IUPAC Strategy Round Table

Representations of Molecular Structure: Nomenclature and Its Alternatives

10-11 March 2000

US National Academy of Sciences, Washington, DC

Summary and Executive Committee Actions

Introduction

The International Union of Pure and Applied Chemistry was formed in 1919, very largely to provide internationally agreed nomenclature and other standards in chemistry. For 80 years, the systematic approach developed by IUPAC Commissions has furnished the basis for international nomenclature standards in organic, inorganic and macromolecular chemistry, and in biochemistry (in collaboration with the International Union of Biochemistry and Molecular Biology). IUPAC nomenclature is used not only by working chemists, journals and archival databases, but also in legal and regulatory matters.

During the last 20 years, the major changes accompanying the widespread use of computers have had substantial impact on nomenclature. Computer programs have used IUPAC rules to provide computer-generated names and to relate names to molecular structures and substructures. At the same time, it has become clear that the power of the computer can be harnessed to provide nomenclature tools, and the emergence of worldwide computer networks gives enormous potential for sharing of information and facilities. The increasingly international nature of scientific research and of commerce makes it even more important than in past decades to provide unambiguous chemical nomenclature.

IUPAC is in the midst of a major reevaluation of its international roles and a reorganization of its structure to meet the needs of the 21st century. This round table discussion brought together experts in organic, inorganic, biochemical, and macromolecular nomenclature; users of nomenclature in academia, industry, the patent, international trade, health and safety communities; journal editors and publishers; database providers; and software vendors -- a total of 41 participants from ten countries. Dr. Alan D. McNaught chaired the meeting. Twenty-seven participants provided pre-meeting discussion documents that described needs for advances in nomenclature and presented proposals for IUPAC activities.

Following the review of perceived needs at the first session, the Round Table turned to proposals for actions that IUPAC might take to facilitate the development of chemical nomenclature, from continuation of some past and ongoing activities to the inception of more novel activities. The participants formulated specific recommendations to IUPAC's governing bodies to guide the development of long-range strategies; to suggest high priority projects that should be undertaken soon; and to propose organizational arrangements for initiating, coordinating and managing nomenclature activities.

This report includes the following sections:

Summary and Executive Committee Actions

Recommendations and Comments

1. PREFERRED NAMES

IUPAC rules for systematic nomenclature often allow multiple names for the same molecule. The Commission on Nomenclature of Organic Chemistry (CNOC) has an ongoing project to recommend "preferred names" for organic compounds. The participants strongly supported completion of this project as soon as possible. In spite of reservations expressed by experts in inorganic and macromolecular nomenclature on the ease with which preferred names might be established in these fields, there was a broad consensus to urge that preferred names be given as soon as feasible in inorganic nomenclature and that macromolecular nomenclature indicate a preference between structure-based and source-based names.

2. LARGER PARENT STRUCTURES

The participants supported the creation of new class types and the adoption of larger parent structures so that simpler names could be generated for complex molecules. It was felt that the lack of simple, useable names, led to the creation of idiosyncratic names by individual researchers. This leads to confusion in the literature.

3. COMPUTER FRIENDLY RULES

CNOC has begun to involve representatives from software developers in order to better frame IUPAC nomenclature for use by computer programs. The group supported the further implementation of this trend by the other IUPAC nomenclature Commissions.

4. IUPAC ON-LINE NOMENCLATURE

Agreement has been reached with Beilstein and ACD to employ limited capability versions of commercial naming software to provide a free naming service via the IUPAC web site. This free service will be restricted to molecules of 50 atoms or less, containing C, H, O, N, S, P, and halogens, with a maximum of three rings. The group supported this service and urged its implementation as soon as possible.

5. NOMENCLATURE ADVICE

To complement the name-generation facilities, the group recommended that IUPAC include on its website links to commercial sources of advice on nomenclature (*e.g.*, the UK Laboratory of the Government Chemist, ACD Labs, TopTerm).

6. DATABASE OF SYNONYMS, TRANSLATIONS, STRUCTURES AND OTHER DATA FOR COMMONLY ENCOUNTERED COMPOUNDS

The discussion on this subject focussed on the need to provide a list that regulators and others could use. IUPAC should coordinate the development of a list covering those chemicals in common use in trade or others commonly encountered, such as in the environment.

7. CHEMICAL IDENTIFIER

A proposal by Dr. Stephen Heller for a major new IUPAC initiative was extensively discussed and clarified. Framed initially in terms of an "IUPAC chemical registry system," the proposal was recast as a "chemical identifier" -- a meaningful alphanumeric text string that can uniquely identify a chemical compound and facilitate its handling in computer databases. This code would be the equivalent of an IUPAC systematic name but would be designed to be easily used by computers. The Identifier could also include other information about the specific substance in question. Since there are several issues to be resolved, the participants recommended that the feasibility of the project and resolution of these issues be carried out as soon as possible by representatives of a wide range of interested parties. Drs. Heller and Stein (NIST) were asked to prepare a list of individuals and groups that should be consulted initially and to propose a framework for addressing the issues.

8. GENERIC/MARKUSH STRUCTURES

Since generic and "Markush" structures play an important role in patents, there was considerable interest in developing a standard for Markush diagrams. Dr. Brennan was asked to survey the patent community to determine the level of such interest. Meanwhile, the Medicinal Chemistry Section Working Party on Combinatorial Chemistry was asked to continue to study standards for generic structures.

9. CHEMICAL MARKUP LANGUAGE (CML)

CML is an application of XML with special ability to handle chemical information. XML is a new standard being adopted by web publishers worldwide. It is expected to replace HTML in many applications over the next few years. The group concluded that CML had reached a point such that it should be sponsored by IUPAC. This would involve IUPAC informing OASIS, the international body coordinating the development of the XML standard, that IUPAC was prepared to name a representative to OASIS and that IUPAC would be willing to coordinate the development of CML by the global chemical community. It is expected that other organizations interested in the communication of chemical information, especially publishers, will also participate in this work.

10. COORDINATION AND PLANNING OF NOMENCLATURE PROJECTS

The group recognized the need for a continuing body to plan and coordinate nomenclature projects, both of the traditional sort and in the broader fields of computer-based structure and chemical identity. Participants pointed out the interrelation of these subjects and the need to overcome the barriers that currently exist between the existing discipline-oriented nomenclature Commissions. While no reasonable sized committee could have the expertise to initiate and review in detail projects in all areas of nomenclature, a small group could manage projects and be sure projects are attracted from the community in areas requiring new nomenclature. The group recommended the formation of such a coordinating committee.

11. NOMENCLATURE AND EDUCATION

The chronic problem of teaching the importance of systematic IUPAC-derived nomenclature and encouraging its widespread use was discussed at some length. Many participants were favorably impressed by the approach adopted by the UK Association for Science Education for handling nomenclature in secondary schools and felt that it might be more widely disseminated by IUPAC. The teaching of nomenclature as a separate subject was felt to be not productive in colleges and universities. Nomenclature should be included as part of standard textbooks, with an emphasis on general concepts and an introduction to the use of naming software. The group recommended that suitable proposals be developed for consideration by the Education Strategy Development Committee.

12. NON-SYSTEMATIC NOMENCLATURE

Many chemical compounds are commonly identified not only by systematic IUPAC names (*e.g.*, propan-2-one) but also by other simple names (*e.g.*, acetone). The IUPAC nomenclature community has long referred to such non-systematic names as "trivial names." Although this term is understood by nomenclature experts, it sounds pejorative to many chemists. The participants agreed that an alternative term would be desirable but reached no consensus on what such a term should be, hence made no formal recommendation on this point. However, the term "common name" (implying commonly used) garnered the most support, with others such as "traditional name," colloquial name," and "short name" also mentioned.

13. ABBREVIATIONS AND ACRONYMS

There was considerable discussion of the value of a list of abbreviations and acronyms used in chemistry and of the feasibility of preparing and maintaining such a list. No consensus was reached.

