Generic Searching by Use of Rotated Formula Indexes*

By EUGENE GARFIELD
Director, Institute for Scientific Information, 33 South 17th St., Philadelphia 3, Pa
Received March 13, 1962

The term, "generic searching," is variously interpreted to mean classified, non-specific, group or categorized searching. For research chemists, generic searches are a valuable tool. Structurally related chemicals with a specified combination of functional groups, rings, or other desiderata used to code or describe chemicals, may be found in a more orderly fashion by generic searching. While many machine searching systems code molecular formulas, this information is rarely used in generic searches. The Chemical Biological Coordination Center (CBCC) file provided molecular formulas which were not used much in searching.

Generic searching often is done on a less sophisticated level. For example, one researcher regularly looks for articles on new uses of lithium. Obviously, as a lithium manufacturer, his company is interested in finding new applications. Similarly, chemists in the missile propellant field are interested in all new boron compounds since many are valuable in the production of propellants.

From my own experience, I have found that the conventional formula indexes of *Chemical Abstracts* and of *Index Chemicus* do not permit a convenient means of searching generically except for certain classes such as steroids (C17), etc. For example, a search for "all compounds in which phosphorus is attached to at least five oxygen atoms" is completely impossible unless you search every page of the index. If L. Pauling and V. Schomaker had been able to make such a generic search, it might have avoided the need to retract statements made in a communication commenting upon the works of E. Ronwin.

In November, 1951, Ronwin published a paper in *J. Am. Chem. Soc.*[1] introducing a new structural formula for both ribo- and desoxyribonucleic acids. The structure had as its core a $(P_2O_5)_n$ polymer chain of phosphoanhydride links. By a theoretical treatment, Ronwin showed the formula to be compatible with available factual data concerned with the structure of nucleic acids. In February, 1952, L. Pauling and V. Schomaker wrote in a communication to the editor[2]: "In the proposed structure for the nucleic acids each P atom has five O atoms attached to it, three of which bind it to adjacent P atoms, and two of which are in a OH group and a sugar ester group, respectively. There is, however, no precedent for a structure in which P is bonded to five O atoms... the ligation of five O atoms about each P atom is such an unlikely structural feature that the proposed phospho-tri-anhydride formula for the nucleic acids deserves no serious consideration."

However, in the July, 1952, issue,[8] they made a most gentlemanly retraction in a second communication: "Dr. Ronwin has now kindly informed us that he has become aware of earlier references in the literature to compounds to which structures have been attributed involving quinquepositive P bonded to five O atoms or to a total of five O atoms and similar atoms. Anschütz[4] prepared four compounds to which he assigned structures involving ligation of five O atoms to a P atom ... our statement that there is no precedent for a structure in which a P atom is bonded to five O atoms must accordingly be withdrawn."

Frankly I can't recall how I even came across this interesting discussion. I was doing a search on a completely unrelated matter. It occurred to me at that time that the retraction would have been unnecessary—if a generic formula index to *Chemical Abstracts* were available. I was intrigued to learn how Dr. Ronwin managed to find the reference to Anschütz's paper and others mentioned by Pauling. It had been impossible for me to find them in a search. I wrote to Dr. Ronwin[5] and asked him how he had managed to find the necessary precedent. Dr. Ronwin replied[6] that a colleague had remembered the paper as an abstractor for *Chemisches Zentralblatt!*

In a subsequent letter, however, Prof. Pauling pointed out that the retraction did not involve a retraction of the criticism of the proposed phospho-tri-anhydride formulas for the nucleic acids. The retraction was limited to the statement that "there is no precedent for a structure in which a P atom is bonded to five O atoms." Prof. Pauling correctly states that there might be some question about the correctness of the structures assigned to some of these compounds, and especially that these compounds are extremely sensitive to moisture.[7]

This, then, is the explanation of why and when I first felt the need for a rotated formula index. I wonder to this day whether there might not have been other examples in the literature which Pauling or Ronwin might have found if a rotated index were available. This example illustrates very well the role that new scientific information services can play in advancing scientific theory. In a short paper such as this, there is little need to belabor the question of the potential value of generic indexes. However, generic formula indexes may not be sufficiently appreciated as a searching tool since they have not, until now, been made available.

* Presented at the 141st National Meeting, American Chemical Society, Divison of Chemical Literature, March 22, 1962, Washington, D. C.

In previous papers, Skolnik[8] and Fletcher[9] have indicated that conventional formula indexes leave much to be desired. They have shown how an "inverted" filing system with carbon and hydrogen given the lowest filing priority can be useful, not only in shortening the search for a specific compound, but also in locating generically related compounds. In the Skolnik system used at the Hercules Powder Co., elements are filed alphabetically with carbon at the end of the formula and hydrogen completely ignored. In this system structural formulas are available on the file cards. At American Cyanamid, the Fletcher system, based on the periodic order of the elements, is used. It has many definite advantages for grouping together structurally related chemicals. Dyson[10] also used a similar method. However, in any of these systems, there is only one file assignment made for each chemical. In order to find a particular compound, it is necessary to anticipate all possible positions where a particular element might occur. For example, if a chemical contains both aluminum and phosphorus, and the chemist is searching for phosphorus compounds, he may not find it under P, but rather under Al.

