## 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. introducing a new structural formula for both ribo- and desoxyribonucleic acids. The structure had as its core a  $(P_2O_5)_{\pi}$  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?: "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 phosphotri-anhydride formula for the nucleic acids deserves no serious consideration."

However, in the July, 1952, issue, 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 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<sup>5</sup> and asked him how he had managed to find the necessary precedent. Dr. Ronwin replied<sup>6</sup> 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.

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.

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

In previous papers, Skolnik<sup>8</sup> and Fletcher<sup>9</sup> 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<sup>10</sup> 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*. <sup>11. 12</sup> 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·   |  | C <sub>1</sub> H <sub>1</sub>  |  | $C_4H_4$               |
|--|--|--|--|------------------------|
| C.  6586-5C 1  7252-3  8619-1  CL2FR N O  6619-3  CL2FR O  6619-3  CL2FR O  6619-3  CL2FR O  6619-3  CL2FR O  6619-2  CL3F O  5640-3  DA CL2 SI  5640-3  DA CL2 SI  5640-4  DA F2 SI  5640-4  DA F2 SI  5640-4  DA F2 SI  5640-8  F2 O  6618-1  F3 G E I3  7049-1  MN N3 O4  6618-2  CL FN O  6681-4  CL FN O  6681-6  F2 NA O3 S  7462-3C 1H 2 GL F2 O  F3 S SI  5640-1  CL FN O  66172-13C 1H 3 BR HG  6172-1  6172-1  6172-1  6168-1  F3 S SI  5479-1  I MN  4683-23  I 3 SN  4682-3  I 3 SN  4683-2  I 3 SN  6635-2  CL2 GE  6756-1  F5 G  F | 7500- 6 MN N6 S2<br>7306- 1 N 03 P S<br>4756- 1 N P 03<br>7500- 3 N6 N1 S2<br>7500- 2 N6 S2 ZN<br>7895- 3 PB<br>6973- 9C 2M 9 B BR N | C.H.  4821- 4C 3H 2 BR CL3 F2 7501-1 CL N 0 7356-8 4802-3 CL3 N3 02 4779-8 53 7892-3C 3H 3 B3 N6 7892-2 7614-20 7303-58 F3 N2 O4 S 7303-30 F3 O4 S 7303-30 F3 O4 S 7356-3 F5 O S 4901-3 I S2 7769-1 N 02 S 4804-4 N3 O4 7122-7 4898-3C 3H 48 N 04 P 2232-1 CL3 N3 S 6914-23 CL2 F4 SI 6923-6 CL2 F4 SI 6923-6 CL2 F4 SI 6923-6 CL2 F4 SI 6923-6 CL2 F4 SI 6923-7 NN 02 6937-2 NN 03 6939-7 NN 03 7305-4 NN 03 7305-4 NN 03 7305-4 NN 03 7305-4 NN 03 6939-7 NA 03 7305-4 NN 03 7305-4 NN 03 7305-4 NN 03 7305-4 NN 03 7305-5 BR C2 6818-4 BR C2 6937-3 N2 O 6937-2 N2 O 6937-2 CL3 NS 60709-1 CL2 NO 6937-1 NN 03 7305-4 NN 03 7305-1 NN 03 7305-1 NN 03 7318-1 NN | 6970-2 4804-7 13 4804-7 13 4804-7 13 4858-9 02 23 5530-2 03 5524-1 03 6973-16C 5524-2 1 8524-1 8524-2 8547-1 8547- | 5352-17C 4H            |
| 6816-1 F4 SE2<br>6737-1 N2 O<br>5910-1 O S   | 5023- 5C 3H1 F6 N O  | Fig. 1.—Old  | format of cumulative m   | olecular formula index |
|  |  |  |  |                        |