14. NOMENCLATURE AND JOURNALS

Journal editors find it difficult to enforce the use of proper nomenclature. The group discussed the desirability of stronger IUPAC efforts to encourage use of systematic nomenclature, but no consensus was reached. The development of the CML standard, along with naming software, should make the establishment of names more transparent to authors and therefore enable greater compliance.

ACTION BY THE EXECUTIVE COMMITTEE

The IUPAC Executive Committee (EC) met soon after the Nomenclature Round Table [San Francisco, March 24-25] and received a summary report of the recommendations. The EC expressed appreciation for the in-depth analysis of nomenclature issues and indicated its support for the recommendations. Most of the recommendations from the Round Table are being implemented or can be implemented without formal action by the EC. However, on the recommendation of the Secretary General the EC approved three actions for immediate implementation, as follows:

1. An *ad hoc* Committee on Chemical Identity and Nomenclature Systems (CCINS) was established to be responsible for developing systems for conventional and computer-based chemical nomenclature; coordinating interdisciplinary activities in the nomenclature field; recommending to the Bureau long-range strategy on chemical nomenclature; cooperating with the four existing Nomenclature Commissions; and ensuring that long-term central planning, management and coordination of chemical nomenclature continues after the current Commissions are discontinued at the end of 2001. Dr. Alan D. McNaught was appointed chairman of CCINS.
2. A feasibility study of the Chemical Identifier project [Recommendation 7 above] was approved, to be managed by the CCINS.
3. The appointment of an official IUPAC representative to OASIS, the coordinating group for development of XML [see Recommendation 9 above] was authorized when and if such a representative is deemed appropriate by the Secretary General and the Chairman of CCINS.

IUPAC Strategy Round Table

Representations of Molecular Structure: Nomenclature and Its Alternatives

10-11 March 2000

US National Academy of Sciences, Washington, DC

List of Paticipants

1. Dr. W. Emmett Barkley
   Director of Safety
   Howard Hughes Medical Institute

2. Dr. Edwin D. Becker
   Secretary General
   IUPAC

3. Mr. Jonathan Brecher
   Software Development Manager
   CambridgeSoft Corporation

4. Dr. John Brennan
   Director in Chemistry
   European Patent Office, DG1

5. Prof. James R. Bull
   Department of Organic Chemistry
   University of Cape Town

6. Mr. Harold E. Cole
   Patent Counsel
   Eastman Kodak Co.

7. Prof. Andreas Dress
   Forschungsschwerpunkt - Math., Struktur.
   Universität Bielefeld

8. Dr. Geoffrey Fairhurst
   Information Manager
   Department of Scientific Information
   BASF Aktiengesellschaft

9. Dr. Lorrin R. Garson
   Director, Information Technology/Publications
   Publications Division
   American Chemical Society
   Unable to attend

10. Mr. John J. Gersic
    Chief, Energy, Chemicals and Textiles Division, Office of Industries
    U.S. International Trade Commission

11. Prof. Leslie Glasser
    Dept. of Chemistry
    University of Witwatersrand

12. Dr. Jonathan M. Goodman
    Royal Society University Research Fellow
    Department of Chemistry
    University of Cambridge

13. Dr. Stephen Heller
    Guest Researcher
    National Institute of Standards and Technology

14. Dr. Rouget F. Henschel
    Patent Agent
    Lyon & Lyon LLP

15. Prof. Bernardo Jerosch Herold
    Química Orgânica
    Instituto Superior Técnico

16. Prof. Reuben Jih-Ru Hwu
    Institute of Chemistry
    Academia Sinica

17. Dr. John W. Jost
    Executive Director
    IUPAC

18. Dr. M. Volkan Kisakürek
    Managing Director
    Redaktion
    Verlag Helvetica Chimica Acta

List of Paticipants

19. Prof. Pavel Kratochvíl
    Ústav Makromolekulární Chemie
    Akademie ved Ceske Republiky

20. Dr. Alexander J. Lawson
    Director of R&D
    Beilstein Informationssysteme GmbH

21. Prof. G. Jeffery Leigh
    School of Chemistry, Physics and
    Environmental Science,
    University of Sussex

22. Dr. David Lide
    Editor-in-Chief
    CRC Handbook of Chemistry and
    Physics

23. Dr. Derek Maclean
    Senior Scientist
    Affymax Research Institute

24. Dr. Alan D. McNaught
    General Manager, Production Division
    RSC Publishing
    Royal Society of Chemistry

25. Dr. W. Val Metanomski
    Senior Scientific Information Analyst
    Chemical Abstracts Service

26. Dr. Gerard P. Moss
    Department of Chemistry
    Queen Mary and Westfield College

27. Prof. Peter Murray-Rust
    Director
    Virtual School of Molecular Sciences
    Nottingham University
    Pharmaceutical Sciences

28. Dr. Rudolph Potenzone
    Sr. Vice President, Product Development
    MDL Information Systems

29. Prof. Jan Reedijk
    Leiden University, Gorlaeus Lab.

30. Prof. Jean-Claude Richer
    Département de Chimie
    Université de Montréal

31. Dr. Mike Ricks
    Managing Patent Attorney
    Intellectual Property Department
    Zeneca Agrochemicals

32. Dr. John Rumble, Jr.
    Chief, Standard Reference Data Program
    National Institute of Standards and
    Technology

33. Dr. Mary Scanlan
    Director, Publishing Operations
    Publications Division
    American Chemical Society

34. Dr. Steve Stein
    Director
    NIST MS Data Center
    National Institute of Standards and
    Technology

35. Dr. Robert Swann
    Director of Information Systems
    Chemical Abstracts Service

36. Mr. T. C. Swinfen
    Editor of 'Signs, Symbols and
    Systematics, the ASE Companion to 16-
    19 Science
    UK Association for Science Education

37. Dr. Kevin Thurlow
    Chemical Nomenclature Advisory
    Service
    LGC (Teddington) Ltd.

List of Paticipants

38. Dr. Matthew J. Toussant
    Director, Editorial Operations
    Chemical Abstracts Service

39. Dr. William G. Town
    Director of Operations
    Chemweb, Inc

40. Dr. Antony Williams
    Chief Science Officer
    Advanced Chemistry Development

41. Prof. Yecheskel Wolman
    Department of Organic Chemistry
    Hebrew University of Jerusalem

IUPAC Strategy Round Table

Representations of Molecular Structure: Nomenclature and Its Alternatives

10-11 March 2000

US National Academy of Sciences, Washington, DC

Agenda

Friday, March 10

9:00 am Introductions

9:30 - 12:00 Discussion of needs for chemical nomenclature from the perspective of various user communities - academic chemists; industry; patent, trade, heath, safety, and other legal and regulatory aspects; publishers; abstract services; database providers; and others

12:00 – 1:00 Lunch

1:00 - 5:30 Discussion of possible activities for IUPAC in chemical nomenclature: desirability, feasibility, methods of implementation

5:30 - 7:00 Reception

Saturday, March 11

9:00 - 12:00 Development of recommendations for IUPAC programs and projects

1. General Comments
2. Proposal for Registry System
3. Preferred Names
4. Larger Parent Structures
5. IUPAC On-Line Nomenclature
6. Computer Friendly Rules
7. Database of Synonyms, Structures and Other Data for Commonly Encountered Compounds Including Translations
8. Nomenclature Advice Network
9. Standard Molecular File Format
10. Generic/Markush Structures
11. IUPAC Sponsorship of CML
12. Abbreviations & Acronyms
13. Alternatives to the Term "Trivial"
14. Nomenclature and Education and Journals
15. Coordination & Planning of Nomenclature Projects

12:00 - 1:00 Lunch

1:00 - Group discussions:
Chemical Identifier Project
Nomenclature Coordinating Committee

IUPAC Strategy Round Table

Representations of Molecular Structure: Nomenclature and Its Alternatives

10-11 March 2000

US National Academy of Sciences, Washington, DC

Meeting Report

1. INTRODUCTION

Dr. A. D. McNaught welcomed the participants to the meeting. He noted that there was only one person absent, Dr. L. Garson, who was recovering from an accident. Dr. McNaught also thanked the National Academy for its hospitality both in allowing the use of their facilities and in helping with the other arrangements. The plan for the first day would be to introduce and discuss the pre meeting submissions. The agenda for Saturday would be based on the discussion, especially proposals, on Friday.

