(27) Brown, M. F.; Cook, B. R.; Sloan, T. E. Stereochemical notation in coordination chemistry. Mononuclear complexes. *Inorg. Chem.* **1975**, *14*, 1273-8.
(28) *C&E News*, July 9, **1990**, 9.
(29) Lipkus, A. H.; Blower, Jr., P. E. Stereochemical substructure searching: Handling of relative configurations. *Abstracts of Papers*, The Second International Meeting on Chemical Structures, Noordwijkerhout, The Netherlands, June 3-7, 1990.

Classification of Chemical Reactions: Potential, Possibilities, and Continuing Relevance†

DAVID BAWDEN
Department of Information Science, The City University, Northampton Square, London, EC1V 0HB England

Received March 18, 1991

An overview is given of the nature of, and methods for, classification of chemical reactions. Particular emphasis is placed on classifications based on formal structural change and on reaction mechanism, and on the role of classifications within computerized reaction retrieval systems. It is argued that classification is of great and continuing value in information retrieval, information discovery, and education.

INTRODUCTION

"The problem of classifying chemical reactions is intrinsically a complex one. None of the methods used to date has satisfied all the requirements of chemists. The reason for this is that there are so many attributes of a chemical reaction in which chemists might be interested." M. F. Lynch[1]

"Many proposals exist for the classification of chemical reactions. Some classifications have been devised simply to provide an author with a convenient framework for some limited purpose; others have been intended to be more rigorous and inclusive. Most have features in common, but the different schemes almost never agree when examined in detail: in short, there exists no consensus about the proper classification of reactions." D. P. N. Satchell[2]

In this paper, we shall consider the nature and meaning of the classification of chemical reactions, its value, and ways in which it may be achieved. We shall also consider how the utility and applicability of reaction classification can be extended in the future, with particular reference to computerized reaction databases. The choice of references is deliberately selective rather than comprehensive, with the aim of covering, and setting into context, all major strands in research into and operational use of reaction classification.

By "classification of reactions", I mean schemes for displaying and understanding the variety of chemical reactions and their interrelations. It is worthwhile at the outset stating what reaction classification is *not*.

First, it is not simply a scheme of nomenclature, notation, coding, or keywording for reaction description, although any classification must necessarily have its own notation. Reaction nomenclature has lagged considerably behind that for chemical structure, although some proposals for a systematic nomenclature, allowing access to *types* of reaction have been made.[3,4] Nomenclatures and similar descriptors, however, do not show the interrelation between reactions, which we shall see to be the essence of any reaction classification.

Second, it is not a computer-searchable form of reaction description. The computerized reaction retrieval procedures pioneered by Vladutz, Lynch, and Willett[1,5-9] and incorporated into operational reaction retrieval systems such as ORAC and REACCS are not in themselves classificatory, although, as we shall see, they may be used to formulate classifications. Nor are the similarity searching routines being incorporated into reaction systems classifications in themselves, though they also may be a tool for classification building.

† This paper is dedicated to Professor Michael Lynch, who first stimulated my interest in reaction retrieval.

VALUE OF REACTION CLASSIFICATION

There are a number of clearly understood benefits of classification, within information systems in general. They are involved in four distinct, though interrelated, processes. The first is the "straightforward" retrieval of information. The second is the less-well-understood process of information discovery, through analysis, correlation, and reasoning by analogy. The third is the teaching and exposition of the variety and scope of chemical reactions. The fourth is the systematization of chemical reaction information for use by other algorithmic systems.

For retrieval, classification confers four particular advantages. First, it can make retrieval from computerized databases more efficient. Second, it makes retrieval simpler and more accessible by encouraging a browsing approach. Third, it gives access to information at precisely the required level of specificity or generality and particularly enables easy access to "generic" types of information. Fourth, it makes it much more convenient to select subsets of related information for subsequent detailed search or analysis.

For information discovery and analysis, the creative and innovative use of the information resource classification allows the identification of new types of entity and the estimation of the qualities and properties of such entities. It is a powerful source of analogical reasoning and a tool for the understanding of entity relationships. In chemistry, the best known example is Mendeleev's use of the periodic table in classifying the elements and making rationalizations and predictions in analogical reasoning.

