Minutes of the IUPAC Chemical Nomenclature and Structure Representation Division (VIII) Committee Meeting

Boston, MA, USA, August 18, 2002

Members Present: Dr Stephen Heller, Prof Herbert Kaesz, Prof Dr Alexander Lawson, Prof G. Jeffrey Leigh, Dr Alan McNaught (President), Dr. Gerard Moss, Prof Bruce Novak, Dr Warren Powell (Secretary), Dr William Town, Dr Antony Williams

Members Absent: Dr. Michael Dennis, Prof Michael Hess
National representatives Present: Prof Roberto de Barros Faria (Brazil)

The second meeting of the Division Committee of the IUPAC Division of Chemical Nomenclature and Structure Representation held in the Great Republic Room of the Westin Hotel in Boston, Massachusetts, USA was convened by President Alan McNaught at 9:00 a.m. on Sunday, August 18, 2002.

1.0 President McNaught welcomed the members to this meeting in Boston and offered a special welcome to the National Representative from Brazil, Prof Roberto de Barros Faria. He also noted that Dr Michael Dennis and Prof Michael Hess were unable to be with us. Each of the attendees introduced himself and provided a brief bit of background information. Housekeeping details regarding breaks and lunch were announced and an invitation to a reception from the U. S. National Committee for IUPAC on Tuesday, August 20 was noted.

2.0 The agenda as circulated was approved with the addition of a report from Dr Moss on the activity on his website.

3.0 The minutes of the Division Committee Meeting in Cambridge, UK, January 25, 2002 as posted on the Webboard
(http://www.rsc.org/IUPAC8/attachments/MinutesDivCommJan2002.rtf and http://www.rsc.org/IUPAC8/attachments/MinutesDivCommJan2002.pdf)
were approved with the following corrections:

3.1 The name Dr Gerard Moss should be added to the members present listing.

3.2 The title Prof should be added to the name Dr Alexander Lawson in the members present listing.

4.0 Matters Arising from the Minutes

4.1 The terms of reference for Division VIII modified according to the discussion at the Cambridge meeting to be presented to the Bureau at their meeting September 14th are attached as Appendix I (see also
(http://www.rsc.org/IUPAC8/attachments/IUPACBureauRecomm2002.rtf or http://www.rsc.org/IUPAC8/attachments/IUPACBureauRecomm2002.pdf)

4.2 Advisory Subcommittee Membership. The current membership is given in Appendix II (see also http://www.rsc.org/IUPAC8/attachments/DivVIIISubcomMembers.rtf or http://www.rsc.org/IUPAC8/attachments/DivVIIISubcomMembers.pdf).

It was recommended that Dr Keith Taylor from MDL be added as an Advisory Subcommittee Member. He will be contacted by A. McNaught.

4.3 Alternatives to "preferred name". Because there had been reported objections to the term "preferred name", a note was posted on the Division Webboard on February 4 asking for opinions and other suggestions. Some eight responses were received and the new terms "regulatory", "canonical", "legal" and "standard" were among those suggested. The solution most preferred was to add IUPAC to preferred name, giving preferred IUPAC name, which yields the convenient acronym PIN. The latter was accepted by the Division Committee.

4.4 Java applet for stereochemical assignment. J. Wisniewski reports that he has identified a person to program the algorithm that has been developed. Accordingly, what is now needed is a project request. Clearly, J. Wisniewski is the person best suited to develop such a request and will be asked to prepare a formal project request.

4.5 Solid-state inorganic nomenclature. R. Metselaar was contacted with regard to the needs in this area and reported that he had nothing to add to the existing chapter in the Red Book (see below under future inorganic projects).

4.6 Computer-based cluster nomenclature. A. Dress was contacted by A. McNaught regarding the status of his work on this project which originated within CNIC. There was no reply. H. Kaesz noted the existence of a paper by D. Coucouvanis and T. Sloan on cluster nomenclature. He will attempt to locate this paper.

4.7 Local Contacts. A request to the Advisory Subcommittee members to put us in contact with nomenclature and education committees in their respective countries produced two responses, from J. Kahovec and P. Fodor-Csanyi.

Further suggestions for establishing local contacts were to contact editors of journals in various countries and to put an article in *The Alchemist*.

4.8 A book on Organometallic Nomenclature is planned. An initial chapter "Nomenclature of Organometallic Compounds of the Transition Elements, IUPAC Recommendations 1999" prepared for publication by A. Salzer was published in *Pure Appl. Chem.* **1999**, *71*(8), 1557-1585. Other topics to be included or expanded include nomenclature for metallacycles and for "ocenes". The issue of the color and design of the book is under discussion. A suggestion that it have a red center surrounded by blue was introduced.

5.0 Liaison with the Committee on Printed and Electronic Publications (CPEP). Steve Heller will be the liaison person with this IUPAC Committee.

6.0 IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN). Division VIII currently has responsibility for JCBN.

6.1 The minutes of the JCBN/Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) meeting in Cologne in May 2002 are given in Appendix III (see also:
http://www.rsc.org/IUPAC8/attachments/JCBNMinutes_Cologne_2002.rtf or
http://www.rsc.org/IUPAC8/attachments/JCBNMinutes_Cologne_2002.pdf).

G. P. Moss emphasized the following points:

6.1.1 Preparation of the Enzyme list began in the late 1950s with the first report approved in 1961. The latest printed edition is *Enzyme Nomenclature*, 1992 with supplements published in *Eur. J. Biochem.* in 1994, 1995, 1996, 1997 and 1999. The complete content of the published book and its supplements is now given on the IUPAC Nomenclature Web Site maintained by G. P. Moss as well as further supplements 6, 7, and 8.

6.1.2 The Enzyme Nomenclature DataBase (END) monitored by T. Kazic is a contributing database to "The Agora" (http://www.the-agora.org). Members of JCBN and NC-IUBMB have been asked to review this website.

6.1.3 A document on the nomenclature of membrane transport proteins has been approved.

6.1.4 Revision of the "Nomenclature of Carbohydrates", probably nonfunded, but IUPAC approved is under discussion.

6.1.5 A comprehensive list of abbreviations and symbols (a glossary) for biochemical compounds is needed. A project proposal is needed. The book by F. Giese, "Trivial Names in Systematic Nomenclature of Organic Chemistry" was suggested as a possible source.

6.2 The Future of JCBN. Following discussions with the Chairman of JCBN/NC-IUBMB, Prof R.Cammack and with the two Committees, A. McNaught prepared a proposal regarding the future of JCBN to be presented to the Bureau at its meeting September 14-15, 2002. It is given as Appendix IV (see also:
http://www.rsc.org/IUPAC8/attachments/BiochemNomComm-DivVIII.rtf or
http://www.rsc.org/IUPAC8/attachments/BiochemNomComm-DivVIII.pdf).

Relative to item 4.4 of the report, it was noted that a de facto group of five people works on enzymes at the present time.

7.0 Division Activities

7.1 Fullerene nomenclature, Part I. " Nomenclature for the C_{60}-I_h and C_{70}-$D_{5h(6)}$ fullerenes (IUPAC Recommendations 2002)" was published in *Pure Appl. Chem.* Vol. 74, No. 4, pp. 629-695 (2002)
(see http://www.iupac.org/publications/pac/2002/7404/7404x0629.html)
This publication is being added to the IUPAC Nomenclature Web Site maintained by G. P. Moss and a German translation of it is being prepared for publication in *Angew. Chem*. by Dr Carlo Thilgen, one of the authors of the IUPAC recommendations.
Angew. Chem. has recently introduced a program for publishing German translations of IUPAC documents.

7.2 Phane nomenclature, Part II. Modification of the degree of hydrogenation and substitution derivatives of phane parent hydrides (IUPAC Recommendations 2002) was published in *Pure Appl. Chem.* Vol. 74, No. 5, pp. 809-834 (2002).
See http://www.iupac.org/publications/pac/2002/7405/7405x0809.html

This publication is being added to the IUPAC Nomenclature Web Site maintained by G. P. Moss and a German translation of it is being prepared for publication in *Angew. Chem*. by Dr K-H. Hellwich.

7.3 Articles about the formation and operation of the Division of Chemical Nomenclature and Structure Representation (Division VIII).

7.3.1 A. McNaught has described the formation, functions, and operation of Division VIII in *Chemistry International*, Vol. 24, No, 2, March **2002,** pp. 12-14. See Appendix V and http://www.iupac.org/publications/ci/2002/2402/newschemical.html and in an interview for *The Alchemist*, a ChemWeb magazine, entitled "What's in a Name" See Appendix VI and http://www.chemweb.com/alchem/articles/1015947151360.html.

7.3.2 K.Thurlow of the Laboratory of the Government Chemist wrote about Division VIII in an article called "IUPAC's New Vision" in VAM Bulletin, 2002, Issue 26, p. 23. See Appendix VII or http://www.vam.org.uk/newswasap/newsdocs/220.pdf

7.4 Articles about the IUPAC Chemical Identifier (IChI) project appeared in *The Alchemist*, entitled "That IChI Feeling" (See Appendix VIII and http://www.chemweb.com/alchem/articles/1015947904091.html), and in *Nature* Vol 417, May 23, 2002 entitled "Chemists synthesize a single naming system" (See Appendix IX or http://www.rsc.org/IUPAC8/attachments/IChI-Nature502.pdf)

7.5 A summary of phane nomenclature principles by H. Favre and W. H. Powell is to be published in the book "Cyclophane Chemistry for the 21st Century" (Research Signpost) scheduled to appear in December of this year.

7.6 A short article entitled "Chemical Nomenclature, Terminology, Units and Symbols" is to be published in Science Editors Handbook (European Association of Science Editors). See Appendix X or http://www.rsc.org/IUPAC8/attachments/IUPAClistforSEH.rtf or http://www.rsc.org/IUPAC8/attachments/IUPAClistforSEH.pdf

7.7 IUPAC Chemical Identifier (IChI). Steve Heller distributed the minutes of the IUPAC IChI project meeting held on June 30 in Columbus, OH. The minutes of the meeting are attached as Appendix XI. He also reported that CAS has done some testing of the IChI algorithm and reported no serious problems. There will be an article on IChI appearing in *Chem. Eng. News* later this year. When Version 1.0 is available, more publicity is needed. Further testing on a variety of files, such as those of CAS, MDL, and ACD needs to be done. Version 1.0 should be tested on a file that contains duplicates.

H. Kaesz noted that there is a need to emphasize the distinction between databases, such as those of CAS and Beilstein and the facilities offered by IChI.

Since the meeting in Columbus, considerable progress has been made in the areas of salts, organometallics, and inorganic compounds, and similar extensions might

enable the algorithm to be applied to alloys, phase diagrams, and other areas. By adding a 'disconnected metal' layer, it looks like all chemicals that can be represented by a drawn structure might be handled by the IChI algorithm. The output will depend on input, for example, Na-O- *vs*. Na^{+} $^{-}$O-. To test this approach, it would be very helpful to have a large file of organometallic and inorganic structures such as is in the Gmelin database. It was agreed to contact Gmelin informally. A. Lawson will talk with Prof Dr Heindirk tom Dieck.

Version 1.0 of the algorithm should be ready by the end of this year.

What to call the algorithm has been a subject of recent discussions. IChI has been used both internally and in publications to refer both to the project and to the algorithm. Whatever is chosen, it must be short and pronounceable. The following were among the suggestions made.

II - IUPAC Identifier
ICID - IUPAC Chemical Identifier
CHILD - Chemical Identifier Label and descriptor
CHESS - Chemical Exchange Structure String
STRING - Structure-Text Reversible IUPAC Notation Generator
EXCITING - Exchange Chemical Identifier Translating IUPAC Notation Globally
STEIN - Structure-Text Exchange IUPAC Notation

[*Secretary's note:* More were suggested at the ACS Nomenclature Committee Open Meeting the following day: USTRI - Universal Structure Identifier and ISTRI or ISID or ISId - International (IUPAC) Structure Identifier]

7.7.1 A CAS/IUPAC Conference on Chemical Identifiers and XML for Chemistry was held on July 1 in Columbus, OH. The abstracts of papers are given as Appendix XII or http://www.rsc.org/IUPAC8/attachments/CAS-IUPACConf-DivVIII.pdf

7.8 Preferred names in the nomenclature of organic compounds (the new Blue Book). A first draft of the new edition of "Nomenclature of Organic Chemistry" was completed earlier this year. It runs to nearly 1000 manuscript pages. H. Favre, W. Powell, and A. McNaught spent two days earlier in the week, August 14-15, discussing various more difficult aspects. An informal working group met on Friday, August 16 (see item 7.9) to deal with aspects of the nomenclature of natural products and related compounds which will be incorporated into Chapter 9. It is hoped that a draft will be ready for review by the end of this year.

7.9 Natural Products and Related Compounds. Provisional recommendations for the nomenclature of natural products were published in *Pure Appl. Chem*. in 1976 and as Section F in the 1979 edition of the Nomenclature of Organic Chemistry. Revised, fully approved recommendations were published as "Revised Section F: Natural Products and Related Compounds, IUPAC Recommendations 1999" in *Pure Appl. Chem*. **1999**, *71*(4), 587-643. A number of problems in these recommendations had been reported, especially by G. Moss in preparing this publication for his IUPAC

Nomenclature Web Site and K-H. Hellwich during his translation of them for *Angew. Chem.* Hence, a group consisting of H. Favre, W. Powell, P. Giles, Jr., K-H. Hellwich, and A. McNaught, met on August 16 to consider the problems that had been found. The results of this meeting will be collated, published as errata, noted as errors on the IUPAC Nomenclature Web Site, and incorporated into Chapter 9 of the new Blue Book.

7.10 Revision of the Red Book (Nomenclature of Inorganic Chemistry). A status report from Prof N. Connelly, coordinator of the Red Book revision project, had been received. It was expected that a complete draft would be available for review by the end of October; it would be posted as separate chapters on the Division Webboard. The need for a good index for the new Red Book was emphasized.

7.11 Alignment of inorganic and organic preferred names. Although there are chemical compounds that clearly are better named by inorganic nomenclature principles and there are chemical compounds that clearly are better named by organic nomenclature principles, there are chemical compounds that can be named by either set of principles equally well. These occur mainly with the metals of Groups 13, 14, 15, and 16. Difficulties also occur with inorganic acids and their noncarbon derivatives. A task group consisting of T. Damhus and R. Hartshorn representing inorganic interests; H. Favre and W. Powell representing organic interests; and G. Moss representing the biochemical interests has met twice, both in Cambridge, September, 2001 and January, 2002. The minutes of these meetings have been posted on the Division VIII Webboard. An initial proposal for dividing chemical compounds into those for which inorganic nomenclature principles would be preferred and those for which organic nomenclature principles would be preferred, also posted on the Webboard, should form the basis for a continuing discussion, which does not seem to be occurring. However, it probably will be the basis for the revised organic Blue Book, and should be considered as the basis for selecting preferred names for inorganic structures when this work begins.

7.12 Fullerenes, Part II. Nomenclature for the C_{60}-I_h and C_{70}-$D_{5h(6)}$ fullerenes and their derivatives was published in *Pure Appl. Chem.* Vol. 74, No. 4, pp. 629-695 (2002) (see above). A project group has been established to develop IUPAC recommendations for naming and numbering other fullerenes. So far, procedures have been developed for naming other fullerenes with axial symmetry and a contiguous numbering pathway, fullerenes with C_s symmetry, and fullerenes with C_1 symmetry. The project group is now working to find procedures for numbering axial fullerenes without a contiguous numbering. The plan is to meet together early next year and hopefully have a draft of a second paper ready for final consideration at a meeting at the Ottawa General Assembly.