In previous papers that I have presented before this Division, I discussed the methods employed in preparing the molecular formula indexes to *Index Chemicus*.[11, 12] Briefly, this involves the preparation of a single punched-card for each formula. The cards are converted to magnetic tape and sorted on an electronic computer and printed on a high speed printer for photo-offset reproduction. Recently, over 150,000 formulas were processed in the preparation of our first two-year cumulation. In this cumulation, we completely redesigned the index format to make it easier to use as well as to save space and lower costs. As a completely by-product operation, we have also prepared the first rotated formula index, a *RotaForm Index*. Let me first describe the *RotaForm Index* and then let me illustrate its use with several examples.

In Fig. 1, you see a typical page from the old cumulative molecular formula index to the *Index Chemicus*. As in *Chemical Abstracts*, the Hill system giving priority to carbon and hydrogen is used. All subsequent elements are filed alphabetically. However, a new heading has been created for each carbon–hydrogen combination to speed up locating a desired compound, simultaneously reducing the size of the index by about 25%. This can be seen by examining Fig. 2 which shows the new format. Notice the considerable amount of wasted white space below the carbon–hydrogen headings in the old format. The second major change in the format is the way that the serial numbers of the individual compounds are listed. Whereas each address was listed on one line on the left, they are now on the right of the molecular formula and one or more serial numbers appear after a dotted line. Notice in most cases that the line containing a new molecular formula is always justified on both margins.

As I mentioned before, one card is prepared for each formula in the molecular formula index used now. To prepare the *RotaForm Index*, the computer essentially duplicates the card as many times as there are different elements. If the compound contains, in addition to C and H, five other different elements, the formula will be repeated in the index five times. Thus, C23 H20 Al2 Br3 F4 Na2 P3 (hypothetical example) would be repeated in the *RotaForm Index* five times, once each under the following arrangements:

(1) Al2 Br3 F4 Na2 P3 C23 H2O
(2) Br3 Al2 F4 Na2 P3 C23 H2O
(3) F4 Al2 Br3 Na2 P3 O23 H2O
(4) Na2 Al2 Br3 F4 P3 C23 H2O
(5) P3 Al2 Br3 F4 Na2 C23 H2O

With this type of index, a search for compounds containing any one of these elements is possible in a quick and orderly fashion. Fig. 3 is a page from the new *RotaForm Index*. This page contains all the formulas containing three phosphorus atoms. Since all of the elements are sorted alphabetically, it is a simple matter to locate, for example, all compounds containing three phosphorus atoms and also containing fluorine. These are found by scanning all of the P3 compound listings until P3 Fl2. There cannot be any fluorine compounds beyond this group because of the alphabetical arrangement. Seven such compounds have been checked. A similar search for P4 compounds containing fluorine is also shown. Four such compounds were located.

There are some obvious advantages of the *RotaForm Index*. It will now be quite easy to determine, for example, all reported compounds containing a particular element such as boron. Much more sophisticated generic searches can also be done, as shown in Fig. 4. (Dr. Joe Clark of Lederle Laboratories kindly supplied the questions and the searches were conducted by members of the *Index Chemicus* staff: Mrs. Eleanore H. Peitsch Baus, Assistant Managing Editor, and George F. Corkery, Indexer, and Miss Sandra Goldman, Secretary).

The *RotaForm Index* is a supplementary tool to the conventional formula index. It permits a limited range of generic searches and is particularly valuable for a search involving the less-frequently occurring elements. It is somewhat more cumbersome for searches involving O, N, and S. Even these, however, can be performed with amazing results considering the low cost involved. The *RotaForm Index* is not a panacea for the problems of generic searching. It is, however, one more tool to add to the chemist's armamentarium. It is a by-product of our regular formula indexing, and completely machine-made. It simply advises the reader which molecular formulas contain a particular combination of elements. In order to keep the size down, we have not repeated serial numbers, which can be found in the molecular formula index.

The searcher will save considerable time if he arranges molecular formulas found in the *RotaForm Index* by the number of carbon atoms before looking up the serial numbers. This is strictly a clerical procedure. In fact, any use of the *RotaForm Index* can be assigned to a clerk until the final screening of the structural diagrams.

Keep in mind that the use of this system in the average company chemical file, as is the case in the Hercules and in the American Cyanamid System, each molecular formula card is accompanied by the structural diagram. Such a card filing system eliminates the chore of going from the compound serial number to a register showing the structural diagram itself.

