Chemical Substructure Searching

ROBERT E. STOBAUGH

Chemical Abstracts Service, Columbus, Ohio 43210

Received February 22, 1985

During the past 25 years, substructure search systems have evolved from small-scale batch operations on punched card databases to commercial on-line systems operating on the CAS Chemical Registry file of over seven million structure representations. Fragment codes, linear notations, systematic nomenclature, and atom–bond connection tables have been used as structural representations. The development of total chemical information systems integrated around substructure search is seen as the important area of the future.

INTRODUCTION

I have been asked to contribute this survey of chemical substructure searching in honor of the silver anniversary of the *Journal of Chemical Information and Computer Sciences*. When my article on editing organic abstracts at Chemical Abstracts Service (CAS) was published in the maiden issue of this journal in 1961,[1] chemical substructure searching was still in a very early phase, using edge-notched and IBM cards derived from a variety of linear notations.[2] The usefulness of substructure search was known, or certainly postulated, at that time. It was widely accepted then, as it is now, that the value of a collection of chemical structural information depends greatly on its versatility in providing for specific and generic searches, that is, for the location of specific compounds and for identification of compounds that have certain atom–bond arrangements in common. For example, as shown in Figure 1, a searcher may be interested in chemical compounds containing the benzene ring linked to the SO_2N group. This group is characteristic of the sulfa family of drugs, which show antibacterial, diuretic, or antihypertensive properties in varying degrees.

Mooers[3] had early suggested an atom-based connection table as a description of the topology of chemical structures. Later, Opler and co-workers[4–6] outlined a method for identification of parts of chemical structures and noted the necessity of incorporating into substructure search systems a total recall of all structural features from coded representations. Waldo and DeBacker[7] used an encoding system based on a topological map as a basis for sorting substances, and Ray and Kirsch[8] suggested sequential enumeration of the atoms in a molecule, including hydrogen, followed by construction of a topological map. However, ideas were forming and plans were developing for more effective, computer-based substructure searching, using structural records on magnetic tape. One of the very early public announcements of such plans was made by CAS in articles in 1959[9] and 1961[10] concerning its internal research activity.

EARLY DEVELOPMENTS (1963–1974)

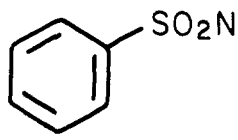
CAS's plans of 1959 and 1961 resulted 2 years later in a program[11,12] for substructure search by means of an atom-by-atom, or iterative, technique, using connection tables generated from IUPAC notations. This program allowed searches for substructures defined as specific groups of atoms linked by specific bonds. Although the program performed adequately when queries were defined in detail, it lacked generality; the deficiency was corrected by allowing increased flexibility in designating atoms and bonds of a search and in posing queries.[13]

The basis for substructure searching was, and is, the structure file of the CAS Chemical Registry System. This computer-based system identifies and records structural representations of chemical substances. It uses computer records of the structural diagrams universally used by chemists to describe molecular configuration. The Registry normalizes the structural representation (which may be recorded by the chemist in a variety of orientations) to a unique form for each different substance and, on the basis of this unique form, assigns an identification number, the CAS Registry Number, to the compound. The topological structure and special characteristics of the compound are recorded in a matrix called a connection table. The algorithm for generation of the unique and unambiguous connection table was developed by Morgan[14] at CAS, on the basis of earlier work at Du Pont.[15] Subsets of the data in this table correspond to chemical fragments (or substructures) for a given compound. An example of this representation is shown in Figure 2.

Each atom in this notation is given a number; the connection of that atom to the atom of lower number is shown; the type of chemical element is given, and the type of bond making up the connection is recorded. In this example, there are no

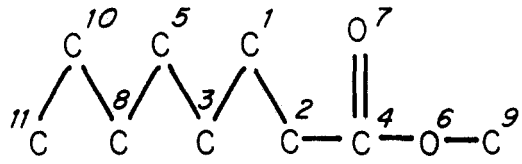
Robert E. Stobaugh received a B.S. in chemistry from Rhodes College in 1947 and the M.S. and Ph.D. in chemistry from the University of Tennessee in 1949 and 1952, respectively. He held postdoctoral positions at The Ohio State University. He is manager of the Information Science Department in the Research and Development Division of Chemical Abstracts Service (CAS), where he is responsible for research and exploratory projects and development work that involves generation of algorithms and functional designs for new and modified information processing systems. Prior to his present position, he was Department Head of the Organic Abstract Editorial Department and Technical Advisor to the CAS Chemical Registry System, where he was involved in the initial development and operation of the Registry System. He joined CAS in 1954.