7.13 Rotaxanes and Catenanes. Responsibility for the project "Nomenclature for Rotaxanes and Catenanes" (35/2000), initiated by the Nomenclature Commission of the Macromolecular Division, was transferred to Division VIII following the dissolution of the Commission at the end of 2001. The project underwent two splits. First, it was divided into nomenclature of macromolecular rotaxanes and molecular rotaxanes and catenanes, that is, discrete molecules. Thus, the project "Nomenclature for Macromolecular

Rotaxanes" (2000-037-1-800) was defined; the project chairman is E. S. Wilks. A revised draft has been circulated for comment.

E. S. Wilks is also the project chairman for the molecular project that deals with discrete molecules: "Nomenclature for Discrete Rotaxanes and Catenanes" (2002-007-1-800). He reported that nomenclature for catenanes has now been removed from this project and that 'glossary', 'history', 'types', and 'rules' sections have been drafted; examples to illustrate the rules still need to be found. Members of the task group are: G. Moss, A. Yerin, A. Harada, and W. V. Metanomski. It is planned that the two projects should be completed at the same time. Accordingly, a catenanes project needs to be initiated.

7.14 Macromolecular projects (reports from Beijing meetings).

7.14.1 Macromolecular Rotaxanes (see also 7.13) Because of lengthy and difficult discussions in the working group a conceptually new draft will be produced by the end of 2002 to be discussed in Ottawa.

7.14.2 Chemically Modified Polymers.(1999-051-1-800). E. S. Wilks project leader. Intensive and controversial discussions resulted in the decision to remove all discussion of multi-stage modifications. B. Novak will look at journal reviews for possible task group members from academia.

7.14.3 Terminology and Nomenclature of Macromolecules with Cyclic Structures (2001-082-1-800). W. Morman project leader. Contributions from external experts have been obtained. A new draft will be prepared by the end of the year.

7.14.4 Terminology and Structure-based Nomenclature of Dendritic and Hyper-branched Polymers (2001-081-1-800). J. Kahovec project leader. Good progress was made and a new draft will be available soon. Drawing of the structures is very time-consuming.

7.14.5 Purple Book Revision. The 1991 edition of the "Compendium of Macro-molecular Nomenclature" has been scanned into an electronic form. The new edition will have 17 chapters, eight of which deal with terminology. Two chapters need to be redone.

7.15 Stereochemistry scoping exercise. R. Hartshorn, convenor of a small working group charged with identifying areas within the area of stereochemistry where projects might be required, submitted the report given in Appendix XIII (see also: http://www.rsc.org/IUPAC8/attachments/DivVIIIStereo02.doc).
It was agreed that the first effort should be a revision of the stereochemistry glossary. R. Hartshorn will be asked to draft a proposal for a project.

It was decided to delay work on providing a clear explanation of the CIP rules pending the completion of the Blue Book chapter on stereochemical nomenclature.

There is a very real need to reconsider the distinction between conformation and configuration; again this may wait until the completion of the Blue Book chapter on stereochemical nomenclature.

Stereochemical nomenclature in bioconjugate chemistry probably should be a JCBN project.

Stereochemical nomenclature for supramolecular compounds such as knots, rotaxanes, and catenanes involves mainly chirality of molecules and should be revisited following completion of projects dealing with their nomenclature (see 7.13 and 7.14.1). At the present time there is no work planned on nomenclature for knots. It was also noted that there are large fullerene cages that probably fall within this category.

7.16 Scoping exercise on structure representation. J. Brecher had been asked to look into the problem of identifying the kind of guidelines for drawing structures that would be helpful to the chemical community. His interim report is given in Appendix XIV (see also: http://www.rsc.org/IUPAC8/attachments/StructScop0802.rtf or http://www.rsc.org/IUPAC8/attachments/StructScop0802.pdf)
E-mail correspondence has stressed stereochemical aspects of drawing structures. Broader perspectives, for example, how to "clean-up" a drawing of a structure such as adamantane, should not be neglected

There is an e-mail based discussion among about a dozen people archived at http://groups.yahoo.com/group/iupacstructures/ Anyone interested in this discussion should contact J. Brecher at jsb@cambridgesoft.com to be added to the mailing list.

It was suggested to contact Dr. Wolf-Dietrich Ihlenfeldt, an Advisory Subcommittee Member, regarding a paper or presentation on structure drawing.

7.17 Organometallic nomenclature. A book on organometallic nomenclature is planned (see item 4.8). A paper "Nomenclature of Metallacycles of the Transition Metals" authored by H. Kaesz, J. Casey, H. Favre, and Y. Yamamoto, is being recirculated for comment. Organometallic nomenclature is a subset of a set of inorganic projects, which are being planned. H. Kaesz, B Novak, and J. Leigh met to plan for a meeting to scope out future inorganic projects. They suggested a meeting involving J. Leigh, N. Connelly, H. Kaesz, B. Novak, T. Damhus, and G. Kanatzidis (Michigan State) around Easter of 2003 to formalize the following projects and to identify others.

Organometallic nomenclature
Solid state nomenclature
Stereochemical descriptors for coordination numbers 7,8, and 9
CIP rules for coordination compounds
Clusters
Boron nomenclature

7.18 Preferred names for inorganic compounds. This will be included in the inorganic project scoping exercise (see 7.17).

7.19 Phane nomenclature, further development. Now that the fundamentals of phane nomenclature have been published, applications in other areas of nomenclature need to be developed. W. Powell received a contents table and some other material from the editor of the book "Cyclophane Chemistry for the 21st Century" which has not yet been analyzed for potential applications of phane nomenclature. This powerful new type of nomenclature needs to be publicized, such as by an article in the "The Alchemist" or by

PowerPoint presentations. A short announcement will appear this fall in *Chemistry International* and a chapter on "Phane Nomenclature" was written by H. Favre and W. Powell for the book "Cyclophane Chemistry for the 21st Century", which should appear by the end of the year.

7.20 Boron nomenclature. Literature search to identify potential task group members has not yet been done. Perhaps this should be placed with the inorganic scoping exercise.

7.21 Nomenclature for miscellaneous compounds of biochemical importance. There was a need for two projects in this area:

- nomenclature advice on miscellaneous compounds not covered elsewhere
- abbreviations and symbols

These areas should be addressed by JCBN.

7.22 Synonyms database for compounds in common biochemical use. See item 6.1.5 and minute 13 of the JCBN minutes (Appendix III). A project proposal is awaited.

8.0 Report from the Committee on Printed and Electronic Publications (CPEP). S. Heller reported that the meeting in Cambridge on July 11-12, 2002 dealt primarily with statistics on publications and books, but that it was intended that future meetings would focus on technical standards for electronic publishing, of which the IUPAC Chemical Identifier would be an integral part. He also noted that the XML project has not yet been finally approved. This project began in Cambridge at a discussion meeting on January 24-25, 2002 with a proposal by T. Davies which was refined by S. Stein as proposed Project Chairman.

9.0 Division Correspondence.

9.1 H. Kaesz reported receipt of two items from Dr. Peter Bayliss relative to the nomenclature proposal "Chemical Adjectival Modifiers on Mineral Species". The first was a compilation of comments from members of the Commission on New Minerals and Mineral Names (CNMMN) of the International Mineralogical Association (IMA) and the second was his reply to the criticisms of these comments.

9.2 A. McNaught received via the IUPAC Secretariat a request from the Secretary of the Nomenclature Committee of the International Union of Pharmacology (NC-IUPHAR) for advice on including chemical structures of receptor agonists and antagonists in a web database of all the receptors in the human genome. For a description of the activities of NC-IUPHAR see
http://www.rsc.org/IUPAC8/attachments/NC-IUPHAR.pdf
or http://www.rsc.org/IUPAC8/attachments/NC-IUPHAR.rtf

After receiving opinions from Committee Members, he had replied, making it clear that we are happy to offer advice on nomenclature problems in this connection, but that IUPAC does not have access to databases containing the kind of material needed. He had suggested that IUPHAR might talk directly to potential suppliers, and also that a message to the listserver CHMINF-L might prove fruitful.

9.3 Nomenclature of Cyclic Organic Compounds (letter from A. A. Durgaryan). W. Powell received the following letter from Prof A. A. Durgaryan, Department of Chemistry, Yerevan State University, Armenia.

> "I would greatly appreciate receiving the opinion of your committee about my proposals concerning the nomenclature of cyclic organic compounds which had been published in *Chem. J. of Armenia* **1999** *52* (1-2), p.180. I am attaching the text of my proposals."
>
> Respectfully yours,
> Prof. A.A.Durgaryan
> Department of Chemistry
> Yerevan State University
> Alek Manukyan 1
> 375025-Yerevan, Armenia

The text of his proposals is given in Appendix XV (see also:
http://www.rsc.org/IUPAC8/attachments/Durgaryan.rtf or
http://www.rsc.org/IUPAC8/attachments/Durgaryan.pdf

The Committee saw no reason for devising simple nomenclature system such as he suggests. Existing systems are too well established. Hence, new systems using fundamentally different principles are simply not useful. A. McNaught will write to Prof. Durgaryan along these lines.

10.0 Administrative Matters. Several items have been forthcoming from the Secretariat.

10.1 Revised Project Guidelines for Divisions/Standing Committees. See
http://www.rsc.org/IUPAC8/attachments/ProjGuide.rtf or
http://www.rsc.org/IUPAC8/attachments/ProjGuide.pdf

10.2 Guidelines on IUPAC Criteria for Retrospective Project Evaluation. See
http://www.rsc.org/IUPAC8/attachments/RetroProjGuide.pdf

10.3 Liaison with National Adhering Organizations, and the role of National Representatives. See
http://www.rsc.org/IUPAC8/attachments/NAOLetter.pdf and
http://www.rsc.org/IUPAC8/attachments/NRHandbook.pdf

10.4 Changes in IUPAC Secretariat personnel and functions. See
http://www.rsc.org/IUPAC8/attachments/SecretariatChanges.pdf

10.5 Publicity

10.5.1 Database of editors of chemistry journals. Should the Secretariat be asked to create such a database? Yes. Perhaps the Bureau should be approached about such a project. It might be possible to create such a database using student help. Even if created in this manner, it would still have to be maintained.

10.5.2 Articles for *Chem. International*. For each publication it would be helpful to prepare an article for CI pointing out changes in previously published recommendations or outlining new recommendations, for instance, in the Red Book I revision or in the new Blue Book.

11.0 Other Business

11.1 G. Moss presented statistics for the Nomenclature World Wide Web Database. Total IP address usage to date is about 1,400,000. Data from 170 countries have been recorded. Summary data for 1996-2001 can be found in Appendix XVI and at http://www.chem.qmul.ac.uk/iupac/usage. For full details on each document see http://www.chem.qmul.ac.uk/iupac or http://www.chem.qmul.ac.uk/iubmb . It was suggested that this data would make a nice article for *Chem. International* (see 10.5.2).

G. Moss noted that the atomic weights for the Periodic Table given in the World Wide Web Nomenclature Database included the 2001 updates to the 1999 recommendations, but those on the IUPAC Web Site under the Inorganic Division publications had only the 1999 recommendations.

11.2 Managing nomenclature programs. With regard to a suggestion that IUPAC should work to make name-generating programs consistent with current nomenclature recommendations, it was noted that IUPAC does not have resources to monitor programs, even nomenclature programs.

11.3 It was noted that there is no longer a live link from the IUPAC Web Site to the free version of Autonom; the only free naming service linked is now that of ACD.

12.0 Next Meeting. According to the draft schedule for the Ottawa meeting from the Secretariat, the Division VIII Committee will meet on August 9-10, 2003. Project Group Chairmen have been contacted with regard to their desire to schedule project meetings August 6,7 or 8. It is planned to have a Division VIII Open Meeting on August 8 in the afternoon.

Respectfully submitted: Warren H. Powell (Secretary) October 14, 2002

Approved: Alan D. McNaught (President) October 17 , 2002

APPENDIX I

IUPAC Division of Chemical Nomenclature and Structure Representation

Recommendations to the IUPAC Bureau, September 2002

1.0 *Division Terms of Reference*

Division VIII Terms of Reference were approved by Council at the 2001 General Assembly in Brisbane. They are reproduced below, showing a minor change recommended by the Division Committee at its meeting on January 26th 2002. Item 6 was considered not entirely realistic, and the modified text reflects this view. Bureau approval for the change is requested

The Division is responsible for maintaining and developing standard systems for designating chemical structures, including both conventional nomenclature and computer-based systems. This responsibility is to be fulfilled by:

1. Identifying the needs of the user community.
2. Generating projects arising from those needs.
3. Identifying project leaders and task groups to carry out the work.
4. Administering approved projects financially, monitoring their progress, and approving resulting recommendations for review by established IUPAC procedures.
5. Identifying new sources of expertise and enabling their involvement in projects.
6. As far as possible, ensuring that nomenclature systems projects and the resulting recommendations are compatible with each other, with established IUPAC recommendations, and with computer-based systems for manipulating chemical names and structures.

2.0 *Division Rules*

The Division Committee has considered the draft Division Rules circulated to other Division Presidents on January 5th 2001, and requests approval for the modifications indicated in the attached copy. Specifically:

2.1 Item 1. Modification of the mission statement to correspond with the Terms of Reference.

2.2 Item 5(a). Change in the composition of the nominating committee from two to three members of the Division Committee and from three to two members from outside IUPAC. The Division VIII Committee felt that its business was more specialised than that of other Divisions, and that it would be more difficult to identify suitable external people. Also some Committee members who had been involved with nominating

committees for other Divisions had found the input from externals not particularly well-informed or helpful.

2.3 Item 5(c). The Committee considered it unnecessary to specify the sectors of the community from which nominees should be drawn. For Division VIII it would be important to have representation from publishing and from professions concerned with regulatory requirements, but is was not thought necessary to spell this out.

2.4 Item 6. It was thought helpful to specify "fair practices" in connection with procedures for elections to be defined by the Secretariat.

2.5 Item 6 again. This relates to ongoing discussions in all Divisions about defining the Division electorate. The Division VIII Committee is concerned to establish an electorate of a size and composition such as to enable a variety of interests to be properly represented and to ensure an acceptable turnover of Division Committee membership. This would be most readily achieved by including the members of the Division's only Subcommittee (the Advisory Subcommittee), consisting of about 40 people representing the whole range of interests in the Division's activities. The Division Committee is concerned to establish a sense of community amongst these individuals, and including them as part of the electorate would be an important contribution to this end. In addition, it would be appropriate to include Chairmen of current Task Groups, although in most cases it is probable that these people would also be members of either the Division Committee or the Advisory Subcommittee. If the IUPAC-IUBMB Joint Commission on Biochemical Nomenclature persists in some form, as a body reporting to the Division Committee, its members should also be included in the electorate.

Alan McNaught
(President, IUPAC Division VIII)

Chemical Nomenclature and Structure Representation Division

DIVISION RULES

1. The mission of the Chemical Nomenclature and Structure Representation Division is to maintain and develop standard systems for designating chemical structures. Terms of Reference are attached.

2. Under the Statutes, Bylaws, and policies of the Union, the Division is managed by its Division Committee. S10 and B4.1 and their subsections are particularly relevant. The Division Committee is responsible for initiating and managing scientific projects, symposia and other activities within its area of responsibility and for cooperating with other Divisions in initiating and managing interdisciplinary projects, symposia and other activities.

