spectacular growth. On the other hand, books published on plastics and elastomers did not increase, reflecting a constant level of interest in polymers in the 1970s, down from the explosive growth of this field in the 1950s and 1960s.

CONCLUSION

Book literature in chemistry and chemical engineering continues to show increased growth. Renewed scientific activity in any chemical discipline is accompanied by an increase not only in journal article and patent literature but also in book literature in that field. Thus, it is a constant challenge to be currently aware of newly published books and reference treatises in one's field of specialization. A convenient source for such information is *Chemical Abstracts*, which covers 6000 book and book-related publications in chemistry and chemical engineering each year. The information can be obtained in a variety of forms and formats to suit the individual needs of the user.

REFERENCES AND NOTES

(1) "CAS Printed Access Tools 1981 Edition"; American Chemical Society: Columbus, OH, 1981.
(2) "CAS Today, Facts and Figures About Chemical Abstracts Service"; American Chemical Society: Columbus, OH, 1980.
(3) "Subject Coverage and Arrangement of Abstracts by Section in *Chemical Abstracts*"; American Chemical Society: Columbus, OH, 1975.

Searches for Polymers in the BASIC Files Derived from the Chemical Abstracts Service Chemical Registry System†

J. HAEUSER and M. HERZ

BASIC, Basel Information Center for Chemistry, CH-4002 Basel, Switzerland

Received August 7, 1981

The methods used for searching the BASIC Substructure Search System for polymers containing specified monomers are described. The BASIC files are based on CAS Registry and REG/CAN data. The searches yield Registry Numbers which are then used to retrieve CA Abstract Numbers.

INTRODUCTION

The Chemical Abstracts Service (CAS) Chemical Registry System[1] contains more than 5 000 000 chemical structures that have been cited in the literature and in *Chemical Abstracts* (CA) indexes since 1965. The file is increasing at the rate of about 350 000 structures per year. The use of this file, as installed at BASIC, for substructure search has been described previously.[2,3] We now report batch search retrieval of polymers of given monomers from the BASIC Substructure Search Files, which contain 170 000 polymers.

Since the structures of polymers are seldom completely regular, polymers are registered by CAS on the basis of their monomers,[4,5] as illustrated in Figure 1. Polymers only registered on the basis of their structural repeating units[5] are not yet retrievable from the BASIC files, so this paper is restricted to describing searches for monomer-based polymer records.

BASIC SEARCH FILES

Five BASIC Substructure Search Files have been generated from CAS data, as shown in Figure 2. They are used in various combinations for batch retrieval of polymers, depending on the type of search.

Connection Table File. The CAS Chemical Registry Structure Standard Distribution Format (SDF) File[6] contains, among other data, the connection table for each monomer for every polymer record in which it occurs. To optimize searching, monomer connection tables are stored only once in the Connection Table File;[3] i.e., only those CAS Chemical Registry System data elements that are necessary for the BASIC Substructure Search System are included. This file is used for the generation of the Fragment Mask File and for iterative (atom-by-atom) topological searches.

Fragment Mask File. The BASIC Fragment Mask File contains all BASIC Fragment numbers for each Registry Number in the form of a bit string.[2] The file is used as an efficient screen prior to searching the Connection Table File.

Multicomponent Registry Number File. This was generated as an additional file for batch retrieval of multicomponent compounds, such as polymers. Each polymer entry contains the Registry Number of the polymer and each monomer as well as an indication of the total number of Registry Numbers. The record for a homopolymer therefore contains the Registry Number of the homopolymer, the Registry Number of the monomer, and the number 2 indicating two Registry Numbers. The record for acrylonitrile-1,3-butadiene copolymer is shown in the following example.

total	copolymer	CH_2=CH–CH=CH_2	CH_2=CH–CN
3	9003-18-3	106-99-0	107-13-1

It is clear from the record description that the homopolymer and all copolymers of, for example, acrylonitrile can be retrieved by means of the Registry Number of the monomer.