ORGANIC COMPOUNDS

C

6586- 5C 1 AG F3 S
7252- 3 B F6 K
7076-19 BR CL2 N S
6819- 1 CL2 F N O
6819- 3 CL2 F2 O2 S
6819- 2 CL3 F O2 S
7207- 1 CL3 F S
5464- 5 D4 CL2 SI
5464- 3 D4 F2 SI
5464- 4 D6 SI
5023- 6 F2 O
6586- 2 F2 S
6678- 1 F3 GE I3
7049- 1 MN N3 O4
6818- 2C 1H 1 BR F2 SE
6819- 4 CL F N O
6568-16 F2 NA O3 S
7462- 3C 1H 2 CL F2 O P
5464- 1 D3 F SI
6568-28 F2 O3 S
6172-13C 1H 3 BR HG
6172- 1 CL HG
6016- 1 CL2 O P S
7569- 1 CL2 O2 P
7039- 6 CL2 P SE
5464- 6 D F2 SI
6586- 1 F3 S SI
5479- 1 I MN
4683-23 I3 SN
4682- 3 I3 SN
6016- 3 O2 P S
7252- 4C 1H 4 B F6 N
6355- 2 CL2 GE
6863- 2 CL2 GE
7568- 1 CL2 N O P
5748- 7 F O3 P
6863- 1C 1H 5 BR GE
6355- 1 CL GE
5464- 2 F SI
5524- 4C 1H 6 B CL4 N
6355- 4 GE
6973- 7C 1H 7 B BR N
6973- 1 B CL N
6973-11 B I N
6695- 1C 1H15 CO N6 O6
7043- 1C 1H16 B10
7063- 2 B10
7064- 2 B10
6695- 2 CO N7 O8
6524-10C 1H34 A2 CL4 O17
6524-11 CL4 KR2 O17
6524-12 CL4 O17 XE2
6524- 7C 1H35 A2 CL3 O17
6524- 8 CL3 KR2 O17
6524- 9 CL3 O17 XE2
6524- 4C 1H36 A2 CL2 O17
6524- 5 CL2 KR2 O17
6524- 6 CL2 O17 XE2
6340- 4C 2 AS CL2 F3
7871- 2 B CL2 F3
7871- 1 B F5
7252- 2 B2 BA F12
5465- 3 B2 D10
7303-26 BA F2 O5 S
6493- 1 BE F2 K2 O4
6493- 3 BE F2 NA2
6853- 2 BR CL F4 SE
6864- 2 BR D3
7076-18 BR2 CL2 N2 S2
7303-27 CA F2 O5 S
7303- 7 CL F3 O3 S
7303- 8 CL F3 O3 S
7303- 3 CL2 F2 O3 S
7303- 4 CL2 F2 O3 S
7303- 5 CL2 F2 O3 S
7303- 6 CL2 F2 O3 S
7303-19 CL2 F2 O6 S2
6340- 5 CL2 F3 P
6853- 4 CL2 F4 S2
7076-20 CL2 N2 S
7199- 1 CL4 F2
7207- 2 CL4 F2 S2
7814- 2 CL5 F3 SI
5023- 7 F3 N
5023- 9 F3 N O
7543- 1 F3 N O3
5023- 8 F4 O
7303- 1 F4 O3 S
7303- 2 F4 O3 S
7303-14 F4 O3 S
7303-15 F4 O6 S2
6340- 7 F5 P
5023-10 F6 N2
6853- 3 F6 SE2
5023- 2 F7 N
7814- 4C 2H 1 CL2 F3
7814- 1 CL4 F3 SI
6818- 3 F2 N SE
7303- 9 F3 O3 S
7303-10 F3 O3 S
7303-24 F3 O4 S
7303-11 F3 O6 S2
7303-21 F5 O2 S
6335- 3C 2H 2 CL N3 O6
6959-13 CL N3 S
7814- 3 CL2 F2
7814- 5 CL3 F3 SI
7814- 6 CL3 F3 SI
6914-26 CL4 GE
6914-27 CL4 SN
7303-25 F2 O5 S
6818- 1 F4 SE2
6737- 1 N2 O
5910- 1 O S

7736- 2C 2H 3 AL2 I7 O
7868- 2 B CL2
7868- 1 B F2
7736- 1 B I4 O
6864- 1 BR CL D
6335- 4 BR CL N O2
6914- 4 CL3 GE
7356- 1 CL3 O S
7356- 2 F3 O S
6586- 4 F3 S
5803-10 K O S2
4705- 2 N3 S2
5859-68 N4 NA S
5196- 5C 2H 4 BA NA O4 P S
7286- 1 F2 S
7225- 3 N2
6582- 7 N4 O
4626-22 N4 S
4706- 1 N4 S
5859- 1 N4 S
5912- 1 O
7575- 1 O
8954- 1 O3 SE
5465- 1C 2H 5 B2 D5
5465- 2 B2 D5
6997- 1 BA O4 P
6172-14 BR HG
6172- 2 CL HG
7569- 2 CL2 O2 P
7039- 7 CL2 P SE
4682- 6 I3 SN
4683-24 I3 SN
6997- 2 NA2 O4 P
6997- 3 NA3 O7 P2
4683- 9C 2H 6 BR2 SN
7076- 3 CL N S
5529- 1 CL O2 P
6016- 2 CL O2 P S
6016- 4 CL O2 P S
7124- 3 CL2 CU N2 O12
7568- 2 CL2 N O P
4683- 8 CL2 SN
5748- 5 F O3 P
5748- 2 F2 O5 P2
4682- 2 I2 SN
5479- 2 MN
6853- 5 O SE
4838- 9 O SI
7068- 3C 2H 7 B CL3 N
5095- 2 BR O2 SI
6355- 3 CL GE
5095- 1 CL O2 SI
5748- 6 F N O2 P
7122-17 N
5170- 3 N O
4858-27 O2 P S2
4858- 2 O3 P S
5196- 6 O4 P S
5524- 3C 2H 8 B CL4 N
7041- 6 B F4 N
7761- 3 B N
6493- 2 BE F2 N2 O4
6031- 3 BR2 N4 S2 TE
7500- 1 CD N6 S2
6031- 2 CL2 N4 S2 TE
7224- 1 CL6 N6 P4
7500- 4 CO N6 S2
7500- 5 FE N6 S2
6355- 5 GE
7500- 6 MN N6 S2
7306- 1 N O3 P S
4756- 1 N P O3
7500- 3 N6 NI S2
7500- 2 N6 S2 ZN
7895- 3 PB
6973- 9C 2H 9 B BR N
6973- 3 B CL N
6973-13 B I N
7730-23C 2H10 N4
6007-11 O SI2
6122- 1C 2H14 B2 N2
7043- 2C 2H18 B10
7043- 3 B10
7063- 1 B10
7063- 3 B10
7063- 4 B10
7064- 1 B10
6695- 3 CO N7 O8
4821- 6C 3 BR CL2 F2
4819- 3 BR2 K N3 O3
7199- 8 CL F2 N3
6165- 1 CL F6 N O
6165- 5 CL F6 N O
7199- 9 CL2 F N3
6165- 2 CL2 F5 N O
6165- 6 CL2 F5 N O
4819- 2 CL2 K N3 O3
4819- 1 CL2 N3 O3
4821- 5 CL3 F2
7204- 1 CO NA3 O9
7303-41 F3 N O3 S2
7199- 7 F3 N3
7303-12 F6 O3 S
7303-20 F6 O3 S
7303-17 F6 O6 S2
7303-18 F6 O6 S2
7856- 1 F7 N
7856- 2 F7 N
6165- 9 F7 N O
6165-10 F7 N O
7856- 3 F9 N
5023- 5C 3H1 F6 N O