For chemical reactions, classification may be seen both as an aid to retrieval and as a discovery and analysis tool. In the former application, a classification will be a complement to structure-based retrieval techniques, to mechanistic nomenclatures, etc. In the latter, it may be used, for example, to identify new types of reactions, whose feasibility may then be investigated in the laboratory, or to show the presence of similar reaction mechanisms in apparently diverse structural environments. As Arens puts it:[10] "generalization can uncover the formal similarity of a wide range of reactions and can lead to their classification and to a strategy for the design of new ones".

For teaching and exposition, the value of classifications as an aid to understanding is well-understood in many subject areas. For chemistry, where much of the skill of a teacher lies in showing the order and interrelation hidden by a mass of facts, classificatory principles are invariably used, explicitly or implicitly, in dealing with concepts of chemical structure. This is less so in the reaction domain, and a major reason for

0095-2338/91/1631-0212$02.50/0

this is the lack of a comprehensive and generally accepted classification of reactions.

For systematization, classificatory principles are a sine qua non. This is particularly evident when the purpose of systematization is to allow large volumes of information to be fed into automatic procedures for analysis, correlation, and prediction. For chemical reactions, the most evident examples of this are the systems for computer-aided synthesis, for which a systematization of reaction information is an essential underpinning. This is particularly so if there is to be a direct link between systems for synthesis planning and for reaction retrieval. Another example of computer systems that could usefully draw upon classified reaction information are those which predict the course and mechanism of a given reaction.

It should then be clear that reaction classifications, far from being displaced or made redundant, by the advent of efficient structure-based reaction retrieval systems, have an increasing amount to offer, operated in conjunction with, or as a part of, such systems. The continuing, indeed increasing, value of reaction classifications is emphasized from very different viewpoint by Vladutz,[5] by Roberts,[11] and by Grethe and Moock.[12] Vladutz argues for the value of hierarchical structurally based classifications in making reaction information systems available to users in the most convenient fashion, especially for "off the cuff" queries. Roberts argues for the value of classifications in teaching and in allowing the interrelations of reaction types and concepts to be readily understood. Grethe and Moock show the usefulness in extracting classified subsets of information from larger computerized files.

FORMS AND ATTRIBUTES OF CLASSIFICATION

Although classification is one of the oldest techniques of scientific thought, indeed of all intellectual endeavor, it is difficult to arrive at an agreed and precise classification. Indeed, Gordon describes it as "a portmanteau term to describe a miscellaneous collection of techniques".[13]

We may take as a working definition the idea that a classification is an intellectual structure, showing relationships among the entities to be classified, based on assessment of interentity similarity. As two leading authorities have put it:

> "classification, in its simplest terms, means putting together things of ideas that are alike, and keeping apart those that are different"[14]
>
> "[classification is] the activity of seeking to assign objects to initially undefined classes in such a way that objects within a class are, in some sense, similar to one another"[13]

Of course, classification may be achieved at very different levels of generality. Different levels within one hierarchical scheme may be very advantageous to the user, but several different schemes at different levels may, as Satchell points out,[2] lead to confusion.

Within this broad definition there are a number of different forms and attributes of classification.

First, classification may either be derived automatically, i.e., algorithmically, or they may be derived by human intellect or intuition. In the former (algorithmic) case, the inter-entity similarities must be explicitly quantified so that some kind of automatic classificatory procedure may operate on them. In the latter (intellectual) case, the similarities may remain implicit and qualitative.

Second, classifications may be exhaustive, providing a niche for every possible entity of the sort being classified, or they may be partial and empirical, and based solely on an examination of a particular set of entities. The term "literary warrant" is used to describe the latter (pragmatic) form of classification applied to a library collection.

Third, classifications may allow overlapping, that is to say they may allow an entity to appear in more than one place in the classification, or they may enforce a unique position for each entity.

Fourth, a classification may be strictly hierarchical, or may be "clustered", with only a partial hierarchy or none at all.

Fifth, a classification may purport to be a single and unique arrangement of entities or may be deliberately proposed as one of a number of alternative, equivalent, and equally valid arrangements.