Link File. For a further simplication of the searches, the Registry Number data of the Multicomponent Registry Number File are divided into multicomponent-component Registry Number pairs and are sorted in component Registry Number order. This procedure is illustrated for polymers in Figure 3 with acrylonitrile homopolymer and acrylonitrile-1,3-butadiene copolymer.

REG/CAN File. This file contains all CAS Registry Number–CA Abstract Number pairs and therefore all polymer Registry Number–CA Abstract Number pairs.

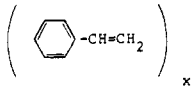
SEARCH PROCEDURE FOR POLYMERS

The simplest way to retrieve polymers, as already mentioned, is through the monomers by searching either the CA indexes or the BASIC Substructure Search Files. The search pro-

† Correspondence should be addressed to D. Ligtenberg, Secretary, BASIC, P.O. Box 273, CH-4002 Basel, Switzerland.

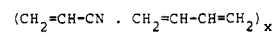
0095-2338/81/1621-0180$01.25/0

Homopolymer of styrene



Registry Number of styrene: 100-42-5
Registry Number of styrene homopolymer: 9003-53-6

Copolymer of acrylonitrile and 1,3-butadiene



Registry Number of acrylonitrile: 107-13-1
Registry Number of 1,3-butadiene: 106-99-0
Registry Number of acrylonitrile-1,3-butadiene copolymer: 9003-18-3

Figure 1. Registration of polymers and copolymers by CAS on the basis of their monomers.

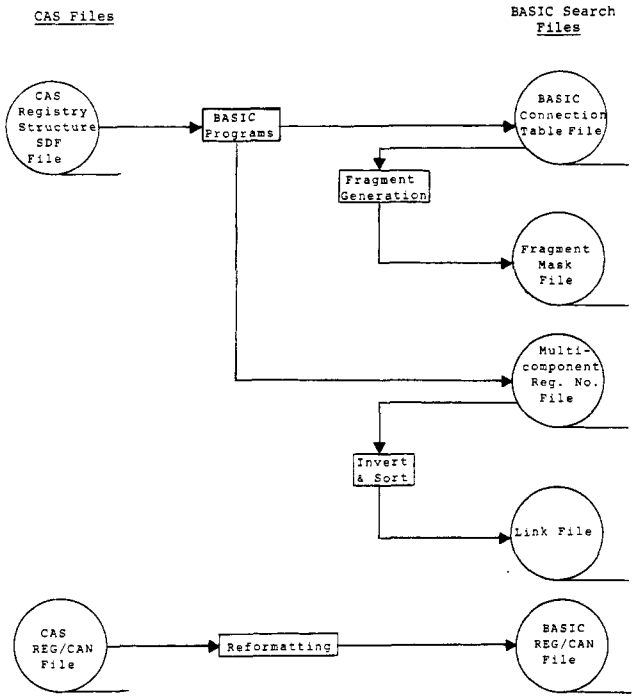


Figure 2. Generation of BASIC Substructure Search Files.

cedures are shown schematically in Figure 4.

All polymers containing the desired monomers are retrieved by searching the Link File for polymer Registry Numbers. This search produces a list of polymer Registry Numbers for each of the monomers. The resulting lists are compared, and only those polymer Registry Numbers that are common to all lists are processed further. This comparison of lists is omitted in searches where only one monomer is defined, e.g., in searches for all acrylonitrile homo- and copolymers or for all styrene homopolymers.

