DEVELOPMENT OF CHEMICAL NOTATION SYSTEMS

PROJECT REPORT SUBMITTED IN
PARTIAL FULFILMENT OF TRAINING
IN DOCUMENTATION AND REPROGRAPHY
CONDUCTED BY INSDOC, NEW DELHI
DURING 1969 - 70

By

MRS. MRUNALINI SUDHAKAR NAIGAONKAR

National Chemical Laboratory,

Poona 8.

ACKNOWLEDGEMENT

Adequate acknowledgement to those who have helped me in different ways in the preparation of this project report would be impossible. However, at the outset, I express my respectful feelings of gratitude to Shri B.Guha, for his immense help and valuable guidance for this report.

Also my grateful thanks are due to Shri S.Parthasarathy, Scientist-in-charge Insdoc, who was very kind enough to grant me more time to submit this project, owing to my illness.

Also I gratefully express my thanks to Dr.J.Gupta, Deputy Director, and Shri M.G.Gokhale, Librarian, who have inspired and encouraged me for exploiting the opportunity of training.

Equally, I owe to Shri R.S.Singh, Documentation Officer for providing me all the facilities for completion of this project at an earlier date.

Last, but not the least to mention my thanks to the Director, N.C.L. Poona, who was very kind enough to depute me for this training.

M.S. Naigaonkar.
18.5.71

(Mrs). M.S.Naigaonkar.

CONTENTS

Pages

- - - - -

ABSTRACT

Within seven years after the birth of structural chemistry in 1861, virtually all of the main ideas relating to line-formula conventions were devised and published in the then leading chemical journals. No basically new practices appeared for another 79 years. Then, within an identically brief period of just seven years (1947-54), virtually all the fundamental features of structure delineating chemical notations appeared in the international chemical literature. The need for these chemical notations was felt when structural formula proved to be an exact means of communication rather than the nomenclature. As the difficulties encountered in arranging the structural formulas in index form were felt, the line formula notation systems have been designed by the workers in the field.

The international recognition to one of the systems, induced several others, to design the new systems compatible with computer use. The vast mass of data to be handled has made imperative to use the systems with computers, and accordingly new systems are being designed and old ones are modified.

CHAPTER - I

NEED FOR CHEMICAL NOTATION

"Dear Sir,

Your paper entitled "The preparation and properties of cyclo hexane-1-carboxylic-2-acetic acid' has not been recommended for publication, because the same acid has already been prepared by A.N.Other & described under the name hexahydrohomophthalic acid".

How many such letters were written before the days of formula indexes, how much unnecessary research work was done, how many exasperated authors exclaimed, "why cannot we have a systematic international nomenclature ?"

For nearly 20 years, chemists of many countries have been trying to devise such a nomenclature. We want a nomenclature based on principles so simple that a few hour's study would enable a chemist to write the name or the formula of any chemical compound of known constitution. We have, however to suffer for the Sins of our forefathers in chemistry; it would be impossible to expunge from the literature all their inconsistencies; in fact, one of the first things agreed upon by the International Union of Chemistry (now International Union of Pure and Applied Chemistry) was that interference with the existing nomenclature should be as little as possible" - says Dr. Clearence Smith,[71] to whom we owe a considerable debt of gratitude for his work on nomenclature.

The need for systematic nomenclature of organic compounds began to be felt about the middle of the last century. A valiant effort was made by The Chemical Society and thanks to the work of Watts, there was issued in 1879, a set of instructions on nomenclature and notation, for the guidance of abstractors and contributors. This was reissued in 1882 almost unchanged in form. Inspection of these instructions makes apparent the serious divergences between the British and Continental practice of the time, especially in relation to aromatic rings.

In the hope of reconciling such differences and of evolving a universally accepted system, the International Chemical Congress of 1889, appointed a commission to report upon the matter. In 1892, the commission issued a set of recommendations from Geneva, and thus founded what was afterwards called the "Geneva System". After much arduous labour over many years, the International Union of Chemistry produced a report from Lie'ge in 1930. This has given us the present system of nomenclature but, unfortunately, by the time of the Lie'ge report, such wide differences had been set up between British, American and Continental usage that any hope of reconciling them fully had vanished, and the commission had to be content with emendations and extensions of the original Geneva System. In other words, in spite of the ungrudging work of most eminent organic chemists over half a century, the system not only leaves many classes of compounds out of its considerations, but has failed in many cases to provide a usable name. As a result, the classification of organic compounds, the indexing of information have been hampered to great extent.

Chemical compounds are being indexed and classified according to[12,31]

1) Alphabetical arrangement of names,
2) Molecular formula,
3) Arbitrary type classification.

In the alphabetical arrangement of names, the systematic names are arranged alphabetically. It should be an ideal system of indexing based on nomenclature and should be logical and to some extent, to group together compounds having similar properties. When a unique name is assigned to each compound, the names can be arranged in alphabetical order and a compound canbe located by the same procedure used to locate a word in dictionary. However in many cases, several names may be applied to a compound with equal correctness.

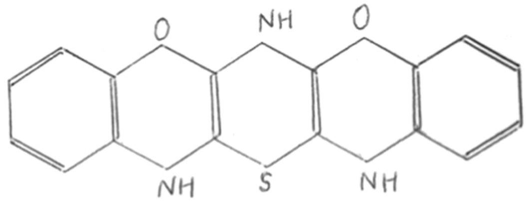
In Chemical Abstracts, for example, where this type of indexing is followed, a relatively small number of compounds are selected as stem compounds, and all others are treated as substituted derivatives of these parents. Each compound is indexed first under the present compound and then according to the modifying radicals arranged in alphabetical order. However, the system is sacrificed for compounds bearing trivial names, which are used wherever possible.

In a compound having several functional groups and rings, it is often difficult, if not impossible, to determine what constitutes the main function and therefore under which stem compound, the structure will be indexed.

However, the principal disadvantages of the Chemical Abstracts' system and, in fact, of any system based on nomenclature lie in 1) the complex nature of the rules governing the assignment of names, and 2) the complexity of the names themselves, which often makes it necessary for a reader to write down a structure before he can grasp what is implied by the name.

Few organic chemists[23] have reason to regard the present position of nomenclature with equanimity. The problems of naming more complex compounds, of enumeration and of reference are becoming so much more difficult that few can claim competence in this field and young chemists coming fresh to the subject are discouraged by the chaos which they find and are dismayed by the fact that there is no simple set of rules defining nomenclature.

The Geneva System of nomenclature appears to be incapable of satisfactory logical development in application to the more complex, fused and heterocyclic ring systems. For example, the structure



is reduced to a truly systematic name as

$$\text{Bis(benzo}\ \genfrac{}{}{0pt}{}{5'\text{-}6'}{5''\text{-}6''}\ [\text{cyclo-azoxa}\ \genfrac{}{}{0pt}{}{1'\text{-}4'}{1''\text{-}4''}\ \text{tetracarbadieno}$$

$$\genfrac{}{}{0pt}{}{2'\text{-}5'}{2''\text{-}5''}\Big]\Big)\ -\ \frac{2\text{-}3}{3'\text{-}2'}\ .\ \frac{5\text{-}6}{2''\text{-}3''}\ \text{cycloazathia} - 1 - 4 -$$

$$\text{tetracarbadiene} - 2 - 5.$$

While this example probably holds pride of place in respect of complexity, it is by no means unique; many examples from the Ring Index are equally unspeakable and are unable to serve as a simple spoken or written reference to the structures they represent.

Few examples from Ring Index referred to by numbers

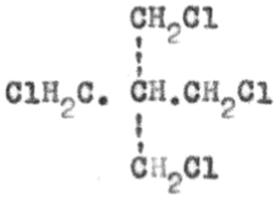
3179. Spiro[3-naphtho[1.2]triazole-2,2'-3'-ox-2' azabicyclo[3.2.2] nona 1'(8')5'(9')6' triene.

3681. 14-1,3-Dioxolo[5,6]indeno[3,1-d][1,3]dioxolo [6,7]isoquinolo]2,3-a] imidazole

3669. Dinaphtho[2,1-C,1',2'-e][2,7]benzo dioxecin.

3716. Dispiro[4-naphtho[1,2-b]pyran-4,4'[1,3]-dithiolane-5',4"-[4]-naphtho[1,2-b] pyran.

Even simple structure such as



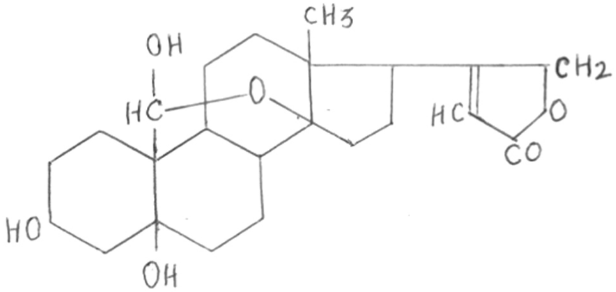
$$\begin{array}{rl} & CH_2Cl \\ & \vdots \\ ClH_2C. & CH.CH_2Cl \\ & \vdots \\ & CH_2Cl \end{array}$$

is described by Beilstein as $1,3,2^1,2^{1'}$-Tetrachloro-2,2 dimethyl propane.

It may be argued that such names can be avoided by the use of trivial names and this principle has been adopted freely by the compilers of the Ring Index. Many of the 'official' names given to fused rings in Ring Index are based on such trivial names as under -

Coronene, Ceramidine, Benzotriazepine,
Anthraparazene, Thebenidine, Phthalazine,
Flavophene, Ceroxene, Perylodipyridazine,
Phthaloperine, Quinolizine, Rubicene.

The use of trivial names is, however, a dangerous and unsatisfactory, as very names in the above sample list are familiar to the average organic chemist and the free introduction of trival names into nonenclature intended for indexing and classification can only lead to confusion. e.g. for the most important compounds derived from the steroid family, systematic names can be given to very few without the expedient of referring to an arbitrary ring structure by a trivial name. If one considers the structure

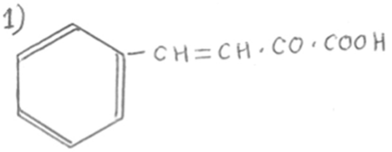


and asks the question "under what name would you look up this compound in an index ?" Unless it is known to the searcher that the trivial name of this compound is 'pseudostrophanthidine', one will be unable even to commence the search.

From this, it is realised, how inadequate systematic nomenclature is to cope with such instances. It is not only in the steroid field that examples of compounds which cannot be named according to any system are encountered, many alkaloids and substances of biological interest are equally unnameable.

The use of trivial names often leads to a multiplicity of names for the same substance. This applies even to simple substances. e.g.

1)



α-Oxo-r-phenyl-β-propylene-α-carboxylic acid.

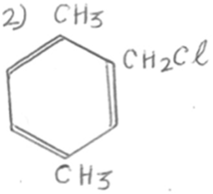
α-Oxo-r-phenyl vinyl acetic acid.

Cinnamoylformic acid

Benzal pyruvic acid.

Styrylglyoxylic acid.

2)



2'-chloro-1,2,4-trimethyl benzene

2'-chloro-ψ-cumol

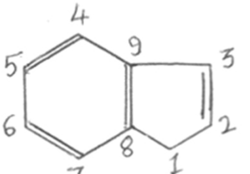
2,5 Dimethyl benzyl chloride

1,4-Dimethyl-2-chloromethyl benzene

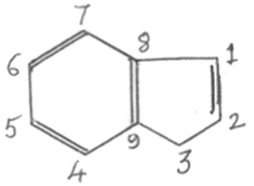
3) $CCl_3.SCl$ Perchloro methyl mercaptan
Thiocarbonyl tetrachloride
Trichloromethyl sulphur chloride
Tetrachloromethyl thiol
Trichloromethyl sulphenyl chloride

The existence of so many alternatives cannot but lead to indecision in selecting heads for searching indexes.

The question of enumeration, so closely bound up with that of nomenclature, especially with fused rings, is a source of much confusion. Different systems have been adopted in Britain, U.S.A. and in Germany. The most simple compound, indene is numberes as (1) in Beilstin and as (2) in the Ring Index.

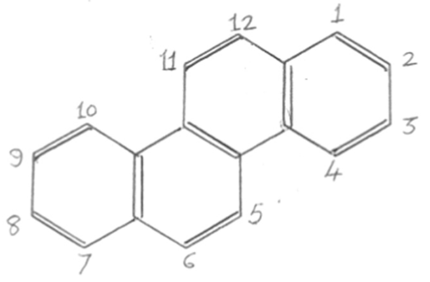


(1)

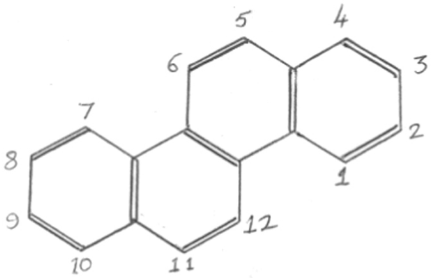


(2)

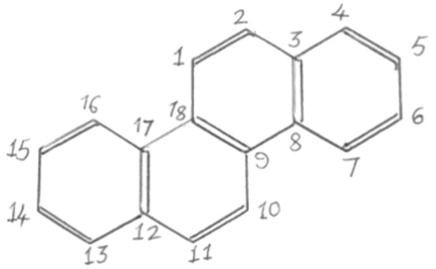
Like is the case with the structure of chrysene which is numbered in three different ways.



(3)



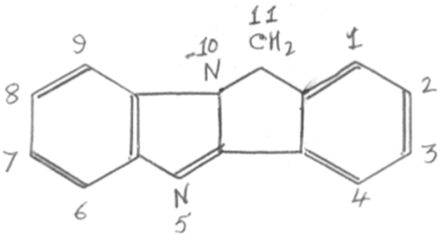
(4)



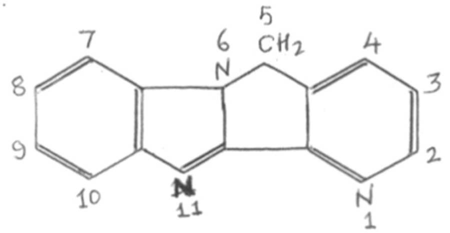
(5)

The Ring Index has put forth, the first of these i.e. (3), and use of the other two is made in Beilstein and in general chemical literature.

