System". J. Chem. Inf. Comput. Sci. 1980, 20, 204-211.
(13) We are aware of two private CIS vendors: Fein-Marquart Associates, Baltimore, MD, and ICI, Inc., Washington, DC. Files for CIS are also available from the National Technical Information Service (NTIS).
(14) Milne, G. W. A.; Heller, S. R.; Fein, A. E.; Frees, E. F.; Margaret, R. E.; 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.
(15) Schafer, E. W., Jr., Wildlife Research Center, U.S. and Wildlife Service, Denver, CO, personal communication.

Implementation of Nearest-Neighbor Searching in an Online Chemical Structure Search System

PETER WILLETT* and VIVIENNE WINTERMAN

Department of Information Studies, University of Sheffield, Western Bank, Sheffield S10 2TN, U.K.

DAVID BAWDEN

Research Information Services, Pfizer Central Research, Sandwich, Kent CT13 9NJ, U.K.

Received August 6, 1985

This paper discusses the provision of nearest-neighbor searching facilities as an adjunct to the retrieval mechanisms of conventional chemical search systems. The facilities are based upon the calculation of a measure of intermolecular structural similarity between a query compound and the molecules in a machine-readable structure file. Examples are presented of the use of nearest-neighbor searching to rank output from substructure searches and to provide a means for carrying out browsing searches in structure files.

DETERMINATION OF INTERMOLECULAR SIMILARITY

Current systems for the retrieval of chemical structure information offer several types of search facility.[1] Registration or structure search involves the comparison of a specified query compound with each of the molecules in a file to identify an exact match, while substructural retrieval involves a partial match search for those molecules that contain the query as a substructure. A further type of search, superstructure search, may be of use in synthesis design programs.[2] A useful adjunct to such search facilities would be the provision of best match, or nearest-neighbor, routines that permitted the ranking of the compounds in a file in order of decreasing similarity to a query structure or substructure, the ranking being based upon some quantitative measure of intermolecular similarity or distance.

The concept of molecular similarity is not a new one. Thus, Adamson and Bush[3] evaluated a range of similarity measures on the basis of the fragment substructures common to a pair of compounds; Wilkins and Randic[4] and Gabanyi et al.[5] have discussed the use of simple topological relationships between compounds while Willett[6] has reported a comparison of hierarchic clustering procedures that are based on different similarity criteria. However, these studies have all been in the structure–property context, involving the use of only small sets of compounds, and until recently, there do not seem to have been any reports in the literature of the use of similarity-based ranking methods as a general search mechanism in computerized structure retrieval systems. Very recently, Carhart et al.[7] have described similarity matching procedures that are closely related to the work reported here, which discusses the implementation of nearest-neighbor searching in SOCRATES, the interactive chemical and biological data search system that has been developed at Pfizer Central Research (U.K.). The main areas of difference between our work and that of Carhart et al. are the types of structural feature and similarity measure that are used and in our adoption of an efficient best match search algorithm that is based upon the inverted file organization.

To explain the approach that we have developed for interactive best match structure searching, a brief description is required of the chemical searching component of the SOCRATES system. SOCRATES includes a chemical graphics module based upon connection table representations for each of the molecules in the file, and these tables are used for the generation of a set of fragment bit strings, one for each molecule. The strings are stored for search as a bit map, B, in which bit $B[I,J]$ is set to one if the Jth fragment screen has been assigned to the Ith molecule: currently, 1315 screens are used for the characterization of a file of over 200 000 compounds. The experiments reported here involved the use of a small subset of this file, containing the 8000 compounds in the Pfizer Stores File.

The bit map can be regarded either as a serial file, in which the bit strings are inspected in sequence one after the other, or as an inverted file, in which access is available to all of the compounds possessing a specific fragment screen. The use of the bit map in this latter form provides sufficiently fast screening for interactive substructure search by the intersection of the inverted file lists corresponding to the fragments in a query substructure; registration and structure search is carried out with the topological search codes that have been described in an earlier paper.[8]

To enable a set of compounds to be ranked in response to a query, some quantitative definition of intermolecular similarity is required,[9] and we have used the work of Adamson and Bush[3] as a basis for the systems developed here. Specifically, it is assumed that each of the structures under consideration is characterized by a set of substructural fragment descriptors and that the degree of similarity between some pair of structures or substructures can be evaluated in terms of the fragments that are, or are not, common to both of them. Given such fragment occurrence data for a pair of molecules, a very wide range of types of measure may be used to determine the degree of similarity between the two compounds. On the basis of simulated property prediction experiments using a small set of compounds with local anaesthetic activity, Adamson and