4821- 4C 3H 2 BR CL3 F2
7501- 1 CL N O
7356- 8 CL2 F4 O S
4802- 3 CL3 N3 O2
4779- 8 S3
7892- 3C 3H 3 B3 N6
7892- 2 B3 N6 S3
7614-20 CL4 N O
7303-58 F3 N2 O4 S
7303-30 F3 O4 S
7356- 3 F5 O S
4901- 3 I S2
7769- 1 N O2 S
4804- 4 N3 O4
7122- 7 NA O2
4898- 3C 3H 4 BA N O4 P
5232- 1 CL N O
5366- 4 CL N3 S
6914-23 CL2 F4 SI
6923- 6 CL2 O2 S
7076-17 CL3 N S
6010-34 CL6 HG3 S3
7879- 5 I2
5147- 7 N NA O2
5147- 5 N NA O3
4804- 1 N2 O
6937- 2 N2 O
6937- 3 N2 O
7305- 4 N2 S3
6959- 7 N4 O3 S2
7669- 2 N4 O4 S
6010-31 S3
7868- 6C 3H 5 B F2
6818- 4 BR F2 SE
5576- 1 BR O2
5045- 5 BR2 N O3
6292- 1 CL HG
6111-15 CL O2
6923- 5 CL O5 S2
5826- 1 CL2 N O
7076- 1 CL2 N O S
6923- 7 CL3 O2 S
6323-12 CL3 SI
5147- 1 N O3
7318- 1 N O3
4802- 2 N3 O2
7505- 1 O6 P
4898- 4C 3H 6 BA N O5 P
5956- 1 BR CL O
6871- 1 CL F3 SN
5826- 3 CL N O
7107- 7 CL N O
7614- 7 CL N O
7044- 2 CL2 N RE S2
6177- 1 CL2 SI
6914-22 CL2 SI
6323- 1 CL4 SI
4625- 1 N NA O4 S
7612- 1 N NA O5 S
7581- 3 N2 S2
4706- 4 N4 S
5859- 2 N4 S
7305- 1 N4 S2
7575- 2 O
5954- 2 O3 SE
6923- 8 O6 S2
6923- 9 O7 S2
4665- 4 S3
6328- 4C 3H 7 B BR2
7868- 5 B F2
4961- 1 BA N2 O5 P
6172-15 BR HG
6172- 3 CL HG
6627- 2 CL HG O
7376- 1 CL O2 S
6787- 4 CL O3 S
7569- 3 CL2 O2 P
5845- 2 K O4 S
6484- 1 N O
5826- 2 N O2
6484- 3 N S
6828- 6 O5 P
5326- 1 O6 P
5529- 2C 3H 8 CL O2 P
7568- 3 CL2 N O P
6711-19 CL2 N2
6914-21 CL2 SI
7225- 2 N SI
4802- 1 N4 O2
7305- 3 N4 S3
4705- 1 N6 S3
4838- 8 O SI
6923- 1 O6 S2
6923- 2 O7 S2
4683- 2C 3H 9 BR SN
7250- 1 BR3 N P
6201-22 CL N O P
4850-28 CL SI
4683- 1 CL SN
6166- 2 CL3 N3 P3
5569- 2 CL4 N ZR
7869- 1 F SN
7630- 1 FE O2 S2
4682- 1 I SN
4683- 3 I SN
5479- 3 L1 MN
5170- 4 N O