0095-2338/85/1625-0271$01.50/0



- isolated or fused ring system
- any substituent on the ring
- any substituent on N

Figure 1. Example of a substructure query.



REGISTRY NUMBER	111-11-5
TEXT	NS
ATOM NUMBER	1 2 3 4 5 6 7 8 9 10 11
CONNECTION	1 1 2 3 4 4 5 6 8 10
ELEMENT	C C C C C O O C C C C
BOND	-1 -1 -1 -1 -1 -2 -1 -1 -1 -1

Figure 2. CAS Chemical Registry connection table of octanoic acid methyl ester.

molecular rings; therefore, each bond is a so-called "chain" bond (indicated by the dash). The numerals indicate whether the bond is a single or double bond. The hydrogen atoms that complete the structure are not shown. The text NS indicates that there is no stereochemistry.

The early substructure search system was tested and demonstrated in 1965,[16] but it remained an iterative search system until an initial screening step, known to be necessary for any large-volume operation, was added. The enhanced system, operating on the CAS Chemical Registry II database, was described in a 1968 presentation[17] and a publication the next year.[18] This experimental system had screen and atom-by-atom search capabilities comparable to those still in use in today's on-line systems, but it was restricted to technology of the time. It was a tape-based, batch-mode operation with queries manually encoded by the searcher and with display of answers consisting only of CAS Registry Numbers. The system showed clearly the feasibility of substructure search procedures on the CAS Registry Structure files but was hampered by lack of sufficiently powerful computer hardware and by the inability to provide structure diagrams for answers. Consequently, it was not promoted as a CAS service.

During the time CAS was building its first operating substructure search programs, ideas were also developing in France for computer-based chemical information handling,[19] including structure–activity correlation, computer-aided design, and substructure searching. In a series of papers in 1966–1967, Professor Jacques-Emile Dubois[20-24] outlined the theories of the DARC System (Description, Acquistion, Retrieval, Correlation), in which the basic role of the substructure in chemistry was stressed. The starting point for the DARC structure file was the DARC topological representation, a code or notation derived from the generation of a graph, i.e., structure diagram, for a chemical compound. In this process, a starting point or focus for generation of the graph, usually a specific atom, was selected by means of a set of formal rules. Starting from the focus, the compound was described on the basis of concentric layers of atom–bond combinations centered on the focus. The DARC code contained topological descriptions of the focus, then the number of concentric layers, then a description of the nature of the bonds, and finally identification of the specific atoms. Features such as ring closures, specific hydrogen, and stereochemistry could also be included. Topological screens (FRELs) were generated as an important component of substructure search. An example of such screens is shown in Figure 3. These were fragments consisting of a central atom with a connectivity of three or more, together with its attached atoms. The DARC Substructure Search System, operating on the CAS Registry Structure File, initially the Registry II File, had screen and atom-by-atom search capabilities. It was first tested in 1973 on a 10 000-structure sample.

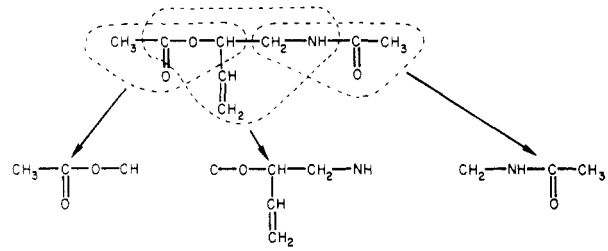


Figure 3. An example of the screens or FRELs generated in the DARC system.