Two possibilities exist in the next search step:

(1) Copolymers with an Undefined Number of Monomers. This possibility is exemplified by a search for copolymers of vinylidene chloride, methyl methacrylate, and acrylonitrile; the copolymers may also contain other monomers. In this type of search, all CANs for all copolymer Registry Numbers are retrieved from the REG/CAN File. The Registry Numbers of the monomers were obtained from the CA Indexes, thus avoiding a substructure search:

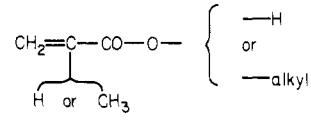
vinylidene chloride	75-35-4
methyl methacrylate	80-62-6
acrylonitrile	107-13-1

The Link File was then searched for all monomer–copolymer Registry Number pairs containing the three Registry Numbers, yielding 954, 7471, and 5572 pairs, respectively. There were 30 copolymer Registry Numbers that were common to all three lists; i.e., there were 30 copolymers that contained all three monomers and possibly other monomers, as exemplified below:

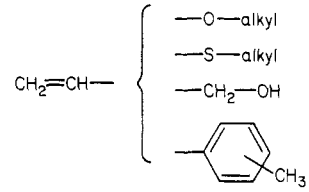
25214-39-5	75-35-4	80-62-6	107-13-1	
69174-18-1	75-35-4	80-62-6	107-13-1	5165-97-9

A search of the REG/CAN File yielded 72 CANs related to the 30 copolymers.

2. Copolymers with a Defined Number of Monomers. This possibility is exemplified by a search for binary copolymers of monomers of general structure A and general structure B.



A



B

Each of the monomer classes A and B gave rise to a series of Registry Numbers. All Registry Numbers of classes A and

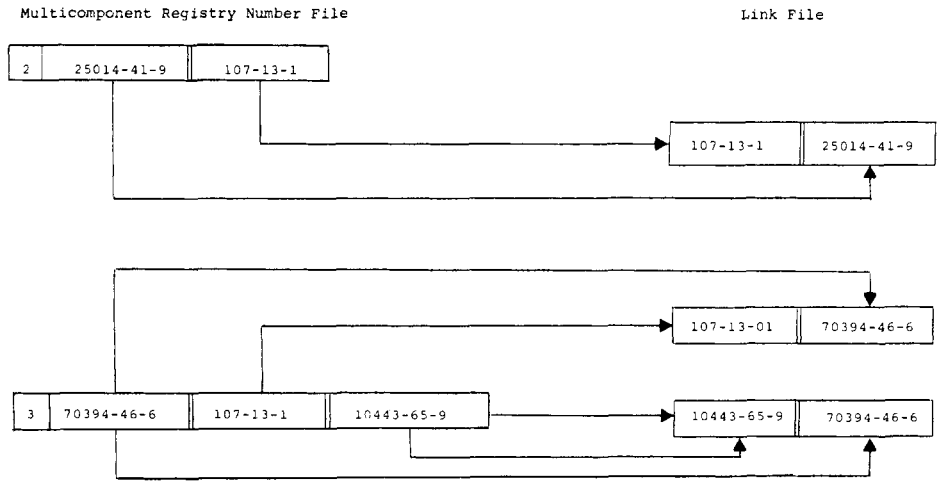


Figure 3. Link File Generation from Multicomponent Registry Number File.