6970- 2 N3
4804- 7 N3 O4
4858- 9 O2 P S
5530- 2 O3 P
5824- 1 O3 P
6063-10 O4 P SI
6973-16C 3H10 B BR2 N
5524- 2 B CL4 N
7041- 1 B F4 N
7041-13 CL N O4
7041-16 F6 N P
6135- 1 N3 O4 P
4858-44 NA P S2
4670-15 O S
6275- 1 O2 SI
7895- 1 PB
6973-10C 3H11 B BR N
6973- 4 B CL N
6973-14 B I N
6449- 1 N2 O3 P S
7137- 9C 3H12 CL6 N4 RE2
6986- 3 F N SI
7630- 2 FE N O2 S2
5569-11C 3H13 CL2 N3 ZR
6007-12C 3H14 O2 SI3
6696- 2C 3H20 B10
7043- 4 B10
7043- 5 B10
6695- 4 CO N7 O8
4850-29C 3H22 O2 SI3
6524- 1C 3H40 A2 O18
6524- 2 KR2 O18
6524- 3 O18 XE2
7871- 3C 4 B CL F6
6853- 1 BR2 F8 SE2
6340- 6 CL F6 P
7599- 4 CL4 O S2
7599- 2 CL4 S2
7599- 5 CL7 S2
7599- 1 CL8 S2
7599- 3 CL8 S2
5472- 5 CO HG N4 S2
5472- 4 CO K2 N4 S4
6864- 5 D7 I
6864- 6 D8
6654- 2 F4 N2
7480- 1 F6
7543- 2 F7 N O3
6340- 8 F7 P
6165- 3 F9 N O
6165- 4 F9 N O
6165- 7 F9 N O
6165- 8 F9 N O
5023- 4 F10 N2
7246- 1 F12 HG O2 P2
5023- 3 F12 N2
6474-16 K3 N3 O3 S
6474-12 K3 N3 O4
86474- 8 N3 NA3 O3 S
6474- 4 N3 NA3 O4
7431- 1C 4H 1 BR N2 O4 S
7199- 4 CL F2 N2
7199- 5 CL2 F N2
7199- 6 CL2 F N2
6654- 1 F3 N2
6474-15 K2 N3 O3 S
6474-11 K2 N3 O4
6474- 7 N3 NA2 O3 S
6474- 3 N3 NA2 O4
6474-28C 4H 2 AG N3 O4
4793- 7 B CL3 F6 O2
7742- 1 BR CL2 N3
7742- 3 BR CL2 N3
7501- 5 CL N O3
7051- 2 CL N3 O4
7303-36 CL3 F3 O4 S
4839- 3 CL5 F3 SI
7599- 6 CL8 S2
5378- 1 D4 O4
7199- 2 F2 N2
7199- 3 F2 N2
5007- 5 F6 N4
6474-14 K N3 O3 S
6474-10 K N3 O4
7753- 1 K N3 O5
4864-24 N2 O
6474- 6 N3 NA O3 S
6474- 2 N3 NA O4
4901- 2 O2 S3
5352- 1C 4H 3 BR F3 N O
7437- 4 BR N2 O
4631- 1 BR N2 S
5352-15 BR2 F3 O
6364- 2 BR2 N3
6711-17 CL N2
7051- 1 CL N2 O2
4999- 2 CL N4 O2
6940- 2 CL N4 O2
7079- 1 CL N4 O3
5879-83 CL O2
4839-13 CL3 F3 I SI
4839- 1 CL4 F3 SI
6864- 8 D5
7356- 4 F7 O S
4999- 1 N3 O4
6523- 7 N3 O4
6940- 1 N3 O4
7310- 7 N5 S

5352-17C 4H 4 AG F3 O4
5352-14 BR F3 O
5352- 6 BR F3 O3
7141- 1 BR N S
6364- 3 BR N3 O
7501- 4 CL N O2
4631- 7 CL N3 S
5695- 1 CL2 HG O4
7614-17 CL3 N O2
7599- 7 CL6 S2
5626- 3 CU2 I2 N2
6378- 2 D2 O4
5352- 2 F3 N O2
5007- 1 F7 N3
6933- 2 N2 O
7437- 3 N2 O
7824- 6 N2 O
4864-25 N2 O
4864-23 N2 O2
6523- 4 N2 O3
6010-25 N2 S
6680- 1 N4
6523- 8 N4 O3
6523- 9 N4 O3
5395-19 N4 S
6465- 1 O
4847- 1 O2
7046- 1 O2
4779- 7 S3
4898- 1C 4H 5 BA F N O4 P
4898- 2 BA F N O4 P
5352- 4 BR F3 N O2
6364- 7 BR N4
4680- 1 BR O
5352-18 BR2 F3 O2
6662- 3 CL
7758- 2 CL F2 O2
7090- 4 CL N2
6367- 9 CL N4
5278- 2 CL O
5232- 4 CL O2
6033- 1 CL O2 S
7758- 1 CL2 F O2
7614-18 CL4 N O
4980- 5 CL5 S
6864- 7 D3
5022- 1 F3 O4
5352- 9 F3 O4
7830-10 F3 O4
7830-11 F3 O4
5216- 1 N
7255-18 N O
6111-14 N O2
5216- 2 N S
5216- 5 N S
6364- 1 N3
6523- 5 N3 O2
6523- 6 N3 O2
7669- 1 N3 O2 S
6523-10 N3 O3
4804- 5 N3 O4
7669- 4 N3 O5 S
6523- 2 N5 O2
7868- 3C 4H 6 B CL
7868- 4 B F
5352-16 BR F3 O2
5812- 3 BR2 O3
5480- 2 BR2 O3 S
5480- 3 CL I O3 S
5052- 2 CL N5
5243- 1 CL3 O4 P
5243- 9 CL3 O4 P
5243- 6 CL5 O4 P
6864- 3 D I
6864- 4 D2
7209-10 F2 N2
5352- 5 F3 N O3
5352- 8 F3 N O3
5972- 7 I N O S
5147- 6C 4H 6 N NA O3
4804- 2 N2 O
6697- 1 N2 O
7025-20 N2 O2
7255-16 N2 O2
5283-10 N2 O2 S2
5283- 2 N2 O3 S
6364- 8 N4
7730-18 N4
6523- 1 N4 O
7159- 1 N4 O S
6523-11 N4 O2
6523-12 N4 O2
6959- 8 N4 O3 S2
7669- 3 N4 O4 S
7669- 5 N4 O4 S
6967- 2 N6 O2
5859-51 N8 S2
5803- 1 NI O2 S4
6712- 1 O
4668-10 O2
5058- 3 O2 S
5480- 1 O3 S
6033- 6 O3 S
5910- 2 O3 S2
6010-18 S3