| C.11 10  | 20  | C 11                                      |  |
|--|---|---|--|
| C2H6 (CONTINUED)                               | 63 U.S. 150UTILUUSSI                            | C3 (CONTINUED)                            | C3 H2 (CONTINUED) C3H4 (CONTINUED) C3 H5 (CONTINUED)   |
|  |   |   | F3 N U3 3 *******************************  |
|  |   |   |  |
| BR2 SN4683-9/8749-8                            | F4 GE \$216616-20                               | BR4 F412846-4                             | Fe N2.168/3-11/1685/-22 CL F5 S. 1009/-3 D3 GC. 1421/-5<br>F6 N4. 1686/-23 CL MG I. 19199-7 F U. 8094-1<br>F6 O3 S2. 11214-3 CL MG N3 06. 11851-1 F D3. 169-3<br>F6 S. 8609-9/16875-4 CL I. 19199-2 F2 N OZ. 11346-1 |
| CL F N O P 986-7                               | GE  | CI 62 N3                                  | 74 S   |
| CL N 04351-1                                   | 12 N4 S2 TE16583-1                              | CL F6 [11193-6.5                          | KZ N4 U82006-1 CL N U410944-2 F3 NZ U.15276-5/168/3-12   |
|  |   |   |  |
| CL O P   | N 04 P  | CL F6 N D214345-3                         | 72 U2 - 12 - 12 - 12 - 12 - 12 - 13 - 12 - 13 - 13   |
|  |   |   | M6 011976-2,3 CL2 D927-1/17072-1 F4 N14347-21<br>\$3956-1 F4 O4 F444444413   |
| CL OZ P S6016-2,4/                             | NZ MAZ 06 PD \$23104-19                         | CL F7 03 S11195-19<br>CL F11 S10098-8,9   | CL2 02 Samman 4973m A F5 C   |
| EL 03 P 13335-7                                | MA 011 Hamana 12209-1                           | C1 2 5 M3                                 |  |
|  |   |   | 8 CL3 N 0  |
| CL2 M O P                                      | 05 F2   | CL2 F61965-9                              | 8 F4 N4 5212891-10 CL3 N 5   |
| CLZ N P  | PB  | CLZ K N3 U34819-2                         | 83 N6  |
| CL2 02 \$113193-2                              | CZ H9   | CL2 0 \$23614-2                           | B3 N6 S3-2698-4/3173-13/ CL6 HG3 S36010-34 N D21300-1/9431-1/  |
| CL2 SI2625-7/3941-5/<br>10139-6                | B 88 86973-9                                    | CL3 F24821-5<br>CL3 F6 N315507-1          | #A F4 04 P2619-15 CU2 D2 S11191-4 N D2 S214757-1   |
| CL3 N T12963-4/12921-3                         | B I N   | CL4                                       | BA 04 Parantinininininininininininininininininini  |
| GL3 P SI2774-4                                 | 8 N NA  | CL4 S2                                    | BR CL F3 M215270-12 F3 N 0214820-7 N 041527-1 BR CL2 MG9199-9 F3 N3 015270-30 N 05340-1  |
| CL5 O S Sb3612-4                               | 82 N 0315658-3                                  | CO NA3 097204-1                           | BR CL3 N D39715-3 F3 N3 D315270-34 N \$2   |
| CL6 N2 P29658-2<br>CL6 S132768-2               | CU NZ \$414674-Z<br>N3 \$128913-4               | D4 D4                                     | BR F414347-17 F4   |
| CO K N2 0616790-16                             |   | F3 M 03 \$27303-41                        | BR2 CL2 F9586-3 F4 M2 D  |
| F K N 02 P986-15                               |   | F5 N2513-3                                | CA F4 Q4 P2619-17 F5 N   |
| F 03 P986-18/1666-1/                           | 82 H211190-14                                   | F6 N2 0313527-3.4                         | CE D614601-Z HG I N3 D61851-3 10923-Z,8,23<br>CL F3 N3 D215270-8 I2  |
| 2855-11/5748-5                                 | CL2 GA N3 0213128-5<br>CL2 N2 PT.8937-2/8938-2/ | F6 03 S2314-4/7303-12,                    | CL F416347-16 K N 08 SU10554-2 N511698-14<br>CL F4 N215270-5/16873- K N3 0616251-1 N5 01981-5/8329-16/   |
| F2 D5 P25748-2                                 | 13786-1   | F6 04 527303-17.18                        | . 16 K2 D11 Unananana 10561-3 9587-6/16033-3/  |
| HG N2 S864-1                                   | CO N7 O4 S213148-1                              | F6 53                                     | CL F4 03 \$14347-7 LI N U22737-3 14296-20,21 CL N2 0310944-3 N15657-1,8,9 N5 021981-10/9579-3  |
| MG 5215931-18                                  | HG N8 D6 S21490-1                               | 14433~3                                   | CL N49949-7 N NA 025147-7 N5 03  |
| 12 SM.151-2/4482-2/8749-                       | NZ NIZ 089728-15                                | F7 N 7856-1.2/9712-9.10                   | CL D6 TI13835-24 NZ12408-9/15196-3 O6 P  |
