been registered with SourceForge.net (see http://sourceforge.net/projects/inchi).

II.3.2 Organic Preferred IUPAC Names (PINs)

Work on the new Nomenclature of Organic Chemistry (IUPAC Blue Book), including recommendations for identifying IUPAC-preferred names, is approaching completion. Final revision of the book (more than 1300 pages) is in progress following ICTNS and public review, and publication is expected in 2006.

II.3.3 Revision of "Nomenclature of Inorganic Chemistry"

The revised IUPAC Red Book is scheduled for publication in August 2005.

II.3.4 Rotaxanes

Recommendations for naming rotaxanes are almost ready to be issued for expert review.

II.3.5 Fullerene nomenclature Part II

This extension of the published Part I recommendations to larger and more complicated molecules, with emphasis on numbering systems, was published in *Pure Appl. Chem*:

http://www.iupac.org/publications/pac/2005/7705/7705x0843.html

II.3.6 Extension of rules for stereodescriptors to include coordination numbers 7-12

The first meeting of the project group took place in March 2005, in London.

II.3.7 Nomenclature of cyclic peptides

The document is undergoing revision following ICTNS and public review.

II.3.8 Graphical representation standards for chemical structures

The first component of this project, consisting of guidelines for two-dimensional representation of configuration, is being revised following ICTNS and public review. Draft guidelines for many other aspects of graphical representation were discussed at a Task Group meeting in Beijing immediately preceding this General Assembly.

Macromolecular Nomenclature Projects

The following projects are managed in full cooperation with Division IV's Subcommittee on Macromolecular Terminology:

II.3.10 Terminology and nomenclature of macromolecules with cyclic structures

This project has been extended to cover polycyclic and spirocyclic macromolecules, and the new draft is expected to be ready for expert review after the Beijing meeting.

II.3.11 Terminology and structure-based nomenclature of dendritic and hyperbranched polymers

This is almost ready for expert review.

II.3.12 Nomenclature for chemically modified polymers

The nature of this project has been reconsidered, and the project group reconstituted. It has a new title: 'Source-based Nomenclature for Modified Polymer Molecules'

II.3.13 Source-based nomenclature of single-strand organic polymers

This is a new project; it is now proposed to change the title to the more explicit "Source-based Nomenclature of Organic Homopolymers and Copolymers".

II.3.14 Nomenclature of macromolecular rotaxanes

This project is on hold pending completion of the non-polymeric rotaxane document (item II.3.4).

II.4 Joint IUPAC-IUBMB Commission on Biochemical Nomenclature (JCBN)

Joint activities of the two Unions are channelled through this Joint Commission. Its main activities are:

II.4.1 Maintenance and updating of the Enzyme List

This is a very substantial and continuous operation drawing on advice from IUPAC participants on chemical names for substrates, reagents and products.

II.4.2 Maintenance and development of specialised naming systems for natural products

Classes of natural product of interest to biochemists, especially steroids, amino acids and peptides, carbohydrates, lipids and nucleic acids require local specialised systems for naming. A project to review and update carbohydrate nomenclature is being developed.

II.4.3 Advice for biochemists on names for specific compounds of biochemical importance

II.4.3.1 Synonyms databases.

Synonyms databases for compounds in common biochemical use are being assembled by various groups, and JCBN members are involved with this work. Examples are the ChEBI database at the European Bioinformatics Institute and the Biochemical Names Database at the University of Missouri.

II.4.3.2 Phosphorus compounds

Recommendations on the naming of phosphorus compounds of biochemical importance are being revised.

II.4.3.2 Small molecules glossary

There are plans to compile advice for biochemists for naming a selection of small molecules of biochemical importance not covered elsewhere

III. Project Development

As noted previously, Division VIII has not received any unsolicited project proposals from the community. The responsibility for developing proposals rests with the Division Committee. The Division therefore carries out scoping exercises from time to time, to establish needs and feasibility. In the most recent such exercise, a group of Division Committee members and other potentially interested people met on September 29th 2004 in Budapest to consider approaches to development of Preferred IUPAC Names for inorganic compounds, bearing in mind the almost-completed similar exercise on organic compounds The consensus was that a Task Group should be assembled and a proposal developed; however it has so far proved difficult to identify people willing to commit the considerable amount of time and effort necessary for this work to proceed..

IV. Division VIII Publications

Since January 2004, the following publications have appeared:

IV.1 Corrections to Revised Nomenclature of Organic Chemistry Section F: Natural Products, *Pure Appl. Chem.*, 2004, **76**, 1283-1292:

http://www.iupac.org/publications/pac/2004/7606/7606x1283.html

A number of errors and inconsistencies in the original Revised Section F had been pointed out in the course of translation, and this substantial Corrections and Modifications document was developed and published both in print and on the web. The changes were incorporated into the web version of Revised Section F, with links to the original text.

IV.2 Numbering of Fullerenes, *Pure Appl. Chem*, 2005, **77**, 801-923:

http://www.iupac.org/publications/pac/2005/7705/7705x0843.html

IV.3 "International chemical identifier goes online", *Chem. World*, 2005, **6**, 7:

http://www.rsc.org/chemistryworld/Issues/2005/June/this_month/International_chemical_identifier.asp

Alan McNaught
16 June 2005