Fig. 1.—Old format of cumulative molecular formula index

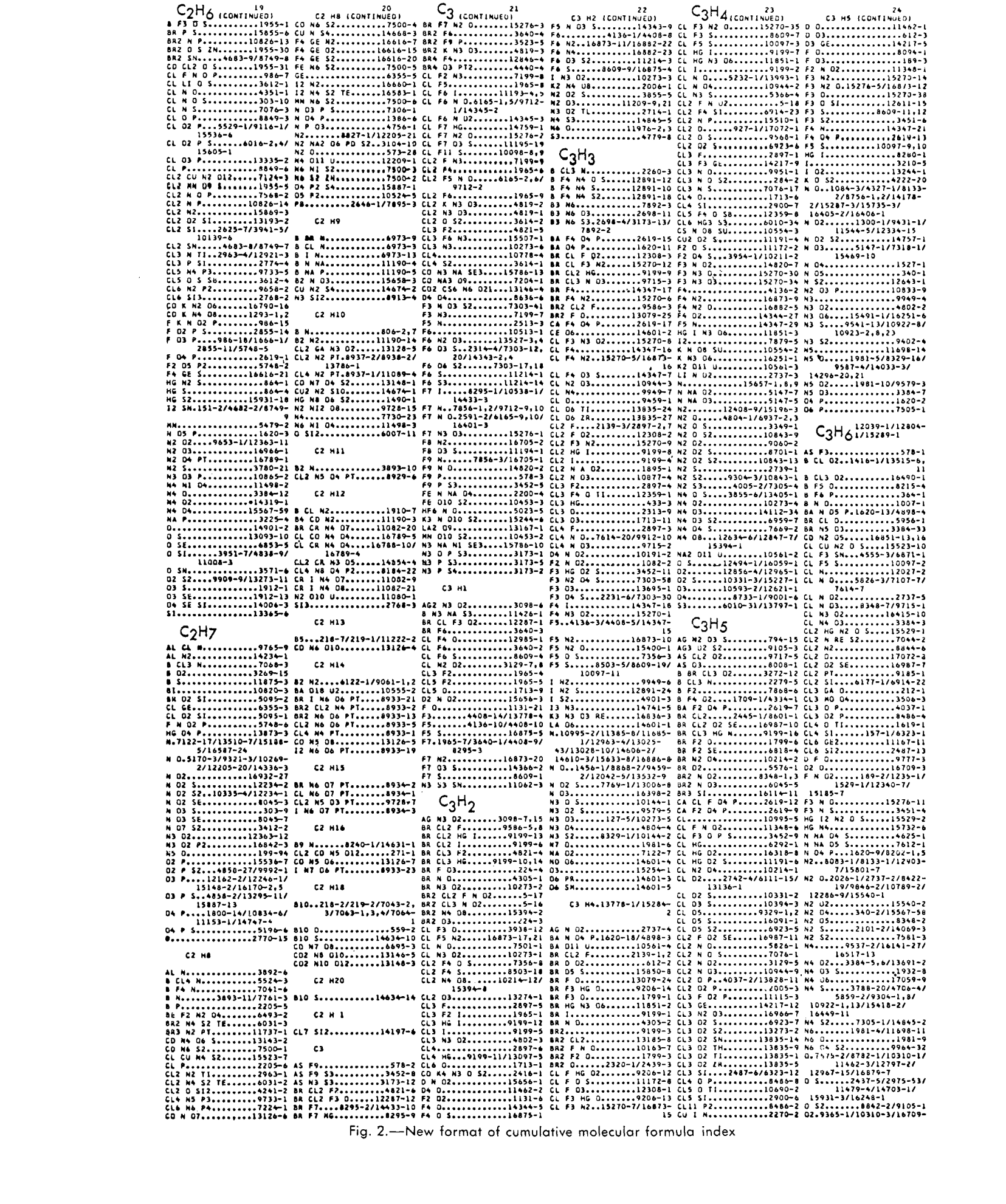


Fig. 2.—New format of cumulative molecular formula index

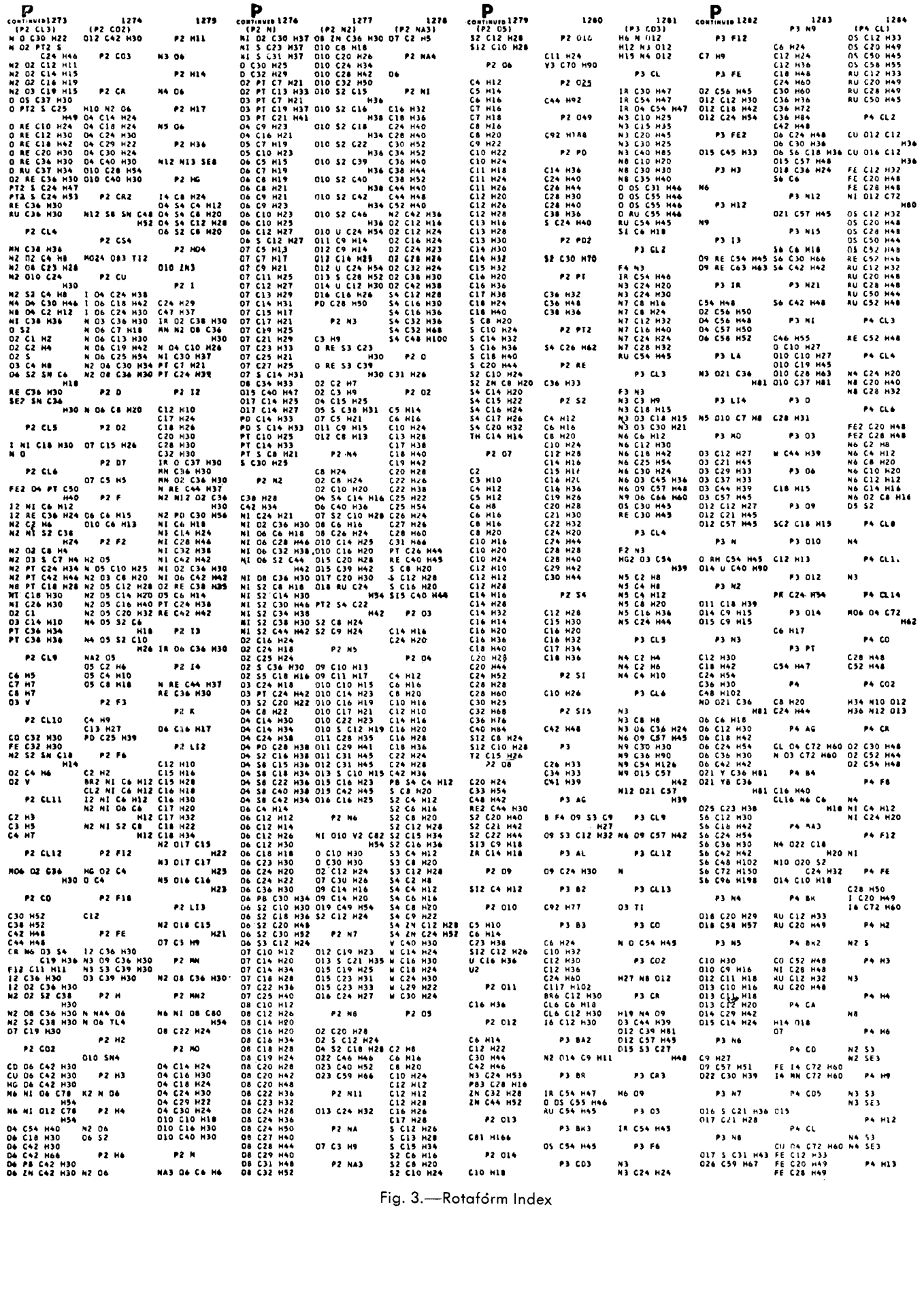
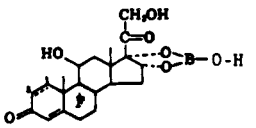


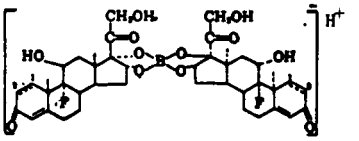
Fig. 3.—Rotaform Index

Search 1: Borate derivatives of steriods. Under B headings examine all compounds containing 19 or more carbon atoms and 3 or more oxygen atoms.

I.C. 8591 STEROID BORATES I. STRUCTURAL CONSIDERATIONS.
L.J. Leeson, J.A. Lowery, G.M. Sieger, S. Muller.
American Cyanamid Co., Pearl River, N.Y. Recd. Mar. 16, 1960.
J. Pharm. Sci. 50(3),193-7(1961).



IIa—Carbons 1,2: CH_2—CH_2
IIb—Carbons 1,2: CH=CH

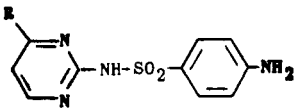


IIIa—Carbons 1,2: CH_2—CH_2 (F-3 Borates)
IIIb—Carbons 1,2: CH=CH (F-4 Borates)

1)	C21 H28 B F O7	(IIa)	(as Na salt)
2)	C42 H55 B F2 O12	(IIIa)	"
3)	C21 H26 B F O7	(IIb)	(as NH_4 salt)
4)	C42 H51 B F2 O12	(IIIb)	"

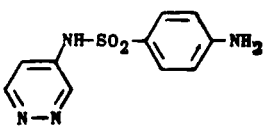
Search 2: Sulfanilamidodiazines (diazine = 6-membered ring containing two nitrogen atoms).

a. With or without alkyl substituents. Under SN_4O_2 search $C_{10}H_{10}$. For alkyl substituents search $C_{10} + {}_nH_{10} + 2n$.