When more complicated instances are considered, it will be found that the method of enumeration set out in Ring Index has been a source of confusion. Thus in (6) below (No.2271 from Ring Index) is enumerated as shown, but in the analogous structure (7) (No.2270, Ring Index), the proposed different method of enumeration can only be a source of confusion.

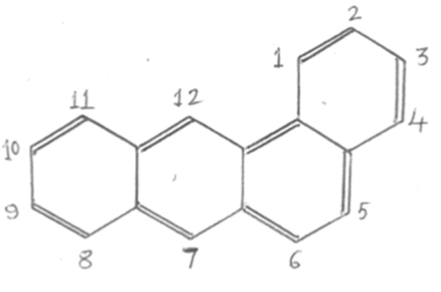


(6)

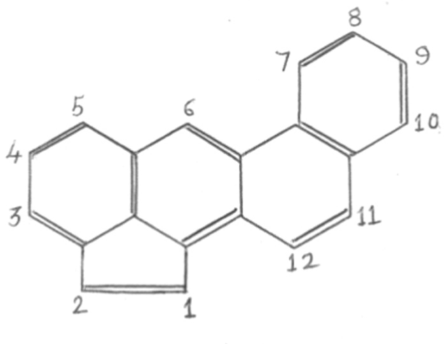


(7)

So also in example (8) (No.2805 from Ring Index) the enumeration has been set out, when a subsdiary ring in the 7-8 position of the structure is added, the whole pattern of enumeration is swept away and an entirely fresh one is created.



(8)



(9)

Moreover chemists in different countries have rules of their own for naming compounds inspite of International Nomenclature and this naturally adds to the confusion of the worker who is, for example, searching for all available information on a given compound.

The 2nd method used for indexing is based on molecular formula. The number of atoms of each element in a compound is calculated and the symbols for the elements, followed in each case by a subscript designating the number of atoms of that element, are set down in some order. Thus, in Chemical Abstracts, for example, the arrangement of the elements is in the alphabetical order of their symbols except for organic compounds, in which carbon and hydrogen are always placed at the beginning of the formula, and the rest of the elements are arranged in the order.

The formulas are arranged alphabetically and according to the number of atoms of each element present.

The accuracy in the calculation of a molecular formula makes for a simple and straight forward system which in one step breaks down all compounds into groups of isomers. Unfortunately, since the molecular formula alone is inadequate to distinguish isomers, the formula must be supplemented by a name or structural formula. Moreover molecular formula of many different compounds may be the same and hence are ambiguous.

Classification of chemical compounds by types is useful for certain purposes, where relatively small number of compounds are involved or where a group of compounds has a common denominator, it is feasible to set up type classes which are useful for indexing purposes. Any attempt to set up a type classification system for all chemical compounds is extremely difficult. For example, if all organic compounds are to be divided into broad classes - the halogen derivatives, the amines, the heterocyclic compounds, etc, immediately a question will be raised as to the disposition of a compound in which two or more of these groups are present. The solution involves setting up priority lists so that halogens come before amines, for example, and all compounds having both halogens and amines, thereby, are classified as halogen derivatives. The consequence is that each of the early classes will contain representatives of all the later groups. The later groups are lost for all practical purposes unless an elaborate cross indexing system is worked out. So this system is also not very satisfactory for indexing.

Classification and indexing of chemical compounds is not only the sole ultimate objective of the chemist. With the mass of scientific data available, it has been often desirable to investigate all compounds having common groups or those derived from one parent compound. Frequently these structural features must be studied in relation to physical, chemical or biological properties.

For this purpose, the name and the empirical formula are useless and some system which will make correlation studies possible, should be devised.

It is apparent that in any organization of compounds chemists think in terms of structure. Structure to the chemists means ions and molecules, rings and chains, functional groups and nonfunctional skeletal material. He indicates the structure of a compound by the highly conventionalized structural formula. There is little doubt that a well written structural formula can be read and understood more rapidly than any other means for conveying the same information. Therefore, the most appealing type of classification system to the chemist must be based on the structural formula.[23]

But the difficulties for using structural formula for indexing or classification are

1) It has no direct speech equivalent,
2) It occupies much space,
3) It cannot be arranged in index or list form.

It is clear, therefore, that, though the ideographs (structural formula) are indispensable to the study of organic chemistry, it must have some "linear equivalent" capable not only of being spoken but also of being arranged in index form.

The first line-formula descriptions[80] were set since when the key idea of structural chemistry was popularized in 1860. Chemists then had identified the molecular configurations in text copy simply by delineating the symbols for the corresponding molecular segments, one after another as connected.

The exact convention relating to "linear expression of formulae" was clearly defined by A.D.Mitchell as : "Each full point (or period mark) is regarded as separating two atoms which are directly linked in the main chain of a compound, and atoms and groups attached to each of these atoms are written immediately after it and before the next full point".

Josef Loschmidt led this era of structural chemistry with his publication of Chemische Studien, which contained 368 remarkably astute graphical diagrams, including scores of benzene ring postulations. His text included a few "rationelle Formel" such as C_2H_5,O,C_2H_5 for ether, C_2H_5,O,CH_3 for methyl ethyl ether, and C_2H_5, O, C_2H_4, O, C_2H_5 for the diethyl ether of ethylene glycol. Emil Erlenmeyer omitted the punctuation with familiar forms like C_3H_7OH, C_3H_7I and $C_3H_7OC_3H_7$.

Meanwhile H. Hübner had used the conventional period punctuation in the first line formula descriptions to appear in a chemical journal, with his discussion of the CH_2CN.CO.Br and CH_2Br.CO.CN isomers.

In 1866, August, Kekule, Henry Debus, H.L.Buff and B.F.Dupp completed this popularization of this line formula technique with scores of examples like $CH_3.CO.OH$, $CH_3.CO.CH_3$, $NC-CH_2-CO_2H$, $C_6H_5-CH_2-Br$ and $C_6H_5.SO_2.OH$. Finally when H. Wichelhaus introduced the word valenz in 1867, he suggested the punctuating refinement of periods to set off 'sidegroups' terminations and hyphens to show directly connected groups as in his CO-OH-COOH, and $CH_3-CHOH-COOH$.

Even with this simplification of structural formula into linear form, the problems increased as the complexity in the compounds increased with the number of complex side chains, functional groups etc. and the representation in the linear form has been very lengthy and much difficult for putting on a paper.

Lately chemists have felt the need of putting the compounds on a paper only by means of structural formulae. However the shortcomings in representing the three dimensional structural formula on paper and day by day increasing number of compounds which led to the mechanization of searching techniques compelled them to develop notational systems.

Within a very brief span of seven years (1947-54) virtually all of the major structure delineating systems of Chemical Notation have appeared in chemical literature.

CHAPTER - II

IUPAC'S ROLE IN CHEMICAL NOTATION SYSTEMS

The important role of speech in human communication makes it certain that chemists will always need names for chemical compounds. Nevertheless, dissatisfaction with previously developed nomenclature (as seen in Chapter I) as a means for expressing molecular structures has developed and the advances in chemistry provided the impetus for the development of notation systems, which express molecular pictographs - structural formulas - as a linear sequence of symbols, most of which are found on an ordinary typewriter keyboard. Although a number of chemists worked independently, their goal was the same - a completely systematic notation based on rules that could be readily learned and easily applied. Such a notation might be used as a substitute for conventional nomenclature in indexes, and particularly in machine searching.

Moreover, the number of organic compounds has tremendously grown. This has rendered acute the difficulties that chemists have long experienced in the classification of compounds and in effecting correlations based on features of molecular structures. Achieving these objectives, it was hoped, might be facilitated by an appropriate notation system. Practical considerations of machine searching suggest that a compound be designated in such a way that any desired aspect of function or structure can be selected at will.

Among the first[3] to propose a substitute for nomenclature and inspire this development of chemical notation was G.M.Dyson. He gave a lecture on "A New Notation for Organic Chemistry and Its Application to Library and Indexing Problems".[19] He further developed the notation in his manual published in 1947 on "A New Notation and Enumeration System for Organic Compounds".[20,23]

At the same time, several groups were approaching the problem from different directions. In connection with the search for antimalarials during World War II, over 15,000 substances were screened for activity. As an aid in keeping track of the tests with these substances F.Y.Wiselogle[12] devised a code for use with punched cards that has been adopted by several industrial firms with similar problems.

Simultaneously D.E.H. Frear of Penn State developed a code that has since been adopted with modification by the National Research Council's Chemical-Biological Coordination Center.[13] This code designates the basic components and functional groups. It does not consider structural isomerism, but the code has been definitive enough for practical machine use in correlating biological activity with chemical structures.

In 1947, F.L.Taylor developed a mathematically elegant method[75] for enumerating topologically similar positions in all ring systems - carbocyclic or heterocyclic, aromatic or saturated.

The G.K.D. ciphers of Gordon, Kendall and Davison appeared in 1947 and 1948 with their publications on "A New Systemization of Chemical Species",[36] and "Chemical Ciphering : A Universal Code as an Aid to Chemical Systematics".[37]

The "Gruber Notation" was first reported in 1949.[40]

J.A.Silk who was inspired by Dyson's lecture of 1946, wrote an article in 1949[80] reviewing methods for ring systems but did not publish it,[12] and later responded to the IUPAC Còmmission's invitation with his "New System of Organic Notation" distributed privately in 1951.[64]

J.G.Cockburn, member of the British Chemical Abstracts Staff, responded at the same time with his "Newcastle System" (1951).[14] Silk met him around that time but did not propose to develop it as in his opinion the system devised was more as a chemical shorthand than a systematic notation.[80]

In 1950, Wiswesser W.J. first described a system of notation that was based on the "Principle of Least Effort"[82] which later was reported in Chemical and Engineering News.[77]

Though all the above systems of chemical notation were developed, yet a completely systematic notation for representing structural pictographs should prove a valuable tool. International Union of Pure and Applied Chemistry (IUPAC) appointed a commission late in 1946 to study the possibilities of notational systems.[7] A milestone in the work of this commission was its meeting in Amsterdam in the summer of 1949 in which the following desiderata were set up for evaluating notation systems.

1. Simplicity of usage.
2. Ease of printing and typewriting.
3. Conciseness.
4. Recognizability.
5. Ability to generate a unique organic chemical nomenclature.
6. Compatibility with accepted practices of inorganic chemical notation.
7. Uniqueness.
8. Generation of an unambiguous and useful enumeration pattern.
9. Ease of manipulation by machine methods e.g. punched cards.
10. Exhibition of associations (descriptiveness).
11. Ability to deal with partial indeterminants.

At this meeting, the commission also decided to invite the inventors of notation systems to submit their systems for consideration.[2,47] They were asked to apply their systems to a list of 700 compounds selected from every 100th page of Beilstein and also to the supplementary lists which included compounds containing radioactive isotopes, organosilicon and organophosphorus compounds, and other presenting difficult notational problems.

By 1951, the following nine notation systems were received.

1. "A Method of Coding Chemicals for Correlation and Classification," developed by Chemical - Biological Coordination Center, National Research Council, Washington. 1950.
2. "Classification of Organic Compounds" developed by Van Weerden, W.J.
3. "A New System for the Classification of Compounds", developed by Wiselogle, F.Y. *et al*. 1946.
4. "A New Notation and Enumeration System for Organic Compounds" devised by Dyson, G.M. 1946.
5. "The Newcastle System of Representation of the Formulae of Organic Compounds" developed by Cockburn, J.G. 1951.
6. "Chemical Ciphering : A Universal Code as an Aid to Chemical Systematics." - devised by Gordon, M., Kendall, C.E. and Davison, W.H.T. 1948.
7. "Pictorially Direct Chemical Structure Symbols", developed by Wiswesser, W.J. 1950.

8. "A New System of Organic Notations", Silk, J.A. 1951.

9. "Gruber Notation", Gruber, 1950.

At MIT, members of the IUPAC's Commission together with members of the American Chemical Society's Committee on Scientific Aids to Literature Searching, the representatives of Beilstein and Gmelin met to review and evaluate the above systems.[60] It was observed that out of the nine systems, three (namely CBCC, Wiselogle and Van Weerden) were not designed to specify structural formulas in full detail, but were intended for correlating and searching procedures using mechanical aids or for classification. These three systems designate constituent groups, both functional and nonfunctional by appropriate code symbols built up from numbers and letters. Without questioning their effectiveness for special purposes (CBCC code for use with standard IBM machines; Wiselogle for correlation using punched cards; Van Weerden for decimal classification), the Commission in September 1952 decided that these three systems were unsuitable for an internationally useful notation as they failed to provide complete satisfaction of molecular structural formulas.

The two other systems (Cockburn's and G.K.D.) were dropped for different reasons. The Newcastle system was described as a highly developed chemical structural short-hand whose symbols and rules appeared unlikely to provide additional useful ideas. The system proposed by Gordon, Kendall and Davison builds up molecular structures by citing individual atoms or groups no larger than CH_3, CH_2 or CH. This according to Commission, results in notational representations whose deciphering is often laborious. The MIT meeting also emphasized the difficulties of designing notation systems that are equally effective for both indexing and searching based on structural features.

The Commission at this meeting decided that the notation system proposed by Dyson came nearest in meeting the eleven requisites and accepted it as a provisional International notation system. As for remaining systems, the commission recommended comparative testing to provide guidance for incorporating into the Dyson System, their various desirable features, especially those of Gruber and Wiswesser's systems.[60,55]

To implement this recommendation in the United States, the National Research Council appointed a committee to conduct practical tests of the systems.[3] This committee has had Walter R. Kirner as its Chairman till 1952 when he was succeeded by Eugene W. Scott, H.A.Geer was the secretary and Alexander M. Moore of Mellon Institute, H.S.Nutting of Dow and J.W.Perry were other members.

The tests were carried out by over 100 chemists who responded to a request for volunteers.[55] The volunteers were grouped according to geographic locations into 15 teams of six or more members. One consisting of a single member was in New Zealand but all of the others were in the United States.

Each team was sent three sets of 200 compounds for encoding and to compare the Dyson system with each of the other three systems. Generally each person coded 100 - 200 compounds by the Dyson system and also by one other system. A record was kept of the time required for coding,

The list of compounds was compiled by selecting every tenth compound from a total of approximately 30,000 in the files of Chemical-Biological Coordination Center. The entire list was sent to the inventor of each notation system under test for encoding the compounds by his system. Although Silk was unable to devote the time required to encode the list, the others did, and their notations were used as standards for checking the work of volunteers.

The number of disagreements in the encoding processes were over 50% for all of the codes so tested.