| MM5479-2                                       | N6 N1 0411498-3                                 | 16401-3                                   | CL2 F2139-3/2897-2,7 NZ O S3349-1  12039-1/12804-  |
| N2 029653-1/12363-11                           |   | F8 N216705-2                              | LLZ P3 NZ13270-Y NZ UZ9080-2   |
| NZ 0316966-1<br>NZ D4 PT16789-1                |   | F8 D3 511194-1                            | CL2 HG I9199-8 HZ OZ S8701-1 AS F3578-1 CL2 I9199-4 NZ OZ SZ10843-13 B CL OZ1416-1/13515-6,  |
| M2 S   | B2 N3893-10                                     | F9 N D14820-2                             | CL2 N A 021895-1 N2 52739-1 11 CL2 N 0310877-4 NZ 529304-3/10843-1 8 CL3 0216490-1   |
| M4 M1 D411498-2                                |   | F9 P \$33452-5                            | CL3 F22897-4 N2 534005-2/7305-4 B F5 08215-4   |
| N4 D   | CS H13  | FE 010 5210453-3                          | CL3 F4 0 TI12359-1 N4 0 S3855-6/13405-1 B F6 P364-1 CL3 HG433-3 N4 02  |
| N4 D4  | B CL N21910-7                                   | HF6 N G                                   | CL3 03   |
| U 19901~2                                      | BK CK NO U711082-20                             | LAZ UY                                    | CL4 Farances 2897-3 N4 D4 Second 7669-2 RR N5 D3 3384-33   |
| 0 S  | CL CO M4 D416789-5<br>CL CR M4 D416788-10/      | MN 010 \$210453-2<br>N3 NA N1 SE315786-10 | CL4 N D7614-20/9912-10 N4 OB12634-6/12847-7/ CD N2 O516851-13,16 CL4 N O39715-2 15394-1 CL CU N2 O S15523-10   |
| 0 \$13951-7/483#-9/<br>1100#-3                 | 16789-4   | M3 D P 533173-1                           | D4 N D210191-Z NA2 011 U10561-Z CL F3 SN4555-3/6671-1 F2 N D2  |
| O SN   | CL4 NB D4 P28184-22                             | N3 P 543173-2                             | F3 HG 02 S3452-11 0212856-4/12965-1 CL N12027-2  |
|  | CR I N4 0711082-9<br>CR I N4 0811082-21         | С3 н1                                     | F3 N2 O4 S7303-58 O2 S10331-3/15227-1 CL N O5826-3/7107-7/<br>F3 O313695-1 O310593-2/12621-1 7614-7  |
| 03 SE1912-13                                   | N2 010 U11080-1                                 | AGZ N3 023098-4                           | F3 04 S2231-6/7303-30 048733-1/9001-6 CL N 022737-5 F4 [   |
| \$113365-6                                     |   | 8 N3 NA \$311426-1                        | F4 N3 0215270-1 C1 N3 0216415-10   |
| C2H7   | CS H13  | BR F6                                     | F54136-3/4408-5/14347- C3H5 CL N4 03   |
|  | B5218-7/219-1/11222-2<br>CD N6 01013126-4       | CL F6                                     | F5 N216873-10 AG N2 D3 S794-15 CLZ N RE 527044-2<br>F5 N2 O15400-1 AG3 U2 S29105-3 CLZ N2  |
| AL M214234-1<br>B CL3 M7068-3                  | C2 H14  | CL F6 S8609-4                             | F5 0 S7356-3 AS CL2 029717-5 CL2 017072-3<br>F5 S8503-5/8609-19/ AS 038008-1 CL2 02 SE16987-7  |
| # 023269~15                                    |   | CL3 F21965-4                              | 10097-11 B BR CL3 D23272-12 CL2 PT9185-1   |
| 81   | BA 018 U210555-2                                | CL5 0                                     | I NZ S12891-24 8 F27868-6 CL3 GA O212-1  |
| BR D2 \$15095-2                                | BR 1 N6 06 PT8933-21<br>BR2 CL2 N4 PT8933-2     | D2 N 0215656-3                            | I \$24901-3 8 F4 021709-1/4334-1 CL3 M0 043506-3 13 M314741-5 BA F2 04 P2619-7 CL3 0 P4037-1   |
| CL 02 SI5095-1                                 | BR2 N6 D6 PT8933-13                             | F34408-14/13778-4                         | K3 N3 D3 RE16836-3 BR CL22445-1/8601-1 CL3 D2 P8486-4  |
| HG 04 P13873-3                                 | CL4 N4 PT6933-1                                 | F5 S16875-5                               | LA 0614601-1 BR CLZ 02 SE16987-10 CL4 0 TI1619-1 M.10995-2/11385-8/11685- BR CL3 HG M9199-16 CL4 SI157-1/6323-1  |
| M.7122-17/13510-7/15186-<br>5/16587-24         | CO N5 OB  |   | 1/12963-4/13025- BR F2 01799-6 CL6 GE211167-11<br>43/13028-10/14606-2/ BR F2 SE6818-4 CL6 SI22487-13   |
| N 0.5170-3/9321-3/10269-<br>2/12205-20/14336-3 |   | F7 N216873-20                             | 14610-3/15633-8/16886-9 BR N2 0410214-2 D F G  |