Sixth, a classification may or may not be designed for direct assimilation and understanding by a human user. At one extreme, the whole purpose of the classification may be didactic and explanatory; this will, of course, require that there is some clear notation and display format, so that the user may appreciate the classification, either on the computer screen or on paper. At the other extreme, the classification may be intended as an interface between a database and a computer program making use of it; in this case there would be no need for the classification scheme to be easily interpreted and understood by a human. An intermediate position would be a classification primarily intended for computer processing, but permitting sections or examples to be viewed by a human user.

As we shall see, many permutations of these forms of classification have some relevance for the classification of chemical reactions.

ATTRIBUTES FOR REACTION CLASSIFICATION

Any classification of a set of entities must be placed on some set of attributes, explicit or implicit, by which interentity similarity, again explicit or implicit, may be judged and a classification constructed. There are several attributes that may be applied for the classification of chemical reactions.

Undoubtedly the most widely applied have been attributes based on the *formal structural change* occurring in the reaction process. This may, of course, be described in different ways, as noted later. An alternative form of attributes for classification is that based on a knowledge of *reaction mechanism*. Most of the discussion below will be based on classification involving these two forms of attributes.

There are other attributes by which reactions may be classified. They include *utility*, i.e., what the reaction may be used for, *type of product*, *type of reactant*, or *type of reagent* or *conditions*. These have all been used for specific purposes, most commonly in printed volumes. They do not compete for generality of application with the two major types of attributes, formal structural change and mechanism, especially when computerized systems are the focus of attention.

MECHANISTIC CLASSIFICATIONS

Overtly mechanistic classifications of chemical reactions date back to the seminal work of Sir Christopher Ingold, who categorized reactions into mechanistic classes, such as SN2 and E1. There have been a number of attempts to develop Ingold's basic concept of a mechanistic reaction typology into a fuller classification; the best known is that due to Littler,[15] which has been accepted as an IUPAC standard. This builds upon the work of Guthrie,[16] which in turn refers to Ingold's concepts. The scheme, reliant upon intellectual coding, provides a linear notation, descriptive of both structural changes and mechanism, and includes chemical changes other than valence bond formation and breaking; weak associations, conformational changes, and bond rotation, for example.

Roberts[11] gives a variant on these schemes, attempting to describe and classify reactions at the lowest possible level, that of essential electron movements, without regard to the nature of the atoms involved. This represents all isoelectronic processes in the same way, i.e., classifying them at the same place.

Clearly valuable in showing up related reactions, it is not directly applicable to representing any reaction in a clear and full fashion. Nonetheless, it points up the value of an extended hierarchy in reaction classification, with the information content at each level geared to the requirements of its users at that moment. Littler's scheme could be modified to allow description at this level.

Bonchev et al.[17] present a somewhat different representation of reaction mechanism based upon a graph theoretic description. It is intended as an exhaustive procedure, so as to be able to indicate potentially novel mechanisms. As described, it requires intellectual coding, though the authors point out the value and feasibility of automation.

Satchell[2] has discussed some of the attributes and possible levels for mechanistic classification, beginning from the most elementary, but without providing any working scheme.

No mechanistic classification has yet been incorporated into a computerized reaction retrieval system, despite the insistence of their originators that this would be both feasible and worthwhile. The closest approach is the elaborate scheme of keyword descriptors in the ORAC retrieval system, many of which are mechanistic in nature.[18]

A link between mechanistic and structural classification, within the ambit of a structure-based reaction retrieval system, could be of particular value for the analysis, correlation, and discovery of mechanistic pathways. This, however, is a problem for the future.

CLASSIFICATIONS BASED ON STRUCTURAL CHANGE

At the risk of oversimplification, we may divide the work carried out on this topic into four categories, accepting that the division is far from watertight.

Literary Warrant Classification. In this form of classification, a given finite set of reactions is categorized according to some criteria. The resulting classification, although strictly limited to the set used for its construction, may, if that set is sufficiently large and representative, serve as a general classification of the types of reactions included.

This is clearly a type of reaction classification that has been often carried out, albeit informally, by researchers considering only those types of reactions of interest to them. The classification will necessarily be intellectual/intuitive and may be based either on structural changes or on mechanistic considerations.