Substructure search systems based on connection tables were in operation in other organizations, for example Du Pont[25] and Walter Reed Army Institute of Research.[26] In Germany, the Internationale Dokumentationsgesellschaft fur Chemie (IDC) Documentation System[27] was based not only on a topological representation of chemical structures but also on the GREMAS fragmentation code. The GREMAS fragmentation code was originally derived manually but was later generated from connection table records. The substructure search systems described above were developed primarily for internal information retrieval systems and continue to be so used.

During this time also, substructure search on files of linear notation was a very active and expanding area. Such systems commonly used text search techniques in which notations were searched for strings of symbols corresponding to the desired substructures. Search techniques and systems were developed for several notations, such as IUPAC,[28] Skolnik,[29] and Wiswesser Line Notation (WLN).[30-33] In addition to the text search approach, other techniques used were expansion of the notation to an atom–bond connectivity matrix, which was then searched topologically,[34] as well as fragment code techniques.[30] The Institute for Scientific Information introduced a printed monthly index of permuted WLN's,[35] which permitted manual searches based on substructure fragments.

LATER DEVELOPMENTS (1975–1984)

The CAS experimental substructure search system, built on Registry II, while not implemented as a CAS service, became the basis of several internal operational systems at other organizations. It was used from 1967 through 1974 by the National Cancer Institute in a chemical information system to support their biological screening programs.[36] Somewhat later, several Swiss chemical firms—Ciba Ltd., J. R. Geigy Ltd. (now Ciba-Geigy Ltd.), F. Hoffman-La Roche and Co. Ltd., and Sandoz Ltd.—decided to develop jointly a computer system for chemical information based on the CAS Chemical Registry System. They formed the Basel Information Center for Chemistry (BASIC).[37] Their system used the experimental CAS substructure search system to search a licensed CAS Registry structure file and private structure files as well. The BASIC system was built and maintained by a system based on CAS Chemical Registry II programs. The BASIC group made a number of significant improvements to the original search system by developing new screen types and revising the

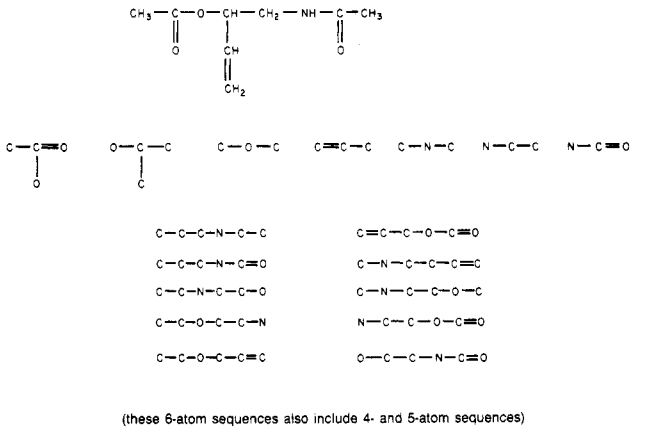


Figure 4. An example of screens generated in the CAS substructure search system.

screen dictionary. These improvements reduced the cost of screen generation and atom-by-atom searching, since the screenout was improved.[38,39] BASIC also upgraded their system with CAS Chemical Registry III programs.[38] An example of the fragment screens is shown in Figure 4.

Another approach to substructure search was developed by CAS in the mid-1970s.[40] The increasing availability of computer-readable files of chemical nomenclature and of text search programs led to the development of methods for performing substructure search in which CA index nomenclature terms were used as search terms. These methods were incorporated into an experimental substance retrieval service.[41] The service offered retrospective and current-awareness searches both for specific compounds and for compounds containing specified substructures. It also provided correlation of the retrieved substances with the corresponding textual information in the CAS database on their uses, activities, and other properties. The search database was derived from files used to produce the printed *CA Volume Chemical Substance Indexes* and the *Parent Compound Handbook*.[42] Both substance records and reference records were present. The substance records contained the CAS Registry Number, CA Index Name, molecular formula, substructure bit screen, and ring descriptions, when applicable. The reference records contained the CA abstract number, CAS Registry Number, and textual index terms. The substructure bit screen was composed of screens based on molecular formula, ring information (elemental composition, number of rings, size), and nomenclature. The nomenclature screen was derived from the alphabetic digrams present in the CA Index Name.