| M 0216932-27                                   |   | F7 \$8609-1                               | 2/12042-5/13532-9 BR2 N 028348-1,3 F N 02189-2/1235-1/   |
| N 02 S210335-4/12234-1                         | CL N6 07 PT8934-1                               | ·   | N D2 S7769-1/13006-8 BR2 N D36045-5 1529-1/12340-7/ N D3   |
| N 02 SE  | CL2 M5 03 PT9728+7<br>1 M6 07 PT8934-3          | С <sub>3</sub> Н <sub>2</sub>             | N3 D S   |
| N 03 SE8045-7<br>N 07 S23412-2                 |   | AG N3 023098-7,15                         | N3 03127-5/10273-5 CL  |
| N3 0212363-12                                  | •   | BR CL2 HG I9199-13                        | N3 S28329-1/10144-2 CL F3 O P S3452-9 N NA D4 S4625-1  |
| H5 D   | CL2 CO M5 012271-1                              | BR CL3 F24821-4                           | N7 D1981-6 CL HG6292-1 N NA D5 S7612-1 NA D27122-7 CL HG D216318-8 N D6 P1620-9/8202-1,5   |
| 02 P15536~7                                    | CO M5 0613126-7                                 | BR CL3 HG9199-10,14                       | NO 0614601-4 CL HG 02 S11191-6 N2.8083-1/8133-1/12403-<br>0315254-1 CL N2 0410214-1 7/15801-7  |
| 03 P12162-2/12246-1/                           |   | BR N D4305-1                              | D6 PR14601-3 CL D22742-4/6111-15/ N2 0.2026-1/2737-2/8422-   |
| 15148-2/16170-2,5<br>03 P S4858-2/13295-11/    | CZ H18  | BR2 CL2 F N 025-17                        | 04 5M14601-5 13136-1 19/9846-2/10789-2/<br>CL 02 510331-2 12286-9/15540-1  |
| 15887-13<br>D4 P1800-14/10834-6/               | 810218-2/219-2/7043-2,                          | BRZ CL3 M 025-16<br>BRZ N4 0815394-2      | C3 H4.13778-1/15284- CL O3 S10394-3 NZ UZ15540-Z<br>2 CL O59329-1,2 NZ 04340-2/15567-58  |
| 11153-1/14747-4                                | 1   | BR2 03                                    | CL 05 S16091-1 NZ 05   |
| e2770-15                                       | B10 S   | CL F5 N216873-17,21                       | BA N 04 P.1620-18/4898-3 CL2 F 02 SE16987-11 NZ SZ   |
| C2 H8  | CD2 N8 01013146-5                               | CL N3 0210273-1                           | BA DII U10561-4 CL2 N O5826-1 N49537-2/16141-27/<br>BR CL2 F2139-1,2 CL2 N O S7076-1 16517-13  |
| AL N   | CO2 N10 01213148-3                              | CL2 F4 0 S                                | BR D 02612-2 CL2 N 023129-5 N4 023384-5,6/13691-2 BR D5 S15850-8 CL2 N 0310944-9 N4 03 S   |
| 8 CL4 H5524-3                                  | CZ HZO  | CL2 N4 O810214-12/                        | BR F 013079-24 CL2 O P4037-2/13828-11 N4 U6  |
|  | B10 S14634-14                                   | 15394-8<br>CL2 0313274-1                  | BR F3 01799-1 CL3 F D2 P11115-3 5859-2/9304-1.8/   |
| B P2205-5<br>BE F2 N2 046493-2                 | C2 H 1  | CL3 F                                     | BR HG N3 0611851-2 CL3 GE14217-12 10922-1,13/15418-2/<br>8R 19199-1 CL3 N2 0316966-7 16449-11  |
| BR2 N4 S2 TE6031-3                             |   | CL3 HG 19199-12                           | BR N D4305-2 CL3 02 S6923-7 N4 S27305-1/14845-2<br>BR29199-3 CL3 02 S213273-2 N61981-4/11698-11  |
| CD N4 06 S                                     |   | CL3 N3 D248D2-3                           | BR2 CL213185-8 CL3 OZ SN13835-14 N6 O1981-9  |
| CD N6 S27500-1<br>CL CU N4 S215523-7           | C3  | CL4 HG9199-11/13097-3                     | BR2 F N 010163-7 CL3 02 TH13635-9 N6 C4 52964-32 BR2 F2 01799-3 CL3 02 F113835-1 Q.7575-2/6782-1/10310-1/  |
| CL P   | AS F9 \$33452-8                                 | CL6 0                                     | BRZ 02320-1/2439-3 CL3 OZ ZR13835-5 11462-3/12797-2/<br>CL F HG 029206-12 CL3 SI2487-6/6323-12 12967-15/16879-7  |
| CL2 N4 S2 TE6031-2                             | AS N3 \$33173-12                                | D N 02                                    | CL F 0 S11172-8 CL4 0 P8485-8 0 S2437-5/2975-53/<br>CL F 0312308-1 CL5 0 Ti10690-2 11479-4/14703-1/  |
| CI 6 N5 P3                                     | BR CL2 F3 D12287-12                             | F2 D2                                     | CL F3 HG Q9206-13 CL5 SI2900-6 15931-3/16248-1   |
| CO N 0713126-6                                 | BR F78295-2/14433-10<br>BR F7 MG8295-9          | F4 0 S16875-1                             | CL F3 M215270-7/16873- CL11 P26486-2 0 S28842-2/9105-1<br>15 CU I M2270-2 02.9365-1/10310-3/16709-   |

