(4) W. E. Cossum, M. L. Krakiwsky, and M. F. Lynch, "Advances in Automatic Chemical Substructure Searching Techniques", *J. Chem. Doc.*, **5**, 33–35 (1965).

(5) H. R. Schenk and F. Wegmuller, "Substructure Search by Means of the Chemical Abstracts Service Chemical Registry II System", *J. Chem. Inf. Comput. Sci.*, **16**, 153–161 (1976).

(6) R. L. Wigington, "Machine Methods for Accessing Chemical Abstracts Service Information", Proceedings of IBM Scientific Computing Symposium on Computers in Chemistry, IBM Data Processing Division, White Plains, N.Y., 1969.

(7) W. Fisanick, L. D. Mitchell, J. A. Scott, and G. G. Vander Stouw, "Substructure Searching of Computer-Readable Chemical Abstracts Service Ninth Collective Index Chemical Nomenclature Files", *J. Chem. Inf. Comput. Sci.*, **15**, 73–84 (1975).

(8) J. F. B. Rowland and M. A. Veal, "Structure-Text and Nomenclature-Text Searching for Chemical Information: An Experiment with the *Chemical Abstracts* Integrated Subject File and Registry System", *J. Chem. Inf. Comput. Sci.*, **17**, 81–89 (1977).

(9) W. Fisanick, "Substructure Searching of Files Derived from the CAS Chemical Registry System", Chemical Abstracts Service Open Forum, Philadelphia, Pa., April 1975.

(10) B. M. Vasta and M. L. Spann, "Chemical Searching Capabilities of CHEMLINE", presented to the Division of Chemical Information, 172nd National Meeting of the American Chemical Society, San Francisco, Calif., Aug 31, 1976.

(11) P. F. Rusch and T. M. Crawford, "Improved Access to the Chemical Literature", presented to the Division of Chemical Literature, 172nd National Meeting of the American Chemical Society, San Francisco, Calif., Aug 31, 1976.

(12) System Development Corporation, *Search Service News*, (5)1, Jan 1977.

(13) J. E. Blake, "Parent Compound Handbook—The Successor to the Ring Index", presented to the Division of Chemical Information, 173rd National Meeting of the American Chemical Society, New Orleans, La., March 22, 1977.

(14) M. D. Harrison, "Implementation of the Substring Test by Hashing", *Commun. ACM*, **14**, 777–779 (1971).

(15) N. Donaldson, W. N. Powell, R. J. Rowlett, Jr., R. W. White, and K. V. Yorka, "*Chemical Abstracts* Index Names for Chemical Substances in the Ninth Collective Period (1972–1976)", *J. Chem. Doc.*, **14**, 3–14 (1974); CA Volume 76 Index Guide, 1972.

(16) These techniques are now a major component of the BASIC Chemical Information System in Switzerland.[5]

(17) The full CAS Registry file is somewhat larger than this retrospective file, primarily because it contains substances which have been added by the registration of special substance files and which have not been indexed in *Chemical Abstracts* since 1967.

Noble Gas Chemistry and the Fluoride Literature—What Influences Research Directions?

DONALD T. HAWKINS* and WARREN E. FALCONER

Bell Laboratories, Murray Hill, New Jersey 07974

A discontinuity was observed in the growth rate of the literature on the free molecular structures of binary fluorides. It is suggested that this resulted from the landmark discovery by Bartlett in 1962 that stable compounds could be made containing noble gases. This discovery has had an influence not only in immediately related areas of chemistry, but also in areas several steps removed.

The question of what influences research directions, and indeed what stimulates research, is often asked. Only rarely is a cause and effect relationship identified. A surge of activity in one field or another is frequently observed. Sometimes this is the direct result of a new technique or a new concept; occasionally more subtle stimuli can be identified.

We have recently compiled a comprehensive bibliography on the molecular structures and force fields of binary fluorides.[1] As part of some statistical information for that bibliography, the number of papers cited in the bibliography for each year between 1955 and 1975 was determined. Figure 1, taken from the Introduction to the bibliography, shows these data.