Like topological substructure search, substructure search based on nomenclature involved a screen search and an "iterative" search, i.e., text search on a character-by-character basis. An additional step, or refinement, called "link search" could be used if necessary. Link searching involved such details of nomenclature construction as the number of occurrences of a term, the order of terms, and the number of characters separating terms. Recall and precision in a nomenclature search were as effective as in the corresponding topological search.

The nomenclature search system was a tape-based, batch-mode operation. Although it could be used to obtain results of very high quality, its effective use required that searches be encoded by persons expert in CAS nomenclature, even when the many trivial names were replaced by systematic names at the beginning of the ninth collective indexing period in 1972. The need for expert involvement resulted in costs high enough to discourage its use, and it was discontinued.

In the late 1970s, the rapid spread of on-line search capabilities led CAS to examine the possibility of a topologically based substructure search service that would be available on-line. Although the earlier work had demonstrated the technical feasibility of using an initial screen search followed by an atom-by-atom search, it became clear that the large size of the file (then already over 5 million substances) would present problems. (The same conclusion was true of the nomenclature-based substructure system.) The result of the investigation was a new approach to large-file searching.[43,44] Under this approach, the file is segmented and the search task distributed over a number of minicomputers operating in parallel, the "search machine". Each minicomputer searches only a portion of the file, using a simple sequential file organization. A "front-end" controller handles all of the system overhead tasks, query input, supervision of the search machine, and answer output.

The segmentation of the search files provides the potential for a very high-speed search. The search machine architecture provides other additional benefits. First, search response time is independent of file size, in that the overall procedure provides for an increase in the number of minicomputers as the file size increases. Second, the "pipeline" flow, in which answers continually move through the system, allows answers to be displayed for review by the user as they are found; there is no need to wait several minutes to see the results of a search. Third, the sequential organization of all the files allows very easy updating. Search files can be updated weekly to provide the CAS ONLINE searcher with almost immediate access to the approximately 7000 new substances added to the CAS databases each week. Finally, multiple users can be accommodated simultaneously without significant degradation of system response times.

During the same period, the DARC Substructure Search System was evolving from initial testing.[45] A file of 500 000 connection tables—taken from the 2 million then comprising the CAS Chemical Registry System—was translated into the DARC code, and samples were studied for screening, file organization, and query performance. The result, the first version of the DARC substructure search system, was operational in 1976 and demonstrated in 1977. Textual data from CBAC (Chemical-Biological Activities) and CA Search were integrated later. In 1980 the system was applied to a large portion of the total CAS Chemical Registry file, including the completely defined organic and inorganic compounds but omitting polymers, coordination compounds, incompletely defined structures, alloys, etc. As of 1984, the DARC system is based on the complete CAS Chemical Registry file, with answers from substructure searches being transferable to bibliographic files to obtain information on articles published about the structures in question.

In both DARC and CAS ONLINE, substructure searching of topological representations of polymers, inorganic compounds, incompletely defined compounds, and Markush formulations is not yet as satisfactory as for completely defined organic compounds. Improvements in both the representation and the search techniques are ongoing.

An on-line substructure search system developed earlier in the 1970s was the NIH-EPA Structure and Nomenclature Search System (SANSS),[46,47] the substructure search being based on programs developed originally by Feldmann.[48] SANSS is a component of the on-line Chemical Information System (CIS), operated for a number of years by EPA. (CIS is now operated by CIS, Inc., a subsidiary of Fein-Marquart Associates, and by Information Consultants, Inc.) The system included not only the search programs but also a collection of databases that contain spectral data, crystallographic data, heats of formation, toxicities, and other data. The CAS connection table, Registry Number, molecular formula, and the CA Index Name are fundamental to the system. The connection table is not used as such, but translated first into

a derived table. Search functions include molecular formula, molecular weight, atom population, ring count/population, fragment search, ring search, iterative search, and structural feature code search.