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

| P   |                             | P   |   |   | Р  |  |  | Р   |  |  |
|---|-----------------------------|---|---|---|--|--|--|---|--|--|
| CONTINUED1273 1274<br>(P2 CL3) (P2 CO2)<br>H O C30 H22 O12 C42 H30 P                      | 2 H11 N                     | 41 DZ C30 H37 C                                       |   | (P2 NA3)                                    | CONTINUED 1279<br>(P2 05)<br>S2 C12 H28  | 1260<br>P2 016                           | (P3 GD3)<br>H6 N O12                         | CONTINUED 1282<br>P3 F12                  | 1283<br>P3 N9                          | 1284<br>(P4 CL)<br>OS C12 H33          |
| N 02 PT2 S<br>C24 H46 P2 C03 N3 O<br>N2 G2 C12 H11  | 36" H                       |   | 310 C8 H18<br>310 C20 H26<br>310 C24 H34      | P2 MA4                                      | \$12 C10 H28                             | C11 H24<br>V3 C70 H90                    | H12 N3 012<br>H15 N4 012                     | C7 H9                                     | C6 H24<br>C12 H24<br>C12 H36           | DS C20 H49<br>DS C50 H45<br>DS C58 H55 |
| H2 02 C14 H15 P   | 2 H14 D                     | C32 H29 C   | 010 C28 H42                                   | 06  | C4 H12                                   | P2 025                                   | P3 CL  | P3 FE                                     | C18 H48<br>C24 H60                     | RU C12 H33<br>RU C20 H49               |
|   | 2 H17 0                     | 02 PT C13 H33 0<br>03 PT C7 H21<br>03 PT C19 H37 0    | 10 SZ C16                                     | PZ MI<br>C16 H32                            | C5 H14<br>C6 H16<br>C7 H16               | C44 H92                                  | IR C30 H47<br>IR C54 H47<br>IR O4 C54 H47    | 02 C56 H45<br>012 C12 H30<br>012 C18 H42  | C36 H36<br>C36 H72                     | RU C28 H49<br>RU C50 H45               |
| H49 04 C14 H24<br>O RE C10 H24 04 C18 H24 N5 G<br>O RE C12 H30 04 C24 H30                 | 3 <b>6</b> 0                | 03 PT CZ1 H41<br>04 C9 H23 — C<br>04 C16 H21          | 310 52 618                                    | C18 H36<br>C24 H40<br>C28 H40               | C7 H18<br>C8 H16<br>CB H20               | P2 049<br>C92 H188                       | N3 C10 H25<br>N3 C15 H35<br>N3 C20 H45       | 012 C24 H54<br>P3 FE2                     | C36 H84<br>C42 H48<br>G6 C24 H48       | P4 CL2<br>CU 012 C12                   |
| O RE C18 H42 O4 C29 H22 P<br>O RE C20 H30 O4 C30 H24                                      | P2 H36 0                    | 5 CT HL9 C  | 10 S2 C22<br>H36                              | C30 H52<br>C34 H52                          | C9 H22<br>C10 H22                        | P2 P0                                    | N3 C30 H25<br>N3 C40 H85                     | 015 C45 H33                               | 06 C30 H36                             | CU 016 C12                             |
| D RU C37 H34 010 C28 H54<br>D2 RE C36 H30 010 C40 H30 P                                   | 0                           | 06 C7 H19   | 110 SZ C40                                    | C36 H40<br>C38 H44<br>C38 H52               | C10 H24<br>C11 H18<br>C11 H24            | C14 H36<br>C24 H40                       | NB C10 H20<br>NB C30 H30<br>NB C35 H40       | P3 H3                                     | 015 C57 H48<br>018 C36 H24<br>56 C6    | FE C12 H32<br>FE C20 H48               |
|   | :8 HZ4 0                    | 06 C8 H21<br>06 C9 H21 (<br>06 C9 H23                 | 110 SZ C4Z                                    | C44 H48<br>C52 H40                          | C11 H26<br>C12 H20<br>C12 H26            | C28 H30<br>C28 H40                       | 0 05 C31 H46<br>0 05 C55 H46<br>0 05 C55 H46 | N6<br>F3 H12                              | P3 N12                                 | FE C28 H48<br>NI D12 C72<br>H80        |
| RU C36 H30 H12 S8 SN C48 C4 S<br>H52 C4 S   | 54 CB H20 0<br>54 C12 H28 0 | 06 CLO H23 C  | 310 SZ C46<br>H36                             | N2 C42 H36<br>D2 C12 H16                    | C12 H28<br>C13 H16<br>C13 H28            | C38 H36<br>S C24 H40                     | D RU C55 H46<br>RU C54 H45                   | N9  | 021 C57 H45                            | OS C12 H32<br>OS C20 H48<br>OS C28 H48 |
| PZ CS4<br>MN C38 H36 P  | 0                           | 06 5 C12 H27 C  |   | 02 C16 H24<br>D2 C24 H23                    | C13 H30<br>C14 H30                       | PZ PD2                                   | 51 C6 H18                                    | P3 13                                     | 56 C6 H18                              | 05 C50 H44<br>05 C52 H48               |
| HZ 02 C4 HB H024 Q83 T12<br>H2 08 C23 H28 010<br>H2 010 C24 P2 CU                         | 2N3 0                       | 17 C9 H21 C   | 712 C14 H29<br>712 U C24 H54<br>713 S C28 H52 | 02 C32 H24                                  | C14 H32<br>C15 H32<br>C16 H20            | \$2 C30 H70<br>P2 PT                     | F4 N3<br>IR C54 H46                          | 09 RE C54 H45<br>09 RE C63 H63            |  | RE C52 H46<br>RU C12 H32<br>RU C20 H48 |
| H30 P<br>H2 S2 C4 H8 1 O4 C24 H38   | 2 1 0                       | 07 C12 H27 C  | 014 U C12 H30                                 | 02 C42 H38<br>54 C12 H28                    | C16 H36<br>C17 H36                       | C36 H32                                  | N3 C24 H20<br>N3 C24 H30                     | P3 1R                                     | P3 H21                                 | RU CZ8 H4B<br>RU C50 H44               |
| NA DA C30 HAG I DG C18 HAZ C24 HA DA C2 H1Z I DG C24 H30 C47 HI C36 H36 H D3 C36 H30 IR D | H37 0                       | 07 C14 H31  | PD C28 H50<br>P2 N3                           | \$4 C16 H30<br>\$4 C16 H36<br>\$4 C32 H36   | C18 H24<br>C18 H40<br>S C8 H20           | C36 H48<br>C38 H36                       | N7 C8 H16<br>N7 C8 H24<br>N7 C12 H32         | C54 H48<br>D2 C56 H50<br>D4 C56 H48       | \$6 C42 H48<br>P3 NL                   | P4 CL3                                 |