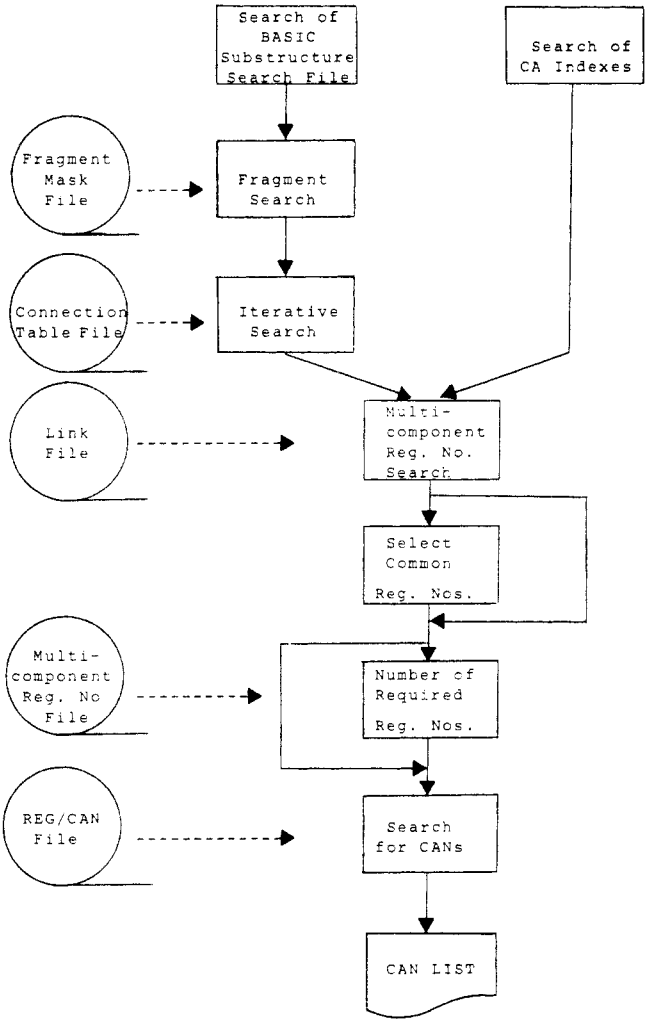


Figure 4. Retrieval of polymers.

B monomers were found by two substructure searches of the BASIC Registry File (Figure 4): a fragment search in the Fragment Mask File followed by an iterative search in the Connection Table File, yielding 185 and 194 Registry Numbers for class A and class B monomers, respectively.

A search of the Link File retrieved 33 191 polymers containing the class A monomers and 1742 polymers containing the class B monomers. When these lists were compared, 214 copolymers were found that contained both class A and class B monomers. Some of these copolymers, however, contained more than two monomers. By restricting the total number of Registry Numbers per polymer to three (see discussion of the Multicomponent Registry Number File), the latter copolymers were eliminated and 57 binary copolymers were identified. A search of the REG/CAN File yielded 81 CANs related to the 57 copolymers.

OTHER MULTICOMPONENT SUBSTANCES

The BASIC Substructure Search Files contain 90 000 alloys and 12 000 mixtures in addition to the 170 000 polymers. Since CAS Registry records of alloys and mixtures, like those of the polymers, contain the Registry Number of each component and the Registry Number of the alloy or mixture, the retrieval procedure described for polymers can also be used for alloys and mixtures.

ACKNOWLEDGMENT

We would like to thank CAS and their technical staff for the continuous support in providing files, programs, and further technical assistance, which enabled the development of the BASIC Substructure Search System. We also thank D. Ligtenberg (BASIC) for fruitful discussions and valuable assistance in preparing this publication.

REFERENCES AND NOTES

(1) Dittmar, P. C.; Stobaugh, R. E.; Watson, C. E. "The Chemical Abstracts Service Chemical Registry System. I. General Design". *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 111–24.
(2) Schenk, H. R.; Wegmüller, F. "Substructure Search by Means of the Chemical Abstracts Service Chemical Registry II System". *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 153–161.
(3) Graf, W.; Kaindl, H. K.; Kniess, H.; Schmidt, B.; Warszawski, R. "Substructure Retrieval by Means of the BASIC Fragment Search Dictionary Based on the Chemical Abstracts Service Chemical Registry III System". *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 51–55.
(4) Loening, K. L.; Metanomski, W. V.; Powell, W. H.; "Indexing of Polymers in *Chemical Abstracts*". *J. Chem. Doc.* **1969**, *9*, 248–251.
(5) "Chemical Abstracts Index Guide"; American Chemical Society: Columbus, OH, 1977; Appendix IV, Paragraph 222.
(6) "Chemical Abstracts Service Registry Structure Standard Distribution Format", Columbus, OH, May 1976.