Use of Wiswesser line notation for substructure search has continued. An example is the on-line substructure search used internally at Hoffmann-La Roche.[49] The use of rotated indexes at Hoffman-La Roche was replaced by a system that, in either on-line or batch mode, expands WLN notations containing contractions and converts them into atom–bond connection tables that are searched by a series of set reductions. The CROSSBOW system developed at ICI Pharmaceuticals Division[31,34] is based on input WLN, which are converted to an atom-by-atom connectivity record. Substructure search is performed on the WLN database by using fragment codes, followed by iterative search of the connectivity record, if necessary. The Institute for Scientific Information's on-line chemical information services[50] make use of WLN searching as previously noted.[32]

In other organizations, for example, at Merck Sharp & Dohme,[51] and Dow Chemical,[52] substructure search of internal files continues, in both batch and on-line modes, by using a two-level technique of fragment screen followed by iterative atom-by-atom search on atom–bond connection tables. In both these cases, biological activity databases are interfaced with the structure database. A number of other organizations perform substructure search on internal files by using software obtained from outside vendors, especially that of Molecular Design, Ltd. or the DARC software. Among the systems offered by Molecular Design, Ltd. is MACCS, which is a system for graphical input, storage, retrieval, and substructure searching of molecular information. Stereochemical features are included, by use of the extended Morgan algorithm.[53]

Searching by fragment codes, the earliest technique used in substructure search, is still in use,[54] particularly for searching Derwent's World Patents Index. The Farmdoc-Agdoc-Chemdoc multipunch code is the foremost fragment code used. There are others for specialty areas.

FUTURE DEVELOPMENTS

The past 25 years have seen substructure search systems evolve from small-scale, batch-mode operations using punched cards to today's commercial on-line systems that provide rapid search of millions of substances. (The CAS Chemical Registry file currently contains over 7 million substances.) However fast and efficient these systems have become, they have continued to be based on the theoretical approaches that were being developed in the early 1960s. The next 25 years can be expected to show continued growth and improvement in the scale and efficiency of structure search systems.

The foundation of the search systems, the structural record database, will be expanded. Registration of the chemical substances from CA indexes prior to 1965 has begun. When complete, the CAS Chemical Registry System will include records of chemical substances back to the early part of the century. The possibilities of an available "Beilstein-on-line" would then include all known substances, perhaps some 10 million, with 350 000 added every year. The structure search systems of the future will handle not only fully defined queries and databases but also combinations of these with generic or Markush queries and databases.

The most important developments of the coming years, however, will be in the development of comprehensive chemical information systems in which substructure searching, including Markush types, is a basic element. These systems will continue to be based on atom–bond records but will also contain such details of stereochemistry that enable stereochemical search and also display of three-dimensional structural images. The systems of the future will provide capabilities for integrating substructure search with searches of files of text, physical properties, biological data, reactions, and other aspects of chemical substances, working with powerful programs for structure–activity correlations, molecular modeling, and other computer applications whose development is only beginning.