| 0 SZ N 06 C7 H18 MN N<br>02 C1 HZ N 06 C13 H30  | 42 08 C36 0<br>H30 D        | 07 C19 H25<br>07 C21 H29 C                            | 3 н9  | \$4 C32 H68<br>\$4 C48 HLOO                 | S C10 H24<br>S C14 H32<br>S C16 H36      | PZ PTZ<br>S4 C26 H62                     | N7 C16 H40<br>N7 C24 H24<br>N7 C28 H32       | 04 C57 H50<br>06 C58 H52                  | C46 H55<br>O C10 H27                   | RE C52 H48                             |
| 02 S N 06 C25 H54 NI C<br>03 C4 H8 N2 04 C30 H34 PT C                                     | 30 H37 0                    | 07 625 H21<br>07 627 H25 - 0                          | RE 53 C23<br>H30<br>RE 53 C39                 | P2 0  | S C18 H40<br>S C20 H44                   | PZ RE                                    | RU C54 H45                                   | P3 LA                                     | 010 C10 H27<br>010 C19 H45             | P4 CL4                                 |
| 06 52 5H C6 N2 OR C36 H30 PT C<br>H18<br>RE C36 H30 P2 D P                                | 0                           |   | H30<br>D2 C2 H7<br>D2 C3 H9                   | E31 H26                                     | 52 C10 H24<br>52 ZN C8 H20<br>54 C14 H20 | C36 H33                                  | P3 CL3<br>F3 N3                              | H3 021 C36                                | 010 C28 H63<br>010 C37 H81             | N4 C24 H20<br>N8 C20 H40<br>N8 C28 H32 |
| SEP SN C34<br>H30 N D6 C8 H20 C12   | H10 0                       | 017 C14 H25 C   | 04 C15 H25<br>05 S C38 H31                    | C5 H14                                      | \$4 C15 H22<br>\$4 C16 H24               | PZ SZ                                    | N3 C3 H9<br>N3 C18 H15                       | P3 L14                                    | P3 0                                   | P4 CL6                                 |
| P2 CL5 P2 D2 C18 C20  | H26 P                       | D S C14 H33 C   | 07 C5 H21<br>011 C9 H15<br>012 C8 H13         | C6 H16<br>C10 H24<br>C13 H28                | 54 C17 H26<br>54 C20 H32<br>TH C14 H14   | C4 H12<br>C6 H16<br>CB H20               | N3 03 C30 H21<br>N6 C6 H12                   | N5 D10 C7 HB                              | P3 03                                  | FE2 C20 H48<br>FE2 C28 H48             |
| 1 MI C18 H30 O7 C15 H26 C28<br>N O C32  | H30 P                       | PT C14 H33<br>PT 5 C8 H21                             | P2 -N4  | C17 H38<br>C18 H40<br>C19 H42               | P2 07                                    | C10 H24<br>C12 H28<br>C14 H16            | N6 C12 H30<br>N6 C18 H42<br>N6 C25 H54       | 03 C12 H27<br>03 C21 H45                  | W C44 H39                              | M6 C2 H8<br>M6 C4 H12<br>M6 C8 H20     |
| PZ CL6 HN C<br>07 C5 H5 HN O  | 36 H30<br>2 C36 H30         | P2 N2 C   | C8 H24<br>D2 C8 H24                           | C20 H28                                     | C2<br>C3 H10                             | C15 HIE<br>C16 HZL                       | N6 C30 H24<br>N6 O3 C45 H36                  | 03 C29 H33<br>03 C37 H33                  | P3 06                                  | N6 C10 H20<br>N6 C12 H12               |
| FE2 D4 PT C50 N RE<br>H40 P2 F N2 N<br>12 NI C6 H12                                       |                             | 38 H28 (  | 02 C10 H20<br>04 S4 C14 H16<br>06 C40 H36     | C22 H3B<br>C25 H22<br>C25 H54               | C4 H12<br>C5 H12<br>C6 H8                | C14 H36<br>C19 H26<br>C20 H28            | N6 D9 C57 H48<br>N9 D6 C66 H60<br>D5 C30 H45 | 03 C44 H39<br>03 C57 H45<br>012 C12 H27   | C18 H15                                | N6 C14 H16<br>N6 DZ C8 H16<br>D5 SZ    |
| 12 RE C36 H24 C6 C6 H15 N2 P<br>N2 C2 N6 O10 C6 H13 NI C                                  | 0 C30 M56 M<br>6 H18 M      | 41 C24 H21 (<br>41 D2 C36 H30 (                       | 07 SZ C10 HZ8                                 | C26 H24<br>C27 H26                          | C6 H16                                   | C21 H30<br>C22 H32                       | RE C30 H45                                   | 012 C21 H45<br>012 C57 H45                | SC2 C18 H15                            | P4 CLB                                 |
| HZ4 P2 F2 N1 C  | 28 H44 N                    | 41 06 C6 H18 C<br>NI 06 C28 H46 C<br>HI 06 C32 H38 C  | 010 C14 H25                                   | C28 H60<br>C31 H66<br>PT C26 H44            | C8 H20<br>C10 H16<br>C10 H20             | C24 H20<br>C24 H44<br>C28 H28            | F2 N3  | P3 N                                      | P3 010                                 | N4                                     |
| N2 PT C24 H34 N 05 C10 H25 NI O   | DE C36 H30                  |   | D15 C39 H42                                   | RE C40 H45<br>S C8 H20<br>-5 C12 H28        | C10 H24<br>C12 H10<br>C12 H12            | C28 H40<br>C29 H42<br>C30 H44            | HG2 03 C54<br>H39<br>N5 C2 H8                | 0 RH C54 H45<br>014 U C40 H90             | F3 012                                 | P4 CL1.                                |
| HB PT C18 H28 N2 D5 C12 H28 D2 R<br>HT C18 H30 NZ D5 C14 H20 D5 C                         | RE C38 H35 N<br>16 H14 N    | NI SZ C8 H18 (  | 018 RU C24<br>H34                             | S C16 H20<br>SIS C40 H44                    | C12 H28<br>C14 H16                       | PZ 54                                    | NS C4 HE<br>NS C4 H12                        | P3 N2                                     | PR C24 H34                             | P4 CL14                                |