In the 2nd phase of the tests, the encoded lists were sent without change to other groups for decoding. Again, records were kept of time required for this. The structures obtained on decoding were checked with the originals by the C.B.C.C. Staff with the help of team members.

No. of Disagreements

Code	Compounds	Errors	Percent disagreement
Encoding			
Dyson	4371	2191	50
Gruber	2742	1780	65
Wiswesser	3465	2395	70
Decoding			
Dyson	6893	1287	19
Gruber	3445	612	18
Wiswesser	3967	1017	26
Silk	1623	425	26

Disagreements in the decoding operation did not exceed 26%. Thus at least half of the notations that were coded incorrectly could still generate the correct structure.

This indicated that "a cipher need not be unique to describe the structure adequately." Evidently a chemist's knowledge of possible structures can aid powerfully in overcoming encoding errors. The high incidence of errors that occurred in the test led the committee to believe that no cipher was ready for immediate adoption.

The committee also pointed out that the differences in the clarity of presentation in the instruction manuals increased the difficulty of comparing the inherent characteristics of the codes. "Whereas the directions for using the cipher are an integrated part of the code and directly determine the success with which it may be used, nevertheless, to compare a code which is not adequately described with one which was excellently presented, as was the Dyson, is extremely difficult."

To compare the time required for encoding, the commission had asked each team to note the time required to encode the 1st 100 compounds and then the 2nd 100 compounds by a given system. The differences were striking.

Time Required for Coding (Hours per 1000 compounds)

	Dyson	Gruber	Silk	Wiswesser
Encoding				
1st 100 compounds	176	180	260	160
2nd " "	65	63	125	64
Overall average	131	121	192	115
Decoding				
1st 100 Compounds	55	47	60	57
2nd 100 "	46	47	45	46
Overall average	52	47	52	52

In most cases, the times for the 2nd 100 compounds were well under half the times for the 1st 100. The times required for decoding were appreciably less than those required for encoding the 2nd 100 compounds.

In their comments on the ciphers, the volunteers expressed many differences of opinion, often diametrically opposed. Certain defects were common to all of the ciphers. For instance, the ciphering of salts, hydrates, addition compounds, stereoisomers was not at that time adequately explained in any of the manuals. The volunteers expressed that the use of a notation system may be important in the future, especially in using extensive files in connection with structure searching, indexing and classification.

Several of the authors have modified their codes over the report of the N.R.C. Survey and at least one new system (developed by Crane and Berry) has been proposed which is reported to include the best features of the previously existing codes.

On the basis of the volunteer survey and the conclusions of the NRC committee, E.M.Crane and M.M.Berry developed in 1954 what they call "A Composite Notation System for Molecular Structural Formulas".[3] In presenting their system at the American Chemical Society meeting in Cincinnati, the authors expressed the feeling that, rather than to choose among the four previous systems, it might prove advantageous to combine the best features of each into a single system. Thus inconsidering each question encountered, they weighed the solutions offered by each system for it, the suggestions of the volunteers and the familier devices of conventional indexes.

The Crane-Berry system is characterized by several types of redundancy and have attempted to combine the Dyson and Wiswesser Systems. It was the need for redundancy to which the NRC Committee called special attention in its report on the volunteer survey. Although this violates the principle of conciseness which was one of the eleven Amsterdam requisites, the committee felt that it would be desirable as a means of reducing errors to incorporate automatic self-checking features. It was on this foundation that they constructed a two-part system made up of a 'Classification Code' and a 'Basic Structural Cipher'. Any part or all of the 'code' could be eliminated without diminishing the specificity of the cipher, but the authors felt that the code will add considerably to the usefulness of the system for classification purposes.

To construct the classifying code for a given compound, Crane and Berry list in order, a ring index, the degree of saturation, number of double bonds, functional groups, and empirical formulas.

The ring index is composed of total number of rings, number of fused rings, number of hetero rings, members in smallest ring, members in largest ring, if any. These are summed up for acyclic compounds by a single zero. Following the ring index is the 1st letter, which may be S (saturated), U (unsaturated) or R (aromatic). The number of double bonds precedes the first comma. A letter between the commas refers to one of 26 classes of functional groups, as C (carboxylic acid), E (ester or lactone), K (hydroxide, alcohol or mercaptan), L (phenol or the like), M (amine), T (nitrous or nitroso), and U (oxide, ether or sulphide).and so on.

The structural cipher is set off from the code by a period. In the cipher, as in the molecular empirical formula, international atomic symbols are retained, but a few special symbols are introduced. Among them are E (methyl), L (methylene), J (methine), G (carbonyl), Q (hydroxyl), parenthesis (ring), and italic lower case letters (ring locants). The end to end approach is retained even for ring systems. Thus the cipher for isopropyl alcohol is not .EJQE. but .EJE.Q.

H.Winston Hayward in November 1961 announced a "New Sequential Enumeration and Line Formula Notation System for Organic Compounds".[41]

By this time, the IUPAC Commission on Codification, ciphering and punched card techniques had adopted the Dyson notation as a provisional international system of notation for organic compounds and a tentative draft[49] was submitted to the organizations adherent to the Union. During 1959, the large mass of material submitted by the organizations adherent to the union, and by others, was considered by the commission, and where desirable, changes were made. The final draft is published in 1961.[50]

IUPAC NOTATION.[50]

The symbols used for this notation comprise the letters of the alphabet, arabic numerals, and some of the common marks of punctuation. They are

1) The accepted symbols for the elements agreed by IUPAC with the following exceptions

   a) Bo is used for boron since, in this notation B is used in a structural sense.

   b) Ch is used for Chlorine to avoid confusion between '1' (unity) and 'l' (12th letter of the alphabet).

   c) Yt is used for Yttrium, since Y is used to denote a triple bond.

   d) Q is used for (i) shared oxygen, that is, an oxygen atom linked to two atoms as in a hydroxyl group or ether and (ii) in the compound symbol EQ for oxygen doubly bound in the carbonyl group $>C = O$.

2) Capital letters are used as follows -

A - for a saturated ring system e.g. A6 - cyclohexane.

B - for a ring system with maximum number of non-cumulative double bonds e.g. B6 - benzene.

D, L, R and S - for stereochemical purposes (printed in italics).

E - for a double bond.

M - for a macrocycle ciphered by assembly notation.

O - for unshared oxygen as in NO, NO_2, SO, SO_2, IO, IO_2 etc. but not for oxygen in $>CO$ and - CHO

X - as in CX for formic acid or after the cipher of a parent compound for a carbonyl group.

Y - for a triple bond.

Z - to indicate hetero atoms.

3) Lower case letters are used as follows

a - after EQ for aldehydes.

b - in a series of locants to indicate the presence of the maximum number of non-cumulative double bonds in a ring.

c and t - for cis and trans arrangements at a double bond as in E_c and E_t, and for configurational relations in cyclic stereoisomers.

h - to designate an implied variable position hydrogen atom as in B65h8.

r - for stereochemical reference.

x - for an unknown number.

4) Punctuation marks and other symbols are used as follows -

a) The semicolon ; - separates the citation of groups attached to the same carbon atom as in $C_4EQ;N$ for $CH_3CH_2CH_2CO.NH_2$ and certain locants in assembly notation.

b) The colon : - segregates certain locants and components in assembly notation as in $B6_2:4C/B6$.

c) Parenthesis ( ) - encloses compound substituents; repetitive components and branch components of assemblies.

d) The stroke / separates components in assemblies. An assembly must contain either a stroke or a colon or may contain both.

e) - Superscript $^{+}$ and $^{-}$ signs are used to denote positive and negative charges; double and triple charges are written as in M^{++} and M^{--}

f) Brackets [ ] indicate molecular or ionic entities and also the common difference in arithmetical series of locants.

g) Underlining i) of letters as in $\underline{NC}$ for the isonitrile group or $\underline{N_2}$ for the aliphatic diazo group, denotes the association (without hydrogen) of a group of atoms among which valency relations cannot be simply expressed, ii) of figures (but not subscript or superscripts) indicates association into numbers greater than 9, as in $\underline{16}$ (= 16).

h) The hyphen - is used to join non consecutive locants as in 3-5.

5) Arabic numerals are used

a) to indicate number (of ring atoms) i) as in B6 - a 'B' type ring of six members, and ii) as locants to describe the point of attachment of stem and substituent.

b) as subscripts, to multiply, thus, $B6_3$ implies an aggregate of three B6 rings, $C_{2/3}$ indicates three ethyl groups.

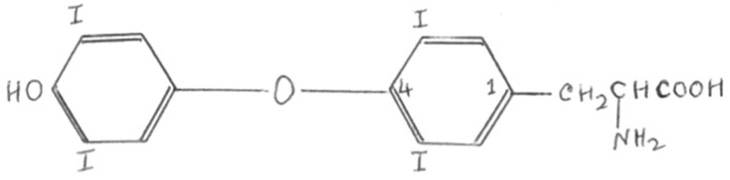
c) as Superscripts placed to the left of an element symbol to indicate the mass number of isotopes, or placed to the right, to indicate abnormal valencies in elements of free radicals. Thus ^{18}O (oxygen - 18) N^2 (bivalent nitrogen), C^3 (methyl radical) etc.

This monograph describes in detail rules with examples for the use of this system in case of

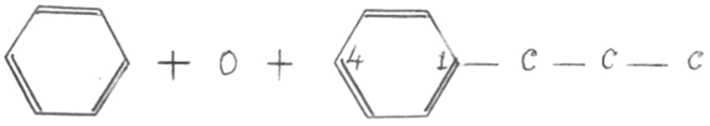
A cyclic hydrocarbons
Substituted Acyclic structures
Aliphatic Heterogeneous structures.
Carboxylic compounds with one ring.
Ring aggregates.
Heterocyclic systems.
Assembly notation of cyclic compounds.
Compounds with elements other than C, H, O, N, S and Halogens.
Stereoisomerism.
Abnormal valencies, salts, isotopes, and persubstitution.

It also shows the mode of ciphering and deciphering

e.g. For ciphering



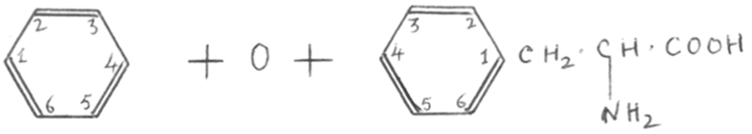
The carbon skeletons of the components are separated from each other and from the links, thus



Ciphering these will result in as

$B6 + Q + B6C_3$

The compound is a derivative of $B6C_3$:4Q/B6 with various substituents in the components. These are best added while the compound is in fragments. Thus



$B6Q1I26 + Q + B6(3C_3X1N_2)1I35$.

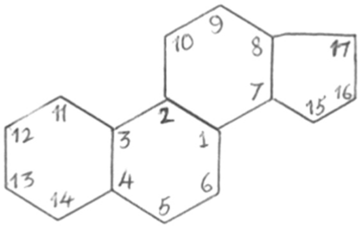
By joining these together, starting with the senior component the cipher obtained would be

B6(3C_3X1N2)1I35 : 4Q/4B6Q1I26.

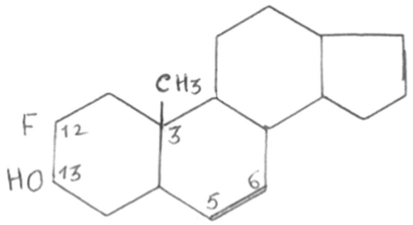
For deciphering, it is essential to start at the left and proceed to the right and to consider the significance of each group of symbols in turn. For example if the following notation is to be deciphered

$$A6_{3}5137C3E5Q\underline{13}F12 : 17/6C_{7}CX3C2$$

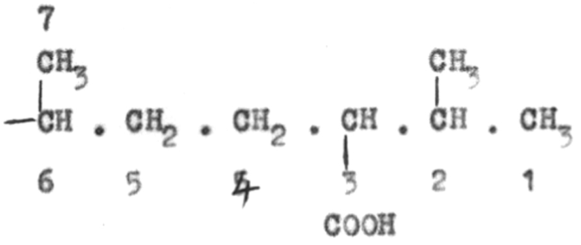
The presence of the colon and stroke denotes that the structure is an assembly. By reconstituting the portion A$6_3$5137, which comprises three 6-membered rings and one 5-membered ring joined at 1-2, 3-4 and 7-8 positions respectively, thus the structure got



The portion C3E5Q$\underline{13}$F$\underline{12}$ merely represents four modifications in the already deciphered structure which would give

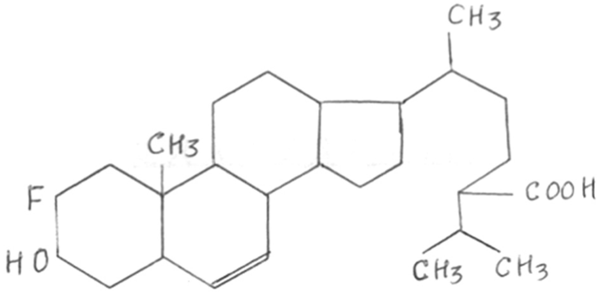


The 2nd component C_7CX3C2 comprises a chain of 7 carbon atoms, with a carboxyl group at '3' and a methyl at '2'. It is joined at its '6' position to the other component at its position '17' -



$$\overset{\overset{7}{CH_3}}{\underset{6}{-CH}} \cdot \underset{5}{CH_2} \cdot \underset{4}{CH_2} \cdot \underset{\underset{COOH}{3}}{CH} \cdot \overset{CH_3}{\underset{2}{CH}} \cdot \underset{1}{CH_3}$$

Putting these two components together, the structure obtained would be



The rules for the arrangement of ciphers in index form are also given. The logic of the IUPAC system is to describe the appropriate hydrocarbon structure and to modify it hierarchically when cited. This system results in the larger fragments congregating towards the left, and means that when a group of ciphers is arranged in order, the probability is that, substances of similar fundamental skeletons will be found together. This constitutes an advantage in indexing.

CHAPTER - III

VARIOUS SYSTEMS OF CHEMICAL NOTATION.