REFERENCES AND NOTES

(1) Stobaugh, R. E. "Editing Organic Abstracts for *Chemical Abstracts*". *J. Chem. Doc.* **1961**, *1* (1), 22–23.
(2) National Academy of Sciences–National Research Council "Chemical Structure Information Handling—A Review of the Literature 1962–1968"; U.S. Government Printing Office: Washington, DC, 1969; NAS/NRC Publication 1733.
(3) Mooers, C. N. "Ciphering Structural Formulas—The Zatapleg System". *Zator Tech. Bull.* **1951**, 59.
(4) Opler, A. "DOW Refines Structural Searching"; *Chem. Eng. News* **1957**, *35* (33), 92–96.
(5) Opler, A. "On the Automatic Manipulation of Representation of Chemical Structures". *Am. Doc.* **1959**, *10* (2), 130–134.
(6) Opler, A.; Norton, T. R. "New Speed to Structural Searches". *Chem. Eng. News* **1956**, *34* (23), 2812.
(7) Waldo, W. H.; DeBacker, M. "Proceedings of the International Conference on Scientific Information", Washington, DC, Nov 16–21, 1959; National Academy of Sciences–National Research Council: Washington, DC, 1959; pp 711–730.
(8) Ray, L. C.; Kirsch, R. A. "Finding Chemical Records by Digital Computers". *Science (Washington, D.C.)* **1957**, *126*, 814–819.
(9) Dyson, G. M. "Research Expansion at the Chemical Abstracts Service". *Chem. Eng. News* **1959**, *37* (36), 128–131.
(10) Dyson, G. M.; Riley, E. F. "Mechanical Storage and Retrieval of Organic Chemical Data". *Chem. Eng. News* **1961**, *39* (16), 72–75; *39* (47), 74–80.
(11) Cossum, W. E.; Dyson, G. M.; Lynch, M. F.; Wolfe, R. N. "Mechanical Searching of Chemical Substructures by Means of Atom-by-Atom Matrices at Chemical Abstracts Service". In "Automation and Communication"; Luhn, H. P., Ed.; American Documentation Institute: Washington, DC, 1963; pp 15–18.
(12) Dyson, G. M.; Cossum, W. E.; Lynch, M. F.; Morgan, H. L. "Mechanical Manipulation of Chemical Structure: Molform Computation and Substructure Searching of Organic Structures by the Use of Cipher-Directed, Extended and Random Matrices". *Inf. Storage Retr.* **1963**, *1*, 69–99.
(13) Cossum, W. E.; Krakiwsky, M. L.; Lynch, M. F. "Advances in Automatic Chemical Substructure Searching Techniques". *J. Chem. Doc.* **1965**, *5*, 33–35.
(14) Morgan, H. L. "The Generation of a Unique Machine Description for Chemical Structures—A Technique Developed at Chemical Abstracts Service". *J. Chem. Doc.* **1965**, *5*, 107–113.
(15) Gluck, D. J. "A Chemical Structure Storage and Search System Developed at DuPont". *J. Chem. Doc.* **1965**, *5*, 43–51.
(16) "Chemical Abstracts Service's Computer Search System Nears Operational Status". *Chem. Eng. News* **1965**, *43* (44), 24.
(17) Dayton, D. L. "Dynamic Applications of Substructure Searching in a Chemical Information System". "CAS Open Forum, 10th"; 156th American Chemical Society National Meeting, Atlantic City, NJ, Sept 10, 1968; American Chemical Society: Washington, DC, 1968.
(18) Wigington, R. L. "Machine Methods for Accessing Chemical Abstracts Service Information". "Proceedings, IBM Scientific Computing Symposium on Computers in Chemistry"; IBM Data Processing Division: White Plains, NY, 1969; pp 97–120.
(19) Dubois, J. E.; Hennequin, F.; Chastrette, M. "Synthesis and Properties of α,α,α-Trisubstituted Aliphatic Ketones. I. Matrix Representation of Alkyl Radicals and Saturated Aliphatic Ketones". *Bull. Soc. Chim. Fr.* **1966**, 3568–3572.
(20) Dubois, J. E.; Laurent, D.; Viellard, H. "System of Documentation and Automation of Correlation Researches. General Principles". *C. R. Seances Acad. Sci., Ser. A* **1966**, 764–767.
(21) Dubois, J. E.; Laurent, D.; Viellard, H. "DARC System. Polymatrix Structural Description. Writing of Formal Matrices". *C. R. Seances Acad. Sci., Ser. A* **1966**, 1245–1248.
(22) Dubois, J. E.; Laurent, D.; Viellard, H. "DARC (Documentation and Automation of Correlation Researches) System. Structural Representation by the Descriptive Method with Limited Environment (DEL)". *C. R. Sceances Acad. Sci., Ser. A* **1967**, 348–351.
(23) Dubois, J. E.; Laurent, D.; Viellard, H. "DARC System. Principles of Research on Correlation and the General Equation of Topo-Information". *C. R. Seances Acad. Sci., Ser. A* **1967**, 1019–1022.
(24) Dubois, J. E.; Laurent, D. "Topo-information Relation. Calculation of the Overall Properties of Molecules Which Lack Preferential Sites. Linear Combination of Defocalized Environments". *C. R. Seances Acad. Sci., Ser. A* **1967**, 780–783.
(25) Hoffman, W. S. "An Integrated Chemical Structure Storage and Search System Operating at DuPont". *J. Chem. Doc.* **1968**, *8*, 3–13.