0095-2338/86/1626-0036$01.50/0

Bush suggested that relatively simple similarity measures performed at least as well as more complex measures, and this conclusion has been supported recently in a much more extended series of experiments.[10] Summarizing the results of these latter tests, it was found that binary characterizations, involving just the presence or absence of fragments in a structure, were less effective for the purposes of property prediction than characterizations in which the number of occurrences of each fragment were available. The two similarity coefficients that were considered, cosine and Tanimoto, gave rather better results than the three distance functions that were used and performed about as well as the more complex correlation coefficient that was also tested; the two coefficients, which have been used extensively in document retrieval search,[11] formed the basis for the systems developed here. Writing $N[I,J]$ for the number of occurrences of the Jth fragment in the Ith molecule, the value of the cosine coefficient for the similarity between two structures A and B is given by

$$\frac{\sum N[A,J]N[B,J]}{(\sum N[A,J]^2 \sum N[B,J]^2)^{1/2}}$$

where the summations are over the entire set of fragments that has been used for the characterization of the compounds in the data set; the corresponding value for the Tanimoto coefficient is given by

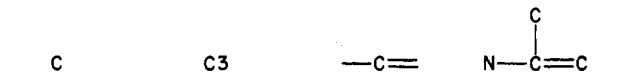
$$\frac{\sum N[A,J]N[B,J]}{\sum N[A,J]^2 + \sum N[B,J]^2 - \sum N[A,J]N[B,J]}$$

In both cases, the magnitude of the coefficient varies from zero to unity, these two extreme values corresponding to a pair of structures having no common fragments at all or having identical fragment characterizations, respectively. No attempt was made to standardize the data in any way since previous work[6,9] has suggested that this is not helpful with substructural occurrence information.

RANKING OF SUBSTRUCTURE SEARCH OUTPUT

The initial experiments were carried out in the context of substructure searching, where the chemist is interested in identifying all of the compounds in a machine-readable file that contain a query substructure. When a chemist carries out such a search, there is usually some group of compounds that the search is intended to retrieve; however, a general or fuzzy query, or the lack of sufficiently discriminating fragment screens, may result in the retrieval of many molecules in addition to those that were expected, and it is not generally possible to predict the number of hits that will be obtained in any given search a priori. In patent searches, for example, complete recall is essential, and the identification of several hundreds of molecules containing the query substructure may well be of importance; more generally, however, the chemist will be interested in a much smaller number of structures than the number that is actually retrieved. In such cases, it would be helpful if some *target structure* could be input, this being a typical representative of the class of molecules that the search is intended to retrieve. The compounds retrieved in the substructure search could then be matched individually against the target structure to obtain a ranked list, with the most similar molecules occurring at the top of the list for display purposes.

In the prototype system that has been developed at Sandwich, the first-level substructure search is carried out in the normal manner with the SOCRATES fragment screen and atom-by-atom routines. The fragments that are used for screening purposes in SOCRATES have been obtained as a result of a statistical analysis of the frequencies of occurrences of a wide range of types of substructure in the Pfizer file: the types considered for inclusion in the screen set include both atom-centered and bond-centered fragments. An automatic screen selection procedure based upon the methodology of Adamson et al.[12] results in the selection of a subset of these fragments that occur neither too frequently nor too infrequently in the file. While such a procedure has been shown to give good levels of screenout for substructure searching, it is not clear that such considerations of equifrequency may be entirely appropriate when the primary requirement is to discriminate between sets of closely related molecules, all of which share some common substructure. Accordingly, once such a set of related compounds has been obtained with the normal screen set, the connection tables corresponding to these structures are used to generate additional fragments that could act as a basis for the ranking procedure. Thus the initial, partial match search is based upon the fragments that have been selected previously for inclusion in the screen set, while the fragments used for the second-level, best match retrieval are generated at search time. A further difference between the two types of search is that all of the best match fragments are at a comparable level of substructural description; specifically, the following four atom-centered fragment types were tested, with one such fragment being generated centered upon each non-hydrogen atom within a molecule:



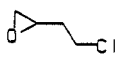
Of these, the augmented atom, which consists of an atom together with the immediately adjacent atoms and bonds, has been widely used for substructure searching and structure–property studies; the three other fragment types represent a steady and progressive broadening of the level of description provided by the augmented atom.