3. The composition of the Division Committee and terms of service of its members are given in B4.103. In addition, under policy established by the Bureau, up to six National Representatives may be elected to the Division Committee.

4. (a) Titular Members of the Division Committee are nominated and elected by the electorate defined in B4.103. Candidates for titular membership are nominated by the Nominating Committee described below.

   (b) Associate Members may be elected by the Division Committee, as provided in B4.103.

   (c) National Representatives are elected by the Division Committee on nomination by National Adhering Organizations, according to procedures defined by the Bureau.

   (d) Interim appointments to fill vacancies on the Division Committee occurring between meetings may be made by the Division President for terms ending at the end of the year in which the next General Assembly is held.

5. Candidates for Titular Member of the Division Committee are named by a Nominating Committee, prescribed by IUPAC policy and procedures defined by the Bureau, as follows:

   (a) The nominating committee consists of five members [subject to an exception by the Bureau], with no more than three members from the existing Division Committee and the other two chosen from outside IUPAC on the basis of the breadth of their expertise. The Division President will not be a member of the Nominating Committee.

   (b) The Nominating Committee is appointed by the Division President with the concurrence of the IUPAC Executive Committee.

   (c) Nominees should be scientists with the expertise and experience necessary to maintain a Division Committee with appropriate scientific stature and breadth. Categories of vacancies may be established by the Division Committee if desired to ensure diversity in subject matter, geographic distribution, or other characteristics. More than one nominee for each vacancy is desirable but not required.

6. Elections shall be conducted by e-mail (or ordinary mail as necessary) under fair practices defined by the IUPAC Secretariat.

7. The Officers of the Division, as prescribed by B4.103 and B4.104, are President, Vice President (designated as President elect), and Secretary, with the immediate Past President serving as a Titular Member. With the advice of the President of the Union, Officers of the Division are elected by the Division Committee, subject to final approval by the Council. The Officers together form an Executive Committee to act for the Division Committee between meetings. The duties of the Officers are as follows:

   (a) The President is the administrative head of the Division, presides at meetings of the Division Committee, and is an *ex officio* a member of all bodies of the Division. The President serves as a member of the Bureau and is the principal representative of the Division within and outside the Union.

   (b) The Vice President acts for the President in his absence, assists the President as requested, and serves on the Division executive committee. He shall assume the office of Division President in the event of the President being unable to perform the functions of that office, without prejudice to the forthcoming period of office as President.

   (c) The Secretary assists the President in carrying out the business of the Division and maintains the records of the Division.

8. (a) The Division Committee may establish and the Division President may appoint subsidiary bodies, such as working parties and advisory groups, which will have the status of Division subcommittees, as described in S 10.6. Task groups may be formed to carry out specific projects. The terms of reference or charge to each group, as well as its lifetime, shall be established by the Division Committee.

   (b) The Division Committee may propose to the Bureau the establishment of Commissions, with terms of reference and lifetimes, under the provisions of B4.301.

   (c) The Division Committee and Division President will exercise responsibility and oversight over all bodies created under parts (a) and (b).

9. These Rules may be amended by the Division Committee, subject to approval by the Council.

Terms of Reference

The Division is responsible for maintaining and developing standard systems for designating chemical structures, including both conventional nomenclature and computer-based systems. This responsibility is to be fulfilled by:

1. Identifying the needs of the user community.
2. Generating projects arising from those needs.
3. Identifying project leaders and task groups to carry out the work.
4. Administering approved projects financially, monitoring their progress, and approving resulting recommendations for review by established IUPAC procedures.
5. Identifying new sources of expertise and enabling their involvement in projects.
6. As far as possible, ensuring that nomenclature systems projects and the resulting recommendations are compatible with each other, with established IUPAC recommendations, and with computer-based systems for manipulating chemical names and structures.

Terms of Reference of the Advisory Subcommittee

1. To advise the Division Committee on the needs of the community with respect to standard systems for designating chemical structures, including both conventional nomenclature and computer-based systems.

2. To propose and participate in projects for the Division, and to advise on project leaders and other suitable participants.

3. The Advisory Subcommittee will not meet in full on a regular basis; the Division Officers will authorise meetings of subgroups as needed to address specific issues, and will call larger meetings when it seems necessary to consider major issues of general interest. Otherwise, discussions will take place via a Web Discussion Board.

4. Members of the Subcommittee are appointed by the Division Officers for two-year periods, renewable as approved by the Division Committee.

APPENDIX II

IUPAC DIVISION VIII

Division of Chemical Nomenclature and Structure Representation

Advisory Subcommittee composition (as of 16/04/02)

Dr Hidetsugu Abe (Toyohashi U of Technology, Japan)
Prof Steven M Bachrach (Trinity U San Antonio, USA; Editor, Internet Journal of Chemistry)
Dr Byron J Bossenbroek (Chemical Abstracts, USA)
Mr Jonathan Brecher (CambridgeSoft, USA)
Dr John Brennan (European Patent Office, Netherlands)
Prof Neil G Connelly (Bristol, UK)
Prof Richard Cammack (Kings, London, UK; Chairman IUPAC-IUBMB Joint Commission on Biochemical Nomenclature)
Dr Ilaria Campagnari (GSK, Italy)
Prof Chong Shik Chin (Seoul, Korea)
Prof Franco Cozzi (Milan, Italy)
Dr Ture Damhus (Novozymes, Denmark)
Prof Bernadette Donovan-Merkert (U of North Carolina, Charlotte, USA)
Prof Andreas Dress (Bielefeld, Germany)
Dr Andrey Erin (ACDLabs, Russia)
Dr Geoff Fairhurst (BASF, Germany)
Prof Henri A Favre (Montreal, Canada)
Dr Piroska Fodor-Csányi (Budapest, Hungary)
Dr Patton M Giles (Chemical Abstracts, USA; ACS Nomenclature Committee)
Dr Jonathan M Goodman (Unilever Centre for Molecular Informatics, Cambridge, UK)
Prof Richard M Hartshorn (Canterbury, New Zealand)
Dr Karl-Heinz Hellwich (Beilstein, Germany)
Prof Bernardo J Herold (Lisbon, Portugal)
Dra. Rita Hoyos de Rossi (Cordoba, Argentina)
Dr Alan T Hutton (Cape Town, South Africa)
Dr Wolf-Dietrich Ihlenfeldt (Computer Chem Center, Erlangen-Nurnberg, Germany)
Prof Aubrey D Jenkins (Sussex, UK)
Prof Jaroslav Kahovec (Prague, Czech Republic)
Prof Alan R Katritzky (Florida Center for Heterocyclic Compounds, USA)
Professor Risto S Laitinen (Oulu, Finland)
Dr Graham F McCann (Royal Society of Chemistry, UK; Editor, Dalton Trans. And J Materials Chem)

Dr W Val Metanomski (Chemical Abstracts, USA)
Prof Ebbe Nordlander (Lund, Sweden
Prof Vincent L Pecoraro (Michigan, USA; Assoc Editor, Inorg Chem)
Prof C Dale Poulter (Utah, USA; Editor, J Org Chem)
Prof Damon D Ridley (Sidney, Australia)
Dr Paolo Righi (Milan, Italy)
Ms Helen Schofield (Manchester, UK)
Dr Steve Stein (NIST, USA)
Dr Sarah Thomas (Royal Society of Chemistry, UK; Editor, ChemComm)
Mr Kevin Thurlow (LGC Nomenclature Advisory Service, UK)
Dr Edward S Wilks (ex-Dupont, USA)
Dr Janusz L Wisniewski (MDL, Germany)
Dr Shen-Gang Yuan (Shanghai, China)

APPENDIX III

DRAFT

Nomenclature Committee of IUBMB (NC-IUBMB) – IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN)

Cologne, May 10-12, 2002

Attendees:

NC-IUBMB and JCBN

Richard Cammack, Chairman (London, UK)
Derek Horton (Washington, DC, USA)
Arnost Kotyk (Prague, Czech Republic)
Keith Tipton (Dublin, Ireland)

NC-IUBMB

Helen Berman (Piscataway, USA)
Gerard Moss (London, UK)

JCBN

Sinéad Boyce, Secretary (Dublin, Ireland)
Keith Elliott (Manchester, UK)
Minoru Kanehisa (Kyoto, Japan)
Nathan Sharon (Rehovoth, Israel)

Others

Rolf Apweiler (Cambridge, UK)	Associate Member of NC-IUBMB
Alan McNaught (RSC, Cambridge, UK)	Associate Member of NC-IUBMB
Dietmar Schomburg (University of Köln, Germany)	Associate Member of NC-IUBMB
Michael Darsow (Cambridge, UK)	Observer
Kirill Degtyarenko (Cambridge, UK)	Observer
Astrid Fleischmann (Cambridge, UK)	Observer
Jiahui Jiang (University of Missouri, Columbia, USA)	Observer
Ida Schomburg (University of Köln, Germany)	Observer

1. Welcome and Apologies for Absence

The meeting began at 10 am on Friday 10 May. Cammack, as Chairman, welcomed everyone to the meeting and thanked Schomburg for making the arrangements. Apologies from Amos Bairoch, Charles Cantor and Alan Barrett were noted.

2. Approval of Agenda

Moss asked that a document on the terminology of guanidine formation, which he had received during the past week, be added as an agenda item. It was agreed that this be added for discussion under Item 20 (Any Other Business).

3. Minutes of the Bethesda Meeting, May 2001

Cammack proposed a different format for the Minutes in future. These would take the form of topics covered and decisions taken rather than a detailed account of all discussions that took place, as had been our previous practice. This was agreed.

The provision of the Agenda and other documents related to forthcoming meetings of the Committees on a temporary web page (http://www.kcl.ac.uk/ip/richardcammack/nomenclature/agendapapers.html), initiated this year by Cammack, was welcomed.

Kotyk pointed out an inaccuracy in Item 5 of the Minutes: he had not indicated that he would resign after the meeting in 2002 but that he would be prepared to step down once the transport protein document had been approved. The minutes of the meeting were approved subject to that change being made. (**Action: Boyce** to amend 2001 Minutes as indicated).

4. Matters Arising

Cammack reported that he had had a complaint from the students at a course on Cell Biology in Moscow concerning the lack of a systematic nomenclature for regulatory proteins. He had prepared a response indicating that the Committees cannot create such a systematic nomenclature until the properties of these proteins are known. It was agreed that Cammack would ask the IUBMB Education Committee to look into this issue. (**Action: Cammack**).

5. Chairman's Report (Cammack)

Cammack congratulated Tipton, Dixon and Boyce for their work on the Enzyme List, Moss for his work on the web site, McNaught for his work on providing future options for the IUPAC part of the JCBN, Kotyk for his transport document and Berman for her document on the geometry of nucleic acids.

Cammack announced that he is editor-in-chief of a new edition of the *Oxford Dictionary of Biochemistry and Molecular Biology* (Oxford University Press). He asked members of the Committees for help in creating definitions of terms. (**Action: Cammack** to contact committee members)

Cammack informed those present that he had attended a meeting on XML in chemistry. IUPAC has adopted the use of XML and will define namespaces that can be used to verify standards and statements.

The Committees learned, with sadness, of the death of the first Chairman of the NC of IUB and IUPAC, Peter Karlsson.

6. Membership of Committees

The membership of the Committees is to remain unchanged. Kotyk suggested that he might retire if his transport document (see Item 10) was fully approved. He anticipated that this would not be before the 2003 Meeting of the Committees, at the earliest.

Cammack also raised the need for new people to join the Committees, but said that funding would be needed for such people. It was agreed not to consider additional full members until we had considered our future activities and initiatives more fully.

6a. Newsletter

It was agreed that newsletters should be published on a regular basis, and that an editor should be appointed. This was agreed and Sharon volunteered to collate newsletter entries. Items for publication in the Newsletter are to be sent to Sharon.

Moss emphasized the importance of alerting the scientific community to changes that we make in enzyme nomenclature. In addition to the *European Journal of Biochemistry*, who published newsletters in the past, it should be offered to a wide range of journals, including the IUBMB journal *Biochemistry and Molecular Biology Education* (BaMBEd). Items of particular importance might be published in *Trends in Biochemical Sciences* (TiBS). The Newsletter will also be published on the web. It was agreed that the title “Newsletter” was not particularly attractive and “Bulletin” was suggested as an alternative. The Committees agreed to reflect upon this.

(**Action: Sharon** to compile Newsletter; **Members of the Committees** to provide Sharon, cc. Cammack, with suggestions).

7. Appointment of a Treasurer

A Treasurer is required to ensure that a central record of all expenditure is maintained. Elliott agreed to act as Treasurer and this was accepted unanimously. Members of Committees are to submit claims in their normal manner but are to send a copy to Elliott. (**Action: All members of Committees** to send copies of expenses claimed to Elliott)

8. Funding (Cammack)

IUPAC funding has been fixed at $13,000 for the biennium 2002-2003 and the IUBMB has provided a budget of $25,000 over the 3 years to 2003. Abolition of the commissions of IUPAC meant that their future support for the Biochemical Nomenclature Committees required a special case to be made. It was agreed that, even if our activities become project-based, we should seek to continue to have an annual meeting to assess progress and decide on future projects. This was discussed further under Item 12.

Cammack pointed out that the 3-year funding provided for by Bill Slater of IUBMB expires at the end of next year. He will produce a report for the IUBMB on the work of the Biochemical Nomenclature Committees and make a case, in that context, for continued (and extended) funding to the Treasurer, Brian Beechey. (**Action: Cammack**)

9. Information on Nomenclature Databases

9.1 Update on the BioBabel project (Fleischmann, Degtyarenko and Darsow)

This EU-funded project began on 1 November 2001 and will run for 3 years. The initial stage of the project is to develop a database to hold the enzyme nomenclature data. Work on this has begun and the data can be seen at http://www.ebi.ac.uk/intenz.

9.2 Progress with Toni Kazic's Enzyme Nomenclature Database (END) (Jiang)

Jiang gave a presentation on the contributing databases in The Agora [END, BND (Biochemical Nomenclature Database), Moirai (a small-molecules database) and UMBBD]. Facilities are provided to query, review, revise, curate and deposit data in the database. Jiang asked that members of the Committees view the website (http://www.the-agora.org) and inform Kazic of any suggestions they have for its improvement. Jiang also asked that members provide answers to questions raised in the document from Kazic on ambiguities and uncertainties in small molecule nomenclature that was distributed at the meeting.

9.3 Update on BRENDA (Schomburg)

Schomburg reported that a new enzyme query engine had been introduced in the database. They now have the ability to review and update 1000 enzyme entries per year. They have a new facility that can extract data by text interpretation, which can be used to link EC numbers to specific diseases. They will also be introducing a substructure editor within approximately two months that can be used to search for a structure using a substructure.

9.4 Update on KEGG (Kanehisa)

This database has introduced a split of EC number information into R numbers for reactions (based on function) and K numbers for orthologs (KO - KEGG Orthology). They also use C numbers for compounds (C0 - chemical ontology). They have approximately 10,000 chemical compounds in the LIGAND database and have a web interface for searching for similar chemical structures. They have developed a method to compare chemical structures using 2D graph comparison and a graphic similarity (correlated cluster) algorithm. Substructure searches, by inputting chemical structures, are now possible on the GenomeNet homepage.

Kotyk raised the subject of 3D structures of substrates and cofactors. Berman said that Cambridge Database (CCD) (a private charity) had 250,000 such structures. She suggested that Schomburg should contact them as they were

now more willing to collaborate. (**Action: Schomburg** to contact Cambridge Database about possible collaboration regarding 3D structures for small molecules).