An interesting recent example of this type of classification is given by Grethe and Moock.[12] They use a similarity searching option to find examples of a particular type of structural transformation and then categorize the resulting set of reactions into a nonhierarchical structural change typology. Their main example is the classification of cyclization reactions, producing five-membered heterocycles from 2- and 3-atom fragments. They point to the value of using similarity searching, particularly in conjunction with "conventional" substructure searching and with keyword access in constructing this sort of classification.

These authors also illustrate the "information discovery" aspects of classification. In classifying the set of Michael reactions found by similarity search, they identified a number of Michael-like reactions; the analogy-based similar but different factor, which is one of the keys to creative use of information resources.

Bond Change/Ring Change Classification. This is the oldest established form of reaction classification. Reaction descriptions in terms such as formation of C–C bond, reduction of C=O bond, or formation of six-membered ring have been used for many years for systematized information in textbooks and printed compendia. A well-known pioneer of this method of reaction classification was Weygand in his *Organische-Chemische Experimentierkunst*. The so-called Weygand code, denoting bonds changed in a reaction, was the predecessor of similar classifications used as the basis of the earliest automated reaction documentation systems, such as GREMAS, IDC, Derwent's CRDS, and Ziegler's Reactiones Organicae.[6] A familiar and convenient means of retrieval, they give classifications which are either nonhierarchical or partly hierarchical and have generally been intellectually produced.[19] Descriptors of this sort may, however, be derived algorithmically from reaction site analyses:[9] the extent to which such algorithmic procedures, without intellectual assessment, could be used directly for classification is a matter for debate.

Somewhat similar generic reaction descriptors, based on formal ring and atomic composition changes rather than on bond changes, are included in the CONTRAST in-house reaction retrieval system.[20]

Reaction Site Classification. Reaction classifications based on reaction site have come to the fore in recent years. This parallels the universal adoption of the algorithmically detected reaction site as the basis for computerized reaction retrieval systems, based on the work of Vladutz, Lynch, and Willett in the comparison of reactant and product structures.[1,5-8] The concept has been used for classification in somewhat differing ways and for different purposes, though all share the algorithmic, structure-based formalism.

Vladutz[5] describes a structurally based hierarchical form of reaction classification, algorithmically generated from comparison of reactant and product structures. He argues that this will provide a particularly intuitive and simple form of access for users of reaction retrieval systems.

Zefirov[21] presents a formal and systematic classification of organic reactions, based reaction center, and bond change description. The resultant classification is used both as input to a synthesis planning program and as a means for "discovering" potentially feasible new reactions. The exhaustive nature of this type of approach allows the identification of individual reactions or categories of reaction for which no laboratory examples have been described.

Gelerneter et al.[22] describe a "conceptual clustering" of reactions into classes. A machine learning procedure clusters and classifies reaction database entries into identifiable reaction species, leading to a uniform and standardized reaction classification scheme.

Blurock[23] describes the algorithmic extraction of reaction classes from the reaction database linked to the RETROSYN synthesis planning system. This is a strictly hierarchical classification, ordered with increasing structural specificity as reaction type: reaction center, expanded reaction center, specific reaction. Blurock emphasizes the value of this sort of classification in providing user-friendly access for high-level queries.

Hendrickson and Miller[24] describe a reaction classification system developed to systematize the reaction information used in the SYNGEN synthesis planning system. The algorithmic reaction site analysis considers the formal change occurring at each reacting carbon atom of the structures. This formal and theoretically based classification is thus entirely exhaustive and may be used independent of any "actual" reaction information. These authors demonstrate, however, that this classification can also be mapped onto the reactions in a computerized database, allowing retrieval by class and subclass. This is demonstrated by application to the REACCS and SYNLIB databases, the clearest example in the literature of a direct link between synthesis planning and reaction retrieval systems.

Reaction Matrix Classification. Closely related to the last category and separated from it largely for convenience of

presentation, this is also an algorithmic structurally based form of classification aimed at giving an exhaustive enumeration and systematization of reaction types. Generally associated with the work of Ugi, Brandt, Gasteiger, and Herges, reaction matrix classification is a highly formalized scheme involving the representation and manipulation of structure and reaction information in the form of bond/electron matrices.