A comparable hierarchy of bond-centered fragments was also used, but tests showed that these gave rankings that were felt to be less generally useful than those obtained from the atom-centered hierarchy, and they will not be discussed further in this paper: full experimental results are given by Winterman.[13]

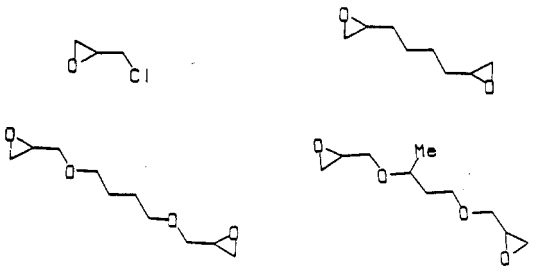
The fragments were generated for each of the compounds retrieved in the substructure search and also for the target compound that was to act as the basis for the ranking. The target was matched against each of the compounds in turn and the similarity calculated with one of the two coefficients above. These similarities were used to rank the molecules and the most similar compounds displayed on a graphics terminal to the chemist in order of decreasing similarity. The similarity measures were evaluated qualitatively on the basis of whether the ranking produced seemed to be intuitively sensible to the chemist, given the particular target compound that had been used. Such a subjective method of evaluation seems not unreasonable given the context in which the search routines were to operate; an alternative quantitative approach, using simulated property prediction, formed the basis of the earlier work on the comparison of similarity measures for structure–activity studies.

It should be noted that the substructure search outputs reported below are all quite small, typically containing only a few tens of compounds. This limitation arises from the need to evaluate the entire ranking that is produced in each case, something that is not feasible in an operational situation; indeed, our approach is designed specifically to allow for the ranking of large outputs so that the chemist need inspect only those few molecules that occur at the top of the ranking.

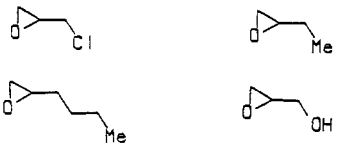
An early finding was that the cosine coefficient sometimes resulted in rankings that seemed rather illogical in character. Thus, when the target structure



was used to rank a set of 20 monosubstituted epoxide rings, the most highly ranked four compounds, according to augmented atom fragments, were

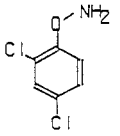


the last three of these being very different from the target structure. The reason for this anomalous behavior is that the cosine coefficient does not sufficiently account for the fragments that are present in one structure but not in the other with which it is being compared. The different denominator in the Tanimoto coefficient would seem to be more appropriate, and this was indeed confirmed since the top four compounds in the ranking when this coefficient was used were the more acceptable

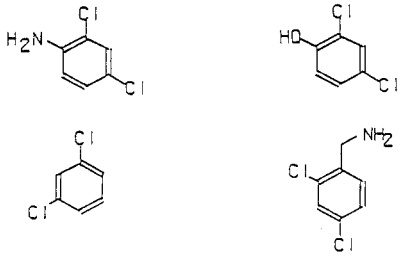


The difference between the two types of coefficient had not been apparent in the earlier property prediction experiments,[10] where the data sets were fairly homogeneous in character, with many of them being sets of analogues; in a typical structure file, conversely, a wide range of structural types will be present, and noncommon fragments will play a much larger role in discriminating between structures when a ranking method is used.

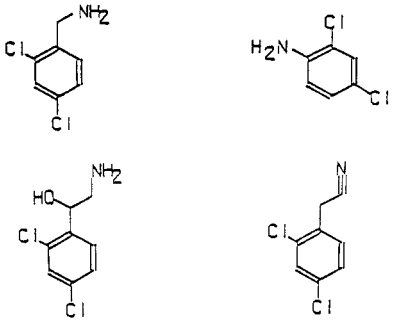
When different fragment types were tested, the rankings obtained were generally found to be more satisfactory the larger the size of the fragments that were used. Thus the best results were obtained consistently with the augmented atom fragments, a result that is in agreement with the findings of the review by Bawden.[14] However, while such large substructures help to identify the detailed bonding pattern within a molecule, the smaller, less specific fragments are useful in characterizing the molecular size and composition, and it was found that generally superior results were obtained with a combination of all four types of atom-centered fragment. Thus, the use of the target structure



on a set of 21 1,3-dichlorobenzenes identified

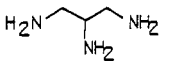


as the four nearest neighbors with just the augmented atoms, whereas the use of all four fragment types gave

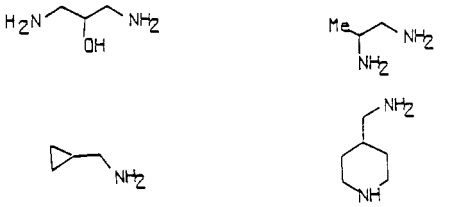


a result that was felt to be more intuitively acceptable.