9.5 Nicholson maps and minimaps (Nicholson)
Nicholson discussed the value of his maps and minimaps for disseminating information to students and said that the use of the internet has made their distribution easier. All proceeds from sales by Sigma of the metabolic pathway map go to IUBMB (not Biochemical Nomenclature Committee). Nicholson has been trying to get the balance redressed. (**Action: Nicholson** to discuss proceeds from maps with Angelo Azzi). Nicholson also demonstrated an animated version of ATP synthesis that was being developed in consultation with John Walker.

10. Kotyk's Transport Protein Document (Kotyk)
The document produced by Kotyk, in collaboration with Milton Saier, was approved and will be published on the web as an IUBMB-Nomenclature Committee document after minor corrections are made. Kotyk asked members of the Committees to give him details of any necessary corrections. (**Action: All members of the Committees**)
Moss and Cammack agreed to work with Kotyk in editing the document and providing a title page and table of contents for inclusion (**Action: Moss, Cammack and Kotyk** to edit; **Moss** to publish it on the web).
Cammack suggested that the Committees establish an independent relationship with Saier so that future developments in the classification are in line with the proposals in the document. (**Action: Cammack** to correspond with Saier)

11. Nomenclature Committees' Website
11.1 Diagrams and metabolic schemes in the Enzyme List (Moss)
A high proportion of new enzymes have associated reaction schemes that are either (1) chemical structures to match the reaction or (2) a metabolic pathway that includes the enzyme. Moss's maps are on secondary metabolites and complement the minimaps. Tipton pointed out that some of the maps also have a mechanism diagram, which can be very useful. Moss said that if people wanted specific maps included in the Enzyme List, they should let him know. Moss was congratulated on his initiative. (**Action: members of the Committees** to inform Moss of any specific maps that they would like).

11.2 Usage statistics (Moss)
Moss circulated usage statistics at the meeting. There are approximately 15,000 users of the nomenclature web sites per week (where user is defined as one computer/24 h period). There are approximately 30,000 hits per day. Usage of the web pages where enzymes are output in blocks of 50 enzymes/page has declined following the introduction of the one-enzyme-per-page format.

11.3 Provision of a glossary of biochemicals based on abbreviations and symbols documents - see http://www.chem.qmul.ac.uk/iupac/misc/abbrev.html, http://www.chem.qmul.ac.uk/iupac/misc/symbols.html and 1978 Compendium p. 212, EJB 2 (1967) 1-2 (Moss)
Moss reminded the Committees that we do not have a comprehensive list of abbreviations and symbols and that these should be documented. As an example, the structure of CoA is not present in any document and it should be provided, with its numbering indicated. Riboflavin and nicotinamide are other compounds that should be included in this document. He also pointed out that the numbering conventions for several compounds, such as those in the nucleic acids document, were somewhat obscure and a document illustrating ring numbering could be very useful. Two separate documents were proposed: (1) on miscellaneous compounds and (2) on abbreviations and symbols. It was believed that IUPAC could be asked to support Item (1). Kotyk pointed out that his publication, *Quantities, Symbols, Units and Abbreviations in the Life Sciences,* might provide a useful basis for such web-based material. (**Action: Moss** to coordinate these activities with assistance from **Tipton**)

11.4 Update on more permanent web address, i.e., iubmb.org (Cammack)
IUBMB have accepted the offer of the web address iubmb.org from Portland Press and Peter Ott has added a number of links to relevant nomenclature sites. Moss indicated that the URL iubmb.org/nomenclature pointed to the Enzyme List rather than to the Nomenclature home page. (**Action: Cammack** to ask Peter Ott to rectify this).

12. Inter-Union Bioinformatics Group (Cammack)
The final meeting of the IUBG will be held on 14/5/02 and Cammack will attend. (**Action: Cammack** to provide report)

13 IUPAC funding and application possibilities (McNaught)
Following extensive discussion, it was decided that McNaught should make an application to IUPAC for the Committees to continue meeting on an annual basis as before. (**Action: McNaught**) In addition, Committee members should try to determine possible areas that could be funded by IUPAC on a project-by-project basis, which is in line with their new format. Moss informed the Committees that the review process for project proposals requesting IUPAC funding is extensive, but if the project is approved, then funding is usually provided at the level requested. Guidelines for submission of projects are available on the IUPAC website (http://www.iupac.org)

Horton proposed that the carbohydrate document (published 1996) should be revised to take account of glycoclusters and glycodendrimers, with a target date of 2006 for completion. He proposed that there should be a meeting every year of carbohydrate experts (at an annual carbohydrate meeting) and has a list of people to invite to form a working group. This project would not require specific funding.

In light of the new organization of IUPAC, Moss said that the IUBMB should have specific responsibility for the Enzyme Nomenclature and that advice could be sought from IUPAC on specific matters. Two projects for which IUPAC funding will be sought were outlined. The first of these is that of a biochemical compound glossary (see Item 11.3) and the other is on a Compendium of synonyms for compounds (small molecules) of biological importance. The latter might be coordinated by Kazic and will include contributions from Tipton, Schomburg, Degtyarenko, Dixon and others. McNaught reported that CAS might be willing to supply their synonyms list as part of this exercise. (**Action: Moss and Kazic** to prepare project proposals for submission to IUPAC and **McNaught** to approach CAS about provision of their synonyms list)

Cammack pointed out that, in future, the Committee may have to define itself in terms of projects. Cammack suggested that we may have to change our constitution if we work in terms of projects in the future. (**Action: Cammack** to investigate if this is necessary).

14. Current activities of the IUPAC Division of Chemical Nomenclature and Structure Representation (McNaught)
McNaught provided a report on activities of the above Division. The document covered activities in the following areas:
Chemical identifiers - algorithm is now available for evaluation and can be provided by McNaught on request. The system will provide a standard way of representing chemical structures on computer that will derive directly from structure representation and can be converted back to a unique structure.
Revision of the *Blue Book* - The final draft will be reviewed in August 2002 and the book should be published in 2003.
The *Red Book* is also being revised and should be published later this year.
Part 1 of a document on fullerene nomenclature was due for publication last month.
Other projects include those on phanes, rotaxanes and chemically modified polymers.

15 Enzyme List
15.1 "Energases"
Following the provision by Dan Purich of a list of possible sub-subclasses (see Agenda) that could be applied to the energase class, it was decided to form a panel to look at the viability of such a classification and whether it was desirable to move any existing entries into it. It was agreed to ask Athel Cornish-Bowden to convene such a panel, which should include Kotyk, Schomburg, Purich and Tipton (**Action: Tipton** to contact Cornish-Bowden).

15.2 New subclasses and sub-subclasses

This item was added to the agenda to inform members of the creation of new subclasses and sub-subclasses during the past year, as follows:

EC 1.14.20	With 2-oxoglutarate as one donor, and the other dehydrogenated
EC 1.14.21	With NADH or NADPH as one donor, and the other dehydrogenated
EC 1.20	Acting on phosphorus or arsenic in donors
EC 1.20.4	With disulfide as acceptor
EC 1.20.98	With other, known acceptors
EC 1.20.99	With other acceptors
EC 1.21	Acting on X-H and Y-H to form an X-Y bond
EC 1.21.3	With oxygen as acceptor
EC 1.21.99	With other acceptors

Cammack had prepared a table outlining substrates and products in class EC 1. It was agreed that this should be published on the web (**Action: Moss**), provided that it was accompanied by documents indicating our conventions as to the direction in which reactions are written. Tipton had previously prepared such a document that also covered classes other than EC 1 and would revise it for the web (**Action: Tipton**). Moss stressed the need for an additional document on how donors and acceptors are classified in EC 1 and Tipton agreed to do this. (**Action: Tipton**)

15.3 Anomaly between classification of monooxygenases and dioxygenases having ferredoxin as donor (Cammack)

Cammack proposed that there be a new class of dioxygenases with ferredoxin as donor. He agreed to produce a list of enzymes involved and where they might be placed in the classification system. (**Action: Cammack**)

15.4 Use of 'A' and 'acceptor' to indicate an unknown acceptor

Toni Kazic had raised a question about our use of 'A' and 'acceptor' to designate an acceptor. Tipton had pointed out to her that the guidelines used are that 'A' and 'AH_2' are used when the acceptor accepts a hydrogen and that 'acceptor' and 'reduced acceptor' are used when we do not know whether the acceptor accepts hydrogen or electrons. Moss has proposed that this be made explicit in the header to EC 1.x.99. (see Item 15.5)

15.5 Update of Rules document

The Rules document requires updating to include items arising from 15.2 and 15.4 above and how we indicate sugars (as explained in the Carbohydrate document at http://www.chem.qmul.ac.uk/iupac/2Carb/). Some changes have already been incorporated (based on document from Dixon that was provided with Agenda) to take account of new subclasses that have been introduced into the Enzyme List.

It was also agreed that the revised Rules should indicate which reaction takes precedence when an enzyme catalyses more than one type of reaction, e.g. oxidoreductase (decarboxylating) and that the definition of dioxygenases and monooxygenases should be changed to remove mention of the specific involvement of O_2.

Moss, Tipton and Cammack agreed to revise the Rules document and hoped that Hal Dixon would also agree to contribute. (**Action: Moss, Dixon, Tipton and Cammack**)

It was also agreed that the Rules document should be given more prominence, perhaps by having it as a link from each EC class. (**Action: Moss)**

15.7 $NAD/NADH_2$ to $NAD^+/NADH + H^+$

The change approved at the 2001 Bethesda meeting from $NAD/NADH_2$ to $NAD^+/NADH + H^+$ will take place for all enzymes in Class 1. Moss pointed out that in some transferase reactions that involve methylation reactions with *S*-adenosylmethionine, the product is NADH rather than $NADH + H^+$. It was agreed that changes in this class should be considered on a case-by-case basis. (**Action: Boyce and Moss**)

15.8 Reference archive

This has been provided for references that have been superceded and have therefore been removed from the Enzyme List, but which are being maintained for historical purposes. This file can be found at http://www.chem.qmul.ac.uk/iubmb/enzyme/refarchive.html.

15.9 Change from 'Recommended name' to 'Common name'
Barrett had indicated that, for the peptidases, the term 'Recommended name' is correct as they are recommended and this is not taken into account by use of the term 'Common name', which had been approved at the 2001 Bethesda meeting. Tipton proposed a rubric as part of the introduction to EC 3.4 indicating that the Common name is in fact the Recommended name in that subclass. This was agreed.

15.10 Ambiguous systematic names (Moss)
Moss raised this as an additional Agenda Item. He indicated that, for some enzymes, the systematic name does not give an indication of the reaction taking place. He agreed to send a list of such systematic names to Boyce, so that they can be amended as necessary. (**Action: Moss and Boyce**)

16. Documents and Panels
16.1 Protein kinases (Berman)
A new protein kinase database is available on the web (http://pkr.sdsc.edu/html/index.shtml). Berman asked if there were any suggestions that the Committees should give them regarding their method of classification. Apweiler suggested that they base their classification on sequence similarity. It was agreed that it would be helpful to have a rough outline of how they intend to classify the protein kinases. (**Action: Berman** to ask panel if they would supply this) Apweiler suggested that we should have a document for protein kinases that is similar to the transport protein document. Tipton would like us to widen our listing of protein kinases in the Enzyme List to include each type of enzyme (but not each individual enzyme).

16.2 Cyclic peptides (Moss)
Nothing to report. Sharon informed the Committees that a colleague of his was prepared to form a panel to look at this issue. (**Action: McNaught** to check if document has already been reviewed; **Sharon** to ask his colleague to contact Moss)

16.3 Modified nucleotides and nucleic acids (Horton)
Horton reported that Jack Secrist had said that his panel is working on the proposal. Horton has not yet received details of their progress.

16.4 Update on carbohydrate database (Horton)
As reported at the Bethesda meeting last year, CarbBank was defunct through lack of funding. Glycominds were prepared to take it over but would have it on a commercial database. Since they have already incorporated CarbBank and developed it from there, Horton felt that it would be better to get them to put the database in the public domain. Horton is hoping to set up a mini-meeting at the Carbohydrate meeting in Australia in August 2002 to discuss the future direction of CarbBank. (**Action: Horton** to provide report after August meeting)

Horton reported that Bohne-Lang et al. have published a paper on LINUCS (Linear Notation for Unique Description of Carbohydrate Sequences) in *Carbohydrate Research*. A pdf version of the paper is attached.

Glycominds, who gave a presentation at the 2001 meeting, are prepared to give free licenses to the academic community. Berman said that the issue of the license is still a problem as the standards they advocate should be in the public domain. She believed that Glycominds should be persuaded to do this and Sharon agreed to facilitate a discussion between Glycominds and Berman on this issue. (**Action: Sharon**)

17. Journal instructions to authors (Elliott)

Elliott has compiled a list of appropriate journals that could be approached and asked to provide a link from their websites to documents on ours. Before proceeding further, he recommended that the Committees prepare a document on the web, containing suitable material to be included in instructions to authors. This would include a brief description of relevant nomenclature documents, containing links to each of the documents. It was agreed that such a guide to our nomenclature documents could be a great help for authors trying to find guidance, if it could be kept down to a convenient size. (**Action: members of Committees** to make suggestions as to content).

18. Follow-up on Items from 2001 Meeting

18.1 Amine oxidases

Tipton reported that Tony Pegg (Hershey, USA) had promised to supply a list of polyamine oxidases that had not yet been classified and that he would await this list before making recommendations to the Committees.

18.2 Cammack to convene a panel on folates/pterins

Alistair McEwan (Australia) has formed a panel to advise on the nomenclature of molybdopterins.

18.3 Ribonucleases

This item arose from correspondence between Tipton and Claudio Cuchillio regarding the classification of pancreatic ribonuclease (see 2001 Minutes). It was agreed that it would not be appropriate to reclassify this enzyme as a transferase and that it should remain classified as a hydrolase. However, a comment might be added to indicate that the reaction was essentially a lyase reaction with subsequent hydrolysis. (**Action: Tipton**).

18.4 Oxidized glutathione

The decision from the last meeting that 'oxidized glutathione' be replaced by the term 'glutathione disulfide' (with a glossary entry indicating that the change has been made and why) has been implemented. In addition, the term 'oxidized thioredoxin' has also been changed to 'thioredoxin disulfide', and analogous compounds have been named similarly.

18.5 Separate EC class of multireaction systems

The suggestion was for a new enzyme class, EC 7, to include multireaction systems. This class would not, in the first instance, be subdivided. (**Action: Moss and Tipton** to create new class, EC 7) These data would be kept for private use until such times as there are a reasonable number of entries. A decision will then be made on whether or not it should be included in the Enzyme List.

18.6 Lyases

At the last meeting it was pointed out that some of the lyases had been classified on the basis of a partial reaction rather than the overall reaction. Some progress has been made in sorting out the lyases, but there is still work in progress on this matter. (**Action: Boyce and Dixon**)

18.7 Removal of enzymes from EC 1.6

At the last meeting, it was agreed that many enzymes currently in EC 1.6 were inappropriately classified. So far, five entries have been transferred out of EC 1.6.99 to EC 1.16.1; 9 from EC 1.6.4 to EC 1.8.1; 10 enzymes have been transferred from EC 1.6.6 to EC 1.7.1 and 2 have been transferred from EC 1.6.8 to EC 1.5.1. Further progress will be reported at the next meeting.

18.8 Common name

The term 'Recommended name' has been replaced by 'Common name' in all enzyme entries.