The various chemical coding systems or the non-conventional chemical notations which have been proposed fall into following four broad categories:-

i) Those which are unique (only one notation is possible for a given compound) but ambiguous (notation generating more than one compound). For example, Frear, CBCC and Wiselogle codes.

ii) Those ~~in~~ which the notations are unique and unambiguous (notation generating only the original compound) such as the Dyson, Silk, Gruber, Wiswesser, G.K.D. codes.

iii) Those in which notations are non-unique (when more than one notation is possible for a given compound) but are unambiguous such as the systems of Opler and Norton, the U.S.Patent Office Code etc.

iv) Those in which the notations are non-unique but are ambiguous such as any fragmentation or classification code, to the extent that their rules are incomplete.

Codes in class (i) have demonstrated their usefulness in correlation studies involving structural features, especially when used with punched cards, and in a limited way in indexing and classifying. For example, the CBCC code has been applied to approximately 70,000 compounds to relate their structure to biological activity; the Frear Code has been used to relate the structure to insecticidal efficiency in 10,000 compounds.

The completely specific codes in class (ii) promise to be useful for structure searching, indexing and classifying. For example, the Wiswesser notation is being used by A. Addelston of the Special Chemicals Department of Winthrop Laboratories, New York, with 10,000 compounds and in Althouse Chemical Company, at Reading, Pennsylvania, for 1,300. The IUPAC-Dyson notation is being tried by Chemical Abstracts Service.

The codes in class (iii) require fewer rules but more complicated machinery. These can be useful where large numbers of compounds are involved.

Codes in class (iv) are not much evolved and are tried only by individual institutes having their particular problems.

A very brief outline of a few more useful notations is followed with examples.

1) The Frear Code (1946)[30,32]

This code for assigning numerical designations to chemical structures was devised by Frear, D.E.H., Seiferle E.J. and King H.L. to enable them to a) order the compounds in such a way that an individual compound could be located in a file containing several thousand other compounds with a minimum of effort and without any possibility of error; b) arrange the compounds, so far as possible in such a way that related compounds would be grouped together in the file, and c) make possible the collection and statistical study of data on the frequency of occurrence of all chemical groupings appearing in the compounds under study.

After a number of trials by the authors with existing classification and indexing systems, including molecular formulas, alphabetical arrangements and several others, it became apparent that each of these failed to fulfill the requirements mentioned above. Accordingly this classification system was devised and was first applied in 1943. Since then it has been used successfully on approximately 8000 different chemical compounds, both organic and inorganic ones.

This classification is based on "code numbers" assigned to each chemical compound. The code number for a particular compound consists of one or more group numbers, depending upon its type and complexity. These group numbers are assigned by referring to a prearranged list in which constituent chemical groupings (not necessarily functional groups) are given numerical designations. The groups are listed in decreasing order of complexity.

To assign a code number to a particular compound, the list of constituent groups is read downward, starting with the most complex; as a group present in the compound is encountered in the list, the number is noted, and the process continued until all constituent groups have been accounted for. For example, in ethyl alcohol, the first group encountered is hydroxyl, which bears the group number 581. The remainder of the compound, the two carbon chain, is farther down the list and bears number 1011. Therefore

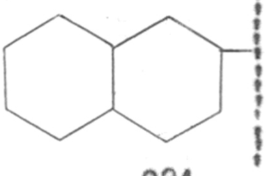
CH_3CH_2OH = 581 - 1011

Similarly, the following compounds are assigned code numbers, as indicated:

(dotted lines separate the constituent groups)

1) $CH_3CH_2CH_2 \;\vdots\; NH_2$ = 671-1003

1003 671

2) 

$\vdots\; SO_2NH_2$ = 56-924

924 56

In the second example, the sulphonamide group - SO_2NH_2, is numbered 56, while - SO_2 - and - NH_2 individually bear higher numbers 264 and 671 respectively, so that the more complex gr is coded as a unit and not split into smaller parts.

The common inorganic anions and cations are coded in the same way so that a simple inorganic salt will have two group numbers.

This very simple code has been reported to have several advantages. The procedure may be learned easily by anyone with a minimum of chemical knowledge. The code numbers may be used on either hand-sorted or machine-sorted punched cards.

However, with a short list of constituent groups, it is obviously not possible to make provision for all possible combinations.

2) The Chemical Biological Coordination Center - Code or National Research Council Code 1950[5,31,43,54,81]

During the war it was necessary to carry out large-scale test programs to find efficient insecticides, insect repellents, rodenticides, fungicides, antimalarial compounds, chemical warfare toxicants, and other applications of chemical compounds. The only possible method was a general screening program to test available compounds for each type of action. In order to carry out correlation studies of chemical structures with biological properties, a system of classifying and inspecting chemical compounds was required. Existing systems of indexing compounds - the alphabetical nomenclature system or the empirical formula classification were not suitable for the purpose. It was found that though the

Beilstein system enabled to index a compound in a unique position - by structure -, yet closely related compounds having slight variation in substituent groups often occurred in widely separated locations.

Because of the large number of chemical compounds involved and the nature of the information desired, the fundamental requirements set up were:

1) the code should be specifically designed for machine sorted punched card use, and

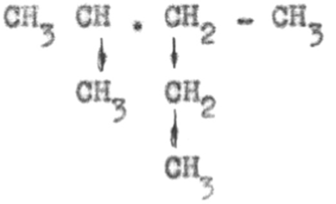
2) it should indicate constituent groups in the compounds. The Chemical Codification Subcommittee of the N.R.C. decided to adapt the code devised by Frear and associates (described earlier).

The basic principles of the Frear code are retained in this code. The code number for a given compound is made up of a series of four-digit group numbers, each representing a specific structure in the molecule. The first character in the group number designates the family to which the structure belongs, i.e. whether it contains, for example, 1) carbon, hydrogen, nitrogen, oxygen, and sulphur or 2) carbon, hydrogen, nitrogen and oxygen, or 3) carbon hydrogen and nitrogen in organic group. The 2nd and 3rd digits in the group number identify the particular chemical structure and the 4th digit denotes the number of times, the group occurs in the compound or ion.

The families are classified in four main divisions: organic groups, organoheteroid groups, inorganic groups and indeterminate structures. In the organic division, the families are arranged according to the number of different elements present in the groups within the family. Only carbon, hydrogen, nitrogen, oxygen, sulphur and halogen are classified in this group. All other elements attached directly to carbon are considered organoheteroid elements.

The degree of branching in noncyclic structures is not considered, but only the total number of carbon atoms linked directly together and the carbon to carbon unsaturation e.g.

1) $CH_3\ CH_2\ CH_2\ CH_2\ CH_2CH_2\ CH_3$ = 06 I.1

2) 
$CH_3\,\underset{\underset{CH_3}{\downarrow}}{CH}\;.\;\underset{\underset{\underset{\underset{CH_3}{\downarrow}}{CH_2}}{\downarrow}}{CH_2} - CH_3$ = 06I.1

3) $CH_3\ CH_2\ CH = CHCH_2\ CH_2\ CH_3$ = 06H.1

In breaking down compounds into separate groups for coding, the groups are ordinarily separated at the point of attachment to a carbon atom.

C_2H_5 ¦ OH = H8M.1 - 089.1
089.1 ¦ H8M.1

In the organoheteroid division, the elements are classified according to their combining power, defined as the number of electrons which the atom in question furnishes for sharing or transfer(s) to other atoms e.g.

($C_6H_5)_3$	AS
NYR.3	PIL.1

In ionic inorganic compounds simple combinations of two elements are coded by two group numbers, one indicating the element and its oxidation state present in the cation, another for the anion. In more complex structures one or more atoms in each portion is considered central and the other atom or atoms coordinated to it. For example, sulphur in SO_4^{-2} is TNJ.1. The oxygen atoms in SO_4^{-2} are coded as U63.4. Thus, the combination of two group numbers TNJ.1 - U63.4, characterizing the SO_4^{-2} ion.

The system does not distinguish between position and stereoisomers. It is also true that the code designations are not unique and occasionally two compounds having dissimilar structures will have the same code numbers. Over 60,000 compounds have been coded by this system at the center.

3) <u>The Wiselogle Code</u>. 1946 [12]

It has been already described in Chapter II. p. 22.

4) The Dyson Notation (1946)[18-24]

This is first notation system developed for a substitute for nomenclature, accepted with modifications as International System of Chemical notation.

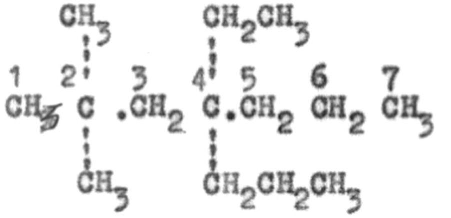
The system uses the upper and lower case letters of the alphabet, arabic numerals 0 - 9 in ordinary and subscript forms, θ, π and simple punctuation marks including parentheses and square brackets.

The notation is considered under the following heads

Saturated acyclic hydrocarbons.
Unsaturated acyclic hydrocarbons.
Cyclic compounds.
Fused rings.
Ring aggregates.
Heterocyclic compounds.
Any of the above with addition of functional groups.

Saturated Acyclic Hydrocarbons

A straight chain hydrocarbon is signified by C followed by a subscript numeral indicating the number of carbon atoms involved, as in C_2 for ethane, C_4, butane, and C_{10} decane. The longest chain is cited first, and the side chains in turn, commencing with the largest etc.



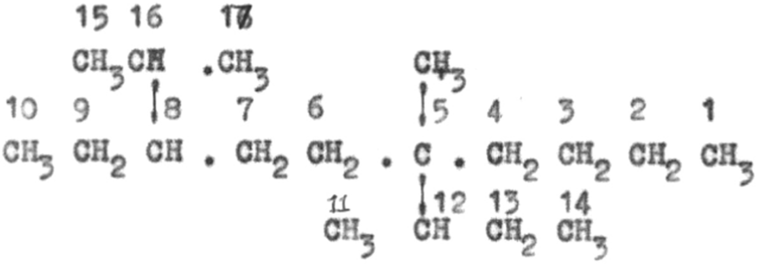
$\overset{1}{CH_3}\overset{2}{\underset{CH_3}{\overset{CH_3}{C}}}.\overset{3}{CH_2}\overset{4}{\underset{CH_2CH_2CH_3}{\overset{CH_2CH_3}{C}}}.\overset{5}{CH_2}\overset{6}{CH_2}\overset{7}{CH_3} = C_7.C_3,4.C_2,4.C_2 2,2$

This shows propyl, ethyl and methyl side chains, The various portions of the cipher, 'operations' are separated by full stops, as in

C_7, C_3,4 . C_2 . 4 . C,2,2

The numerical subscript indicates the number of atoms in straight chain association, the normal figures, after the commas at the end of each operation, are locants and serve to locate the position of the side chains e.g. '4' and '2, 2'.

The principle of longest chain is carried to side chain as in



$\overset{10}{CH_3}\overset{9}{CH_2}\overset{\overset{15\ 16\ \ 17}{CH_3CH.CH_3}}{\overset{8}{CH}}.\overset{7}{CH_2}\overset{6}{CH_2}.\underset{\overset{11\ \ 12\ \ 13\ \ 14}{CH_3\ CH\ CH_2\ CH_3}}{\overset{\overset{CH_3}{5}}{C}}.\overset{4}{CH_2}\overset{3}{CH_2}\overset{2}{CH_2}\overset{1}{CH_3}$

C10.C_4, 5-12. C_3, 8-16. C,5

The hyphenated locants '5-12' and '8-16' indicate the points of attachment. Every structure cited by the notation is assumed to be fully saturated unless otherwise specified and hence no account is taken of hydrogen atoms in citations.

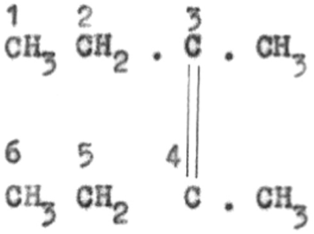
Unsaturated Acyclic Hydrocarbons

The presence of unsaturation is signified by 'E' for a double bond or 'E3' for a triple bond. The '3' following the 'E' is termed as a modulant. Thus

Ethylene	C_2.E
Butane - 2	C_4.E,2
Butadiene	C_4.E,1,3
Acetylene	$C_2.E_3$
Butyn - 2	C_4.E3,2

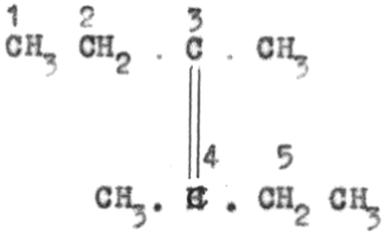
The modulated forms E1 and E2 are used to delineate the cis and arrangements at the double bond.

cis



C_6.C,3,4. E 1,3

trans



C_6. C,3,4.E2,3.

Cyclic Compounds

Rings are ciphered as A - saturated and B - aromatic.

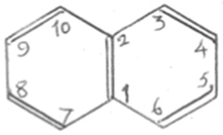
A5	cyclopentane
A6	cyclohexane
B6	benzene
B8	cyclooctatetrene.

Any adjustment of hydrogen is made, either by adding an operation in H to the B ring, or by adding an E operation to an A ring. Where the same compounds would result from either an A or B ring, the choice of method is governed by the number of locants produced, the method giving the fewer being chosen. Thus A6,E is selected for cyclohexene and not B6.H, 1,2,3,4 where either method gives the same number of locants, B is selected. Thus

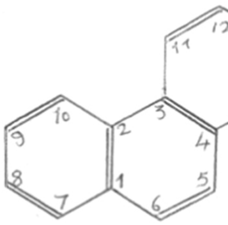
A6	cyclohexane
A6,E	cyclohexane
A8	cyclooctane
A8.E	cyclooctane
A8.E,1,3	cyclooctadiene
B8.H,1,2	cyclooctatriene
B8	cyclooctatetrene.

Fused rings

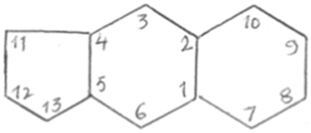
These are delineated from A or B according as they are saturated or aromatic, the number and size of the individual members of the network are stated, followed by a series of locants which indicate the mode of attachment. Thus:



$B6_2$



$B6_3,1,3$



$A6_2 5,1,4$

The rings are cited in decreasing order of size.

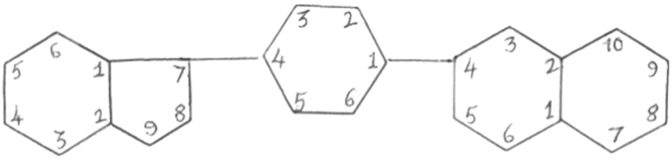
Ring aggregates

Aggregates of similar rings are ciphered with θ as:

B6.θ_2	diphenyl
B6.θ_3	terphenyl
B6.θ_4	tetraphenyl
B6.θ_5	pentaphenyl

The locants after θ become necessary as there are several ways in which more than two benzene rings can be coupled.