(26) Jacobus, D. P.; Davidson, D. E.; Feldman, A. P.; Schafer, J. A. "Experience with the Mechanized Chemical and Biological Information Retrieval System". *J. Chem. Doc.* **1970**, *10*, 135-140.
(27) Meyer, E. "The IDC System for Chemical Documentation". *J. Chem. Doc.* **1969**, *9*, 109-113.
(28) Dammers, H. F.; Polton, D. J. "Use of the IUPAC Notation in Computer Processing of Information on Chemical Structures". *J. Chem. Doc.* **1968**, *8*, 150-160.
(29) Skolnik, H. "A Chemical Fragment Notation Index". *J. Chem. Doc.* **1971**, *11*, 142-147.
(30) Bowman, C. M.; Landee, F. A.; Lee, N. W.; Reslock, M. H.; Smith, B. P. "A Chemically-Oriented Information Storage and Retrieval System. III. Searching a Wiswesser Line Notation File". *J. Chem. Doc.* **1970**, *10*, 50-54.
(31) Eakin, D. R.; Hyde, E. "Evaluation of Online Techniques in a Substructure Search System". In "Computer Representation and Manipulation of Chemical Information"; Wipke, W. T.; Heller, S. R.; Feldmann, R. J.; Hyde, E., Eds.; Wiley: New York, 1974; pp 1-30.
(32) Crowe, J. E.; Leggate, P.; Rossiter, B. N.; Rowland, J. F. B. "The Searching of Wiswesser Line Notations by Means of a Character-Matching Serial Search". *J. Chem. Doc.* **1973**, *13*, 85-92.
(33) Garfield, E.; Revesz, G. S.; Granito, C. E.; Dorr, H. A.; Calderon, M. M.; Warner, A. "Index Chemicus Registry System: Pragmatic Approach to Substructure Chemical Retrieval". *J. Chem. Doc.* **1970**, *10*, 54-58.
(34) Thomson, L. H.; Hyde, E.; Matthews, F. W. "Organic Search and Display Using a Connectivity Matrix Derived from Wiswesser Notation". *J. Chem. Doc.* **1967**, *7*, 204-209.
(35) Granito, C. E.; Rosenberg, M. D. "Chemical Substructure Index (CSI)—A New Research Tool". *J. Chem. Doc.* **1971**, *4*, 251-257.
(36) Richman, S.; Hazard, G. F., Jr.; Kalikow, A. K. "The Drug Research and Development Chemical Information System of NCI's Developmental Therapeutics Program". *ACS Symp. Ser.* **1978**, *No. 84*, 200-221.
(37) Schenk, H. R.; Wegmuller, F. "Substructure Search by Means of the Chemical Abstracts Service Chemical Registry II System". *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 153-161.
(38) 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.
(39) Graf, W.; Kaindl, H. K.; Kniess, H.; Warszawski, R. "The Third BASIC Fragment Dictionary". *J. Chem. Inf. Comput. Sci.* **1982**, *22*, 177-181.
(40) Fisanick, W.; Mitchell, L. D.; Scott, J. A.; Vander Stouw, G. G. "Substructure Searching of Computer-Readable Chemical Abstracts Service Ninth Collective Index Chemical Nomenclature Files". *J. Chem. Inf. Comput. Sci.* **1975**, *15*, 73-84.
(41) Dunn, R. G.; Fisanick, W.; Zamora, A. "A Chemical Substructure Search System Based on Chemical Abstracts Index Nomenclature". *J. Chem. Inf. Comput. Sci.* **1977**, *17*, 212-219.
(42) Blake, J. E.; Brown, S. M.; Ebe, T.; Goodson, A. L.; Skevington, J. H.; Watson, C. E. "Parent Compound Handbook—Successor to The Ring Index". *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 162-167.
(43) Farmer, N. A. "The Proposed Chemical Abstracts Service's Substructure Search System". In "Proceedings of the Technical Information Retrieval Committee of the Manufacturing Chemists Association", Arlington, VA, Aug 1977; McNulty, P. J.; Smith, R. B., Eds.; Manufacturing Chemists Association: Washington, DC, 1977.
(44) Zeidner, C. R.; Amoss, J. O.; Haines, R. C. "The CAS ONLINE Architecture for Substructure Searching". In "Proceedings of the National Online Meeting, 3rd"; Learned Information Inc.: Medford, NJ, 1982; pp 575-586.
(45) Attias, R. "DARC Substructure Search System: A New Approach to Chemical Information". *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 102-108.
(46) Feldmann, R. J.; Milne, G. W. A.; Heller, S. R.; Fein, A. E.; Miller, J. A.; Koch, B. "An Interactive Substructure Search System". *J. Chem. Inf. Comput. Sci.* **1977**, *17*, 157-163.
(47) Milne, G. W. A.; Heller, S. R.; Fein, A. E.; Frees, E. F.; Marquart, R. G.; McGill, J. A.; Miller, J. A.; Spiers, D. S. "The NIH-EPA Structure and Nomenclature Search System". *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 181-186.
(48) Feldmann, R. J.; Heller, S. R. "An Application of Interactive Graphics—The Nested Retrieval of Chemical Structures". *J. Chem. Doc.* **1972**, *12*, 48-54.
(49) Tomea, A. V.; Sorter, P. F. "Online Substructure Searching Utilizing Wiswesser Line Notations". *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 223-227.
(50) Sarkisian, J. "Introduction to ISI's New Online Chemical Information Services—An Integrated Search System for Chemical Information"; 188th ACS National Meeting, Philadelphia, PA, Aug 31, 1984; American Chemical Society: Washington, DC, 1984.
(51) Brown, H. D. "Substructure Searching of Chemical Files". In "Proceedings of the Technical Information Retrieval Committee of the Manufacturing Chemists Association", Arlington, VA, Aug 1977; McNulty, P. J.; Smith, R. B., Eds.; Manufacturing Chemists Association: Washington, DC, 1977.
(52) Bond, V. L.; Bowman, C. M.; Davison, L. C.; Roush, P. F.; McGrew, R. D.; Williams, D. G. "Online Storage and Retrieval of Chemical Information. II. Substructure and Biological Activity Searching". *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 231-234.
(53) Wipke, W. T.; Dyott, T. M. "Stereochemistry Unique Naming Algorithm". *J. Am. Chem. Soc.* **1974**, *96*, 4834-4842.
(54) Kaback, S. M. "Chemical Structure Searching in Derwent's World Patents Index". *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 1-6.