Dr. T. M. Wong welcomed the group on behalf of the National Academy. She pointed out that the Academy had as one of its purposes the promotion of international cooperation in science. This meeting was an excellent example of scientific cooperation at the global level. She wished the meeting participants well, and hoped that their discussions would be fruitful.

Dr. McNaught then asked the participants to introduce themselves and to give brief descriptions of their interest in nomenclature.

Dr. E. D. Becker welcomed the participants to this meeting. In some ways, this meeting could be considered a successor to the founding meeting of IUPAC in 1919. One of the major concerns of that meeting was nomenclature, especially as it affected international scientific cooperation and trade. Dr. Becker then described the changes in the management of the scientific work of IUPAC that had been approved at the recent General Assembly, and would take full effect in 2002. As part of the reorganization, all existing commissions, including the nomenclature commissions, will be discontinued. This is, therefore, a good opportunity to consider not only the question of how the Union's nomenclature activities should be structured, but also what activities the Union should be supporting. Dr. Becker congratulated Prof. Richer and Dr. McNaught for their report, which had proved to be invaluable in organizing this meeting and setting its agenda.

IUPAC needs to allocate its limited resources as effectively as possible. The purpose of this meeting is to develop a sense of what the global chemistry community's needs are in the nomenclature area and what IUPAC can do to satisfy those needs. The meeting was called a roundtable to express the desire on the part of the organizers to have full participation by all. This group is not a decision making group, however, its recommendations will be taken seriously by IUPAC governance. Dr. Becker noted that he had no doubt that nomenclature would continue to be important to IUPAC and to the global chemical community.

2. INDIVIDUAL PRESENTATIONS

Dr. McNaught referred to the written submissions that had been circulated and indicated that it should be assumed that these had been read by the participants. To initiate discussion, he would ask a few participants to highlight issues and views within their communities.

2.1. ACADEMIC COMMUNITY, ORGANIC CHEMISTRY, BIOCHEMISTRY

Dr. G. P. Moss noted that chemists often communicate by drawing structures rather than by using names, IUPAC or not. Biochemists communicate using acronyms or abbreviations, often with no intention to provide structural information. IUPAC names are considered useful only when publishing in a journal. In many areas, especially new areas, involving complex molecular structures, the IUPAC rules do not cover all the cases. Dr. Moss also noted the need for access to older nomenclature rules in order to understand the older literature. One should be cautious in changing nomenclature because of the danger of rendering a large part of the current literature unintelligible to future researchers. Dr. Moss noted that the UK chemistry education standards for the use of nomenclature, to be described later by Mr. T. C. Swinfen, have the effect of sending students to University who do not know many of the traditional, "trivial", names for common chemicals.

Dr. Moss then made some comments on types of nomenclature. Systematic names are exemplified by IUPAC nomenclature. The name conveys structural as well as functional information. The CAS and Beilstein naming systems can be viewed as dialects of the IUPAC system. Some areas of nomenclature still require development, such as fullerenes. Semi-systematic names are commonly used in natural product chemistry. Examples are the naming systems for steroids and carbohydrates. "Trivial" names are often used in biochemistry, natural products, and pharmaceuticals. A special sort of nomenclature is represented by abbreviations. These can be either informal or codified. The latter is exemplified by the abbreviations used to specify amino acids in proteins, nucleic acids in DNA, etc. Class names are in high demand by practicing chemists and IUPAC could do more to satisfy this demand. Dr. Moss then noted that the naming system for enzymes was not structure based at all but rather was based entirely on the enzyme's action.

Dr. Moss pointed out that conventions and rules regarding consistent ways of drawing structures are very important in areas such as steroids and carbohydrates. He noted that the representation of stereochemistry is still difficult with the literature on the subject confused. He also pointed out that terminology, the meaning of words, is an important part of nomenclature. The distinction between conformation and configuration is one that not all chemists appreciate.

2.2. ACADEMIC COMMUNITY, INORGANIC CHEMISTRY

Prof. G. J. Leigh began his remarks by noting that inorganic nomenclature was not as highly developed as organic. There are several systems of inorganic nomenclature in use. This can lead to multiple names for the same molecule, depending on the way in which the person naming it has chosen to approach it. For this reason, preferred names are not easy to agree on and will probably not be a significant aspect of inorganic nomenclature. Prof. Leigh noted that there were different approaches, even within IUPAC, to the nomenclature of molecules that can be viewed as either organic or inorganic. Prof. Leigh emphasized the importance of continuity in the people working on nomenclature to ensure that consistency is maintained as rules are developed. Prof. Leigh noted that the proposed nomenclature advisory service did not appear to be necessary. He also pointed

out that "trivial" names do not fill the same role in inorganic chemistry that they do in organic.

2.3. MACROMOLECULAR CHEMISTRY

Prof. Kratochvil noted that macromolecular chemistry uses two nomenclature systems, source-based and structure-based. Each has problems. Source-based nomenclature usually gives simpler names but source-based names can be ambiguous. Structure-based names eliminate the ambiguities found in source-based names, but they are often complex, and do not lend themselves to rapid verbal communication. Prof. Kratochvil commented that he expected both systems to continue in use.

Prof. Kratochvil then pointed out the importance of terminology for noiseless communication. Interdivisional communication in IUPAC can be very difficult due to different meanings of the same term. He also noted the importance of the dissemination of IUPAC rules, especially to textbook publishers. There is a need for active translation groups to ensure that IUPAC rules are made available to all the world's chemists.

Discussion

Dr. McNaught noted that while terminology is an important subject, it is not part of the remit of this group. He then asked Dr. Moss what the time scale was for the project on IUPAC preferred names. Dr. Moss replied that the preferred names project was part of a complete revision of the "Blue Book" and was expected to be completed by the end of 2000. Dr. McNaught asked Prof. Leigh why he felt that the preferred name approach was not feasible in inorganic chemistry. Prof. Leigh replied that it would be very difficult to impose preferred names in cases where alternative names had long histories and support in different parts of the community. Dr. McNaught commented that the benefits might justify the need to impose one name.

A general discussion then followed on the use of systematic names in scientific journals. The ACS has no requirement to use systematic names. They have found that this is too difficult to enforce, in terms of both time and the necessary skill of the editorial staff. The RSC asks for systematic names, but does not enforce this requirement. *Helvetica Chimica Acta* does require systematic names and enforces this requirement. They are able to do this because they have the necessary staff resource and expertise in house. Authors can be persuaded to use correct nomenclature, but it is difficult. Dr. Fairhurst asked how one knows that a name is correct. Prof. Hwu commented that currently available software often fails in naming complex molecules. Students have developed the practice of generating an informal name for the core of the molecule and naming it by adding substituent names to the name of the core.

Dr. Lawson commented that names should be both unique and unambiguous. There is also a need for a basic terminology. He suggested that we might learn from the design of computer algorithms. He then concluded that unambiguous is more important than unique.

Dr. Gersic commented that the confusion of names has been a problem for some time. He asked if anyone had compiled, or was compiling, a list (database) of synonyms.

Mr. Brecher noted that a compendium of trade names, etc. existed, but that a compilation of all possible names was impossible.

Prof. Dress noted that a name refers to a specific object in a given context. The best name can be different for different contexts. A canonical description is necessary to provide a basis for communication, especially for patents and databases. There is also the need to identify substructure. All this requires a computer interface to implement.

Dr. Toussant noted that an equivalence table is available via Chemical Abstracts Service. Dr. Stein commented that it would be possible to uniquely name all compounds, but that there needed to be a central source for storage of these names. Dr. Brennan noted that combinatorial libraries are defined in patents as a process byproduct.