		R=
20)	C11 H12 N4 O2 S	Me
21)	C12 H14 N4 O2 S	Et
22)	C13 H16 N4 O2 S	Pr
23)	C14 H18 N4 O2 S	i-Bu
24)	C15 H20 N4 O2 S	n-amyl

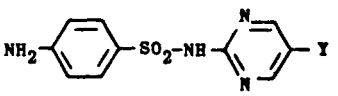
I.C. 4345



m.f. = C10 H10 N4 O2 S

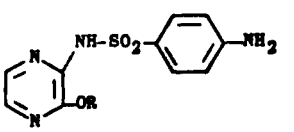
I.C. 10610-3

b. If ether substituents are desired, under SN_4O_3 search $C_{11} + {}_nH_{12} + 2n$.



		Y=
1)	C11 H12 N4 O3 S	MeO
2)	C12 H14 N4 O3 S	EtO
3)	C13 H16 N4 O3 S	PrO
4)	C14 H18 N4 O3 S	BuO
5)	"	sec-BuO

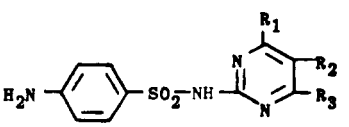
I.C. 10035



		R=
5)	C11 H12 N4 O3 S	Me
7)	C12 H14 N4 O3 S	Et

I.C. 6360

c. For diethers, under SN_4O_4, search $C_{12} + {}_nH_{14} + 2n$.



		R_1=	R_2=	R_3=
1)	C12 H14 N4 O4 S	MeO	MeO	H
2)	C13 H16 N4 O4 S	EtO	"	"
12)	"	MeO	Me	MeO
13)	C15 H20 N4 O4 S	EtO	"	EtO

I.C. 12534

Fig. 4.—Two searches using the RotaForm Index

The user must in each case decide which is less costly—additional filing space and card processing (repeating structural diagram under each element) or additional clerical time in looking up diagrams from a serial number. In the former case one also has to consider cost of refiling cards if they are to be shown to chemist or cost of reproducing them. In the latter case it is the time to write down serial numbers.

REFERENCES

(1) E. Ronwin, *J. Am. Chem. Soc.*, **73**, 5141 (1951).
(2) L. Pauling and V. Schomaker, *J. Am. Chem. Soc.* **74**, 1111 (1952).
(3) L. Pauling and V. Schomaker, *J. Am. Chem. Soc.*, **74**, 3172 (1952).
(4) L. Anschutz, *Annalen*, **454**, 71 (1927).
(5) E. Garfield, private communication, September 1, 1954.
(6) E. Ronwin, private communication, September 24, 1954.
(7) L. Pauling, private communication, August 4, 1962.
(8) H. Skolnik and J. K. Hopkins, *J. Chem. Educ.* **35**, 150 (1958).
(9) J. Fletcher, *Chem. Eng. News*, **34**, 5888 (1956).
(10) G. M. Dyson, *Chem. Ind.* (London), 676 (1952).
(11) E. Garfield, *Index Chemicus*, First Cumulative Author Index, pp. 1, 33 (1961).
(12) E. Garfield, "Molecular Formula Indexing with Computers," presented at the 139th National Meeting, American Chemical Society, Division of Chemical Literature, March 21, 1961, St. Louis, Missouri. See abstracts, p. 1-Q.

A Unified Method of Delineating Polymeric Species*

By K. J. LISSANT
Tretolite Company, St. Louis 19, Missouri

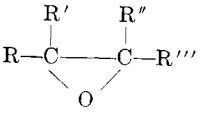
Received January 8, 1963

Much work is currently being done in the field of data handling as applied to chemistry. There is obviously a great need for methods of organizing, manipulating, storing, and retrieving chemical data. One of the areas in which many chemical data are being generated and which is particularly difficult to systematize is the field of polymer chemistry. The problem of an unambiguous method for encoding chemical formulas for machine handling is far from solved. The problem of delineating individual members of a polymeric series has hardly been considered.

The need for a method of organizing and delineating polymeric species is particularly apparent to one who reads the patent literature. Hundreds of patents are being issued on uses for, methods of making, or compositions of matter involving polymers. It is becoming increasingly difficult to determine the scope of claims, or the novelty of particular compositions. This paper presents a method of defining polymeric species, and of differentiating between species and between individual members of a species. The method has been applied with particular success in the field of oxyalkylates but is sufficiently general to be applicable to many if not most fields of polymers.

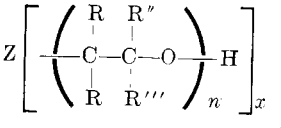
In the mathematical discussion of the development and use of this method, specific examples will be taken almost solely from the alkylene oxide polymers and for this reason a brief discussion of the chemistry of this class is given.

Alkylene oxides have the general formula



where R, R′, R″, and R‴ may be, for example, hydrogen, an aliphatic radical, a cycloaliphatic radical, an aryl radical, etc. The R's may also be joined to form a cyclic structure. In cases where one or more of the R's contain an epoxide group, a diepoxide or a polyepoxide results. For the purposes of simplicity this discussion will consider only the materials with one epoxide group.

Alkylene oxides react with active centers of other organic or inorganic molecules to build up polyether chains of considerable length. The generic reaction product may be written:



where n is the number of monomer units in the chain and x is the number of reactive sites in the starting molecule.

The relative reactivity of the chain terminal groups and of the original reactive sites with respect to the monomer determines the positions and relative lengths of

*Presented before the Division of Chemical Literature, ACS National Meeting, Cincinnati, Ohio, January 14, 1963