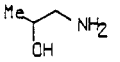
It was found useful in some cases to differentiate between cyclic and acyclic bonds during the generation of the fragments. Thus, the use of the target structure



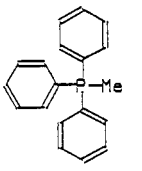
on a set of 24 amine derivatives produced



as the four nearest neighbors with the combined set of four atom-centered fragment types; when bond discrimination was used, the nitrogen-containing ring structure was replaced by



In other cases, however, cyclic and acyclic structures were strongly differentiated in the rankings even without the use of this refinement. Thus, the use of the target structure



with a combination of all four fragment types on a set of 40 structurally diverse organophosphorus compounds gave a ranking consisting of three discrete subsets, these consisting of triphenyls, monophenyls, and acyclics. For some purposes, of course, it may be appropriate to consider ring and chain bonds as synonymous, and thus, ring/chain differentiation is available as a run-time option for specification as a particular query may dictate.

Many substructure searches are sufficiently specific in nature to result in the retrieval of only a few molecules that satisfy the constraints of the query; when this is not so, however, ranking procedures of the sort described here can provide a simple means of focusing upon the structures that are of most interest. Our experiments have been carried out solely within the context of internal chemical files, but with the increasing availability of downloading facilities, there is no reason in principle why the ranking methods could not be applied to the output from substructure searches of public chemical data bases such as CAS ONLINE.

A BROWSING MECHANISM FOR CHEMICAL STRUCTURE SEARCH SYSTEMS

It is often the case that a chemist may be interested in compounds that are closely related to some specific query molecule in structural terms but may not wish, or may not be able, to specify precisely the nature of these relationships. An example of such a requirement might be at an early stage in a drug development program where a potential lead compound is available but where the substructural requirements for activity have not been located precisely: such a query could be handled in current systems only by performing a whole series of substructure searches, each of which would correspond to a possible topological pharmacophore. A best match retrieval system, conversely, would allow the molecule of interest to be matched against the entire file to obtain either some number of the most similar structures or all structures whose similarities with the query compound were greater than some threshold value.

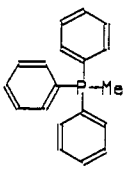
The major problem associated with the design of a retrieval system to handle this sort of structural query is that of ensuring a sufficiently rapid response to permit interactive searching. The approach described in the previous section for the ranking of substructure search output is efficient in operation only because the fragment generation and similarity matching procedures are applied to the relatively small number of molecules that are usually retrieved in a substructure search: such methods could not be applied to a whole structure file for interactive searching however.

The ranking system that has been developed is based upon the nearest-neighbor algorithm used by Willett for the calculation of intermolecular similarity coefficients in chemical classification experiments[9,15] and that had been originally described for applications in document retrieval by Noreault et al.[16] This algorithm assumes the presence of an inverted file to the compounds that are to be searched and is based upon the *addition* of the inverted file lists corresponding to the fragments in the query, rather than their intersection or union as in conventional substructure searching. The addition results in a vector, the elements of which contain the number of fragments in common between the query structure and each of the structures in the file. Given this information, it is trivial to evaluate the corresponding similarity coefficient and, hence, to rank the compounds in order of decreasing similarity with the query structure. The procedure is sufficiently fast in operation to allow for interactive best match searching of large files of structures, with response times comparable to those for conventional inverted file substructure searching. A PASCAL-like formulation was presented in a recent paper that describes the use of the algorithm for chemical superstructure searching.[2]

Two points need to be made at this point about our approach. First, since it is to be applied to an inverted file that contains only binary fragment data, the resulting intermolecular similarities are likely to be a less accurate reflection of the similarities between file compounds and the chosen target structure than if full fragment occurrence data were available: the algorithm can be used with such information, but its inclusion in the bit map would result in unacceptable overheads with present-day storage media. Second, and unlike the procedure described in the previous section, the methods can be used only where ready access is available to the inverted file postings: while this will generally be true of internal structure files, the procedure would not seem to be directly applicable to the searching of public chemical data bases.

The nearest-neighbor search is carried out in two stages as with the substructure search procedure described in the previous section: however, both of the stages in this case involve the ranking of compounds.