2.4. PRIMARY AND SECONDARY EDUCATION

Mr. T. C. Swinfen began his discussion of the UK Association for Science Education's initiative in nomenclature by noting the difficulty of teaching the language of science. He briefly reviewed the systems of examination in use in the United Kingdom. The ASE decided that in its recommendations to science teachers they would recommend the use of one name for each compound students were likely to encounter in the course of their primary and secondary education. It was also decided to use systematic names as much as possible. This decision was made in order to enable students to recognize the structure of rarely encountered chemicals from the systematic name, rather than having to memorize long lists of common names for compounds likely to be encountered only once in a school career. Mr. Swinfen noted that these recommendations had been widely adopted by science teachers in the UK.

Dr. Fairhurst commented on the difficulties faced by the new graduate beginning work in industry, faced with the history of that industry and that company. That history is often embodied in the names of the materials used. Dr. Williams suggested that globalization would push world education to commonality in names.

2.5. LINGUISTIC CONSIDERATIONS

Prof. Herold began his discussion of language considerations in nomenclature by noting that one of his goals was to avoid the importation of "archaic" names into languages that do not have a long history of chemistry. He gave a number of examples demonstrating how the current rules change not only the order of substituents, but also their numbering, when a name is expressed in different languages. The current IUPAC rules require that in order to translate an IUPAC name, the structure must be generated from the name in the source language, and then a name generated from that structure in the target language. Attempts to simply translate the IUPAC name can lead to nonsense. The translatability of rules should be a consideration when they are developed.

Prof. Glasser commented that this reinforces the fundamental nature of the graphical representation. Prof. Leigh commented that the Commission on Nomenclature of Inorganic Chemistry had taken the view that they were writing rules in English and that the rules in other languages could require differences in detail.

2.6. INDUSTRIAL COMMUNITY

Dr Fairhurst noted that systematic names do not play a major role in industry because they are a cost factor and their use is avoided whenever possible. European authorities often request the IUPAC name, but CAS names are often accepted; these can be readily obtained if a CAS Registry Number exists. BASF integrates commercial nomenclature software into its working environment. In tests of currently available software, they found that 20% of the candidate compounds couldn't be automatically named. This is due to program limitations, which possibly arise because of inconsistencies or imprecision in some of the IUPAC rules. Dr Fairhurst commented that the existence of multiple registry systems increases the cost of information and makes life difficult for users. He suggested that there should be some cooperation among the groups that have proprietary systems, with IUPAC taking the lead in arranging this cooperation.

Dr. Williams commented that in his experience if a molecule can be drawn, it can be named providing systematic rules exist to generate the name for that structure. He suggested that a standard line structure could be helpful. He also noted that there were some issues about the standard way to display various classes of compounds.

Dr. Lawson commented that we should distinguish between labels and names. Names must communicate information directly. IUPAC should try to write its rules to lead to unambiguous assignment of names. There was then a discussion of what authors wanted in this area. The comment was made that authors wanted whatever the journals required. The earlier discussion on journal requirements was then recalled. It was pointed out that there is a difference between a recommendation by a journal and a requirement. Journals may recommend the use of IUPAC names, but usually do not require them, since it is difficult to obtain compliance by the authors, especially by the leading authors. If compliance is made easier through software, then compliance will be easier to require. It was also noted that, while this discussion has focussed on organic compounds, these represent only a small fraction of the substances assigned CAS numbers.

2.7. PATENT COMMUNITY

Mr. H. E. Cole discussed the nomenclature needs of the patent community. Nomenclature is important because it is necessary when searching the patent literature to find prior art. This can be very difficult. In many cases, searches by different organizations will find different prior art. If a search does not find a key patent, a patent can be filed and disallowed by an examiner based on prior art found by the examiner. The worst outcome is to have a patent allowed and be subject to interference claims by the owners of a patent that is prior art but was not found. The ability to properly define and find Markush structures, used to define general classes of compounds, is also important and could be improved by establishing rules. In summary, the patent community is interested in consistent definitions of terms and in unambiguous nomenclature rules. Dr. Brennan noted that in designating generic compounds, nomenclature is less significant than structure. Dr. Ricks pointed out that in patents there is a tendency to name a range of compounds based on the same core structure with substituents, even if correct application of the IUPAC nomenclature rules would require a change in the part of the molecule

defined as the core after the addition of a substituent. Dr. Henschel suggested that the three major patent offices could cooperate in this area.

2.8. JOURNALS

Dr. M. V. Kisakürek described the need to standardize acronyms and abbreviations. If a table of approved acronyms and abbreviations could be made available to authors, this would greatly simplify the task of the journal editor. In the field of supermolecular chemistry, there is a great need to give names to large parent structures, e.g. dendrimers. Dr. Kisakürek reiterated the point expressed earlier, that many journals do not have the time or the expertise to help authors with nomenclature. This leads in many cases to a request for the use of IUPAC nomenclature being ignored in favor of acceptance of any consistent system used by an author. This results in idiosyncratic variants of IUPAC nomenclature adopted by authors in a particular field as a matter of convenience.

Dr. Thurlow informed the group that the ISO committee on plastics is compiling a list of abbreviations and acronyms. Prof. Kratochvil suggested that instructions for authors should always require that they define acronyms when they are first used. Dr. Moss commented that editors should insist on the use of proper definitions. Prof. Leigh noted that while a table of abbreviations might be necessary, recommendations on how to make abbreviations should also be provided. Prof. Herold noted that given the frequency of the occurrence of letters in particular languages, acronyms would be duplicated. Dr. Moss noted that use of acronyms in chemistry should be considered in the broader context of science. Prof. Wolman suggested that acronyms should be based only on English. Dr. Rumble noted that trying to agree on a list of approved acronyms could consume a lot of time in discussion.

2.9. HANDBOOKS

Dr. Lide described the contents of the *CRC Handbook of Chemistry and Physics*. The *Handbook* covers ~35 000 substances. IUPAC recommendations on the use of symbols, terminology, and nomenclature are followed as much as possible. There is a need to balance the use of systematic names with the need to communicate with a diverse audience. Dr. Lide commented that many people would prefer a name be assigned unambiguously, even if arbitrarily, rather than balancing and carefully compromising alternative views. He suggested that IUPAC should be less sensitive than it has been to being seen to be imposing a particular view on the community. He suggested that the number of commonly encountered compounds for which names needed to be assigned might be of the order of 100 000. Most of the compounds that appear in CAS are referenced only once or twice in the literature. He noted that different CAS registry numbers are assigned by different sources to the same compound. Similarly, crystalline polymorphs and optical isomers are not handled consistently by sources. Dr. Williams suggested that an identifier directly related to structure rather than an arbitrarily assigned code would be useful.

Prof. Wolman noted that CAS numbers have an error identification algorithm. He also noted that some materials cannot easily be specified, e. g. nonstoichiometric compounds such as superconducting oxides. Prof. Kratochvil described the difficulties in specifying

polymeric materials. Molecular structure is often not sufficient. Dr. Rumble noted that there are different issues of nomenclature for compounds and complex materials. Prof. Dress commented that there are various levels of description, from the level of the molecular graph to the folding of DNA. Dr. Moss noted the problem of tautomerism.

2.10. The Secondary Literature (Abstracting Services)

Dr. Toussant described the CAS registry system. He noted that it was designed to point from materials to references and to avoid the need for users to name compounds. Input to the registry database requires only a name, not even a structure. Last year CAS spent USD 30 million to maintain the system. The total expenditure to create the system was USD 750 million. It is probably not rewarding to attempt to create an alternative registry database. Dr. Toussant then noted that CAS would like to bring the nomenclature systems of CAS and IUPAC closer together. There is also a need to define data storage standards. Dr. R. Swann commented that there was a need to define standards for structure descriptions. Prof. Dress asked how the CAS database related to commercial sequence databases for genomes. Dr. Metanomski noted that ~70% of compounds are referenced only once in the database. Mr. Brecher asked why CAS registry numbers are proprietary. Dr. Toussant replied that they are the basis of the CAS system. He also noted that the use of registry numbers is licensed for use at low levels. Prof. Murray-Rust commented that a database provides authority for information. He then noted that the medical field is developing a meta-thesaurus on the Internet.