Ugi et al.,[25] for example, show how this approach can be used to produce a hierarchical reaction classification suitable for application within retrieval systems for automatic classification of reactions and subsequent searching capabilities. The reaction matrix formalism leads to the concept of the reaction category, a generic reaction class corresponding to the reaction matrix.[26]

Brandt et al.[27] demonstrate a three-level hierarchical reaction classification based upon bond changes represented in the matrix formalism. At the highest, most general level, the nature of the atoms in the reaction site and the unchanged molecular environment are ignored. This means that such apparently disparate reactions as the Cope and pinacol rearrangements and the Diels–Alder reaction are classed together. The next and more specific level of the hierarchy includes the molecular environment, while subsequent, still more specific levels include atom types and similar information.

Arens' formalism[10] is similar in concept, though it does not employ the same form of matrix description and, as described, is limited to one class of reactions.

OTHER FORMS OF REACTION CLASSIFICATION

Compared with classifications based on mechanism or on structural change, other forms of classification have been little used in computerized systems.

Classification by *similarity of reagents or catalysts* is one example of clear if somewhat limited utility. An example has been given of the use of a computerized reaction retrieval system as a basis for the construction of literary warrant intellectual classifications based on these criteria, for reactions with Lewis acids as catalysts and for those with crown ethers as reagents.[12]

In the realm of printed volumes, the seminal work by Fieser and Fieser, *Reagents for Organic Synthesis*, has been widely used. Elaborate coding of reagents following this pattern is also a feature of the keywording in the ORAC reaction retrieval system.

Classification by *type of product* is of similar status. It has not been used in any computerized information system, although classification of reactions according to the functional groups or compound types obtained has been the norm in textbooks and encyclopedias. Perhaps the best-known example is the Houben-Weyl series of printed volumes in which reactions are categorized according to the functional group produced.

Classification by *type of reactant* has been little-used, certainly not in computer systems, although the categorization of reactions in the well-known *Cahiers de Synthese Organique* printed volume series was according to reacting functional group.

There may well be scope for extending the application of these types of classification in computer files. Although not of major and general importance, they would prove valuable for specific purposes.

CONCLUSIONS

Reaction classifications are still a valuable tool for the chemist and information scientist. Far from being supplanted by computerized reaction retrieval systems, the existence of these systems provides a dual spur to classification: by providing increased scope for its application and by providiing convenient tools for its development.

Classification should be seen as essentially a complement to "precise" retrieval tools, such as reaction site analysis and retrieval. Classifications provide generic access and promote a browsing style of access, which is particularly user-friendly. They fullfil the linked objectives of information retrieval, information systematization and discovery, and education/explanation.

There is a clear and continuing need for internal or hidden classifications imposed on a database for the purposes of a "user" computer program. Examples of such programs could be synthesis planning systems, retrieval and display programs, mechanistic prediction systems and tutorial, or computer-based training programs.

For didactic and educational purposes, it would be valuable to evolve one or more standard classifications of reactions for use in printed and computerized educational and training materials. The IUPAC standardization on Littler's scheme may point the way for mechanistic classification. A structure-based scheme could standardize on a intuitively clear definition of reaction site, such as that proposed by Wilcox and Levinson,[28] from which could be generated equally intuitively clear bond or ring change attributes. This could be the beginning of the formation of links between mechanistic and structural classifications.

It should be possible to allow users to construct their own classifications, largely by an intellectual typing of reactions according to need, based on searching in computerized reaction databases. The availability of similarity matching in reaction databases, which allows the user to draw in material with the appropriate degree of interrelation, may be particularly valuable in setting the boundaries for the classification exercise.

The potential also exists for algorithmic aids in the production of user-oriented classifications. The readiness with which chemical structures may be classified, by techniques of numerical taxonomy, is well established.[29] The classifications produced either may be hierarchical and monothetic (i.e., each structure has its unique place)[30] or may be nonhierarchical and permit overlapping.[31] Equivalent procedures for chemical reaction information are entirely feasible, although none has yet been described in the literature. The attributes for classification would be structure-based, reflecting either the reaction site structure alone or the reaction site in conjunction with features of overall reactant and product structure. The feasibility of this has been established in experiments with CONTRAST inhouse reaction retrieval system.[20]

Automatically generated classifications of this sort would still be under user control to some extent by variation of the parameters used in the classification process. The resultant classification could be either used as it stood or taken as the basis for an intellectual typing and categorization. Other multivariate analysis and display techniques could be used as an alternative to cluster analysis for this purpose.[32]

In short, the ability to provide for a variety of classificatory approaches and procedures is a necessary attribute of a mature reaction retrieval system. Research and development toward this end is a continuing need.