18.9 Inositol enzymes

Good progress has been made by Dixon, with the help of Robin Irvine, in modifying *myo*-inositol enzymes to take account of the numbering system that should be used. These enzymes now form part of newenz.html.

18.10 Transfer of enzymes out of 1.x.99 where the acceptor is known but does not fall into any of the other sub-subclasses

A list of such enzymes has been prepared by Boyce, and Cammack agreed to prepare draft entries for these enzymes. (**Action: Cammack**)

18.11 Provision of Enzyme List on CD-ROM

Although this idea was welcomed last year, it was reported to have been a costly procedure for PDB. Because of doubts about the viability of such a venture, it may not be pursued. Cammack reported that he had communicated with the IUBMB Treasurer, Brian Beechey, about this. While Beechey thought that the idea was a good one, he did not offer financial support for this project. Before deciding whether to abandon the idea or not, Cammack agreed to write to the IUBMB Publications Secretary, Angelo Azzi, to ascertain his views. **(Action: Cammack** to communicate with Azzi).

18.12 Document on nucleic acid geometry

Moss asked Berman to provide a statement of approval. Berman indicated that she had already provided this to Cammack.

19. Discussion: Future Projects

These are outlined under Item 13.

20. Any Other Business

A document on guanidine formation from John Jones was circulated by Moss. McNaught objected to the term 'guanylation', which Jones appeared to use to refer to the conversion of an amino group into a guanidino group. McNaught thought that, by analogy with formation of an ester (esterification), the term should be guanidinification. (**Action: McNaught** to write to Jones outlining his views).

Kotyk proposed that Kazic be made an Associate Member of IUBMB. This was seconded by Cammack. (**Action: Cammack** to write to the Secretary of the IUBMB, Jacques-Henry Weil, to indicate that this is our intention).

21. Date and Place of Meeting in 2003

The next meeting will be held in Dublin on 2-3 May 2003.

The meeting concluded at 15.30 on 11 May, 2002.

APPENDIX IV

The Future of the IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN)

1.0 The Present Situation.

As part of the restructuring of IUPAC at the beginning of 2002, Division VIII assumed responsibility for the IUPAC component of the IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN). Although all other IUPAC Nomenclature Commissions were dissolved, the IUPAC Bureau agreed that JCBN should be maintained in its present form for the biennium 2002-2003, while the people involved consider alternative ways of working which may be desirable in view of the restructuring of IUPAC and its adoption of the project system. The IUPAC Bureau has asked for a proposal on the future management of joint IUPAC-IUBMB nomenclature activities for consideration at its next meeting. It will then set up a small Committee to review the situation, liaise with IUBMB, and make a final proposal to IUPAC Council in 2003. The options were discussed at the May 2002 meeting of JCBN and the Nomenclature Committee of IUBMB (NC-IUBMB), and input from the IUPAC Division VIII Committee is now requested. The following text is being sent out with the Bureau agenda papers, and I will submit any comments from Division VIII at the Bureau meeting.

Present terms of reference of JCBN are attached.

2.0 Future Needs

What will be required from joint activities of the two Unions in future? It seems reasonable to expect a continuing need for the following:

2.1 Maintenance and updating of the Enzyme List. This is a very substantial and continuous operation needing a permanent body to run it, with access to advice on chemical names for substrates, reagents and products. Classification systems for other types of protein (e.g.transport proteins) also need to be maintained, probably by the same body.

2.2 Maintenance and development of naming systems for classes of natural product of interest to biochemists, especially steroids, amino acids and peptides, carbohydrates and nucleic acids. Activities in this area could be carried out on a project basis, revisited for regular updating; indeed this is the way that JCBN has managed this work in the past.

2.3 Advice for biochemists on names for specific compounds of biochemical importance. A project to develop an online database source for advice of this kind can be envisaged; such a database would require regular updating.

2.4 Maintenance and development of standards for symbolism to be used in databases for biopolymers. Again this could be carried out as project work, with suitable maintenance arrangements.

2.5 Development of general terminology for biochemists, in the form of glossaries in specialist subject areas. There is an increasing need for such material, for application in standards for mark-up languages (e.g XML "data dictionaries"). Again, these could be developed as projects but would require continual maintenance if provided online (e.g. as authoritative XML "namespaces"). Liaison with similar work elsewhere within IUPAC would be needed.

Of the five tasks listed above, only the first clearly requires a regular ongoing commitment of resources, incompatible with the project system. It is also currently the most time-consuming part of JCBN's business.

Items 2.1 to 2.4 are all relevant to IUPAC Division VIII. However, with respect to item 2.5, glossary projects may be carried out by any IUPAC Division, and work on XML (eXtensible Markup Language) for chemistry is the responsibility of the IUPAC Committee on Printed and Electronic Publications. IUPAC will need to ensure appropriate cross-representation.

3.0 Options

3.1 Dissolve JCBN, retaining a standing committee to manage the Enzyme List, and carry out other activities as limited-lifetime projects.

This could work only if projects were assigned specifically either to IUPAC (managed by Division VIII) or IUBMB (managed by NC-IUBMB?). If joint ownership of projects is required, then a joint body is needed to administer them. In the absence of JCBN an Enzyme List Committee would need to be run by IUBMB either as a separate body or as a subcommittee of NC-IUBMB, but would need access to IUPAC nomenclature expertise.

3.2 Replace JCBN by a (smaller?) administrative group with representatives from both Unions, with responsibility for joint projects, and with an Enzyme List Subcommittee including IUPAC nomenclature advisers. (Alternatively the Enzyme List group could be run separately by IUBMB, as in option 3.1.)

The present JCBN is to some extent an administrative body anyway, and annual JCBN meetings consume most/all of its (IUPAC) budget. However, it could be helpful for a Joint Committee to have, in addition to a meetings budget, the option to apply separately for project money from IUPAC Division VIII and any designated IUBMB contact. Also

it could be argued that an Enzyme List Subcommittee should be funded entirely by IUBMB.

3.3 Merge JCBN and NC-IUBMB, giving a single (smaller?) joint committee for administering projects funded by either Union alone or jointly, and managing the Enzyme List (through a subcommittee or not).

Other than funding source, there is currently no distinction made between members of JCBN and those of NC-IUBMB. Effectively there is only one Committee. It could make administration simpler if this were recognised by merger, but IUBMB might prefer to retain a nomenclature group over which it had complete control.

3.4 Retain JCBN in its present form.

In practice, this arrangement does not differ markedly from option 3.2, except for the suggested budgetary arrangements and the separation of the Enzyme List committee. The present annual JCBN meetings take place concurrently with meetings of NC-IUBMB (the groups do not meet separately), and outstanding questions relating to the Enzyme List are handled by the whole group or subgroups as convenient. However, there has not been much project work carried out recently, and separating administrative from project budgets might encourage more activity.

4.0 The Way Forward.

The above options were discussed by JCBN and NC-IUBMB at the Cologne meeting in May 2002, and the following comments were made:

4.1 In order to administer work in which both IUPAC and IUBMB have an interest, a joint body is needed.

4.2 It would be simpler to have a single body responsible to both Unions rather than to maintain two groups (JCBN and NC-IUBMB).

4.3 An increased level of activity in chemical biology means that IUPAC may have more interest in the Enzyme List than hitherto, and it might make sense for this activity to be managed jointly.

4.4 The Enzyme List work requires a permanent body (5-6 people meeting annually) but other work could be handled on a project basis. However, it would be necessary to have a permanent proactive group to generate initiatives in areas other than enzymes. Furthermore, there is an additional need for a permanent source of advice on biochemical nomenclature in general, and for cross-disciplinary input.

4.5 It would be helpful to discuss with IUBMB whether a project-based system could be introduced, and if so what funding could be made available and what proportion of funds might be allocated specifically to projects as opposed to operational expenses.

In the light of this discussion (and about ten years' attendance at JCBN meetings, currently as Associate Member of NC-IUBMB), I suggest that a permanent Joint Committee is needed, consisting of representatives from both Unions (perhaps three people from each), to initiate and manage projects, with responsibility for an Enzyme List Subcommittee.

Alan McNaught
July 16th 2002

IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN)

Current Terms of Reference

(i) To consider the nomenclature and symbolic representation of natural products and related compounds including synthetic analogs and, where appropriate, to make recommendations in conformity with those issued by other nomenclature Commissions of IUBMB and IUPAC.

(ii) Before recommending any material for publication as an IUBMB-IUPAC document, to ensure that the fullest possible consultations have taken place and the widest possible consensus has been reached with the appropriate bodies of the two Unions. For IUPAC, these are the Interdivisional Committee on Nomenclature and Symbols (IDCNS), or, through IDCNS, the appropriate divisional nomenclature Commissions. For IUBMB, these are NC-IUBMB and journal editors.

(iii) Approval to publish any material as an IUBMB-IUPAC document is to be obtained, in the case of IUPAC from IDCNS, and in the case of IUBMB from the Executive Committee of IUBMB.

(iv) The Commission shall normally hold an annual meeting, concurrently with NC-IUBMB, and with the approval of the President of IUPAC and the Executive Committee of IUBMB.

(v) Associate Members will receive all documents of the Commission and their opinion will be sought by correspondence. An Associate Member may attend any meeting of the Commission, but his or her expenses will not be defrayed by the respective Executive Committee unless he or she has been invited as an observer under item (vi).

(vi) The Commission shall be entitled to invite observers from other Commissions and experts in special fields to attend the meeting. The respective Executive Committee shall be asked in advance to defray any expenses that would arise from such an invitation.

Current Composition and Terms of Office

(i) There shall be a standing IUBMB-IUPAC Joint Commission on Biochemical Nomenclature, composed of eight Titular Members, one of whom shall serve as the Chairman and one as the Secretary, and up to eight Associate Members.

(ii) The Chairman and the Secretary shall each be appointed for a term of four years by the Presidents of the two Unions, subject to whatever ratification is imposed by Union Statutes and Bylaws. JCBN may propose names of persons suitably qualified for appointment.

(iii) Four of the Titular Members shall be appointed by IUBMB and four by IUPAC. JCBN may itself appoint Associate Members.

(iv) The Titular Members appointed by IUBMB shall also be Members of the Nomenclature Committee of IUBMB (NC-IUBMB).*

(v) Candidates for Membership may be proposed by the Commission and shall be appointed by the respective Unions.

(vi) The periods of service of the Titular Members and of the Associate Members shall be in accordance with the Statutes and Bylaws of the appointing Union. The sum of the years of service as a Titular Member and as the Chairman or the Secretary shall not exceed ten years.

---

*NC-IUBMB has four other Members and (currently) four Associates

APPENDIX V

Chemistry International

Vol. 24, No. 2

March 2002

IUPAC News

Chemical Nomenclature and Structure Representation

by Alan McNaught

Alan McNaught

This year sees the birth of a new division of IUPAC. Its conception dates from a survey carried out in 1998 for the Organic Division, in which the chemistry community was asked for opinions on future nomenclature requirements. Comments received highlighted the increasing need for a body to oversee IUPAC nomenclature development across all disciplines, to ensure compatibility with previous work, and to coordinate related activities. The resulting report[1] drew attention in particular to the need to integrate nomenclature standards with computerized facilities, and to push ahead with efforts to define for each unique structure a single preferred IUPAC name, correlated with other names in common use.

This report stimulated further consultation. A strategy roundtable in March 2000, involving people from many professions with a need for standard chemical identifiers, reinforced the views from the original survey and added some important new items.[2] In particular, the need for a IUPAC standard for computerized representation of a chemical structure was recognized.

The roundtable recommendations led IUPAC's Executive Committee to establish a temporary Committee on Chemical Identity and Nomenclature Systems,[3] which developed plans for future management of IUPAC's nomenclature work and launched the IUPAC Chemical Identifier project.[4] The work of this ad hoc Committee culminated in a proposal for a new IUPAC division. The IUPAC Council endorsed this proposal, stressing the continuing importance to IUPAC of nomenclature and related activities. And so, the Division of Chemical Nomenclature and Structure Representation[5] was established on 1 January 2002.

So What Will the New Division Be Doing?

Most importantly, it will bring together work on nomenclature of chemical compounds with development of other methods of designating chemical structures. The Chemical Identifier

project is a first step in this direction. It will also be tackling interdisciplinary issues that have been difficult to deal with hitherto. For example, we have a project group considering divergent recommendations arising from the development of preferred names for organic compounds in parallel with revision of the inorganic nomenclature rules. Also, we will study the applicability of naming systems developed for polymers with particular reference to macrocycles, rotaxanes, and catenanes. Additionally, we will be assessing to what extent recent developments in conventional organic nomenclature and phane nomenclature can allow us to deal conveniently with these and other structures.

Other areas of activity will include the following:

- Extension to all chemical compounds of procedures for identifying IUPAC-preferred names
- Organometallic nomenclature
- Computerassisted cluster nomenclature
- Stereochemical nomenclature
- Fullerene nomenclature
- Dendritic and hyperbranched polymers
- Databases of synonyms for compounds in common use

How Will the New Division Operate?

A division committee has been assembled consisting of 12 members plus National Representatives, including people with extensive experience in developing conventional nomenclature recommendations and others with expert knowledge of computerized systems for designating chemical structures. The division has an advisory subcommittee of about 40 people, charged with advising the division committee on the needs of the community, and developing project proposals. This subcommittee contains many individuals with experience in nomenclature work, as well as chemical software developers, journal editors, and a range of other users of IUPAC recommendations. It is expected that these people will lead or otherwise participate in projects as well as provide advice to the division committee. Apart from meetings of task groups, most of the subcommittee's work will be carried out via electronic communication. A Web discussion board has been set up, to which drafts of new recommendations and comments on them are to be posted. Where possible, meetings of task groups will be organized to take place concurrently, to enable informal discussions between members of different groups. At suitable intervals (probably about every 5 years), we will convene a roundtable meeting with the user community to review results and define future requirements.

Of course, this way of working is quite new to IUPAC, and we shall need to adjust our procedures as we gain more experience. However, this scheme will allow better use of resources than previously, and will expedite the development of the standards that the community needs.

This new arrangement also accommodates the IUPAC/IUBMB Joint Commission on Biochemical Nomenclature, as a commission attached to the new division. This commission meets jointly with the Nomenclature Committee of IUBMB and the combined committee acts essentially as a single body. Its main responsibility is the upkeep and development of the Enzyme List, a very substantial and ongoing project.[6] It is also increasingly involved with standards for bio-informatics, needed to accommodate the explosion of information on bio-polymers. In addition, JCBN has traditionally dealt with specialist biochemical nomenclature systems (e.g., carbohydrates, lipids, and polypeptides), and this is where a close link with

the new Division of Chemical Nomenclature and Structure Representation is important. For nomenclature work, this commission has always operated through project task groups, something we envisage for future work managed by the new division.

We should not lose sight of the fact that IUPAC can retain a credible role in nomenclature development only by paying close attention to the needs of the community and responding to them. We need to give wide publicity to the fact that IUPAC's project system now allows ideas for future work to be developed by anyone. My colleagues and I welcome project proposals[7] from any source, and are happy to discuss suggestions informally. Only if we have a clear view of what our "customers" want can we hope to make the best use of IUPAC's resources.

References

1 http://www.iupac.org/projects/1998/310_28_98.html
2 http://www.iupac.org/news/archives/2000/NRT_Report.html
3 http://www.iupac.org/organ/ad_hoc_cmt/ccins.html
4 http://www.iupac.org/projects/2000/2000-025-1-050.html
5 http://www.iupac.org/divisions/VIII/index.html
6 http://www.chem.qmw.ac.uk/iubmb/enzyme/
7 http://www.iupac.org/projects/form_guide.html

Alan McNaught is President of the IUPAC Division of Chemical Nomenclature and Structure Representation. He is General Manager of the Production Division of the Royal Society of Chemistry, Cambridge, UK.