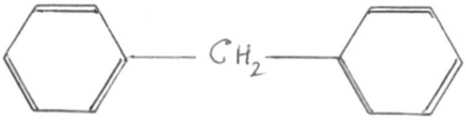
Aggregates of dissimilar rings are ciphered by treating each ring as a separate component in square brackets, e.g.



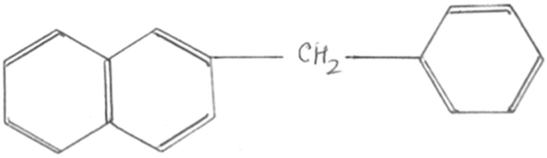
A6.[A6$_2$.4] . [A65.7], 4

The senior of two components is always cited first.

When similar groups are joined through a carbon atom, it is included after θ. However when dissimilar rings are joined through a carbon, the latter is held to be associated with the senior ring e.g.



B6.θ_2 C

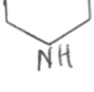


B6, . C,4 . [B6], 11

Heterocyclic compounds

The international symbols are used for elements other than carbon and hydrogen except that 'Q' is used for oxygen to avoid confusion with zero. The heteroatoms are shown by 'Z' preceding the symbol for the atom. Thus:

A5.ZQ



A5.ZN

B6.ZN

A6.ZQ,1,4

Functional groups

The use of 'Q' to denote oxygen leads to Q for hydroxyl, EQ for carbonyl, whilst X is used for carbonyl as:

CH_3CH_2OH = C2.Q

CH_3CH_2CHO = C3.EQ

CH_3COOH = C2.X

CH_3COCH_3 = C3.2EQ

$COOH.CH_2CH_2COOH$ = C4.1,4X.

The notation of all alcohols, glycols, aldehydes, ketones, acids would be according to the above rule. Ethers and esters are ciphered from the side containing the larger hydrocarbon residue thus -

$CH_3CH_2OCH_2CH_3$ = C2.Q[C2]

$CH_3CH_2COOC_2H_5$ = C3.X[C2]

$CH_3COOC_3H_7$ = C3.[XC2]

When an ester is ciphered from the acid side, X is cited outside the square bracket, when from the alcohol side, X is included within the bracket, thus distinguishing the two modes of treatment.

The following modulated forms of N are used for different nitrogenous compounds.

N - Amino - NH_2

N5 - Pentavalent nitrogen

N1 - Nitroso - NO

N2 - Nitro - NO_2

N3 - Azido - N_3

N4 - Azo - N - N -

N6 - Alidiazo nitrogen.

These symbols are cited in the order set down as above, but after operations in X and Q. Where both oxygen and nitrogen are present in the same symbol as ENQ for the oxime grouping = N.OH; EQ : N for the acid amide, etc. such symbols are cited after X, and Q but before N. The sulphur analogues of Q and EQ are S and ES; the various thio acids are ciphered in terms of the composite symbols Q : ES, EQ : S and S : ES, S.2 and S3 are the - SO_2 - and SO_3H groups respectively, as in sulphones and sulphonic acids. The S citations follow those of Q and precede those of N. Halogens are denoted by their usual symbols: I, Br, Cl and F and are cited in this order after the nitrogenous substitents e.g.

CH_3NH_2	Methylamine	C.N.
$C_6H_4(OH)NO$	p - nitrosophenol	B6.Q.N1,4
$[C_6H_5N(CH_3)_3]Br$	Trimethyl phenylammonium bromide	$B6.N5[C]_3[Br]$
$F_3C.COOH$	Trifluoroacetic acid	$C_2.X.[F]_3,2$
$H_2NC_6H_4SO_2NH_2$	Sulphanilamide	B6.52N.N,4

The following order of citation is always used

1. Rings
2. Heteroatoms

3. Chains
4. Adjustments of H : unsaturation.
5. Functional groups.

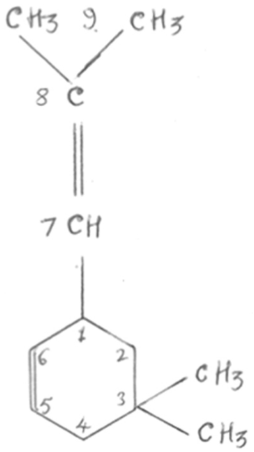
The monograph gives detailed exposition of rules for the delineation of all possible structures.

Dyson has claimed that both the ciphers and the related enumeration patterns are unique and unambiguous; no structure can have more than one cipher; no cipher can refer to more than one compound. This means that the ciphers can be used as a means of reference to structures entirely unambiguously and independently of existing nomenclatures. The ciphers can be indexed readily and have a powerful classifying effect when arranged in list form, which will be evident from appendix.

He also claims[21] that this notation fulfills the desiderata set out by IUPAC's commission.

1) It is easy to learn and simple to use, so that both coding and decoding become automatic for the simpler instances. The user soon begins to think in terms of C_2.Y for acetic acid; B6.N_2 for nitobenzene, B6.ZN for pyridine etc. This ease of use extends to indexing since the ciphers are amenable to alphabetical arrangement.

2) The notation is concise, always shorter than the systematic name and often shorter than the trivial name. Thus



1,1-Dimethyl-3-(3^2-metho
propen (3)yl) cyclohexen-4
(Beilstein)

A6.C_3.C,3,3,8.E,5,7

3) The principle of using separate operations for each structural features and of citing these in a specific order has the effect of producing recognizable sequences such as A6.C_3, 1-8.C,4 in the terpene family for the familiar hexahydro-p-cymene structure.

4) Use of the accepted symbols for heteroids and of the subscript numerals for enumeration of quantity makes the system quite compatible with inorganic conventions. Thus a compound in which four molecules of triethylarsine combines with two of cuprous iodide cannot be logically portrayed as a valence structure, but readily be catalogued as

$$\left\{ As.[C_2]_3 \right\}_4 \; \left\{ Cu_2I_2 \right\}_2$$

The composition is immediately apparent and the structure would fall automatically into the arsenic section of the index.

5-6) The ciphers are unique and generate unique enumeration patterns.

7) The notation has been designed with the possibility of applications of punched card techniques in mind. It is to be noted that the recognizability associated with ciphers built in the notation is carried over into punched cards, and the machines can be adopted to search for such associations as may be desired.

8) The various associations which may interest the chemist are set in the notation with each operation. All the associations could be searched for optically or mechanically. It is possible to search for certain combinations of associations, which is not normally feasible.

9) It is essential that a notation shall have ~~a~~ proper means of indexing and describing substances in which the structure is not fully revealed e.g. a nitro-ethyl-anthracene in which the position of either or both substituents is unknown. These possibilities are dealt with thus:

$B6_3,1,4.C_2,7.N2, ?$

$B6_3,1,4.C_2, ?.N_2, 3$

$B6_3,1,4.C_2, ?.N_2, ?$

10) The system is independent of language since it is presumed that all scientific workers in the chemical field may be expected to be familiar with the English and Greek alphabets and Arabic numerals.

Moreover the author says that the cipher can be used to generate a universal and unambiguous nomenclature. In fact, the cipher can be regarded as the abbreviated written form of an exactly parallel expanded form which can be printed or spoken in exactly the same way as existing nomenclature, with the enormously enhanced value which the cipher confers upon it. As a simplest example, the connotation of C_1 C_2, C_3 .... C_{10} etc. the methane, ethane, propane ..... decane etc. are preserved intact.

5) Silk notation 1951.[64,66]

"A Linear Notation for Organic Compounds" uses first four letters of the alphabet to cipher the fundamental chains and ring systems of organic molecules and these are referred to as basic symbols.

A - An aliphatic carbon atom, together with the appropriate number of hydrogens.

B - Simple (unfused) benzene ring.

C - General symbol for a cyclic structure. The number following it gives the number of rings in the structure. Ring systems are understood to contain the maximum number of noncumulative double bonds and rings to be six-membered in the absence of other information.

D - Individual heterocyclic rings in a ring system. The number following it shows the size of rings and it is followed by the symbols for each of the hetero atoms.

A, C, D and also H, are the main letter symbols for which numbers immediately following serve as multipliers. Multiplication of most other single letter symbols is shown by repeating the symbol. These letters are chosen for the advantage for indexing and classification for these basic structures.

A double bond is denoted by E and a triple by EE. The symbol H is used in those situations where hydrogen atoms require to be specified, namely, on ring systems, and with elements other than oxygen, nitrogen and sulphur.

The rules for ciphering branched carbon chains such as

1) the locants for substituents precede the symbols to which they refer, and multipliers follow the symbols,
2) In a series of locants no punctuation is used between numbers upto nine, but every two-digit number, including a single one, is preceded by a comma.
3) Enumeration is determined by considering the attached substituents, one at a time, in order of seniority, and assigning each the lowest available locant until the enumeration has been uniquely determined.

4) Stops are used to separate parts of the cipher and break it up into more readily recognizable portions.

The functional groups are denoted as follows:

Q - for - O - i.e. divalent oxygen linking two other atoms, one of which may be hydrogen. This hydrogen is implicit, so that AQ denotes methanol.

O - denotes = O, or → O, i.e. divalent oxygen joined to only one other atom, so that formaldehyde is ciphered AO or it denotes oxygen in a higher valency state.

T - denotes - S - i.e. divalent sulphur joined to two other atoms, methane thiol is diphered as AT.

S - denotes = S, or → S i.e. divalent sulphur joined to only one other atom, or suáphur in higher valency state.

M - denotes amino-type nitrogen and is used to cipher nitrogen in all groups related to NH_3, NH_2OH, NH_2SH and NH_2NH_2 (but never - N = N -)

N - represents nitrogen in all other circumstances.

F - Fluorine atom

J - Iodine atom

K - Bromine atom

L - Chlorine atom.

P - Phosphorus atom.

Multiplication of single atoms in functions is shown by repeating the symbol, thus peroxide is written as QQ, a disulphide as TT etc. The nitro group is written NOO and not NO_2, because O is terminal. Similarly, a sulphone is SOO, and a sulphonamide is SOOM. The carbonyl group is denoted by AOQ.

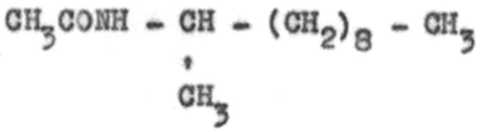
The enumeration in ring systems starts in the most central ring of a system, regardless of its size and proceeds throughout in one direction, clockwise or anticlockwise. It proceeds in the manner which gives the lowest total of fusion point locants.

The rules for analysis of structures and the seniority principles are given exhaustively.

6) Gruber System 1949 [31,40]

In this system aliphatic compounds are coded by numbers indicating the number of carbon atoms in the chain, plus designations for the attached groups : 1 - dodecanol for example, is C12.OH. The symbol Z is used to denote an interrupted carbon chain or hetero structure.

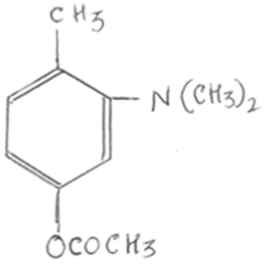
To code the compound



$$CH_3CONH - \underset{CH_3}{CH} - (CH_2)_8 - CH_3$$

Gruber uses the symbols Z13[C2(=O)NC10,4C]. The figure 13 refers to the total number of carbon atoms, the oxygen in the acetyl group is enclosed in parentheses to indicate its position as a substituent group. The methyl group is located by the designation 4C.

Gruber indicates a benzene ring by 6], so that the compound



is coded as 6]C.2Z2[N(C)C]4Z3[O(O=)C2], the dimethylamino structure being represented as Z2[N(C)C]. The acetate group is coded as Z3[O(O=)C2]. The groups of symbols are preceded by 2 and 4 respectively, indicating attachment to the benzene ring in these positions.

In the trial coding tests conducted by the N.R.C.[16] it was noted that the order of citation is frequently not clear including the functional groups and rings, and the enumeration of rings other than the parent ring is often difficult.

More recently Gruber has contributed a number of suggestions to the Wiswesser system and these have been incorporated into the latter. As far as is known, this system has not been used to any ~~d~~extent.

7) Wiswesser notation 1950.[77,78,73]

First proposed in 1950, to describe complex ring structures in a simple yet logical manner, and to calculate hydrocarbon isomers in terms of familiar mathematical series. In 1945, F.D.Rossini and his associates published a report showing that hydrocarbon properties could be calculated precisely only in terms of the primary, secondary, ternary and quaternary carbon atom components. These and other properly correlations with structure convinced Wiswesser that the basically significant hydrocarbon segments are not the 'carbon skeleton' abstractions of the Geneva nomenclature but the pictorially obvious alkyl chains and branched carbon atoms of the much older 'Line formula' notation. Thus he solved the isomer calculating problem in 1955 through this natural 'branch group' analysis of structures.

This system of notation has been systematized by 1) extending the longest chain descriptive technique to cyclic systems, of all degrees of complexity and 2) by supplementing the customary atomic symbols with a few new symbols in order to retain the versatility and graphical directness of the method. Functional groups are emphasized with distinctive single letter symbols to facilitate visual and mechanical scanning. The following examples show how Q is used to denote the OH - group, Z the NH_2 - group, V the keto - CO - group, N the - NH group and R the benzene ring.

Arabic numerals are reserved to denote the n-alkyl groups and unsubstituted alkylidene segments as shown.

Typical Organic Structure Notations

Name	Line Formula	Notation
Acetanilide	CH_3-CO.NH,-C_6H_5	1VMR
Ethyl vinyl ether	C_2H_5 - O.- CH: CH_2	201:1
Ethyl acrylate	C_2H_5 - O - CO.CH:CH_2	20V1:1
Fluorobenzene	F - C_6H_5	FR
Acrylonitrile	NC - CH : CH_2	NC1:1
Allyl cynnate	NCO - CH_2CH:CH_2	NCO_2:1
Cinnamaldehyde	O:CH CH:CH - C_6H_5	O:2:1R
Diethylene glycol	OH-C_2H_4-O-C_2H_4-OH	Q2O2Q
Diazoacetic acid	OH-CO.CH:NN	QV1:NN
Diazoaminobenzene	C_6H_5-N:N - NH-C_2H_5	RN:NMR
Phenyl isothiocyanate	SCN - C_6H_5	SCNR
Biurea	NH_2-CO.NH-NH-CO.NH_2	ZVMMVZ

Hydrogen atoms joined to atoms other than C, N and O in open chain segments are indicated with the appropriate number of H letters (avoiding subscripts) but cited immediately after the adjoining symbol. Thus hydrogen sulphide is denoted as SHH and arsine AsHHH.