2.11. Government and Regulatory Needs

Dr. Thurlow described the needs of governments and regulatory agencies. He noted the need for unique and unambiguous names for both commercial and regulatory purposes. Authorities use multiple sources and want to be sure a material is correctly identified. Many groups use old IUPAC rules for consistency with older regulatory databases because it is an overwhelming task to update older information with new names. He noted that there are ~30 000 chemicals in world trade. A database of names is what the non-chemist user community needs.

2.12. Web Communities

Dr. Town pointed out the difference between formal and informal communication. Formal communication involves systematic names and relates to publication and databases. Informal communication is the norm on the web as well as in person to person communication. The structural diagram is a natural language for chemists. Trivial names are used when diagrams are not available. Chemists need to be able to connect structure to trivial names. Structure representation needs to be standardized. Nomenclature rules should be adjusted to facilitate structure to name conversion. He endorsed the concept of an IUPAC registry system as described in the proposal from Dr. Heller.

Prof. Dress commented that it was now possible to create a virtual database on the Internet using a signifier or compound identifier to find material on a particular substance. Prof. Glasser commented that authority is not easily identified on the web.

2.13. CHEMICAL MARKUP LANGUAGE (CML)

Prof. Murray-Rust described the status of Chemical Markup Language. CML is a specialized application (Data Table Definition) of XML, which is under development as a successor to HTML, the markup language currently in general use on the web, in many applications. Markup languages are a way of presenting information for use by reading programs, such as web browsers. XML, and CML as an application of XML, is more flexible than HTML as it enables an author to specify more information about the data contained in the document. CML contains definitions for data types that are of special interest to chemists. Efforts are also underway to develop applications of XML for use by mathematicians, physicists, and other groups of users with special interests.
Prof. Murray-Rust proposed that IUPAC should adopt CML and develop it. Prof. Leigh asked what level of effort would be required to bring CML to a satisfactory situation. Prof. Murray-Rust replied that viewers, creation software, and editors were needed. He then noted that there was a need to better define certain chemical concepts in a way that would allow these concepts to be included in CML.

2.14. SOFTWARE DEVELOPERS

Dr. R. Potenzone noted that the molfile format for structure files started as a proprietary format and was put into the public domain by its owner. MDL has helped create hundreds of registration systems for commercial clients. As a result, they have developed a product to create registration rules. The removal of ambiguity from the nomenclature rules would help these kinds of systems. Definitions of terms are also necessary for the creation of registration rules. Some issues to be resolved in creating a registration database are: how compounds should be stored; what is the subset of compounds for which there is data and which are of commercial or other significance; what are commonly encountered compounds.

Mr. Brecher commented that experience with the ChemFinder database indicated that searching by structure yields more hits than searching by name.

Dr. McNaught noted the plans being implemented to have naming software, with limited capability, available on the IUPAC web site. Dr. Williams then gave a brief demonstration of the ACD software for generating IUPAC names from structures. A limited capability version of this software and comparable software from Beilstein is what would be made available on the IUPAC web site. Mr. Brecher commented that a flow chart would help to decide on how to assign a name. He also suggested that more work should be done to integrate the updated versions of IUPAC rules with the older versions. Prof. Richer commented that had been done in the upcoming revised version of the Blue Book. Mr. Brecher then gave a demonstration of the ChemFinder system.

Dr. McNaught concluded the first day’s session by thanking everybody for their contribution. He and Dr. Becker would develop an agenda for the next day’s session based on the concepts discussed in today’s session.

3. SECOND SESSION

Dr. McNaught introduced the second session by asking if there were any comments on the previously distributed agenda. There were no changes suggested. Further comments were then offered by a number of speakers.

3.1. GENERAL COMMENTS

Prof. Hwu began by noting the need to teach students how to use computer programs to do nomenclature. He then described how IUPAC nomenclature has been adapted for use in Chinese. It is necessary to adapt nomenclature to the Chinese writing system rather than to simply translate the rules. Prof. Hwu recommended the book *Chemistry Through the Language Barrier* by Reid for an understanding of how difficult it can be to translate chemistry to languages other than the major scientific languages.

Prof. Bull emphasized the need to make nomenclature more relevant to students. In this connection, he noted the value of preferred names when teaching nomenclature. Prof. Bull then suggested that an on-line dialog process might be a useful way to develop nomenclature recommendations. Recommendations should be made available for public comment at an earlier stage than in the past. He noted that the documents generated for this meeting could form the basis for a special issue of *Pure and Applied Chemistry*. Prof. Moss noted that IUBMB enzyme nomenclature recommendations are developed entirely on-line. Dr. Thurlow added that the comments on ISO plastics names are made on-line.

Dr. Goodman noted that while software can connect structures to names, software cannot assign a preferred name in cases where a choice between alternative systems is needed (e.g. steroid vs. full systematic). IUPAC should concentrate on the assignment of preferred names. He asked if preferred names filled the same function as registry numbers. The preferred name need not be used for teaching. Prof. Richer commented that the preferred name text for organic compounds would be done at the end of this year. He then noted that not having preferred names was a choice formerly made by IUPAC to not impose its preferences on the community. The discussion today clearly indicates that the broader user community is less concerned with being free to make choices and more concerned with simple, unambiguous rules. Prof. Murray-Rust noted that for many applications, especially in databases on the web, there is a great need for a unique identifier. Structures, names, or preferred names do not satisfy this need. Dr. Moss noted that there are still areas of nomenclature that need development.

3.2. IUPAC REGISTRY SYSTEM PROPOSAL

Dr. Heller circulated a summary of comments he had received to his proposal. He noted that, based on some of the discussion at this meeting and other comments he had received, there was a misunderstanding of his proposal by some people. The proposed Registry System would be a set of rules to generate a unique identifier for molecules (structures). The algorithm to apply these rules would be public and open source. Dr. Goodman asked if this proposal implied a database of identifiers and structures or not. Dr. Heller replied that there might or might not be a database. Prof. Murray-Rust

commented that registering a structure and annotations could be expensive. Dr. Heller replied that annotation would be done by the submitter. Dr. Swann asked if a database was necessary to guarantee uniqueness. Dr. Heller replied that if the rules are well defined then the structure could be constrained to lead to a unique identifier. Dr. Swann observed that attempts had been made by others to do this with no success. Prof. Dress commented that the algorithm developed would have to include all information that is available. Dr. Williams observed that a key to such a process is the way the molecular structure is drawn. Dr. Stein suggested that the problem could be addressed in stages starting with simple and well-defined molecules.

Dr. Potenzone asked how the project would start. Dr. Heller replied that a small working party, interacting over the Internet, would set priorities and assign groups to work on different aspects of the project. Dr. Lawson commented that the problems were soluble by working from simple to more complex issues. Dr. Swann asked why do we need a new registry number. Mr. Brecher replied that the purpose of a Registry Number is to allow connection of disparate databases. There followed a discussion of the role of CAS and the CAS Registry Number.

Dr. Rumble asked if generation of the Registry Number would be reversible and Dr. Heller replied that it would not. Prof. Goodman asked why not, if it is unique. There followed a discussion of what would be required to ensure uniqueness. The point was made that the system for enzyme registration has certain similarities to this proposal. Prof. Murray-Rust observed that a unique chemical identifier offered the opportunity to hyperlink information in different locations on the Internet. Prof. Glasser commented that generation of the Registry Number should be reversible. He also supported the concept of dividing the work up among a number of groups. Prof. Leigh noted that there were two issues, is this project feasible, and is it worth doing. He felt that the answer was yes to both questions. Dr. Lide advised caution in relying on an algorithm to make all decisions. Prof. Richer commented that his preference would be to have CAS Registry Numbers connected to IUPAC names. Why try to duplicate an existing database.