From 1958 through 1961, the number of papers per year was constant at about 30 citations per year. There was a large increase to nearly 60 papers in 1962, and again to 85 papers in 1963. Since 1966 the yearly publication rate has been approximately steady near 100 papers per year. (The data for 1975 and 1976 are incomplete because of the time lag in the abstracting services used.)

We ascribe the discontinuity in the publication rate in 1962, and the increased number of publications since that time, to a surge of interest in and familiarity with reactive fluorides following the landmark discovery by Bartlett in 1961 that compounds of the noble gas xenon (previously thought to be chemically inert) could be synthesized.[2] The first order effect of this discovery was the well-known flood of research activity on noble gas chemistry itself which began in 1962 and is continuing actively to this day. A secondary effect could also be observed in the chemical literature; much more active synthesis and characterization in the whole spectrum of inorganic fluorides ensued in the immediate wake of Bartlett's synthesis of xenon hexafluoroplatinate. Free molecular
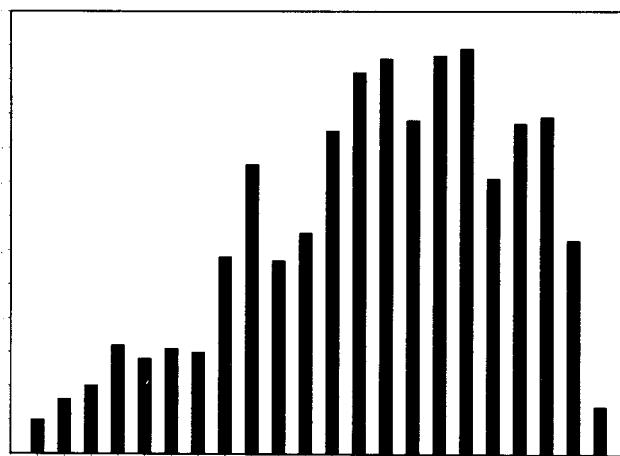

Figure 1. Papers cited per year.

structures, however, are at least one step further removed. Here we are dealing with physical measurements on a class of molecules which were generally not handled by chemists making structural measurements prior to the advent of noble gas chemistry in a real sense. The great interest in characterizing the structures and bonding of the noble gas fluorides, and the experimental and theoretical experience gained with these molecules, apparently spilled over to encompass a much wider selection of binary fluorides and related molecules.

It is rare that one can point to a single event which influences a broad area of chemistry. More often, one observes slow growth in a field, a major discovery, and then further steady growth as the effects of the discovery spread.[3] In this

case, however, we observe a sudden spurt of interest in the structural properties of binary fluorides following a major discovery some steps removed. The interest, as evidenced by the rate of publication, has been sustained ever since. This observation exemplifies the wide and profound influence Bartlett's discovery has had. It is a noteworthy correlation of the stimulating effects of an exciting landmark discovery in chemistry.

REFERENCES

(1) D. T. Hawkins, L. S. Bernstein, W. E. Falconer, and W. Klemperer, "Binary Fluorides: Free Molecular Structures and Force Fields. A Bibliography 1959-1975", NSRDS Bibliographic Series, IFI/Plenum, New York, N.Y., 1976, 238 pp.
(2) N. Bartlett, "Xenon Hexafluoroplatinate", *Proc. Chem. Soc. (London)*, 218 (1962).
(3) B. C. Bennion and L. A. Newton, "Epidemiology of Research on "Anomalous Water"", *J. Am. Soc. Inf. Sci.*, **27**, 53-56 (1976).

An Efficient Design for Chemical Structure Searching. III. The Coding of Resonating and Tautomeric Forms[1]

ALFRED FELDMAN

Walter Reed Army Institute of Research, Washington, D.C. 20012

A simple method for coding resonating and tautomeric structures is described. The method does not require the introduction of artifical bonds and facilitates the recovery of the original input structures.