Only eleven new letter symbols are used to denote commonly encountered atoms or functional groups.

G - denotes chlorine
E - denotes bromine.
W - denotes O_2 group joined to a third atom as in NO_2 or SO_2 group.
Y - denotes Y forked C(H) group.
X - denotes quarternary C atoms
K - denotes N^+ atoms.
Q - hydroxyl group.
Z - NH_2 group.
M - NH group
V - CO group
R - benzene.

The characteristic feature of this notation is that the atomic symbols are cited "end to end" in a pictorially direct connecting order. This practice follows the oldest structure delineating tradition and eliminates much unnecessary enumerating labour.

Another characteristic feature is that all otherwise equal seqances of symbols, are resolved through their self-evident alphabetic order. Preference is always given to the highest ranking sequence of open chain symbols, in order to preserve the traditional emphasis on terminal functions rather than the carbon chain symbols.

Punctuation marks show modes of connection or disconnection, such as the branch-terminating period mark. Thus butyl ethyl propyl methane is 4Y3.2

Colon : denotes a single unsaturation in the familiar manner of use.

Double Colon :: - denotes the doubly unsaturated acetylenic link with quantum-mechanical accuracy as a double double bond. Thus ethylene is 1:1 and acetylene is 1::1

Lower case letters are used in a distinctive manner to locate relative ring positions. The first two, a- and b-positions are not specified when the corresponding atomic group symbols are next to the benzene ring symbol. Thus the ortho-phenylene connective is denoted as - R -, the meta as - Rc -, and the para as - Rd - . Thus aspirin is QVROV1 resorcinol is QRcQ, and benzidine is ZRdRdZ.

Numerals have been used to denote the number of atoms in a given ring. These are enclosed in parentheses with the other heteroatomic or heterocyclic symbols that are a part of the ring description. The ring numerals are punctuated with a stroke mark if the ring is saturated or "nonaromatik". Thus (3/) is cyclopropane and (6/) is cyclohexene, (5/M) is pyrrolidine and (6/M) is piperidine. Heterocyclic numerals are punctuated with a period mark if the ring is dehydrogenated to the aromatic limit, thus (5.0) is furane, (5.5) is thiophene, and (6.N) is pyridine.

Within the ring parentheses, all otherwise equal delineations are resolved by selecting the sequence that gives the lowest possible measure of every kind - lowest set of heterocyclic symbols, lowest set of locants, and the like. Thus pyrazole is (5.MN) and thiazole is (5.NcS). The ring branches are cited in simple alphabetic order of their locants. Thus proline is (5/M)bVQ and niacin is (6.N)cVQ.

Bicyclic systems are pictured through the use of prime marks that are inserted between the ring numerals to show the atomic bridges. No saturation stroke is necessary with prime marks. Thus norpinane is (4'6) and quanuclidine is (6 "6/cN). The stroke mark is added to show saturation in unbridged systems, as with (6 6/) for decalin and (66/) for tetralin.

It has been claimed by the author of the system that the system can be very efficiently used,

1) to make lucid two-dimensional structure diagrams with printable symbols.
2) to prepare long listings of the corresponding delineations with ease.

3) to correlate physical, chemical and biological properties with structures.
4) to sort and list such information with standard tabulating machinery or with the new IBM scanning machinery.

The revision[73] of the system done by E.G.Smith has grown out of the practical experiences of chemists who have been using this notation in a variety of operating and experimental systems for the retrieval of chemical structure information. The Chemical Notation Association formed out of the above chemists now governs the rules of this notation system.

This monograph[73] presents the Standard Notation Rules which have been tested and approved by the Association which cover the great bulk of organic compounds. It also describes certain tentative and proposed procedures that are still being tested for denoting such structure types as chelates and inorganics, and for denoting special problems such as stereoisomerism, uncertainties, Markush structures, mixtures and the like.

8) Code of Gorden, Kendall and Davison 1948 [31,37,16]

The basic concept of this code is the chemical species, which is defined as "a set of atoms individually given, given pairs of these atoms being linked together by directed bonds. The net charge of the set and of any discrete ion contained in it must be specified." Thus no distinction is made between the various types of chemical bonds.

Commonly used chemical symbols are used for the elements, with the following additions:

1) CH_3, CH_2 and CH are represented by letters J, L and M respectively.

2) X is used to indicate ring closure.

3) E and G indicate negative or positive net charges.

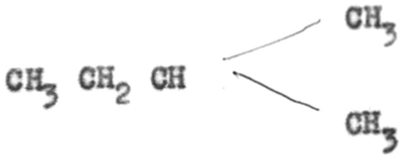
4) Isotopes are indicated by the letter T.

5) Repeating (polymeric) units are indicated by Q.

6) Structures designated for purposes of classification are indicated by R.

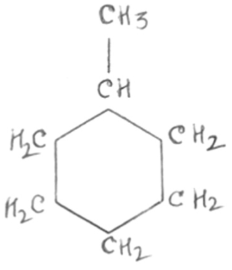
7) Discrete portions of the molecule are separated by oblique /.

In translating chemical structure to the cipher, the symbols are listed in sequence as the elements occur in the compound. Rules for the seniority of symbols are stated so that every one will begin the cipher at the same point. To the symbols representing the elements are added superscript numbers indicating their coordination numbers. Branching in carbon chains is indicated as :



$$CH_3\,CH_2\,CH\begin{matrix}CH_3\\ \\ CH_3\end{matrix} = J^{1}LM^{3}J^{1}2$$

Ring compounds are ciphered using the symbol X to indicate the point of closure, e.g.



$= J^1M^3L5X$

methyl cyclohexane.

Ionic compounds are ciphered using the appropriate symbols E and G. For example

$NaCl = ECl^0/GNa^0$

$BaBr_2 = EBr^0/2/G2Ba^0$

$(CH_3COO)_2Ca = EO^1C^3O^1J^1/2/G2Ca^0$

The seniority rule also determines the order of indexing. The principle is that of the dictionary, except that the ciphering seniority of successive symbols is used rather than alphabetical seniority. The repetition figures are treated as symbols in their own right, being junior to other symbols. For example all L2 ...'s are indexed after all single L.....'s.

The system is fairly easy to learn and a chemist could be trained in its use in about an hour as is claimed by the authors. This system can be adapted readily to computers, for the presence or absence of any structurally defined chemical group. Such a group is defined in terms of atoms and bonds, with provision for alternatives and varying degrees of specificity. It is estimated that searches could be carried out with 100% efficiency and no redundancy.

9) Opler and Norton System 1957.[31,59]

The basis of this coding system is actually one used for encoding networks. Chemical compounds are regarded as networks of chemical groups. Using intuition and experience, a judicious selection of 332 structural elements are assigned arbitrary three-digit numbers. The mode of attachment of these groups one to another is indicated by three additional digits, and a seventh digit serves to number the group. No special order of listing the groups is necessary.

For coding 2-chloro-4-isopropyl benzoic acid, the compound is divided into four groups as follows:

1) Benzene ring
2) Chloro (- Cl)
3) Acid (- COOH)
4) Propyl (- C_5H_7)

The three digit numbers assigned to these groups are entered as 3rd, 4th and 5th digits in the seven digit unit as follows -

_ _ 1 0 6 _ _ benzene
_ _ 0 6 1 _ _ chloro

_	_	3	0	4	_	_	acid
_	_	0	0	3	_	_	propyl

Then the groups are numbered sequentially, starting with any group, (digit 7).

_	_	1	0	6	_	1	benzene
_	_	0	6	1	_	2	chloro
_	_	3	0	4	_	3	acid
_	_	0	0	3	_	4	propyl

For each group, the number of previous group to which it is attached is listed. Benzene, having group 1 in this case has no number under this step. The previous group to which chloro is attached is benzene group 1. This digit is placed in the 6th position.

_	_	1	0	6	_	1	benzene
_	_	0	6	1	1	2	chloro
_	_	3	0	4	1	3	acid
_	_	0	0	3	1	4	propyl

The position in each group by which it is attached to the previous group is given in the 2nd position. Chloro and acid groups are attached to the benzene at its 1 position to benzene and the propyl group is attached to at its 2 position, and the 1st digit indicates the position in the previous group to which the group being coded is attached. For example, chloro is attached to the 2 position on the benzene ring.

-	-	1	0	6	-	1	benzene
2	1	0	6	1	1	2	chloro
1	1	3	0	4	1	3	acid
4	2	0	0	3	1	4	propyl

The completed code for 2-chloro-4-isopropyl benzoic acid becomes

/ --106-1/2106112/1130413/4200314/

Few conventions are followed, such as the use of the numeral 9 in the 1st position of the seven digit code number to indicate indeterminate structures; bridged, spiro and fused ring compounds are identified with a two-digit (instead of a normal three-digit) number in the 4th and 5th positions. This number begins with 7, so that the numeral 7 occurring in the 4th position automatically indicates this type of structure. Stereo- and optical isomers are indicated by numbers between 900 and 999.

The advantages claimed for this system are its simplicity and adaptability to high speed computing machines. It is reported that coding compounds of average complexity required less than 2 minutes per compound. Once the compounds have been coded - reduced to their digital equivalents - and key punched on cards, they are ready for storage and search by a high speed computer. The designers of the system have prepared a manual for programming for computers for the use with this system, which includes a list of characteristics as well as detailed directions for carrying out the searching operations.

Other systems developed by particular group for a particular problem include Zatopleg System devised by M Mooers;[31] The Nodal Index for branch structures developed by H.P.Luhn;[31] Coding on the basis of the atomic number,[1] Code for carbohydrate derivatives,[48] and so on.

CHAPTER - IV

DEVELOPMENT OF CHEMICAL NOTATION SYSTEMS FOR MECHANIZATION

The large number of chemical compounds now known, makes it imperative that any chemical notation system be adaptable to mechanical manipulation. Punched cards offer a ready means of sorting, classifying and correlating large numbers of individual pieces of data. Electronic searching and calculating machines have been developed for the purpose of computing and correlating mass of data. These complex machines, when supplied with suitable information can compare structural units and determine relationships between such units in molecules.

The present trend seems to be towards the development of specific coding systems for particular applications. The choice of the code to be used depends on whether the purpose is indexing, classifying, correlating or structure searching. It also depends on the availability of mechanical equipment, such as punched card sorters, digital computers etc.

"Ease of manipulation by machine methods e.g. punched cards" was one of the eleven requisites of the desiderata set up by the Commission of IUPAC in 1946. With the advent of computers the new systems coming up are being designed to be used by the computers for indexing,classifying or correlating the structures. The old systems are being remodelled to suit the recent trend of machanization.

In reviewing the potentialities of organic ciphering systems with respect to indexing and related topics F. Richter[63] says:-

"The comprehensive and unique ciphering systems of the type at present under review in different countries seem to open up interesting possibilities in the field of literature searching. Major applications suggesting themselves are 1) replacement of official nomenclature in indexes, 2) automatic generation of an arrangement approaching to some extent a system of organic chemistry, and 3) sorting of chemical structures as to details of special interest."

Recent development of the notation systems enlighten the above applications, mostly with the help of computers.

As early as 1942-45, Frederich Y. Wiselogle devised "A New System for the Classification of Co pounds"[12] which was used for correlating the activities with the antimalarials by using punched cards.

Further Arendell F.H.[4] designed a 3 symbol code based on the Wiselogle classification for pharmaceutical chemicals. The functional units present in the compounds are described in descending order of importance. The 3 symbol code consists of division, class and descriptor. The 1st symbol consisted simply of chemical symbol for the atom of principal importance. The 2nd symbol was a number classing the functional unit in one of six ways, in descending order of importance.

6 - as a heteroatom in an aromatic ring.

5 - as a heteroatom in a pseudoaromatic or five membered ring.

4 - as a functional unit attached to an aromatic ring.

3 - as a functional unit attached to a pseudoaromatic ring.

2 - as a functional unit attached to an aliphatic chain.

1 - as an inorganic functional unit.

The third symbol was a 2nd number designating one of the functional unit descriptors selected from a list provided for each class.

Moreover Geer H.A. et al.[34] had tried to use the Wiselogle classification system with computer.

The G - K - D ciphers of Gorden, Kendall and Davison had designed their system for early versions of computers. Subsequently Davison W.H.T. reported on "Programs and Equipment for sorting G-K-D punched cards for any structurally defined group".[16] He gave programs and schematic designs for a special purpose electronic structural computer, which would sort a pack of cards, punched with G-K-D ciphers only, for the presence or absence of any structurally defined chemical group. Such a group is defined in terms of atoms and bonds, with provision for alternatives and varying degrees of specificity. The statistical analysis of the probable number of provisional assignments in the course of a search enables to show that searches could be carried out with 100 % efficiency and no redundancy.

The Frear Code - N.R.C. - CBCC - Code was also designed with the use of machine sorted punched card with one card for each compound. The card having 80 columns is divided into 5 sections - serial number, empirical formula, structure, relevant physical properties and biological data. In order to select all amino sulfonamide compounds, the empirical formula section can be sorted to select all compounds containing two or more nitrogen atoms and one or more sulphur atoms.

John A. Silk had responded to the IUPAC Commission's invitation with his "New System of Organic Notation" in 1951. After a decade, he published additional details and improvements in his "Linear Notation for Organic Compounds".[66] The Dyson system develops the systematic description and enumeration of ring structure in which individual rings are built up one by one. Silk presented the modifications in this system which yield improved enumeration patterns, (reticular systems), show the characteristic types of ring fusion clearly and enable molecular formulas to be calculated from the systematic expressions without any additional information.[65]

He has further developed a notation based on fragment code[67] for chemical patents.