Dr. Town summarized his view of the previous discussion by noting that there were four items under discussion:

1. A set of structure converters
2. A method of generating a unique label
3. A facility to generate labels
4. Utilization of these labels to access existing data

Dr. McNaught noted that this group had done all it could. If IUPAC is to proceed, it needs to set up small task groups. Dr. Fairhurst commented until item 1 above is done, a registry is not doable. He then noted that CAS covers all materials, whether or not they can be represented by structures. This discussion has covered only “structurable” materials. Dr. Heller replied that structurable materials would be done first. The group agreed that IUPAC should be setting standards for communication by modern technology among chemists globally. Dr. McNaught then asked for suggestions on ways forward. It was agreed that those interested would meet after the plenary session was completed.

3.3. PREFERRED NAMES

Dr. McNaught introduced the subject by noting that for organic chemistry progress is being made; completion of this work should be encouraged. The other disciplines should be encouraged to proceed along the same lines. The subject of preferred names should be talked through in all areas. Prof. Leigh commented that after the previous day's discussion he now feels that preferred names are both desirable and desired in inorganic chemistry. Dr. Moss observed that progress on preferred names in organic chemistry did not seem to be a problem. While all cases seem to have been covered, there may be some ambiguity in deciding what the parent structure is. He identified alkaloids as a potential problem. Prof. Herold suggested that once the algorithmic approach is in place, preferred name assignment could go faster because the name no longer has to be a unique identifier. Dr. Lide asked how pharmaceuticals fit into the preferred name project. Dr. Moss reviewed the lists of approved names provided by international organizations. These lists include IUPAC or CAS names. Mr. Brecher asked what the boundary is between the organic and biochemical names. Dr. Moss noted that CAS has gone with systematic names to a high level. A similar problem exists at the boundaries with inorganic and with macromolecular chemistry. Prof. Reedijk asked how boundary issues should be resolved. Prof. Leigh suggested that the system should allow synonyms. Prof. Glasser asked if the proposed Registry Number would include links to synonyms. Dr. Lawson proposed that there should be not only a set of preferred names, but also a preferred syntax. This should specify the placement of numbers, brackets, commas, etc. Dr. Moss noted that the 1993 Guide to Organic Nomenclature had moved in the direction of defining syntax.

3.4. CLASS AND PARENT STRUCTURE NAMES

The group recommended more work in this area. This would greatly simplify names by naming large collections of atoms. This should include supermolecular assemblies.

3.5. IUPAC ON-LINE NOMENCLATURE

Two alternative software programs for generating IUPAC names, from ACD and Beilstein, will be available on the IUPAC web site early in the second quarter. The naming service will be free and be able to name compounds of up to 50 atoms. The atoms can be H, C, N, O, S, P, and halogens, with a maximum of three rings.

The issue of IUPAC certifying naming programs as being compliant with IUPAC rules was raised. This would be very difficult to do. Dr. Moss noted that on a number of recent recommendations under consideration by CNOC Beilstein and ACD had been consulted on the implications for their naming software. Dr. Henschel suggested that IUPAC make its nomenclature documents available on-line. Dr. Jost reviewed the plans of the IUPAC Secretariat in this area. He first noted that most nomenclature recommendations for organic and biochemistry were on the web site of Commission III.1, maintained by Dr. Moss. As new nomenclature books appear, they will be made available on the web site, probably as Adobe Acrobat pdf files.

3.6. COMPUTER FRIENDLY RULES

The group encouraged the more active consideration by IUPAC nomenclature Commissions of the needs of software based naming.

3.7. DATABASE OF SYNONYMS, STRUCTURES AND OTHER DATA FOR COMMONLY ENCOUNTERED COMPOUNDS INCLUDING TRANSLATIONS

Prof. Herold supported work on this subject. He suggested that it include multilingual synonyms. He noted that the European Union's EINECS database would be a good place to begin. Funding might be available from the EU and other international organizations and from governments. Dr. Becker noted that this would have to be a living document. Prof. Leigh commented that such a database should include preferred names. Dr. Rumble pointed out that publication of such a document could cause liability claims and involve the Union in competition with commercial vendors. Mr. Brecher commented that IUPAC could develop a base set of synonyms and could also provide data on commonly encountered chemicals. He noted that the ChemFinder database currently has ~70 000 – 100 000 entries. Dr. Lawson commented that this could be a large and complex problem. Dr. Potenzone observed that he could see value in collecting data not readily available. Dr. Williams commented that if IUPAC provided a database, software vendors could provide tools for its use. Dr. McNaught summarized the sense of the discussion as being in favor of a feasibility study of this project.

3.8. NOMENCLATURE ADVICE NETWORK

Dr. McNaught introduced the subject by noting the recommendation in the report of Richer and McNaught circulated to the group. Should IUPAC provide advice? The range of possible activities is from increased staff at the Secretariat to the establishment of a network of IUPAC Fellows. Prof. Glasser commented that the computerized name generation proposal would make this unnecessary. Dr. Metanomski suggested that there might still be a need for a referral service for advice on nomenclature rules rather than on providing names. Dr. McNaught asked if the group thought IUPAC should refer people to commercial vendors. The consensus was that this should be done and the price for the services should be listed on the IUPAC web site. Prof. Wolman suggested waiting until the naming service is available and the registry project is further along. Prof. Kratochvil noted the importance of terminology to many Divisions and proposed the convocation of a meeting similar to the current one to consider terminology. Dr. Becker pointed out that IUPAC provides recommendations on a wide range of terminology that applies not only to chemical structure but also to such areas as spectroscopy, electrochemistry, clinical chemistry, etc., with limited common interest. It is expected that the IUPAC Divisions will continue work in various aspects of terminology, with IDCNS overseeing the consistency of the resultant recommendation.

3.9. STANDARD MOLECULAR FILE FORMAT

It was pointed out that the group that considers the computerized naming proposal would need to deal with this item as part of its work.

3.10. GENERIC STRUCTURES AND MARKUSH DIAGRAMS

Dr. Brennan commented that this subject had been discussed by many bodies with no resolution. Dr. Town suggested that this should be a part of the computerized naming project. Dr. Maclean noted that the Working Party on Combinatorial Chemistry of IUPAC Commission VII.M.1 would be pursuing the issue of generic structures. Dr. Fairhurst commented that there is a desire in industry for a standard Markush file format. There are many errors in current databases and a standard could help reduce the prevalence of errors. Dr. McNaught asked if IUPAC was the proper body to be concerned with this problem. Dr. Fairhurst replied that IUPAC was the proper body. Dr. Potenzone suggested that the topic of generic structures should be considered as part of the computerized naming project, while Markush structures should be considered separately.

Dr. Brennan noted that a meeting had been held in Washington as part of an ACS meeting in 1990 on the subject of Markush structures. The European Patent Office and United States Patent Offices had sent delegates. A report was published in the *J. Chem. Inf. Comput. Sci*., **31**, 1991, p.1-68, but there were no other outcomes. Prof. Richer commented that people tended to feel free to use generic and Markush structures as they wished because the rules were so idiosyncratic. Dr. Maclean noted that a paper had been published in the ("Definitions of templates within combinatorial libraries" by Alan R. Katritzky, John S. Kiely, Norman Herbert, Christophe Chassaing. *Journal of Combinatorial Chemistry* **2**, 2000, p.2-5) on the use of generic structures in combinatorial chemistry. Dr. Maclean also supplied the following after the meeting: the *Journal of Chemical Information and Computer Sciences* dedicated an entire issue (**37**, Jan/Feb 1997) to the proceedings of the 4th International Chemical Structures Conference, Noordwijkhout, Netherlands, 1996, which includes a number of papers dealing specifically with generic structures. Mr. Brecher suggested that a person with a patent background be added to the IUPAC Working Party on Combinatorial Chemistry. Dr. Brennan was asked to ask the major patent authorities about the need to establish better standards for Markush diagrams.