1. INTRODUCTION

In retrieval systems using atom-by-atom codes, chemical structures are treated as finite graphs. Classical valence bonds and atoms are easily equated with edges and nodes. Exceptions are the hybridized bonds of resonating groups and the shifting structures of tautomers. These are not so readily equatable and risk complicating an otherwise simple coding process.

The difficulty might be avoided by the heroic means of generating, coding, and storing all possible resonance and tautomer isomers of the input compounds, or by having every submitted query undergo a similar transformation. Such a solution would greatly burden processing and storage facilities, and thus far no system has implemented it.

Most current systems have solved the problem by creating artificial bonds to replace the hybrid and shifting ones. This requires a number of new bond types, for example, "delocalized", "alternating", and "tautomeric" bonds.[2] Using these, structures can be normalized.

The use of artificial bonds does not represent an ideal solution. The determination, in a search, as to which of these bonds are equivalent to which, requires more complex programming and lengthier processing. Furthermore, the replacement of the hybrid and shifting bonds by normalized ones results in a loss of information. No longer will it be possible to recover the structure originally used for input. For this reason, normalization methods must be used with caution. For example, the most common type of tautomerism, the keto–enol tautomerism, is not normalized by *Chemical Abstracts.*

2. GOPPELT CODES

The coding method, implemented in the WRAIR system since *ca.* 1962, avoids the need for artificial bonds in resonating structures. The method, developed by Richard Goppelt, is based on the use of a connection table which is insensitive to differences among optional representations of alternating single and double bonds.

Conventional connection tables for chemical structures normally consist of three lists: one for the atoms, one for the bonds, and one for the connections. Figure 1 shows a connection table obtained according to Morgan's method.[3] The connection table developed by Goppelt differs in that only two lists are present. The information on atoms and bonds is

combined. For each atom, the element code is followed by an inventory of all the bonds, charges, hydrogen atoms, etc., that are attached to it (Figure 2). In this inventory, only classical valence bonds are recognized.

From Figure 3, it is seen that this method produces identical codes for the two resonating isomers of benzene. Furthermore, the fragments shown in Figure 4 will match these benzene isomers. This is because, in a substructure search, the categories DBN, SBN, HCT, etc., of the atom codes of the query are matched inclusively with the corresponding codes of the file compounds. That is, the C on line 2 in Figure 3 is a match for the C on line 2 in Figure 4a, as the former possesses at least two SBN. It will also be a match for the C on line 2 in Figure 4b, because it possesses at least one DBN and one SBN.

The original Goppelt method could not take odd ring atoms into the resonating system (Figure 5). Further, it could not distinguish certain cases of known isomerism, e.g., the cyclooctatetraenes. The extension of Goppelt codes to other classes of shifting structures, described below, suggested, however, a solution to these deficiencies.

3. ADAPTATION OF GOPPELT CODES TO TAUTOMERS

In the redesigned WRAIR system,[4] Goppelt codes have been adapted to tautomers, avoiding the need for artificial tautomer bonds in the connection table.

Prior to applying Goppelt codes to tautomers, the tautomers must be detected. This is accomplished by means of appropriate algorithms (Figure 6). An effort was made to keep these algorithms similar to those used by CAS.

The tautomer detecting algorithms each define a central atom (Q) and two terminal atoms (M and Y for algorithm A; N and Z for algorithm B). Codes are made insensitive to tautomeric variations in structure by overcoding these terminal atoms. The necessary rules are shown in Figure 7.

For example, in the structure shown in Figure 8, the terminal atom on line 1 has its categories SBN and HCT incremented each by one. The terminal atoms on lines 3 and 7 are similarly treated. The terminal atom on line 9 has its category DBN incremented by one. The central atoms, located on lines 2, 4, and 8, are not altered.

As a result of the application of the rules of Figure 7, different tautomers of a compound will have identical codes.