F.R. Benson showed how Wiswesser Notation could be used for "Recording and Recovering Chemical Information with Standard

Tabulating Equipment."[6] E. G. Smith started using this notation with standard punched card equipment in 1952 and in 1954 wrote a faculty report on its use in "A Punched Card Catalog of the Physical Properties of Some Common Organic Chemicals." In 1960, he gave a detailed account of his experience in encoding some 50,000 structures in "Machine Searching for Chemical Structures,"[72] and claimed that in many cases the search could be made within few minutes. He completed a manual that incorporates a number of his own ideas and provides more comprehensive coverage.[73]

Opler A and Baird N.[58] showed how this notation could be used and fed to digital computers. Bonnett H.T. and Calhoun, D.W. in 1961 described the "Application of a Line Formula Notation in an Index of Chemical Structures."[8] They gave the details of the card format assigning columns for a four digit 1) classification number of which each digit is itself a code-showing the number of rings, types and number of hetero atoms, and the type and amount of nonaromatic unsaturation present in a compound, 2) Reference field, 3) Author field and 4) Notation field.

This design of the system provides a total of eight indexes of which they have printed seven with the help of tabulator into a list form. These are 1) structure sequence, 2) numeric sequence, 3) author sequence, 4) ring index, 5) hetero atom index, 6) heteroatom count, and 7) unsaturation index. The structure sequence can be achieved by sorting the file of cards alphabetically beginning at column 79 and moving toward and including column 39 of a 80 column card.

Their attempts also have shown the desirability of indicating how the fragments are connected including such fairly detailed information as : 1) attachment to rings, 2) attachment to non-carbonyl carbon atom, 3) attachment to carbonyl carbon atom, 4) attachment to non-carbon atom, 5) multiple occurrence, 6) positional relation of hetero atoms in rings, 7) ring sizes, 8) aromatic rings, 9) hetero-atoms and the like.

From their experience it seemed that the notaton 1) provides a means of recording structures in unmodified machine language in punched cards and tabulated listings, and 2) the notation can be scanned much like a structural formula, and can be organized alphabetically.

At the same time, Gelberg, A; Nelson, W.; Yee, S. and Metcalf, E.A. reported on "A Program Retrieval of Organic Structure Information via Punched Cards,"[35] using the same notation in collaboration with E.G.Smith. The author of the notation called the attention of the American Chemical Society's Division of Chemical Education to the fact that the notation could be used as "A New Tool for Teaching Structural Chemistry,"[79] for accelerating studies by better understanding of the logical sets of structures and organization of structural knowledge.

Horner J.K.[45] and Lefkovitz D.[52] have tried to use the Wiswesser's line notation with computers at low cost storage and retrieval system.

Moreover Hyde E, Mattews F.W., Thomson L.H. and Wiswesser W.J.[46] and Lynch M.F.[53] have tried the computer program for this notation for the conversion to a connectivity matrix and vice versa.

Hoffmann E.[44] has tried to use the modified version of this notation for the encoding of proteins. The protein chain is treated as one uniform structural entity consisting of a defined number of repeated peptide linkages. The side chains of the amino acids which make up the protein are treated as substituents of this entity and their order of appearance in the notation, describes the sequence of the amino acids in the chain.

Revesz G.S.[61,62] and Garfield E. et al.[33] have designed a one letter code for peptides with the combinationof IUPAC notation with Wiswesser's which is being used in Index Chemicus Registry System.

This above system has announced an ICRS Substructure Index which is an alphabetical listing of Wiswesser line notation descriptions of each of the 200,000 new chemical compounds reported in Current Abstracts of Chemistry and Index Chemicus each year. It has been claimed to have been organized for quick efficient searching for all types of compound fragments.

The "Survey of Chemical Notation Systems"[56] in 1964 revealed that Wiswesser Notation is being adopted by more users rather than the IUPAC notation. The reasons given by the users are

1) It can be handled by standard IBM machines.

2) It requires, on the average, fewer symbols per structure, thus leaving room on an IBM card for more information of other kinds.

3) It emphasizes functional groups rather than the less significant carbon skeletons.

4) It is less complex and/or easier to learn.

5) It was clearly shown (in user's opinion) to be superior by the 1952 NRC volunteer study.

A few further reasons are also cited by one of the users:

1) The blank spaces divide most Wiswesser notation into "words", thus making them easier to read.

2) The method of handling ring systems is superior.

3) Having less punctuation and no subscripts or superscripts, it is easier to typewrite.

4) Generic searches are performed without adding a functional group field to the auxiliary classification code used with the notation.

Hayward H.W. in 1961 announced a "New Sequential Enumeration and Line Formula Notation System for Organic Compounds." He later on reported[42] on some experiences with his notation as the definitions and rules of the system are such that each atom and bond in a structure falls into one and only one of the following five set-based structure classes:-

1) ring,
2) bridge,
3) link,
4) chain,
5) non-skeletal.

The procedure for ranking and enciphering structural features is such that complex networks of rings and chains can be uniquely and unambiguously delineated without the inclusion of seniority numbers (locants) to show how the component parts are interconnected.

In the meantime Dyson G.M. and Riley E.F.[29] tried to shift the use of IUPAC's notation from standard IBM punched cards to magnetic tapes, which are searched by the IBM 1401 digital computer. This notation is being tried and experimented on for computerization and for the production of notational indexes at Chemical Abstracts. Dyson has cited[56] the following reasons for adopting the IUPAC rather than the Wiswesser notation for this experiment:

1) It has received official approval by IUPAC for international use.

2) It is easier to use.

3) It is more readily adapatable to computer searches of magnetic tape - an important desideratum not considered by IUPAC.

4) It does a better job of grouping related compounds together in an index.

5) It can generate a systematic nomenclature.

6) It is simpler to encode and decode, even though it does not differe greatly from the Wiswesser notation in brevity or in encoding time.

7) It is hierarchical.

8) Its numerical enumeration of ring systems is superior to Wiswesser's alphabetical enumeration.

The survey has mentioned that the Wiswesser notation can be handled on standard typewriters and on the ordinary punched card machines then commonly available, while the IUPAC and Hayward notations cannot. The IUPAC notation, in particular, contains so many symbols not present on an ordinary IBM machines that Chemical Abstracts had to install specially built machines to deal with this notation.

Since then this system is being modified[25-27] for the handling of organic structures in the form of connection tables and ciphers by computers as the original notation was conceived only as a means of translating structures into exact linear algebraic forms and without much of knowledge of present computer systems. A fully modified version is given by Dyson G.M., Lynch M.F. and Morgan H.L.[28]

Simultaneously, Dammers H.F. and Polton D.J.[15] have been using the computer operated storage and retrieval system for chemical structures based on the use of the IUPAC notation at Shell Research Ltd. Sittingbourne, since mid of 1965. Use of the notation for a file of nearly 50,000 structures provides, the advantage of speed and low cost. The cipher is converted by an input program into an internal representation, more suited for search purposes, and requiring 26 characters per cipher on average. The computer programs now enable the conversion of ciphers into atom-connection tables, and also assists in the calculation of empirical formulae. Programs for generating a fragmentation code from the cipher have been tried. This enables on one hand for manual searching and on the other it provides the means of setting up inverted files in the computer system for rapid on-line consultation of the structure file.

Meanwhile Sorter, Granito, Gelberg and their associates developed encyclopedic notation indexes made by 'permuting' the symbols, of the Wiswesser Line Notation with computer programs.[38-39,74] A functional group search would require the scanning of the line-notation for each compound in the entire list. To overcome this deficiency, it has been necessary to adopt functional group codes for use with IBM punch card systems. If the functional groups could be located by means of the line-notation, the need for such a special code would be obviated.

A list of permutations of chemical line notations alphabetized on individual symbols could be used to locate readily all compounds containing any specified functional group as well as specific compounds and specific classes of carboxylic or heterocyclic structures. The figure shows the permutation of the notation T6N$\overset{J}{\text{H}}$ BG (2-chloropyridine).

						44						
						T	6	N	J		B	G
					T	6	N	J		B	G	
				T	6	N	J		B	G		
			T	<u>6</u>	N	J		B	G			
		T	<u>6</u>	N	J		B	G				
	T	6	N	J		B	G					
T	6	N	J		B	G						

Column 44 of 80 column punched card could be used for setting up an index. If all of the columns from 44 to 79 are placed in alphanumeric order, then this compound could be found in 35 locations in a permuted index. However, only six entries in this example can be considered to be of value, since column 44 is blank on the remaining 29 cards. Exclusion rules are given to avoid indexing symbols for which search would be done rarely. This study has shown that tabulated lists can be used to rapidly locate specific compounds, classes of compounds or all compounds having some functional group.

Landee F.A. in 1964[51] have tried computer programs for handling chemical structures expressed in the Wiswesser notation and recorded on magnetic tape as a part of file system including structure in notation form, molecular formula, systematic name, properties etc. By computer analysis of the notation and molecular formula portions of the file, an inverted group and formula file index is created. By appropriate programming, searches of varying depth are possible even atom by atom.

He with his associates[9-11] has developed notation programming further in creating the correct Wiswesser notation for polycyclic structures. The program developed requires as input a definition of the ring atom connections. It examines every possible path through the network and selects the one which meets the requirements of the notation rules. It then generates the correct notation and identifies all the atoms in the network which are multicyclic, bridges, spiro points etc. The program's greatest value is its speed and accuracy in encoding complex polycyclic structures.

The Wiswesser Line Notation has been tried for various purposes for different computers. Ofer K.D.[57] has programmed in FORTRAN to be used to prepare a key symbol out of context index, to disseminate notices on additions to the file to interested chemists, to search the file for compounds with certain structural elements and list the selected items, or to convert from a linear to a fragmentation code.

Skolnik H.[58] has devised "A New Linear Notation System based on Combinations of Carbon and Hydrogen", which he has further modified for NMR Data[69] and for indexing polymers.[70]

Duffin W.M.[17] has developed a code consisting of 3 parts α / β / γ, where α - indicates senior ring present, β - other rings and groups and γ - total number of carbon atoms present.

So also Waldo W.H.[76] has designed a method for searching the structures by computer. The structure is written on cross hatched paper with numerals written for chemical bonds and single character letters for the elements. Thus

CH_3CH_2SH becomes

C	I	O	I	5

The development so far seen reveals the fact that any new system being developed or the old ones are modified keeping in view of its use with mechanical devices. The chemical notation systems will be there as long as chemists continue to read and write about the substances they are creating and studying.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Amster, A.B., Beckett, C.W. and Johnson, H.L. The use of punched cards for the coding of inorganic and simple organic compounds. J. Chem. Educ., 26, 304-06, 1949.

2. Anonymous. Ciphering of organic structural formulae. Nature, 165, 301, 1950.

3. Anonymous. Which notation ? Chem. Eng. News., 33, 2838-2843, 1955.

4. Arendell, F.H. A three-symbol code for searching chemical structures. J. Chem. Doc., 1, 47-57, 1961.

5. Bailar, J.C. Jr., Heumann, K.F. and Seiferle, E.J. The use of punched card techniques in the coding of inorganic compounds. J. Chem. Educ., 25, 142-144, 1948.

6. Benson, F.R. Recording and recovering chemical information with standard tabulating equipment. Remington Inc., South Norwalk Conn., 1953.

7. Berry, M.M. and Perry, J.W. Notational systems for structural formulas. Chem. Eng. News., 30, 407-410, 1952.

8. Bonnett, H.T. and Calhoun, D.W. Application of a Line Formula Notation in an index of chemical structures. J. Chem. Doc., 2, 2-6, 1962.

9. Bowman, C.M., Landee, F.A. and Reslock, M.H. A chemically oriented information storage and retrieval system. I. Storage and verification of structural information. J. Chem. Doc., 7, 43-47, 1967.

10. Bowman, C.M., Landee, F.A., Lee, N.W. and Reslock, M.H. A chemically oriented information storage and retrieval system. II. Computer generation of the Wiswesser Notations of complex polycyclic structures. J. Chem. Doc., 8, 133-137, 1968.

11. Bowman, C.M., Landee, F.A., Lee, N.W., Reslock, M.H. and Smith, B.P. A chemically oriented information storage and retrieval system. III. Searching a Wiswesser line notation file. J. Chem. Doc., 10, 50-54, 1970.

12. Buhle, E.L., Hartnell, E.D., Moore, A.M., Wiselogle, L.R. and Wiselogle, F.Y. A new system for the classification of compounds : A contribution from the survey of anti-malarial drugs. J. Chem. Educ., 23, 375-91, 1946.

13. Chemical Biological Coordination Centre. A method of coding chemicals for correlation and classification. National Research Council, Washington, 1950.

14. Cockburn, J.C. The Newcastle system of representation of the formulae of organic compounds. Private communication. 1951.

15. Dammers, H.F. and Polton, D.J. Use of IUPAC notation in computer processing of information on chemical structures. J. Chem. Doc., 8, 150-160,1968.

16. Davison, W.H.T. Programs and Equipment for sorting GKD punched cards for any structurally defined group. Abstracts of Papers, 120th meeting. American Chemical Society, New York, Sept. 1951.

17. Duffin,W.M. Chemical coding for information retrieval, J. Chem. Doc., 1, 44-46, 1961.

18. Dyson, G.M. A notation for organic compounds. Nature, 154, 114-115, 1944.

19. Dyson, G.M. A new notation for organic chemistry. Royal Institute of Chemistry Lecture, published jointly with the Chemical Society and the Society of Chemical Industry, London,1946.

20. Dyson, G.M. A new notation and enumeration system for organic compounds. Longmans Green, London, 1947.

21. Dyson, G.M. A new notation for organic chemistry. Research, 2, 104-14, 1949.

22. Dyson, G.M. Some applications of the Dysonian notation of organic compounds. J. Chem. Educ., 26, 294-303, 1949.

23. Dyson, G.M. A new notation and enumeration system for organic compounds. 2nd ed. Longmans, Green, London, 1949.

24. Dyson, G.M. Chemical ciphers or notations: their nature and use. ICSU Review, 4, 110-16, 1962.

25. Dyson, G.M. Inf. Stor. and Retr., 2, 59, 1964.

26. Dyson, G.M. Inf. Stor. and Retr., 2, 159, 1964.

27. Dyson, G.M. Modifications and abbreviations recommended for computer and visual handling of the IUPAC notation. Abstracts of Papers, 155th meeting, April 1968.

28. Dyson, G.M., Lynch, M.F. and Morgan, H.L. A modified IUPAC - Dyson notation system for chemical structures. Inf. Stor. and Retr., 4, 27-83, 1968.

29. Dyson, G.M. and Riley, E.F. Mechanical storage and retrieval of organic chemical data : Identification of structural features. Chem. Eng. News, 39, 74-76, 78,89, 1961.