REFERENCES AND NOTES

(1) Lynch, M. F.; Nunn, P. R.; Radcliffe, J. Production of printed indexes of chemical reactions using Wiswesser Line Notations. *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 94–96.
(2) Satchell, D. P. N. The classification of chemical reactions. *Naturwissenschaften* **1977**, *64*, 113–121.
(3) Bunnett, J. F. Systematic names for substitution reactions. *Chem. Eng. News* **1954**, *32*, 4019.
(4) Bunnett, J. F. Nomenclature for straightforward transformations. *Pure Appl. Chem.* **1981**, *53*, 305–321.
(5) Vladutz, G. Do we still need a classification of reactions? In *Modern Approaches to Chemical Reaction Searching*; Willett, P., Ed.; Gower: Aldershot, 1986.

(6) Willett, P. The reaction indexing problem: a historical viewpoint. In *Modern Approaches to Chemical Reaction Searching*; Willett, P., Ed.; Gower: Aldershot, 1986.
(7) Harrison, J. M.; Lynch, M. F. Computer analysis of chemical reactions for storage and retrieval. *J. Chem. Soc. C* **1970**, *23*, 2082-2087.
(8) Lynch, M. F.; Willett, P. The automatic detection of chemical reaction sites. *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 154-159.
(9) McGregor, J. J.; Willett, P. Use of a maximum common subgraph algorithm in the automatic identification of the ostensible bond changes occurring in chemical reactions. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 137-140.
(10) Arens, J. F. A formalism for the classification and design of organic reaction. I. The class of $(\pm)_n$ reactions. *Recl. J. R. Neth. Chem. Soc.* **1979**, *98*, 155-161.
(11) Roberts, D. C. A systematic approach to the classification and nomenclature of reaction mechanisms. *J. Org. Chem.* **1978**, *43*, 1473-1480.
(12) Grethe, G.; Moock, T. E. Similarity searching in REACCS. A new tool for the synthetic chemist. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 511-520.
(13) Gordon, A. D. *Classification*; Chapman and Hall: London, 1981.
(14) Vickery, B. C. *Classification and indexing in science*, 3rd ed.; Butterworth: London, 1975.
(15) Littler, J. S. An approach to the linear representation of reaction mechanisms. *J. Org. Chem.* **1979**, *44*, 4657-4667.
(16) Guthrie, R. D. A suggestion for the revision of mechanistic designations. *J. Org. Chem.* **1975**, *40*, 402-407.
(17) Bonchev, D.; Kamenski, D.; Temkin, O. N. Graph theoretical classification and coding of chemical reactions with a linear mechanism. *J. Comput. Chem.* **1982**, *3*, 95-111.
(18) Johnson, A. P.; Cook, A. P. Automtic keyword generation for reaction searching. In *Modern Approaches to Chemical Reaction Searching*; Willett, P., Ed.; Gower: Aldershot, 1986.
(19) Schier, O.; et al. A system for the documentation of chemical reaction. *Angew. Chem. Int. Ed. Engl.* **1970**, *9*, 599-604.
(20) Bawden, D.; Wood, S. I. Design, implementation and evaluation of the CONTRAST reaction retrieval system. In *Modern approaches to chemical reaction searching*; Willett, P., Ed.; Gower: Aldershot, 1986.
(21) Zefirov, N. S. An approach to systematization and design of organic reactions. *Acc. Chem. Res.* **1987**, *20*, 237-243.
(22) Gelernter, H.; Rose, J. R.; Chen, C. Building and refining a knowledge base for synthetic organic chemistry via the methodology of inductive and deductive maching learning. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 492-504.
(23) Blurock, E. S. Computer-aided synthesis design at RISC-Linz: automatic extraction and use of reaction classes. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 505-510.
(24) Hendrickson, J. B.; Miller, T. M. Reaction indexing for reaction databases. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 403-408.
(25) Ugi, I.; et al. Chemical similarity, chemical distance, and computer-assisted formalized reasoning by analogy. In *Concepts and applications of molecular similarity*; Johnson, M. A., Maggiora, G. M., Eds.; Wiley: New York, 1990.
(26) Herges, R. Reaction planning: prediction of new organic reactions. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 377-383.
(27) Brandt, J.; et al. A classification of reactions by electron shift parameters. Application of the Dugundji-Ugi model. *Chem. Scripta* **1981**, *18*, 53-60.
(28) Wilcox, C. S.; Levinson, R. A. A self-organized knowledge base for recall, design, and discovery in organic chemistry. *ACS Symp. Ser.* **1986**, *No. 306*, 209-230.
(29) Willett, P. *Similarity and clustering in chemical information systems*; Research Studies Press: Taunton, 1987.
(30) Adamson, G. W.; Bawden, D. Comparison of hierarchical cluster analysis techniques for the automatic classification of chemical structures. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 204-209.
(31) Willett, P.; Winterman, V.; Bawden, D. Implementation of nonhierarchic cluster analysis methods in chemical information systems. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 109-118.
(32) Bawden, D. DISCLOSE: an integrated set of multivariate display procedures for chemical and pharmaceutical data. *Anal. Chim. Acta* **1984**, *158*, 363-368.