3.11. IUPAC SPONSORSHIP OF CHEMICAL MARKUP LANGUAGE

Dr. McNaught noted that this subject had been discussed at various meetings of the Committee on Printed and Electronic Communications. He then asked Prof. Murray-Rust to explain how he saw IUPAC's involvement. Prof. Murray-Rust began by explaining the current situation with regard to the development of XML. XML (Extensible Markup Language) is a new standard being adopted by web publishers worldwide. It complements HTML (HyperText Markup Language) by supporting the transport of complex structured documents and data. CML is an XML DTD (Document Type Definition) developed by Dr. Murray-Rust and collaborators with special ability to handle chemical information. It will take IUPAC nomenclature and structure representation as its default vocabulary (ontology) but can interoperate with existing conventions.

Dr. McNaught asked how we would move from an expression of interest to authoring tools and standards. Prof. Murray-Rust replied that the next step would be development of what are known as schema. These would allow the software developers to write tools for use by authors and web page designers. Mr. Brecher commented that the software

community would welcome a standard CML. Dr. Becker asked if IUPAC should have an official representative to OASIS (Organization for the Advancement of Structured Information Standards). Prof. Murray-Rust replied that there should be an official IUPAC representative to OASIS. He noted that the development of XML had brought out important issues in the area of ontology, open forums, and electronic materials development.

The group concluded that CML had reached a point such that it should be sponsored by IUPAC. This would involve IUPAC informing OASIS, the international body coordinating the development of the XML standard, that IUPAC was prepared to name a representative to OASIS and that IUPAC would be willing to coordinate the development of CML by the global chemical community. It is expected that other organizations interested in the communication of chemical information, especially publishers, will participate in this work. IUPAC's role will be to facilitate and coordinate the development.

3.12. ABBREVIATIONS AND ACRONYMS

Dr. Becker commented that the task of compiling a database of all acronyms seemed to be large and ill defined. Dr. Thurlow suggested that acronyms could be part of the synonyms database discussed earlier. Dr. Lide reminded the group about the document mentioned earlier on rules for the formation of acronyms (*Pure and Applied Chemistry* **52**, 2229-2232 (1980), see also *Chemistry International* **8**(2), 7-8 (1986)).
Dr. Metanomski noted that many glossaries contain acronyms. The *IUPAC Handbook 2000-1* contains reference to the 1980 IUPAC recommendations for the use of acronyms. Mr. Brecher suggested that a list of generally accepted acronyms would be useful.

3.13. ALTERNATIVES TO THE TERM "TRIVIAL NAME"

Dr. Becker noted that to people outside the nomenclature community the term "trivial" name has a derogatory connotation. Is there an alternative term that could be introduced? This would help in communicating about IUPAC nomenclature to the general chemical community. Dr. Thurlow suggested the term "common name." Dr. McNaught supported this proposal. He also commented that the term "retained name", which is being used in the draft revised Blue Book has too specialized a connotation and is not clear to those outside the nomenclature community. A number of other people supported the use of the term common as in "commonly used" name. This would not be synonymous with preferred name. Dr. Lawson asked if chemists really do not understand the meaning of the term trivial. If common is adopted, we will have to explain that trivial and common mean the same thing. Prof. Dress suggested as an alternative the term informal name. Dr. Thurlow noted that when he was considering the alternative spellings of sulfur he consulted the editors of the Oxford English Dictionary. Could that be done in this case?

3.14. NOMENCLATURE AND EDUCATION AND JOURNALS

Dr. Becker described for the group the examination of the Union's education activities being carried out by the *ad hoc* Education Strategy Development Committee. The presentation by Mr. Swinfen on nomenclature for secondary schools had shown an

interesting way of treating this area. Dr. Becker suggested that the group recommend to the ESDC that it look at the UK approach for teaching nomenclature for broader dissemination. Mr. Swinfen suggested that the US National Science Teachers Association should be contacted. Prof. Leigh commented that nomenclature should be connected to the chemistry being taught rather than being taught as a separate topic. Mr. Swinfen noted that the UK standards suggest that nomenclature (names and rules) not be tested as such, but rather should be incorporated in the curriculum as something students need to know to do chemistry.

Prof. Reedijk noted the problem of different nomenclature requirements of journals. Dr. McNaught noted that publishers find it difficult and expensive to enforce correct nomenclature usage. Prof. Glasser noted the related issue of different reference and citation structures for journals. Dr. McNaught commented that standard XML schema could resolve both of these issues by making them transparent to authors.
Prof. Kratochvil suggested that the President of IUPAC send a letter to journal editors asking them to follow IUPAC recommendations on nomenclature. Dr. Becker commented that such a message would almost certainly be ineffective and would have the effect of diluting the value of a message from the President. Dr. Lawson suggested that IUPAC post a list indicating which journals support IUPAC nomenclature.
Prof. Hwu commented that nomenclature is a service and should not impose a constraint on the community. He also suggested that an announcement should be sent to chemistry list servers and "web communities" regarding the availability of the on-line naming service on the IUPAC web site.

3.15. Coordination And Planning of Nomenclature Projects

Dr. Becker reviewed for the group the recent changes in the way IUPAC manages its scientific work. Commissions will be terminated at the end of 2001 and a project-based system will be used. Division Committees and Standing Committees will manage projects. The transition to this structure is underway, with the new project approval system in place and functioning. He noted that there seems to be a need for a group to provide continuity and planning in the nomenclature area, in addition to the projects recommended above. Prof. Richer and Dr. McNaught in their report recommended a committee to manage the nomenclature area. Dr. Becker suggested that the group might wish to recommend the need for some such body.

Dr. Metanomski suggested that such a group should be responsible not just for nomenclature but also chemical structure standards. Chemical Identity might be an appropriate term. Prof. Leigh agreed that such a group was necessary. He then noted it would be difficult to achieve the required breadth of expertise in a small committee. He also commented that there would need to be a mechanism to attract and train new people for nomenclature work. Prof. Herold noted that an IUPAC body is the only place where the various experts can meet. He commented that if the people who undertake translations of nomenclature recommendations were not involved in the discussions that lead to recommendations they would have difficulty in doing proper translations. Translation is more often a question of adaptation to a language than simply finding words to translate from English to another language. Dr. Jost commented that the adoption of the earlier

proposal to make recommendations available for public comment at an early stage would help this situation.

Prof. Bull asked how the proposed committee would relate to the work of IDCNS. Dr. McNaught commented that it was difficult for the same body to be a project management and a quality control body. Dr. Moss noted that some work would continue in the current way. The work of the Enzyme Nomenclature group is now done on an on-going basis, with regular advice from JCBN. Prof. Richer proposed the creation of a Division of Nomenclature. Prof. Leigh commented that whatever form is decided, the decision should be made quickly. Dr. Stein commented that that the differences between conventional nomenclature and computer nomenclature should be recognized. Dr. Moss commented that the role of such a committee should be defined. Is the committee a management group or a group for discussion of the results of a nomenclature project? Different sets of expertise would be necessary for coordination and for oversight of specific areas of nomenclature.

4. CONCLUDING REMARKS

Dr. Becker thanked the participants for their active participation in the meeting. The results have exceeded his expectations. This was due to the excellent preparatory material provided by the participants and by the free and open discussion during the past day and a half. Dr. Becker noted that the recommendations of the meeting would be presented to the IUPAC Executive Committee at its meeting in two weeks in San Francisco. He expected that the recommendations presented would be considered seriously by the Executive Committee and acted on positively.

Dr. McNaught concluded by adding his thanks to those expressed by Dr. Becker. The meeting had been a very successful one. He expected that it would have a significant impact on IUPAC and the field of Nomenclature for many years. Dr. McNaught asked that all those who were interested stay for discussions of the major recommendations in small groups.

5. ADDENDUM – GROUP DISCUSSIONS

A number of participants were able to remain after completion of the plenary sessions to help define the needs and suggest courses of action for the computerized naming proposal and for coordination of IUPAC nomenclature activities. Although the discussions were informal and were carried out by small groups sometimes meeting concurrently, the major outcomes are described here.