| NI G26 H30 N2 05 C16 H40 PT C<br>02 C1 N2 05 C20 H32 RE C<br>03 C14 H10 H4 05 S2 C6       | 42 H4Z N                    | WI S2 C30 H46   | H42   | P2 03                                       | C14 H28<br>C14 H32<br>C16 H14            | C12 H26<br>C15 H30                       | NS C8 H20<br>NS C16 H34<br>NS C24 H44        | 011 C18 H39<br>014 C9 H15<br>015 C9 H15   | P3 014                                 | MO6 04 C72                             |
| PT C36 H36 N4 05 SZ C10   | 13 K                        | 41 52 C44 H42 5<br>32 C16 H24                         | S2 C9 H24                                     | C14 H16<br>C24 H20                          | C16 H20<br>C16 H36                       | C16 H32                                  | P3 CL5                                       | P3 N3                                     | C6 H17                                 | P4 C0                                  |
| PZ CLO NAZ OS   | Č                           | 02 C24 H18<br>02 G25 H24<br>02 S C36 H30 (            | P2 N5<br>09 C10 H13                           | P2 04                                       | C18 H40<br>C20 H28<br>C20 H44            | C17 H34<br>C18 H36                       | N4 C2 H4<br>N4 C2 H6                         | C12 H30<br>C18 H42                        | P3 PT<br>C54 H47                       | C28 H48<br>C52 H48                     |
| C6 M5 05 C4 H10<br>C7 H7 05 C8 H18 N RE   | E C44 H37 0                 | 02 55 C18 H16 C<br>03 C24 H18 C                       | 09 Cll H17<br>010 Cl0 H15                     | C4 H12                                      | C24 H52<br>C28 H28                       | P2 51                                    | N4 C4 H10                                    | C24 H54<br>C36 H30                        | 74                                     | P4 CO2                                 |
| 03 V P2 F3  | C                           | 03 PT C24 H42 (<br>03 S2 C20 H22 (<br>04 C8 H22       |   | C8 H20<br>C10 H16<br>C12 H10                | C28 H60<br>C30 H25<br>C32 H68            | PZ 515                                   | P3 CL6                                       |   | C8 H20<br>C24 H44                      | H34 H10 012<br>H36 H12 013             |
| PZ CL10 C4 H9 C13 H27 D6 G C0 C3Z H30 PD C25 H39  | 16 H17 C                    | 34 C14 H34 (  | D10 C22 H23<br>D10 S C12 H19<br>D11 C28 H35   | C14 H16<br>C16 H20<br>C16 H28               | C36 H76<br>C40 H84<br>S12 C8 H24         | C42 H48                                  | N3 C8 H8<br>N3 U6 C36 H24<br>N6 O9 C57 H45   | 06 C6 H18<br>06 C12 H30<br>06 C18 H42     | P4 AG                                  | P4 CR                                  |
| FE C32 H30 P<br>H2 S2 SH C18 P2 F6  | 2111                        | 04 PD C28 H38 C                                       | D11 C29 H41<br>D11 C31 H45                    | C18 H36<br>C22 H24                          | \$12 C10 H28<br>T2 C15 H26               | <b>P3</b>                                | N9 C30 H30<br>N9 C36 H90                     | 06 C24 H54<br>06 C36 H30                  | CL 04 C72 H60<br>N 03 C72 H60          | DZ C52 H44                             |
| H14 C12<br>D2 C4 H6 C2 H2 C15<br>D2 V BR2 NI C6 H12 C15                                   | H16 0                       | 04 S8 C15 H36 C<br>04 S8 C18 H34 C<br>04 S8 C22 H36 C | 013 S C10 H15                                 | C24 H28<br>C42 H36<br>PB S4 C4 H12          | 22 08<br>C20 H24                         | C26 H33<br>C34 H33<br>C41 H39            | N9 C54 H126<br>N9 D15 C57<br>H42             | 06 C42 H42<br>021 Y C36 H81<br>021 YB C36 | P4 B4                                  | 02 C54 H48                             |
| 65 CF37 15 HI CO HIS C10  | H30 0                       | 04 58 C40 H38 C<br>04 58 C42 H34 C                    | 015 C42 H45                                   | S C8 H20<br>S2 C4 H12                       | C33 H54<br>C48 H42                       | P3 AG                                    | H12 D21 C57                                  | H <b>61</b>                               | C16 H40<br>CL16 N6 C6                  | N4<br>HI C4 H12                        |
| N2 NI CG CG C17<br>C2 H3 H12 C17<br>C3 H5 N2 NI SZ CB C18                                 | H32 0                       | 36 C4 H14<br>36 C12 H12<br>36 C12 H14                 | P2 N6   | S2 C6 H16<br>S2 C8 H20<br>S2 C12 H28        | REZ C44 H30<br>S2 C20 H40<br>S2 C21 H42  | B F4 09 53 C9                            | P3 CL9                                       | 56 C12 H30<br>56 C18 H42                  | F4 543                                 | HI C24 H20                             |
| C4 H7 H12 C18 N2 O P2 CL12 P2 F12   | 117 C15                     | 36 C12 H30  | NE 010 V2 C82<br>H54<br>D C10 H30             | \$2 C15 H34<br>\$2 C16 H36<br>\$3 C4 H12    | SZ CZZ H44<br>SI3 C9 H18<br>ZR C14 H18   | P3 AL                                    | NA 09 C57 H4Z                                | \$6 C24 H54<br>\$6 C36 H30<br>\$6 C42 H42 | N4 022 C18                             | P4 F12                                 |
| HO6 02 636 HG 02 64   | 117 C17 0                   | 36 C23 H30 C  | 0 C30 H30<br>02 C12 H24                       | 53 C8 H20<br>53 C12 H28                     | P2 D9                                    | 09 C24 H30                               | N SELEC                                      | 56 C48 H102<br>56 C72 H150                | N10 020 52<br>C24 H32                  | P4 FE                                  |
| H30 0 C4 N5 0   | H23 0                       |   | 07 C3U H26<br>09 C14 H16<br>09 C14 H20        | 54 C2 H8<br>54 C4 H12<br>54 C6 H16          | 512 C4 H12                               | P3 82                                    | P3 CL13                                      | 56 C96 H198<br>P3 N4                      | 014 C10 H18                            | C28 H50                                |