Division Committee

Alan D. McNaught, *UK*
Division President

Warren H. Powell, *USA*
Division Secretary

Michael Dennis, *USA*
Michael Hess, *Germany*
Herbert D. Kaesz, *USA*
G. Jeffery Leigh, *UK*
Gerard P. Moss, *UK*
William G. Town, *UK*
Antony Williams, *USA*
Stephen Heller, *USA*
Alexander J. Lawson, *Germany*
Bruce M. Novak, *USA*

National Representatives

Roberto de Barros Faria, *Brazil*
Jiasong He, *China*
Jean Marie François Toullec, *France*
Yohsuke Yamamoto, *Japan*
Osman Achmatowicz, *Poland*
Bernardo Jerosch Herold, *Portugal*

APPENDIX VI

Alchemist
ChemWeb.com

What's in a Name?

21 March 2002

Alan McNaught was previously Journals Manager at the Royal Society of Chemistry and is now General Manager of RSC Publishing's Production Division. Over the past 25 years he has worked with the International Union of Pure & Applied Chemistry (IUPAC) to try to ensure chemists are all speaking the same language when it comes to names of compounds. He discussed the formation of a new IUPAC nomenclature division (of which he is the first President), cryptic crosswords, and horses for courses with David Bradley.

David Bradley: is The Alchemist's Catalyst columnist. He is a chemist by training but sidestepped the laboratory in favour of science writing in 1989. He was winner of the Telegraph Young Science Writers Award in 1992 and a runner up in the CIA Awards for 1995 as well as a Merit award winner in the British Medical Journalism awards in 1997. His Reactive Reports website was runner-up in the Pirelli InterNETational awards in 2000. He has written for numerous publications including New Scientist, Science, The Telegraph, The Guardian and many scientific trade magazines. He is a fine guitarist and when not strumming strings or tapping keys spends time with his two children, Matthew and Bethan.

IUPAC has created a Division of Chemical Nomenclature and Structure Representation, what was the rationale behind that?

The new Division arises from a need to bring together divergent approaches to naming problems in different chemical disciplines, and to integrate work on naming systems with the development of standards for computerised representation of chemical structures.

So, does that mean IUPAC is responding to the needs of the chemical community more actively?

During the past few years IUPAC has been fundamentally reassessing its mission and its way of working, aiming to respond more directly to the needs of the chemistry community and to use resources more effectively. One of the results of this exercise was the abolition of all IUPAC's Nomenclature Commissions at the end of 2001, with the intention that all future work on nomenclature will be carried out by limited-term project groups. This reorganisation has provided the opportunity to consolidate responsibility for nomenclature work in a single Division, and the long-standing need to consider computerised representations, as alternatives to conventional names, was easily accommodated.

Why is nomenclature still such an important issue today when electronic structure searches are possible or a CAS registry number available?

We need to tackle the broader issue of how one specifies a chemical compound. People need tools to distinguish the 19 million substances (and 17 million sequences) listed by Chemical Abstracts. Faced with this sort of problem, we need all the identification methods we can muster, including both systematic and trivial names, registry numbers, and structural identifiers. The use to which the name or other identifier is to be put determines the choice. For some purposes a nonsense number is fine, for others a short but meaningful verbal label or a string of characters derived from a connection table is needed.

Which can presumably be useful to non-chemists too?

We should remember that the need for chemical identifiers is not confined to people actively engaged in chemical research; we also need to provide identifiers for regulatory and legislative needs, *e.g.* for patents, international trade and tariff specifications, and health and safety.

So, trivial and systematic names can live in harmony despite the ongoing debate between chemists?

A colleague of mine is fond of the expression 'horses for courses'. There is a common misconception that IUPAC nomenclature equals systematic nomenclature and that one can derive a unique IUPAC systematic name for every substance. In fact, IUPAC manuals have always offered choices between various kinds of systematic and trivial names, and it is up to the user to decide which option is fit for purpose. Arguments over which name is ultimately 'best' are pointless. That being said, however, people who require standard names for regulatory purposes would like IUPAC to make choices for them, and we have therefore been working recently on rules to identify a IUPAC-preferred name for any organic compound. These rules will appear in the next edition of *Nomenclature of Organic Chemistry*, but not with any implication that alternative names are illegitimate. In the absence of regulatory needs, the variety of names available will persist.

With all these various identifiers how will chemists and non-chemists cope?

I think that people need help to find their way through the variety of both systematic and trivial names available, and synonym databases can help them to do this. Chemical Abstracts has a need to develop unique names 'on the hoof' for indexing purposes, and inevitably there is some divergence from IUPAC recommendations, which need to serve a variety of different needs. But, I must emphasise that IUPAC nomenclature is not a 'system'; IUPAC codifies a variety of nomenclature systems and recognises many trivial names, often without specifying choices between alternatives.

Presumably, there will be a way to make chemical identification more transparent through electronic means, such as hyperlinking?

I am sure that this is technically possible, and I will be encouraging people to develop databases along these lines.

Will nomenclature software play a role in this?

We have recently begun work on what we are calling the IUPAC Chemical Identifier. This is to be a set of algorithms for the standard representation of chemical structures that will be readily accessible to chemists in all countries at no cost. The intention is that a standard graphical representation of a chemical structure will be used as input into existing and newly developed computer programs to generate an IUPAC name and a unique IUPAC identifier. Such a non-proprietary identifier for chemical substances could be used in printed and electronic data sources thus enabling easier linking of diverse data compilations. So far, the work has been confined to relatively straightforward organic structures that can be expressed as valence-bond representations, and a beta algorithm is now available from NIST for evaluation. The problems raised by more complex structural types will be addressed subsequently.

Who is involved in this endeavour?

The work on the IUPAC Chemical Identifier is being carried out at the US National Institute of Standards and Technology (NIST). However, a consultation meeting with various interested parties, including people from commercial software companies, was held before the project was properly established. But, obviously, if the results of this work are to be widely accepted, they must not be seen to be associated with any particular commercial interest.

When new types of molecule appear, such as the fullerenes and the rotaxanes, should we be leaving chemists with a little art in the naming process?

Much of so-called systematic nomenclature relies on systematic manipulation of 'trivial' parent names. There is no virtue in trying to apply to large molecules systems devised for small ones: you end up with a name extending over several lines of text that needs an expert to interpret. Manageable names for big molecules need to be based on short (trivial) names for big parents. Very often suitable trivial names emerge from the inventive constructions of the people working in the field.

We still see 'sulfur' spelled 'sulphur' in many publications because many editors say there is no 'f' in sulphur, does that annoy you?

The origins of the divergence between 'sulphur' and 'sulfur' may stimulate the juices of etymologists, historians and pedants, but for most people arguments over the rights and wrongs of established spellings represent no more than a little harmless fun. However, having standard English spellings for the names of the elements makes searching the literature a little easier.

It all looks like a lot of work, what's your personal motivation in making sure the nomenclature schemes are right?

It is a lot of work, and it is all carried out by volunteers. People are funded by IUPAC to attend face-to-face meetings if needed, but much of the work is carried out in 'spare' time by electronic communication. My own participation is made somewhat easier by the fact that I am a member of the Royal Society of Chemistry's publishing staff, and RSC is the UK National Adhering Organisation for IUPAC. In helping with this work, my wish is to facilitate communication about chemistry, not only in research reports but also in the many other fields where chemical information is needed.

Is there an element of personal satisfaction in the work too?

I must admit to a certain fascination with the intellectual challenge of producing comprehensive, coherent, consistent rules for naming. The appeal is similar to that of a cryptic crossword puzzle.

APPENDIX VII

CHEMICAL NOMENCLATURE

IUPAC's New Vision

Kevin Thurlow
LGC

The International Union of Pure and Applied Chemistry (IUPAC) has had a number of functions since its inception in 1920. Amongst these has been the provision of rules and recommendations for chemical nomenclature. IUPAC split its various work functions into different areas or 'Commissions', but eventually there was a feeling that some of these Commissions attended the biennial meetings faithfully but never published anything. This was certainly not the case with nomenclature, but IUPAC took the somewhat draconian step of announcing that the Commissions would be dissolved in 2001, and replaced with specific projects. These would be set up as and when necessary, with appropriate funding and milestones. This is not a bad idea in theory, but one of the problems in the field of nomenclature is that a general overview is required. If groups work in isolation on different problems of nomenclature, relating to similar compounds, then they might take completely different views on how to name the compounds. Analytical chemists and biochemists have become very inventive at making new compounds. If someone produces a compound with a fullerene attached to a sugar attached to a steroid, deciding which rules to use in naming the beast is a bit tricky. So it is nice if the rules bear some sort of similarity of approach. Accordingly the nomenclature experts made it clear to the powers that be at IUPAC that some sort of general overview was essential or chaos would ensue. IUPAC recognised the force of this argument and set up a 'Division of Chemical Nomenclature and Structure Representation'. This will oversee all work on chemical names, not just systematic names, but approved 'trivial' names. The 'structure representation' part in the title refers to the project to produce a chemical identifier[1].

It is particularly impressive that IUPAC has carefully considered the needs of the chemical community. The new Division committee has experts in nomenclature of course, but also experts in software, especially that related to nomenclature. In the past, IUPAC has been accused of bumbling along producing rules without considering what the users really need. IUPAC has now made it very clear that that they are aware that if you produce rules that people do not like, then the rules will just be ignored. IUPAC did well to sort out the rather unseemly wrangle that developed over the names of recently created chemical elements, where self-interest rather overtook scientific endeavour. Hence elements 104 and 105 appear under a variety of names in the literature, which is most unsatisfactory. There has always been some confusion over systematic nomenclature, because both IUPAC and CAS (Chemical Abstracts Service) produced their own rules. Although these rules sprang from the same seed, there were two entirely different approaches. IUPAC liked to consider new rules and eventually publish what they considered was the definitive set of rules. However, if CAS saw a new chemical in the literature, then

they had to give it a registration number, along with a name, and this had to be done quickly! Naturally, this has led to major differences in the two forms of systematic nomenclature. Once CAS has a system in place, changes are (not surprisingly) unwelcome, because it will affect previous decisions. IUPAC and CAS are working together to try to iron out these difficulties, and it is good to see active cooperation between the two bodies.

The new division is backed up by an 'Advisory Subcommittee', which will suggest and take part in projects, as well as comment on other draft rules. This subcommittee communicates mainly by electronic means, which is just as well, because the 36 members come from 17 different countries, with South America the only continent not represented so far. There is an enormous breadth of expertise. Apart from nomenclature specialists, there are experts in patents, publishing, computers, linguistics, and representatives of chemical companies. The 17 professors, 17 doctors and 2 misters share a great deal of expertise. It is clear that IUPAC have made a very serious attempt to produce the best nomenclature systems possible. The current format is experimental and will be amended as necessary.

It is encouraging to see that trivial names will also be considered. For many years, IUPAC primarily concerned itself with systematic names. Trivial names are popular because they are easier to use, but if they are not properly regulated, problems arise. People glibly refer to 'AA' or 'Smith's Reagent', which is fine if only a few people are involved. But names like that turn up in the literature without being defined! There is a very real risk that the wrong reagent will be used, and that could destroy the quality of the work – or even the laboratory itself! If IUPAC can advise on trivial names as well, there are splendid opportunities for improved quality of analysis. Abbreviations and trivial names are particularly confusing if you think you know what they mean. ABS means acrylonitrile-butadiene-styrene, which happens to be a polymer used in the manufacture of cars. A former author of this column once pointed out that it was rather strange that advertisements were ranting on about cars having 'ABS'. "Who cares what plastic they use?" he would say. Then one day, 'ABS' was actually defined as 'Anti-lock Braking System', and it began to make sense. This is one of those cases where being an expert works against you.

IUPAC has produced an exciting new plan and it is good to see so many different people and organisations trying to work together to produce optimum results.

For advice on chemical nomenclature, contact The VAM helpdesk.

REFERENCES

1. Thurlow, K, *VAM Bulletin 23*, 32-33

APPENDIX VIII

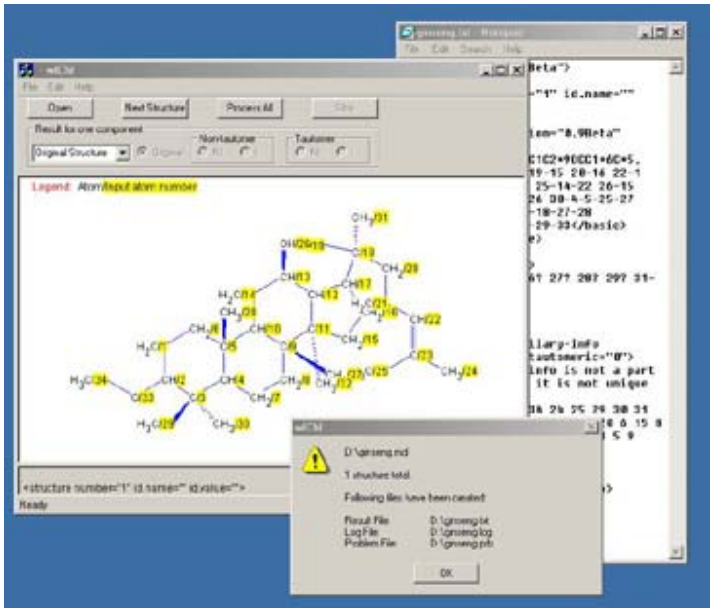
Alchemist

ChemWeb.com

That IChI feeling

Identifying chemicals has never been easy. From isolation to structural identification there can be many arduous steps and afterwards there is the need to represent each compound uniquely. This point has become increasingly critical as chemical numbers rise and chemical information streams widen.

Steve Stein, Steve Heller, and Dmitrii Tchekhovskoi of the National Institute of Standards and Technology (NIST), in Gaithersburg, Maryland, presented a putative solution to the IUPAC General Assembly in Brisbane, July 2001 — the IUPAC Chemical Identifier (IChI), a transparent and portable means to fingerprint every chemical. Since then, NIST programmers have been endeavouring to implement the IChI in software form and March saw the release of the beta version of the resulting package. This first version covers only well-defined, covalently bonded organic molecules.



The version of IChI devised by NIST expresses a five-layered sandwich of information, with each layer representing a distinct class of chemical information. The basis of the software is then to allow a user to input the relevant information from a compound into a 'connection

table' and to resolve this into a unique species — the IChI. The beta software, of course, does not require a chemist to physically enter the data. This can be read-in by the software, thanks to careful design, from the (almost standard) mol file format.

The package uses a set of rules to convert the mol file into a sequence of bytes based on the chemical structure represented. The first set of rules normalizes the input file so that unnecessary information is removed. The second, canonicalization, produces all the atom labels based on the underlying structure and not on any artefact of the drawing process. The third rule set carries out a serialization, which is essentially the digital conversion — the transformation of labels into the IChI string. The five layers within the IChI then represent structural information, basic connectivity, overall charge, tautomerism, isotopic composition and stereochemistry.

That string of information serves as what the NIST team describes as a 'precise digital signature' for the compound. Inherent in this IChI, is the fact that no two compounds can have the same IChI. A single compound can have only one IChI, which does not depend on how its structure has been drawn. The user can choose which layers to use in the final representation, although the unique connectivity layer must be retained.