Digital Computers in Electrochemistry

PEIXIN HE and LARRY R. FAULKNER*
Department of Chemistry, University of Illinois, Urbana, Illinois 61801

Received April 1, 1985

The history and current status of digital computing in electrochemistry are reviewed. Six different roles are distinguished: (a) sample preparation, (b) experimental control, (c) preparation of data for interpretation, (d) model building, (e) interpretation, and (f) tactical and strategic decision making. Due to the small amount of work in area a, it is not considered in detail. Each of the remaining five areas is discussed separately.

INTRODUCTION

The advent of computing equipment has had a massive impact on almost every aspect of scientific life over the past quarter century, and electrochemistry has shared in those changes. In fact, the effects on electrochemistry predate those in most other areas of chemistry for interesting reasons. The modern era began in the middle 1950s, when potentiostats and function generators based on operational amplifiers were first invented. They gave electrochemists access to a much wider range of wave forms with which to excite an electrochemical system, and they introduced much shorter time scales than had previously been investigated with any seriousness. The new instrumentation made possible the very extensive development of new electrochemical methods that took place throughout the 1960s. Since operational amplifiers were themselves invented for analog computing, one can say entirely accurately that the advent of computing equipment gave birth to modern electrochemistry and has driven its development for 30 years.

Of course, the intent of this issue is to celebrate the improvements wrought by digital computing, not analog computing, on chemistry. Aside from this little reminder about deep roots, we will be true to that theme. Digital computing has had the most sophisticated of influences on electrochemistry: It has changed the way electrochemists think about doing electrochemistry, and it has widened the spectrum of

0095-2338/85/1625-0275$01.50/0