5.1. CHEMICAL IDENTIFIER PROJECT

The discussion of this proposal began by reviewing what had been discussed earlier during the plenary session. It was concluded that rather than creating a database or registry of chemical names, the project should aim to develop a set of rules to allow the generation of an unambiguous text string, suitable for use by computer software, from a chemical structure – a “chemical identifier." Prof. Murray-Rust and Dr. Town in separate presentations developed the ideas that had been expressed by various participants. After

the meeting, these concepts were further developed and a preliminary project description was written. What follows is that description.

The aim of the Chemical Identifier project is to establish a unique label, the IUPAC Chemical Identifier (IChI), which would be a non-proprietary identifier for chemical substances that could be used in printed and electronic data sources thus enabling easier linking of diverse data and information compilations.

IChI will not require the establishment of a registry system. Unlike the CAS Registry System, it will not depend on the existence of a database of unique substance records to establish the next number for any new chemical substance being assigned an IChI. It will use a, yet to be defined, set of IUPAC structure conventions, and rules for normalization and canonicalization of the structure representation to establish the unique label. It will thus enable an automatic conversion of a graphical representation of a chemical substance into the unique IChI label which can be performed anywhere in the world and which could be built into desktop chemical structure drawing packages (such as ChemDraw, ISIS/Draw, etc.) and online chemical structure drawing applets (such as ACD/Draw).

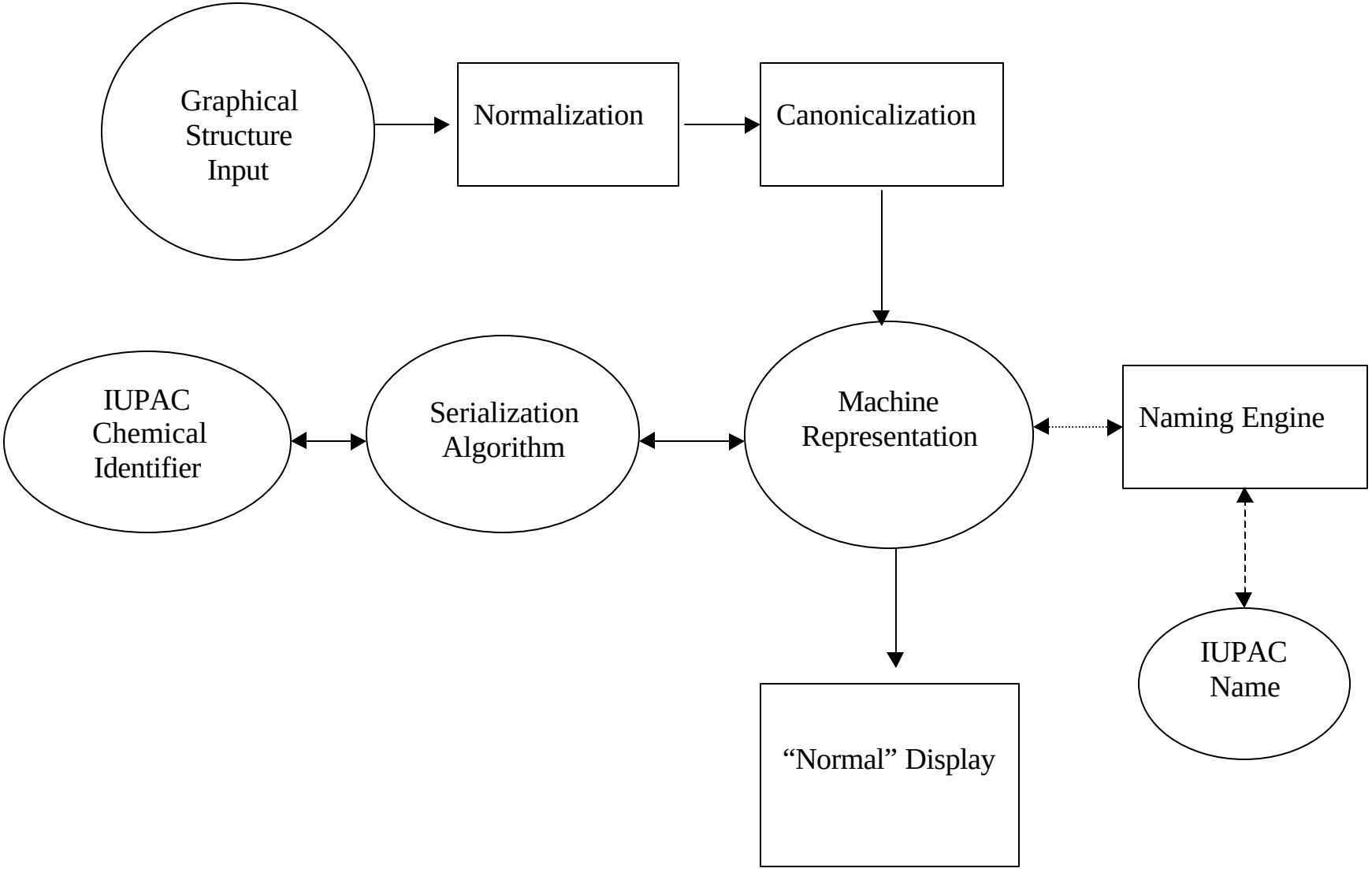
The brainstorming session after the IUPAC Strategy Round Table in Washington, suggested a number of mutually incompatible attributes for IChI:

1. It should be short and easily usable by humans and contain a check digit that could detect typing errors such as transposition of characters (as in the CAS Registry number)
2. It should be fully reversible (*i.e.*, a computer should be able to convert IChI back into a structure representation that can be displayed) which is likely to result in a representation too long to be used by humans
3. It should contain intelligence (*i.e.*, it should group families of salts, stereo isomers, etc.) and thus be a type of classification system. Consensus needs to be obtained by the working group set up to investigate the feasibility of the project [or imposed by IUPAC management] before the IChI project can proceed to establish working groups to look at structure conventions and the rules for normalization and canonicalisation.

A diagram follows which shows the process flow from input by a chemist of a structure, using drawing software, to the creation of the Identifier. IUPAC would define three aspects of this process: the rules for drawing a structure, the form of the in-memory representation of the structure diagram, and the mathematical algorithm for converting the in memory representation to a text string, the Identifier. The steps to convert from the drawn structure to the machine representation, normalization and canonicalization, would be done by vendor developed software. This part of the process has been implemented by a number of software developers. The new aspect to be introduced by IUPAC would be a standard data table structure.

The definition of a standard data table structure is necessary to allow the next step, the conversion of the data table to an alphanumeric text string. IUPAC would define the mathematical algorithm; the implementation of the algorithm would be left to interested software developers. The output from application of the algorithm would be the IUPAC Chemical Identifier.

As indicated on the diagram, the process would be reversible if criterion 2 above is met, so that the Identifier could be used to recreate the machine representation. This would then allow either display of the structure or generation of the IUPAC name. The Identifier would thus serve as the computer equivalent of the IUPAC name for a molecule. This would facilitate searching the Internet and labeling information in electronic documents with the name of the chemical substance in question.


Graphical Structure Input
Normalization
Canonicalization
IUPAC Chemical Identifier
Serialization Algorithm
Machine Representation
Naming Engine
IUPAC Name
"Normal" Display

5.2. NOMENCLATURE COORDINATING COMMITTEE

Dr. Moss and Prof. Leigh supported the concept of a central IUPAC committee to coordinate nomenclature work (including computer-generated nomenclature identifiers) but expressed serious concern about obtaining the expertise in a small committee that could substitute for the existing Nomenclature Commissions when they are discontinued. Dr. Becker suggested one model that would consist of a small core committee to deal with strategy, coordination, and management of projects, but with additional members (Associate Members?) expert in specific branches of chemical nomenclature and computer aspects. Much work could probably be carried out by e-mail, but any one of these ancillary groups could be convened as needed, separately or in combination with the core group. This model was thought to be worth pursuing. In any event, considerable further discussion on a permanent structure will be needed. At various times during the plenary sessions suggestions had been made for a new Nomenclature Systems Standing Committee; an expansion of the role of IDCNS to include the functions of nomenclature project proposal and management; and a new Division of Chemical Nomenclature.