AUTONOM: System for Computer Translation of Structural Diagrams into IUPAC-Compatible Names. 2. Nomenclature of Chains and Rings†

L. GOEBELS, A. J. LAWSON,* and J. L. WISNIEWSKI
Beilstein Institute, Varrentrappstrasse 40-42, D-6000 Frankfurt/Main 90, West Germany

Received January 25, 1991

The IUPAC recommendations for naming organic chains, carbocyles, and heterocycles are discussed with respect to the implementation in AUTONOM Version 1.0. The general criteria used in the nomenclature approach are given in detail, along with limitations and extensions of the current rules.

INTRODUCTION

The general subject of this series of papers[1] is AUTONOM, a fully automatic and practical computerized system for the generation of IUPAC systematic nomenclature direct from the input of structure diagrams of organic compounds. Our first paper outlined the general programming approach without significant detail of the actual nomenclature principles involved. This paper is concerned with some fundamental aspects of nomenclature, in particular with the implementation of the IUPAC recommendations of the standard work on this subject, the "Blue Book".[2]

This paper takes the form of a page-by-page documentation of those particular aspects of the Blue Book (Sections A and B) which are implemented in the algorithm, and also those which are not. This systematic approach is necessary, since the IUPAC recommendations (generally regarded as "rules") were consciously formulated to allow considerable freedom in their application, and in many cases are not fully defined to their logical conslusion. Thus, any given structure does not necessarily relate to one uniquely correct name, which is simultaneously the greatest strength (flexibility in passive usage) and greatest weakness (imprecision in active usage) of this language. We have attempted to follow the recommendations as closely as possible. Nevertheless, we are aware that several "dialects" already exist in wide usage in the major reporting media and reference works, which differ slightly not only in style but also deviate from the IUPAC recommendations in some fundamental aspects. AUTONOM Version 1.0 does not cater to all these dialects, but we feel we cannot ignore them since they reflect current usage. Nomenclature experts will recognize in the following the elements of the dialect of IUPAC as used in the Beilstein Information System, which in our opinion follows "pure" IUPAC very closely indeed. We have taken pains in this paper to draw attention to those aspects which differ from other major IUPAC dialects, such as that practiced by Chemical Abstracts Services. It is our hope that ultimately these small variations may be resolved and that our work may contribute in achieving a consensus.

We have chosen to concentrate in the first phase on single component, uncharged organic species of less than 100 heavy atoms (atoms other than H), neglecting for the moment salts, mixtures, polymers, alloys, coordination compounds, and inorganic materials.

AUTONOM is not intended to produce "the" uniquely correct name for structures; on the other hand, the names

† For Mike Lynch, pioneer and gentleman.
* Author to whom correspondence should be addressed.

0095-2338/91/1631-0216$02.50/0