| C30 H52 C12   | 2 L13 d                     | 36 SZ C10 H30 (                                       | 019 C49 H54                                   | \$4 C8 HZO<br>\$4 C9 HZZ                    | PZ 010                                   | C92 H77                                  | 03 TI  | D16 C20 H29                               | RU C12 H33                             | 16 C72 H60                             |
| C42 H48 P2 FE<br>C44 H48 07 C   | H21 C                       | 06 SZ CZO H48<br>06 SZ C30 H52<br>06 S3 C1Z H24       | P2 N7   | 54 24 C12 H28<br>54 28 C24 H52<br>V C40 H30 | C5 H10<br>C6 H14<br>C23 H38              | P3 B3<br>C6 H24                          | P3 C0<br>N O C54 H45                         | 018 C58 H57<br>P3 H5                      | RU C20 H49<br>P4 BHZ                   | P4 H2<br>H2 S                          |
|   | 2 MN 5                      | 07 C10 H12 C  | 012 C19 H23<br>013 S C21 H35                  | W C14 H24<br>W C16 H30                      | \$12 C12 H26<br>U C16 H36                | C10 H32                                  | P3 CO2                                       | C10 H30                                   | CO C52 H48                             | P4 H3                                  |
| 12 02 C36 H30   | 38 C36 H30 C                | 37 C18 H28 (  | 015 C19 H25<br>015 C23 H31<br>015 C23 H33     | W C18 H24<br>W C24 H30<br>W C29 H22         | PZ 011                                   | C12 H36<br>C24 H60<br>C117 H102          | H27 NB 012                                   | 010 C9 H16<br>012 C11 H18<br>013 C10 H16  | NI C28 H48<br>RU G12 H32<br>RU C20 H48 | N3                                     |
| H30   | C                           | D7 C25 H40 (<br>D8 C10 H12<br>D8 C12 H26              | 016 C24 H27<br>P2 N8                          | W C30 H24<br>P2 O5                          | C16 H36                                  | BR6 C12 H30<br>CL6 C6 H18<br>CL6 C12 H30 | P3 CR<br>H19 N4 09                           | 013 C11 H18<br>013 C12 H20<br>014 C29 H42 | P4 CA                                  | P4 H4                                  |
| N2 S2 C38 H30 N 06 TL4<br>D7 C19 H30 08 C   | H54 C                       | D8 C14 H20<br>D8 C16 H20 (                            | 02 C20 H28                                    | FE 47                                       | P2 012                                   | 16 C12 H30                               | 03 C44 H39<br>012 C39 H81                    | D15 C14 H24                               | H14 018<br>07                          | P4 H6                                  |
| 010 SN4   | 2 MO 0                      | D8 C18 H28 (  | D2 5 C12 H24<br>D4 52 C18 H28<br>D22 C46 H46  | C2 H8<br>C6 H16                             | C6 H14<br>C12 H22<br>C30 H44             | P3 BAZ<br>N2 014 C9 HLL                  | 012 C57 H45<br>015 53 C27<br>H48             | P3 N6<br>C9 H27                           | P4 C0                                  | NZ 53<br>NZ 5E3                        |
| CD 06 C42 H30   | 14 H24 C                    | 08 C20 H28 (<br>08 C20 H42 (                          | 023 C40 H52<br>023 C59 H64                    | C8 H20<br>C10 H24                           | C42 H46<br>N3 G24 H53                    | P3 BR                                    | P3 CR3                                       | D9 C57 H51<br>D22 C30 H39                 | FE 14 C72 H60<br>14 MN C72 H60         |  |
| N6 H1 06 C78 K2 N D6  | 24 H30 (                    | D8 C20 H48<br>D8 C22 H36<br>D8 C23 H32                | P2 N11  | C12 H12<br>C12 H12<br>C12 H28               | PB3 C28 H16<br>2N C32 H28<br>2N C44 H52  | IR C54 H47<br>O DS C55 H46               | H6 09  | P3 N7                                     | P4 CD5                                 | N3 \$3                                 |
| H54 010   | 30 H24 C                    | D8 C24 H28 (<br>D8 C24 H36<br>D8 C24 H50              | 013 C24 H32                                   | C16 H26<br>C17 H28<br>S C12 H26             | P2 013                                   | RU C54 H45<br>P3 BK3                     | P3 03<br>IR C54 H43                          | 016 5 C21 H36<br>017 C21 H28              | 015<br>P4 CL                           | P4 H12                                 |
| 06 C18 H30 06 52 D10  <br>06 C42 H30  | C40 H30 C                   | D8 C27 H40<br>D8 C28 H44 (                            | 07 C3 H9                                      | S C13 H28<br>S C15 H34                      | C81 H166                                 | OS C54 H45                               | P3 F6  | P3 N6                                     | CIJ 04 C72 H60                         | N4 53<br>N4 SE3                        |
| 06 PB C42 H30   |                             | D8 C29 H40<br>D8 C31 H48<br>D8 C32 H52                | PZ NA3  | S2 C6 H16<br>S2 C8 H20<br>S2 C10 H24        | P2 014<br>C10 H18                        | P3 C03                                   | N3<br>N3 C24 H24                             | 017 5 C31 H43<br>026 C59 H67              | FE C12 H33<br>FE C20 H49<br>FE C28 H49 | P4 H13                                 |
|   |                             |   |   |   |  |  |  |   |  |  |

Fig. 3.—Rotaform Index

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).

1)

2)

3)

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

a. With or without alkyl substituents. Under  $SN_4O_2$  search  $C_{10}H_{10}$ . For alkyl substituents search  $C_{10} + n^{H}_{10} + 2n^{\bullet}$ 

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

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

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, 1965

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

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

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:

$$Z \left[ \begin{array}{ccc} R & R'' \\ C & C & O \\ R & R''' \end{array} \right]_{n}$$

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