30. Frear, D.E.H. Punched cards in correlation studies. Chem. Eng. News, 23, 2077, 1945.

31. Frear, D.E.H. Comprehensive coding schemes for chemical compounds. Chapt er 22 in Casey, R.S., Perry J.W., Berry M.M. and Kent, A.(Eds.). Punched cards : their applications to science and industry. 2nd ed. Reinhold, New York, 1958.

32. Frear, D.E.H., Seiferle, E.J. and King, H.L. A new classification system for chemical compounds. Science, 104, 177-178, 1946.

33. Garfield, E., Revesz, G.S., Granito, C.E., Dorr, H.A., Calderon, M.M. and Warner, A. Index Chemicus Registry System : pragmatic approach to substructure chemical retrieval. J. Chem. Doc., 10, 54-58, 1970.

34. Geer, H.A., Moore, A.M., Howard C.C. and Eady, C.E. The Parke Davis Code for chemical structures. J. Chem. Doc., 2, 110-113, 1962.

35. Gelberg, A.W., Nelson, G., Yee, S. and Metcalf, E.A. A program retrieval of organic structure information via punched cards. J. Chem. Doc., 2, 7-11, 1962.

36. Gordon, M., Kendall, C.E. and Davison, W.H.T. A new systematisation of chemical species. Proc. 11th Int. Congr. Pure and Appl. Chem. London. 2, (3), 115-32, 1947.

37. Gordon, M., Kendall, C.E. and Davison, W.H.T. Chemical ciphering : a universal code as an aid to chemical systematics. Royal Inst. Chem. Monograph, London, 1948.

38. Granito, C.E., Gelberg, A., Schultz, J.E., Gibson, G.W. and Metcalf, E.A. Rapid structure searches via permuted chemical line notations. II. A key-ounch procedure for the generation of an index for a small file. J. Chem. Doc., 5, 52-55, 1965.

39. Granito, C.E., Schultz, J.E., Gibson, G.W., Belberg, A., Williams, R.J. and Metcalf, E.A. Rapid structure searches via permuted chemical line notations. III. A computer produced index. J. Chem. Doc., 5, 229-233, 1965.

40. Gruber, W. Die genfer nomenclatur in chiffren und vorschlage fur ihre erweiterung auf ringverbindungen. Angew. Chem., 61, 429-31, 1949., Beihefte Nr. 58, 1950.

41. Hayward, H.W. A new sequential enumeration and line formula notation system for organic compounds. U.S.Pat. Office, R. and D. Rept. No.21, 1961.

42. Hayward, H.W. Molecular structure, connected sets, topological linguistics, and chemical notation. Abstracts of Papers, 155th meeting. A.C.S. April 1968.

43. Heumann, K.F. and Dale, E. Statistical survey of chemical structure. Chapter 14 in Peakes, G.L., Kent A and Perry, J.W. Progress report in chemical literature retrieval. Interscience, New York, pp 201-214, 1957.

44. Hoffman, E. Use of a modified Wiswesser notation for the encoding of proteins. J. Chem. Doc., 9, 137-140, 1969.

45. Horner, J.K. Low-cost storage and retrieval of organic structures by permuted line notations : small collections. J. Chem. Doc., 7, 85-88, 1967.

46. Hyde, E., Mattews, F.W., Thomson, L.H. and Wiswesser, W.J. Conversion of Wiswesser notation to a connectivity matrix for organic compounds. J. Chem. Doc., 7, 200-204, 1967.

47. International Union of Chemistry. Codes invited. Chem. Eng. News. 27, 2998, 1949.

48. Isbell, H.S. System for classification of structurally related carbohydrates. J. Res. N.B.S., 57, 171-178, 1956.

49. IUPAC. A proposed International Chemical Notation, tentative version. Commission on codification, ciphering and punched card techniques. IUPAC. Longmans, Green, London, 1958.

50. IUPAC. Rules for IUPAC notation for organic compounds. Commission on codification, ciphering and punched card techniques of the IUPAC. Longmans, Green, London, 1961.

51. Landee, F.A. Computer programs for handling chemical structures expressed in the Wiswesser notation. Abstracts of Papers, 147th meeting, A.C.S. April, 1964.

52. Lefkovitz, D. A chemical notation and code for computer manipulation : the mechanical chemical code. Abstracts of Papers, 155th meeting, A.C.S. April 1968.

53. Lynch, M.F. Conversion of connection table descriptions of chemical compounds into a form of Wiswesser notation. Abstracts of Papers, 155th meeting. April 1968.

54. Morgan, J.A. and Frear, D.E.H. Coding and sorting chemical compounds by means of punched cards. J. Chem. Educ., 24, 58-61, 1947.

55. Murphy, W.J. Ed. $C_{12}.Q - C_{12}.OH - QC_{11} - Q_{12}$. Chem. Eng. News, 30,403, 1952.

56. NAS - NRC. Survey of chemical notation systems. National Academy of Sciences. National Research Council Publ. 1150, 1964.

57. Ofer, K.D. A computer program to index or search linear notations. J. Chem. Doc., 8, 128-130, 1968.

58. Opler, A. and Baird, N. Display of chemical structural formulas as digital computer output. Am. Doc., 10, 59-63, 1959.

59. Opler,A and Norton, T.R. New speed to structural searches. Chem. Eng. News, 34, 2812-2814,2816, 1956.

60. Patterson,A.M. Words about words. Chem. Eng. News, 29, 4078. 4116, 1951.

61. Revesz, G.S. Simplified notation for peptides in computer compatible format. Nature, 219, 1113, 1968.

62. Revesz, G.S., Granito,C.F. and Garfield, E. One letter notation for calculating molecular formulas and searching long-chain peptides in the Index Chemicus Registry System. J. Chem. Doc., 10, 212-215, 1970.

63. Richter, F. Potentialities of organic ciphering systems with respect to indexing and related optics. Abstracts of Papers, 120th meeting. A.C.S. New York, Sept. 1951.

64. Silk, J.A. A new system of organic notation (Private communication). 1951.

65. Silk, J.A. An improved system for the enumeration and description of ring systems. J. Chem. Doc., 1, 58-62, 1961.

66. Silk, J.A. A linear notation for organic compounds. J. Chem.Doc., 3, 189-195, 1963.

67. Silk, J.A. A notation based fragment code for chemical patents. J. Chem. Doc., 8, 161-165, 1968.

68. Skolnik, H. A new linear notation system based on combinations of carbon and hydrogen. J. Heterocyclic Chem., 6, 689-695, 1969.

69. Skolnik, H. A correlative notation system for NMR Data. J. Chem. Doc.,10, 216-220, 1970.

70. Skolnik, H. and Hays, J.T. A new notation system for indexing polymers. J. Chem. Doc., 10, 243-247, 1970.

71. Smith, C. Modern chemical nomenclature : a lecture delivered before the Chemical Society on May 14th, 1936. J. Chem. Soc. 1067-78, 1936.

72. Smith, E.G. Machine searching for chemical structures. Science, 131, 142-146, 1960.

73. Smith, E.G. The Wiswesser line-formula chemical notation. McGraw-Hill, New York,1968.

74. Sorter, P.F., Granito, C.E.,Gilmer, J.C.,Gelberg,A. and Metcalf, E.A. Rapid structure searches via permuted chemical line notations. J. Chem. Doc., 4, 56-60, 1964.

75. Taylor, F.L. Enumerative nomenclature for organic ring systems. Ind.Eng. Chem.,40, 734-738, 1948.

76. Waldo, W.H. Searching two-dimensional structures by computer. J. Chem. Doc., 2, 1-2, 1962.

77. Wiswesser, W.J. The Wiswesser line formula notation. Chem.Eng. News, 30, 3523-3526, 1952.

78. Wiswesser, W.J. A line formula chemical notation. Crowell Co., New York, 1954.

79. Wiswesser,W.J. A new tool for teaching structural chemistry. Abstracts of Papers, 139th meeting A.C.S.1961.

80. Wiswesser, W.J. 107 years of line-formula notations. (1861-1968). J. Chem. Doc., <u>8</u>, 146-150, 1968.

81. Wood, G.C. A detailed biology code for storing, retrieving and correlating chemical-biological data. Am. Doc., <u>8</u>, 168-180, 1957.

82. Zipf, G.K. Human nature and the Principle of Least Effort. Addison-Wesley Press, Cambridge, 1949.

APPENDIX

NUMBERS 2401-2700 OF THE RING INDEX CIPHERED AND ARRANGED IN ALPHABETICAL ORDER

AsZ, 7, $B6_4$, 1, 3, 8, ZN, 5, 10	2600
AsZ, 7, $B6_4$, 1, 3, 8, ZN, 10	2602
AsZ, 7, $B6_4$, 1, 3, 8, ZN, 10, 11	2599
AsZ, 7, $B6_4$, 1, 3, 8, ZQ, 10, H, 7, 10	2604
AsZ, 10, $B6_4$, 1, 3, 8, ZN, 7	2601
AsZ, 10, $B6_4$, 1, 3, 8, ZQ, 7, H, 7, 10	2603
$A6_3$5, 1, 13-13, 3, E, 1(2)9	2575
$A6_3$5, 1-3, 11, 2-2, ZN, 5, 8, E,10(2)16	2528
$A6_3$5, 1-4, 10, 2-2, ZQ, 5, 8, E, 9(2)15	2541
$A6_4$, 1-1, 2, 5, ZQ, 4, 7, ZS, 10, E, 2,5, 12(2)18	2678
$A6_4$, 1-1, 2, 5, ZQ, 7, ZS, 10, ZN, 4, E, 2, 5, 12(2)180	2677
$A6_4$, 1-1, 4-4, 9-9, ZQ, 3, 5, 8	2683
$A6_4$, 1-1, 4-4, 9-9, ZS, 3, 5, 8	2690
$A6_4$, 1-1, 4-4, 9-9, ZS, 3, 5, 8, 10, 14, 19	2688
$B6_3$5, 1, 3, 1, ZQ, 16, H, 1, 2, 15, 16, 17	2497
$B6_3$5, 1, 3, 5, H, 15	2564
$B6_3$5, 1, 3, 5, H, 16	2565
$B6_3$5, 1, 3, 5, ZN, 15, H, 15	2506
$B6_3$5, 1, 3, 5, ZN, 16, H, 16	2454
$B6_3$5, 1, 3, 5, ZQ, 15, H, 15	2489
$B6_3$5, 1, 3, 5, ZQ, 15, 17, H, 15, 16, 17	2406

B6$_3$5, 1, 3, 5, ZQ, 16, H, 16	2490
B6$_3$5, 1, 3, 5, ZS, 16, H, 16	2506
B6$_3$5, 1, 3, 5 - 14	2567
B6$_3$5, 1, 3, 5 - 14, ZN, 7, 10	2520
B6$_3$5, 1, 3, 5 - 14, ZN, 7, 16	2439
B6$_3$5, 1, 3, 5 - 14, ZN, 10	2539
B6$_3$5, 1, 3, 5 - 14, ZN, 14, H, 11, 14	2438
B6$_3$5, 1, 3, 5 - 14, ZN, 16	2457
B6$_3$5, 1, 3, 5 - 14, ZQ, 16, H, 15, 16	2495
B6$_3$5, 1, 3, 7, H, 15	2561
B6$_3$5, 1, 3, 7, H, 17	2562
B6$_3$5, 1, 3, 7, ZN, 14, H, 17	2537
B6$_3$5, 1, 3, 7, ZN, 14, 15, H, 15	2435
B6$_3$5, 1, 3, 7, ZN, 15, H, 15	2452H
B6$_3$5, 1, 3, 7, ZQ, 10, H, 10	2547
B6$_3$5, 1, 3, 7, ZQ, 15, H, 15	2486
B6$_3$5, 1, 3, 7, ZQ, 16, H, 16	2485
B6$_3$5, 1, 3, 7, ZQ, 17, H, 17	2484
B6$_3$5, 1, 3, 7-7, ZQ, 15, H, 7, 8, 15, 18	2498M
B6$_3$5, 1, 3, 8, H, 15	2559
B6$_3$5, 1, 3, 8, ZN, 7, 10, H, 15	2519
B6$_3$5, 1, 3, 8, ZQ, 7, H, 7	2544
B6$_3$5, 1, 3, 8, ZQ, 15, 17, H, 15, 16, 17	2404
B6$_3$5, 1, 3, 8, ZQ, 15, 17, ZN, 4, H, 3, 4, 15, 16, 17	2402F

B6$_3$5m 1, 3, 8, ZQ, 15, 17, ZN, 4, H, 4, 14(1)17	2402
B6$_3$5, 1, 3, 8, ZQ, 15, 17,ZN, 5, H, 15, 16, 17	2403
B6$_3$5, 1, 3, 8, ZQ, 15, 17, ZN5, 4, H, 15, 16, 17	2402M
B6$_3$5, 1, 3, 8, ZQ, 16, ZN, 7, 10, H, 16	2460
B6$_3$5, 1, 3, 8-8, H, 7, 8, 15(1)18	2571
B6$_3$5, 1, 3, 9, H, 17	2563
B6$_3$5, 1, 3, 9, ZN, 7, H, 17	2538
B6,5, 1, 3, 9, ZN, 7, 16, H, 17	2436
B6$_3$5, 1, 3, 9, ZN, 17, H, 17	2452R
B6$_3$5, 1, 3, 9, ZQ, 15, H, 15	2488
B6$_3$5, 1, 3, 9, ZQ, 15, 17, H, 15, 16, 17	2405
B6$_3$5, 1, 3, 9, ZQ, 16, H, 16	2487
B6$_3$5, 1, 3, 9, ZQ, 16, ZN, 7, H, 16	2471
B6$_3$5, 1, 3, 9, ZS, 15, H, 15	2505
B6$_3$5, 1, 3, 10, H, 15	2570
B6$_3$5, 1, 3, 10, ZQ, 15, H, 15	2496
B6$_3$5, 1, 3, 10, ZSe, 15, ZN, 7, 14, H, 15	2509
B6$_3$5, 1, 3, 10, ZS, 15, ZN, 7, 14, H, 15	2500
B6$_3$5, 1, 3-10, 1-6, ZQ, 9. H, 9	2549
B6$_3$5, 1, 3-10, 4, ZN, 4, 6, H, 4, 14	2419
B6$_3$5, 1, 3-10, 4, ZN, 16	2458
B6$_3$5, 1, 3-10, 4, ZQ, 8,15, H, 7, 8, 9, 15	2475M