The reverse process could then be used to generate the compound uniquely in any format one might consider — from traditional print to a fully hyperlinked Web document. Interestingly, the IChI might also provide chemists with the much sought-after means to automatically generating names even for the most complex of chemical substances that are not catered for by the present nomenclature programs. But, more than that IChI will provide the elusive transparent and portable system for representing chemistry.

The beta test program can be requested from Steve Stein and a demonstration of IChI generation in a structure-drawing program, working in conjunction with the beta test program, is available from Alan McNaught at the Royal Society of Chemistry. ■

David Bradley

APPENDIX IX

Chemists synthesize a single naming system

David Adam, London

An international team of chemists is working on something that chemistry sorely lacks — a consistent and comprehensive way of labelling all chemical compounds. The new technique will apply computer algorithms to molecular structures, generating a unique digital signature for any chemical compound. The new labels are not intended to replace common chemical names, but to allow easier linking to compounds in online chemical databases and journals.

"The hope is that all organizations that handle information on chemicals will be able to use a single format to say what the chemical is," says Alan McNaught, general manager of the production division at the Royal Society of Chemistry in Cambridge, who coordinates the project for the International Union of Pure and Applied Chemistry (IUPAC).

Right now there is no single international standard for identifying chemicals. The IUPAC and the American Chemical Society use different rules. Some drug companies, as well as different branches of chemistry, have their own chemical-naming systems. Even simple structures can cause confusion. For example, the formal name for acetic acid, the main ingredient in vinegar, is ethanoic acid.

IUPAC believes that its new system — which would be freely available to all — could unify the different approaches. Tentatively known as IChI, for IUPAC chemical identifier, its development is being led by a team at the US National Institute of Standards and Technology (NIST) in Gaithersburg, Maryland.

A preliminary version of the software covering well-defined, covalently bonded organic molecules was released this year to let other chemists test the idea. It labels each atom in a compound in a way that does not depend on how the structure is drawn, and converts the label to a string of characters. The format has not been finalized, but at present ethane is 'C3C3, 2-1', for example, and acetone is 'C3C3OC, 4-1-2-3' (the labels are easily converted to structures using the algorithm). The process is reversible, so molecular structures can be generated from the identifiers.

The next step is to extend the system to include more complex organic compounds, such as polymers, and ultimately to tackle inorganic compounds. By adding it to software packages commonly used to draw chemical structures, the NIST team hopes that IChI will enter into widespread use.

In effect, the IChI number will provide each chemical molecule with a digital object identifier (DOI) — a concept increasingly being applied to everything from scientific papers to individual genes. Jonathan Goodman, a chemist at the University of Cambridge, says chemistry suits this approach well. "Molecules are a wonderful unit of information to treat in this way," he says. "They are complex enough to have lots of interesting features and difficulties but simple enough to represent quite a small subset." ■

www.iupac.org/projects/2000/2000-025-1-800.html

APPENDIX X

Chemical Nomenclature, Terminology, Units and Symbols

The International Union of Pure and Applied Chemistry (IUPAC) is the body responsible for developing and disseminating internationally agreed recommendations for deriving chemical names, for chemical terminology in general, and for the use of units and symbols in chemistry. Until the beginning of 2002, this work was carried out by semi-permanent Commissions; it is now done by project groups of experts in the relevant subdisciplines, managed by one or another of the eight subject-oriented Divisions of IUPAC. In particular, recommendations for deriving chemical names are the responsibility of the new Division of Chemical Nomenclature and Structure Representation. For units and symbols used in common with other sciences there is liaison with other International Unions, and with international standards bodies (e.g. ISO and the Bureau International des Poids et Mésures). More information on IUPAC's objectives and way of working is given on the IUPAC website http://www.iupac.org .

The following list of IUPAC recommendations was compiled for inclusion in instructions for authors of articles for the journals of the Royal Society of Chemistry, and provides a valuable source of information for editors. It is also available, with live links to those documents that are accessible online, on the RSC website http://www.rsc.org/pdf/general/append.pdf . Probably the most useful items for those with only a superficial knowledge of chemistry are:

- IUPAC Compendium of Chemical Terminology (item 1.8); a searchable online version is freely available at http://www.chemsoc.org/goldbook
- Quantities Units and Symbols in Physical Chemistry (item 1.9)

A useful introduction to chemical nomenclature is given in:

- Principles of Chemical Nomenclature: a Guide to IUPAC Recommendations, G J Leigh, H A Favre and W V Metanomski, Blackwell Science, Oxford, 1998

Details of how to obtain IUPAC publications are given on the IUPAC website at http://www.iupac.org/publications/).

(Dr) Alan McNaught
President, IUPAC Division of Chemical Nomenclature and Structure Representation
General Manager, Production Division, RSC Publishing

APPENDIX XI

Minutes of the IUPAC IChI meeting, Columbus, OH

June 30, 2002

Attendees:

Steve Bachrach (sbachrach@trinity.edu)
Jonathan Brecher (jsb@camsoft.com)
John Brennan (jbrennan@epo.org)
Steve Heller (srheller@nist.gov)
Sandy Lawson (alawson@mdli.com)
Alan McNaught (mcnaughta@rsc.org)
Peter Murray-Rust (pm286@hermes.cam.ac.uk)
Warren Powell (wpowell2@juno.com)
Henry Rzepa (h.rzepa@ic.ac.uk)
Steve Stein (steve.stein@nist.gov)
Dmitrii Tchekhovskoi (dmitrii.tchekhovskoi@nist.gov)
Matt Toussant (mtoussant@cas.org)
Bill Town (bill.town@chemweb.com)
Tony Williams (tony@acdlabs.com)

Summary:

Steve Stein reviewed the progress made by NIST in developing the test version of the IUPAC Chemical Identifier – the IChI. The test version handles simple organic molecules. To date, in all of the testing (almost 70 copies have been distributed) there are no known examples of chemicals that the program does not handle. A number of suggestions (described below) were made regarding testing and output. The overall view was that the project is progressing considerably faster than expected. A lecture by Steve Stein on the project was given the following day at the CAS/IUPAC Conference on Chemical identifiers and XML for Chemistry and a copy of the slides presented can be viewed at:

http://www.hellers.com/steve/pub-talks/columbus-702/frame.htm

Background on the project can be found at:

http://www.iupac.org/projects/2000/2000-025-1-800.html

Steve Stein presented the results of the work being undertaken at NIST in support of the IUPAC Chemical Identifier (IChI) project, which is the result of the programming efforts of Dmitrii Tchekhovskoi. He started by reviewing how NIST evolved the algorithm, which was based on existing work of others and did not involve any new principles. A beta-version of the algorithm was distributed starting in March 2002. While almost 70 people have requested and received the beta version, the feedback has been minimal. The most useful comment from a beta tester has

been a suggestion to consider being able to represent molecules with both defined and relative stereochemistry centers. (Articles in *Nature* and *The Alchemist* have stimulated some requests for the test version, but have not resulted in any feedback to date.) No examples of chemical structures which the program cannot handle have yet been found. The testing has been done on a few databases, such as the NIST and NCI structure databases. A large file (perhaps 1 million) structures from MDL is expected to be tested shortly. While testing millions of structures is possible, there still is the question of needing to actually examine the output to be sure it is what was input. With a 700 MHz PC, average processing time per structure is 2 milliseconds, but faster PCs will reduce this time.

Steve Stein indicated there were two major problems found: Chemists and Chemicals. Chemists are a problem as they have different ideas on how to represent chemicals. This is a human problem not likely to be resolved. Chemicals are a problem since the chemical structure depends on conditions – such as temperature, pH, and so on. This is also a problem not likely to be resolved.

The assumptions being made for the IChI algorithm are:

1. Throw away all electron density
2. Free rotation around all single bonds
3. Always a basic connectivity layer, then an isotopic layer, then a stereochemistry layer (*Z/E* and sp^3), and then a tautomer layer.

Issues not yet resolved in the current beta version include stereogenic centers (an atom and its bonding partners which is not superimposable on its mirror image) and zwitterions. NIST expects to have all normalization rules defined and programmed by the end of 2002. A second beta version is expected to be released at that time. It was suggested that a feature to compare multiple structures and show their differences be included in the next test release.

The issue of what the final and actual output should look like was discussed. At present the plan is to have a number of parts to the output:

1. Molecular formula
2. Connectivity listing (i.e., connections between atoms)
3. Isotopic and stereochemistry

It was suggested that Sandy Lawson look into creating some hash-code identifier with some degree of chemical intelligence. Sandy agreed to consider this project.

Jonathan Goodman, Cambridge, (who was unable to attend) is working on a Java version of the IChI. The best way NIST has found to test the IChI is to renumber the structures and see if the results are the same.

The meeting attendees were all very pleased with progress made at NIST. The hope was expressed that others could be brought into the project for expert help in deciding on how to handle proteins, polymers, and other chemical classes of compounds, so that the IChI could be a true, pure, and complete chemical identifier.

Steve Heller, Secretary

July 8, 2002

APPENDIX XII

CAS/IUPAC Conference on Chemical Identifiers and XML for Chemistry
July 1st 2002, Columbus , Ohio

Meeting Program and Abstracts

9:00 – Introduction

9:10 - Matthew J. Toussant (CAS): *CAS chemical identifier systems*

Abstract: Chemical compounds, their syntheses, their properties, and their applications, are the core of chemistry. Recording, storing, and retrieving information on chemical substances have been paramount to the progress of chemistry. The challenge to CAS has been to provide its users and itself with efficient and effective means of identifying substances reported in the world's chemical literature. This presentation will describe the components of the CAS chemical identification systems with a primary focus on the CAS Registry System. The foundation of the CAS Registry is the computer-based connection table with its threedimensional structure representation. Complementing the structure-based representations are chemical substance names, including names systematically assigned by CAS according to a set of rigorously based rules, plus other systematic and semi-systematic names and trade names compiled from the chemical literature. Linking each set of structure and nomenclature information for a particular substance is the CAS Chemical Registry Number, a concise and unique identifier that has become widely used as a standard for chemical substance identification. A recently introduced thesaurus capability by CAS provides users with additional links to specific substance information from general subject and class terms. The CAS MARPAT service provides a means of identifying generic substances reported in the patent literature, thus complementing and extending the range of substance identification systems offered by CAS.

9:50 - Stephen E. Stein*, Dmitrii Tchekhovskoi, Steve Heller (Physical and Chemical Properties Division, NIST): *The IUPAC Chemical Identifier*

Abstract: IUPAC has long been a recognized source of rules for naming chemical substances. However, names generated by these rules are designed primarily for human communication and are not optimal digital representations of chemical identity. In view of the ever-increasing volume of digital communication in chemistry, IUPAC has undertaken a program to establish a digital signature for a compound derived algorithmically from its digital structure representation (connection table). It is hoped that this IUPAC Chemical Identifier (IChI) will one day become an accepted standard representation of chemical substances.
Following discussions at IUPAC meetings, a test version of the Identifier was developed at

NIST and distributed in March, 2002. The principal objective of this version is to begin a community-wide discussion of the form of the Identifier.
The first implementation of IChI was designed for covalently bonded structures only. It processes an input connection table in three steps: 1) Normalization - all structural information unnecessary for identification is ignored. This, for example, eliminates ambiguities arising from different representations of pi-electrons, such as occur in the depiction of aromatic and zwitterionic structures. 2) Canonicalization - each unique atom is given a unique label. This is a mathematical procedure applied individually to distinct "layers" that describe connectivity, tautomerism, isotopes, stereochemistry (presently includes sp^3 and Z/E) and charge. 3) Serialization - a string of characters derived from labels produced by canonicalization. This generates the observable output form of the IChI, which may be viewed as a series of ordered connection tables, one for each "layer".
In addition to current features of the IChI, this discussion will examine still unresolved structure representation issues as well as ideas for extension to other classes of chemical compounds.

10:50 - Alexander J. Lawson (MDL Information Systems GmbH): *Nomenclature practice and post-Postman factors*

Abstract: The history of the communication of concepts (as influenced by the technical development of the available medium) was dramatically summarized by Postman in the 1980's. The phases of relative importance of the graphic, spoken and written traditions have been a constant companion to the development of civilisation in general, always involving deep consequences for the societies involved. The current general trend in the technically developed world involves an accelerated transition to the graphic representation at the expense of the spoken word in particular.
This general phenomenon can be argued to apply also to the learned sciences, none more so than mainstream organic chemistry.
Some possible consequences and opportunities for the specialist field of chemical nomenclature will be explored, with particular emphasis on organic chemistry.

11:30 - Jonathan Brecher (CambridgeSoft Corporation): *From chemical name to structure: finding a noodle in the haystack*

Abstract: Of all ways to identify a chemical, the one with the longest history and widest use is the simple chemical name. On the one hand, the broad acceptance of chemical names brings several advantages, including that they are convenient and easy to use in many environments. On the other hand, those strengths bring with them several serious drawbacks when chemical names are used as chemical identifiers -- what is the molecular structure of "glucose", let alone "sugar"? This presentation will highlight the state of the art in interpreting textual chemical names to produce chemical structure diagrams. Practical uses of such automated conversions will be demonstrated, with special emphasis on the strengths and weaknesses of using chemical names as chemical identifiers.

1:45 - Antony J. Williams (Advanced Chemistry Development): *Unifying chemical nomenclature standards - the roundabout of names and structures*

Abstract: Systematic Nomenclature is predisposed to software generation since rules-based systems are ideal tasks for computers to handle. In an ideal world there would be a single static, non-language specific systematic nomenclature accepted by chemists and in general usage. With general acceptance, rigorous application of systematic rules would produce fully reversible chemical names from which chemical structures could be generated. Of course there are multiple systematic nomenclature systems and chemical names found in the literature often are only close approximations to the correct names. The ability to generate systematic names from structures would lead us to conceive that systematic names can be reversed to structures using software. Of course this is possible with trivial names, synonyms, IUPAC names and CAS Index names being able to be reversed to the structure based world. Systematic naming and its reversal are by their very nature demanding of quality and the adherence to naming standards is a true test during software development. This issue will be reviewed during this examination of systematic nomenclature software development.

2:25 - Janusz L. Wisniewski (MDL Information Systems GmbH): *Computer-based naming service for very large chemical databases: from AutoNom in the Beilstein File to AutoNom in the ISIS system*

Abstract: Design and practical implementation of algorithms and routines in the worldwide first computer-based system for generation of the systematic IUPAC-sanctioned nomenclature directly from connection tables of organic compounds is discussed. Detailed overview of the performance, accuracy, and reliability of the system is presented. Practical issues and obstacles encountered and solved during inclusion of the program package into the established traditional production of large chemical databases such as Beilstein Handbook at the beginning, the Beilstein Database later and finally of the current MDL Crossfire system are described. The seamless integration of the nomenclature software into a company compound database registration and production using the ISIS platform is discussed. Advantages of the AutoNom TT package as DLL for DBMS independent general Naming Services are illustrated and analyzed.

3:25 - Henry S. Rzepa (Department of Chemistry, Imperial College of Science): *The vision of a chemical semantic web*

Abstract: The increasing trend enabled by the Web is of fusion between the sources of primary data (Instruments, modelling and simulation, databases) and the repositories of terms, dictionaries and peer-reviewed publications in a multi-disciplinary environment. Much of this fusion currently has to be achieved with a significant injection of "human perception", both on the part of the creators and authors of the information and knowledge, and of second and tertiary publishing resources. Part of this process involves establishing "trust" and common

semantics within a domain such as chemistry. XML is essentially a remarkably powerful infra structure built up over the last six years which provides a set of guidelines for introducing appropriate elements of machine processing capability to the overall process, and where the previously expensive need to create software and tools to achieve this is ameliorated by the ability to re-use a vast communal toolkit. These themes will be illustrated in the context of a chemical vision of Berners-Lee's Semantic web, in which the "datument" (data+document) and the information objects it contains plays a central role. In such datuments, the chemistry can be identified via specific namespaced components such as CML (Chemical markup language) seamlessly integrated with other components such as MathML, STMML, SVG. Demonstrations of these concepts will be included in the presentation.

4:05 - Peter Murray-Rust (Unilever Centre for Molecular Informatics, University of Cambridge): *The chemical semantic web: a common infrastructure for chemistry*

Abstract: An XML infrastructure for a domain must be built from carefully designed components, which have an Open architecture and which can interoperate. The components describe agreed subdomains of the discipline, such as molecules, reactions, spectra, computations, analytical and theoretical. They are defined by XML Schemas, which take advantage of re-use of existing designs, e.g. reactions can be based in part on molecular components. Schemas allow very precise validation of chemical information objects - it is possible to prevent invalid input to systems - and this will lead to a degree of quality missing in current practice but essential for the Semantic Web. A key resource is metadata which must be systematised for Chemistry and must use a universal architecture. Metadata is required for discovery of resources, validation, and for descriptions and annotation (e.g. the history of a piece of chemical information as it passes through the community). Allied to this is an XMLbased
query system designed for chemical applications.
We shall describe an Open system, including a toolset, on which groups can layer their applications.

4:45 - Stephen E. Stein (Physical and Chemical Properties Division, NIST): *An XML namespace for IUPAC*

Abstract: IUPAC has long served as a source of standard terminology in Chemistry. However, these "standards" have generally been expressed in conventional publications for use in conventional publications. A project intended to transform a portion of these standard definitions into an IUPAC "XML namespace" to aid the digital transmission of chemical information is under consideration within IUPAC. The scope, goals, methods and challenges of this potential project will be discussed.

APPENDIX XIII

Report on the Stereochemistry Scoping Exercise

7 August 2002

A small working group was established in order to identify areas within the nomenclature specialty of stereochemistry where projects might be required. The members of the working group were:

Dr Richard Hartshorn (Convenor, Inorganic and Bioinorganic Chemistry)
Prof. Bruce M. Novak (Organometallic and Macromolecular Chemistry)
Dr. Gerard P. Moss (Organic Chemistry and Biochemistry)
Dr. Byron J. Bossenbroek (Chemical Abstracts Service (ret.))

A letter inviting comment and/or ideas was sent to over 80 editors of scientific journals. A copy of the letter is appended to this report. In order to facilitate future contact with journal editors, IUPAC may wish to consider establishing and maintaining a data base of names and addresses. A further 120 similar letters were sent by e-mail to the registrants of the 37th Euchem Conference on Stereochemistry. A posting was also made to the web-board of the Division of Nomenclature and Structure Representation, with the intention of inviting comment from the members of advisory committee of the division. RMH also took the opportunity to visit Jonathan Brecher (leader of a scoping exercise on structure representation), GPM, and BJB while travelling to and from a conference in Germany.

Overall the response was somewhat disappointing. Only 12 responses were received and a number of these contained only a comment to the effect that the person concerned had no comment to make. Of the remaining responses, several related to areas outside stereochemistry, but in view of the value of any outside comments in the area of nomenclature, these suggestions have been included in this report to the Division Committee.

The responses raised the following areas where work may be required:

- Updating and consolidating definitions and nomenclature in stereochemistry, including comments on outdated terms.
- Providing a clear explanation of the CIP sequence rules, with examples, particularly of the more complex kind.
- Naming and rendering of structures with restricted rotation, or structures that contain absolute and relative stereogenic centres, where there may be many members in a relative group and more than one relative group.
- Bioconjugate Chemistry
- Supramolecular Chemistry

A further response was received relating to the standards and preferences for representation of the results of x-ray structural studies in publications. However, we suspect that this is something that falls within the purview of the International Union of Crystallography and individual journal editors.

The members of the working group are of the view that the first bullet point item is one worth pursuing. The establishment of the new Division makes it timely to consolidate and update the recommendations on stereochemistry that appear in the various IUPAC publications. There would also be some considerable benefit in including a glossary of outdated terms that provided their current equivalents and perhaps also gave some indication of the reason for the change. The scope of such a project would need to be considered in some detail as, for example, there are a large number of rather specialised nomenclature systems for particular classes of compounds. This is particularly true in the area of biochemical nomenclature.

The structure of the publication resulting from such a project might include several large chapters that address the general aspects of stereochemical nomenclature. These would then be followed by chapters on the more important classes of compounds that have their own nomenclature systems. A glossary would also be included, and this might be used to direct the reader to publications that deal with the less important nomenclature systems as well as to appropriate points in the text. RMH and GPM are interested participating in such a project, should the Division Committee feel that it is something worth pursuing.

Provision of a clear explanation of the CIP rules along with examples both of a basic and advanced nature is something we would also recommend. The examples should be numerous and drawn from organic, inorganic and organometallic chemistry. In principle, this might form part of the consolidation project outlined above, but it is of sufficient importance and general utility to require separate publication. Presumably the items under the third bullet point could also be incorporated into such a project.

While they undoubtedly do have stereochemical aspects to them, the more general areas of Bioconjugate Chemistry and Supramolecular Chemistry fall outside the area of interest of this working group. In neither case did the people suggesting these topics (Professors Claude Meares and Jean-Marie Lehn, respectively) identify particular problems that needed to be solved. Rather, it seems to be a case that the fields have grown to a stage where they feel that current practise should be evaluated, documented, and systematised. These topics should be considered for projects under the auspices of the Division. In the latter case, Professor Lehn suggested that either Professors Georges Wipff and/or Alexander Varnek could be approached on this matter (e-mails: wipff@chimie.u-strasbg.fr and varnek@crypt.u-strasbg.fr).

Dr R. M. Hartshorn

Department of Chemistry
University of Canterbury
Private Bag 4800
Christchurch New Zealand

Telephone: +64-3-364 2874
Facsimile: +64-3-364 2110
Email: r.hartshorn@chem.canterbury.ac.nz

24 April 2002

Dear

The Division of Chemical Nomenclature and Structure Representation has been established by IUPAC as part of the recent reorganisation of its activities to fit within a project driven framework. A small working group has been put together by the Division and given the task of identifying problems and areas for development in **stereochemistry and structure representation**. An early step in this process is to contact journal editors and others who may have an interest in this general area in order to invite comment.

We are particularly interested in identifying topics in the area of stereochemistry and structure representation where work is needed, and in mapping out requirements for the future so that projects can be established to address these issues. Please reply to this letter, or contact a member of the working group (details overleaf), if:

- a difficulty has arisen in this area as a result of your work
- you can foresee a difficulty arising
- you have any thoughts on the needs for the future and the role of IUPAC in addressing them
- you have any other comments.

Yours sincerely

Dr R. M. Hartshorn

Convenor, IUPAC Stereochemistry Working Group

APPENDIX XIV

Structure representation: scoping exercise

A new "scoping exercise" has been launched to assist the IUPAC Division of Chemical Nomenclature and Structure Representation in identifying what kind of guidelines for drawing chemical structures might be helpful to the community. Many people require maximum flexibility in order to be able to express in a diagram whatever chemical concepts are appropriate to the case in question. However, there are also situations (e.g. data collections) where some degree of standardization is helpful, and others (e.g.

stereochemistry) where the variety of styles in current use is confusing. Some discussion of structure drawing conventions already exists in previously published IUPAC recommendations, but it is scattered and incomplete. It might be appropriate for IUPAC to publish a cohesive set of recommendations to provide some guidance that has been lacking to date.

So far, discussion has focused on issues relating to depiction of stereochemistry. Despite the varied backgrounds of the participants in the discussion, there has been a remarkable lack of discord. This bodes well for the ultimate success of the exercise. Some partial (and tentative) new recommendations have already been written as a means of facilitating further discussion and of documenting consensus as it forms. Those materials are available at http://www.angelfire.com/sc3/iupacstructures/.

Moving forward, we will continue to work on additional tentative recommendations, and we will also be collating and documenting other conventions that already exist. Some conventions have been published by IUPAC and other standards bodies and other organisations. Other conventions are "obvious" to a practicing chemist, but seem never to have actually been documented. We will be looking for commonality of approach, and identifying areas where guidance from IUPAC could be useful. We would regard this work as a scoping exercise that would enable us eventually to define a project or projects to be funded by the new Division, although it is also likely that the scoping project will produce a number of documents that would be carried directly into that future project as well.

At this writing, about a dozen people are participating in the email-based discussion, which is also being archived at http://groups.yahoo.com/group/iupacstructures/. Anyone interested in contributing to the discussion is asked to contact Jonathan Brecher at jsb@cambridgesoft.com to be added to the mailing list.

Jonathan Brecher

19 July 2002

APPENDIX XV

The Chemical Nomenclature of Cyclic Organic Compounds

A.A. Durgaryan

Principally new chemical nomenclature for cyclic organic compounds has been offered. This nomenclature is simpler and more accessible, especially, for multicyclic compounds. The proposed chemical nomenclature is based on the principle, that cyclic compound is created from saturated hydrocarbon, by connection of hydrocarbon chain to non neighboring carbons. It is suggested that the letter "c" be added to the end of hydrocarbon's name and after the compound's name indicate in the brackets the number of linked carbon atoms forming the ring. Thus, cyclohexane is named hexanc(1-6) or hexane($c1^6$), methylcyclopentane – hexane($c1^5$), ethylcyclobutane – hexane($c1^4$), perhydronaphthalene – decane ($c1^6$, 1^{10}), perhydroanthracene – tetradecane($c1^6$,1^{14}, 8^{13}), bicyclo[3.2.1]octane – octane ($c1^7$, 4^8), tricyclo[5.3.1]dodecane – dodecane ($c1^6$, 4^9,7^{11}), cubane(pentacyclo[4.2.0.0.$^{2.5}$0$^{3.8}$0$^{4.7}$]) – octane($c1^4$,1^6,2^7,3^8,5^8), spiro[3.3]heptane – heptan($c1^4$,4^7), [2.2]-propellan – heptane($c1^6$, 3^6, 3^7), tetrasterane – dodecane ($c1^6$, 2^9, 3^{10}, 5^{12}, 7^{12}) etc.

For aromatic compounds we propose to replace the suffix "an" of the hydrocarbon name by suffix "ar", e.g. benzene is named hexar(1^6), naphthalene – decar(1^6,1^{10}), anthracene – tetradecar(1^6,1^{14},8^{13}), coronene – tetracosar(1^{18},3^{20},6^{21},9^{22},12^{23},15^{21},19^{24}), diphenyl – dodecar(1^6,7^{12}).

As another approach for aromatic cycles it is suggested to leave the name of alicyclic hydrocarbon and indicate the aromaticity in brackets by letter "a" before the numbers which indicate the carbon atoms linked together. Thus, benzene will be hexane ($a1^6$). This approach is more preferred in cases when substance contains both aromatic and nonaromatic cycles. In that case tetrahydronaphthalene will be called decane ($a1^6$, $c1^{10}$), 9,10 – dihydroanthracene – tetradecane (a 1^6, $c1^{14}$, $a8^{13}$).

New nomenclature for compounds containing functional groups is also easy-to-use. Thus, cyclopropanol will be called propanol ($c1^3$), hydroxymethylcyclopropane – 4-butanol ($c1^3$), cyclobutanone – butanone ($c1^4$), phenol – hexarol, α-naphthol – 2-decarol (1^6, 1^{10}), o-cresol – 1-heptarol, benzyl alcohol – 7-heptarol, benzoic acid – heptaroic acid, nitrobenzene – nitrohexar etc.

Heterocyclic compounds can be named using substituting nomenclature.

APPENDIX XVI

Nomenclature World Wide Web Database – Statistics

Statistics based on log of IP addresses used each day. Total usage to date about 1490000. Data on 170 countries recorded so far. Summary data for 1996-2002 at http://www.chem.qmul.ac.uk/iupac/usage/ For full details of each document see http://www.chem.qmul.ac.uk/iupac/ or http://www.chem.qmul.ac.uk/iubmb/

Average use **per week**

Year	1996	1997	1998	1999	2000	2001	2002	max
Total usage	296	650	1476	2786	5515	9813	14195	16549
Search Facility	-	-	-	204	1663	4169	7284	8834
Bibliographic Data	-	61	142	235	325	470	578	704
Map of Usage	-	7	8	29	37	58	78	138
IUPAC Nomenclature								
Class Names Glossary	138	157	430	693	1039	1504	2013	2578
Physical Org Chem Glossary	29	36	136	343	751	1089	1694	2212
Atomic Weight	23	48	95	144	310	651	845	1199
Bioinorganic Glossary	-	-	61	108	201	391	633	774
Stereochemical Glossary	-	32	85	135	231	392	565	733
Medicinal Chemistry Glossary	-	-	56	87	150	316	521	630
Section F (Natural Products)	-	-	-	14	121	321	413	541
Periodic Table	-	-	-	17	155	291	406	585
Fused Ring	-	-	64	73	110	198	229	370
Ions and Radicals	-	-	-	-	72	150	180	264
Gold Book	-	-	-	-	80	127	151	197
Numerical Term	-	18	27	35	54	99	138	182
von Baeyer	-	-	-	29	61	106	124	153
Hantzsch Widman	12	14	31	46	56	89	115	145
Spiro	-		-	26	47	90	110	139
Delta Convention	8	9	19	30	54	82	108	163
Phanes	-	-	31	42	56	80	88	111
Section H (Isotopic Label)	-	-	26	34	46	73	85	128
Lambda Convention	6	8	17	28	40	60	77	105
Element Name > 100	-	-	-	20	45	78	73	109
Guide Errata	-	-	-	20	21	25	28	48

Fullerenes (in preparation)	-	-	-	-	-	-	-	-
Phane II (in preparation)	-	-	-	-	-	-	-	-

IUPAC/IUBMB Nomenclature

Carbohydrates	46	72	144	237	453	835	1135	1486
Steroids	12	21	87	93	396	811	1072	1387
Amino Acids & Peptides	31	62	135	186	359	670	948	1122
Vitamin B-12	-	-	-	49	69	146	208	306
Tetrapyrroles	-	-	-	-	-	124	203	308
Lipids	-	-	-	29	70	132	188	255
Folic acid	-	-	-	60	58	210	178	269
Glycoproteins	-	-	20	32	71	134	170	219
Vitamin B-6	-	-	-	34	95	155	170	213
Nucleic Acid Abbreviations	-	-	-	45	77	136	167	215
Cyclitols	-	-	21	51	72	113	149	186
Polypeptide Conformation	-	8	14	34	61	111	143	210
Biochemical Phosphorus	-	-	-	-	62	103	141	202
Carotenoids	-	-	-	-	46	84	125	174
Lignans and Neolignans	-	-	-	-	-	71	122	155
Glycolipids	-	-	15	35	65	91	119	188
Tocopherol	-	-	21	33	48	80	119	186
Polysaccharide Conformation	-	8	14	26	49	82	113	180
Vitamin D	-	-	-	-	47	69	110	150
Retinoids	-	-	-	-	35	71	91	124
Polynucleotide Conformation	-	7	15	27	44	68	89	117
Quinones with Isoprenoid Chain	-	-	-	-	-	47	86	119
Polymerised Peptides	-	-	-	-	34	56	85	130
Prenols	-	-	-	19	33	55	73	97