In the first stage, the inverted file lists corresponding to the query screens are added together as described above so as to identify the numbers of screens in common between the query molecule and each of the structures in the file. These numbers form the basis for a display that shows the numbers of molecules that contain a certain prespecified percentage of the query screens, thus allowing the chemist who is carrying out the search to obtain a feeling for the distribution of similarities that is present. To illustrate this, the query structure



is assigned a total of 34 different screens and yields the following output:

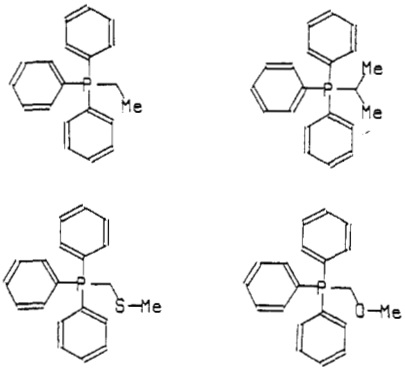
P:	100	90	85	80	75	50	25
$N(P)$:	12	35	45	47	451	3818	4648

where $N(P)$ is the number of structures in the file that contain at least P percent of the query screens. This display, which is typical of those obtained for a wide range of types of query molecule, reveals a highly skewed distribution of similarities with relatively few molecules having a high degree of similarity with the query but with the numbers increasing very rapidly as P is decreased so that by the time one reaches $P = 25$ $N(P)$ corresponds to over half of the file. For this particular example, a reasonable number of compounds of potential interest to the chemist may be obtained from the 451-member subset of the file that contains at least 75% of the query screens. This subset may then be used for ranking purposes in the second-stage search, which may be carried out in one of two rather different ways, both of which involve the use of the Tanimoto coefficient. Since only binary molecular characterizations are used here, the expression given earlier for this coefficient may be simplified to some extent. For a structure containing SIZE screens, C of which are in common with a query that has been assigned a total of QSIZE screens, the coefficient may be calculated as $C/(\text{SIZE} + \text{QSIZE} - C)$.

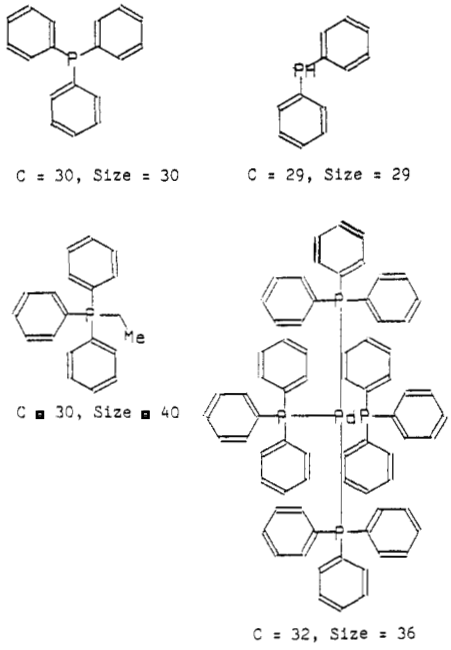
The simpler of the two approaches, which we shall refer to hereafter as a type A search, involves taking each distinct value of C and then calculating the Tanimoto coefficient for all compounds having that number of screens in common with the query: this corresponds to a ranking of the structures containing C of the query screens in ascending order of SIZE. The overall ranking is then obtained by concatenating the rankings for C = QSIZE, QSIZE − 1, QSIZE − 2, etc. until all of the compounds in the chosen subset, e.g., $P = 75$, have been ranked. It may be noted that the compounds associated with C = QSIZE are those that are obtained in a conventional substructure search: however, the rankings obtained here will be different from those resulting from the use of the ranking procedure described in the previous section since the similarities are calculated upon the basis of the bit string matches rather than upon the basis of the subsequent fragment matches.

The second type of search, a type B search, involves calculating the Tanimoto coefficient for all compounds in the subset identified by the chemist as being of potential interest and using the coefficients to obtain just a single ranked list, rather than a series of them which are then concatenated. The inclusion of all of the compounds may result in a rather broader sort of search in some cases since it can bring molecules to the top of the ranking that are structurally dissimilar, although still related, to the query compound. This arises from the nature of the Tanimoto coefficient, which allows a compound to occur high in the ranking, even if the value for C is significantly lower than the largest observed C value.

The difference between these two methods for identification of molecules similar to the query can be illustrated by reference to the query compound listed above, which was also used to test the ranked substructure search. The type A ranking for this compound identified the following four compounds as the most similar:

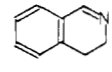


Each of these structures contains all 34 of the screens that had been assigned to the query compound: in all there were 12 such compounds in the file that exactly matched the query at the bit string level. The corresponding type B search identified



with the $P = 75$ subset. In each case in this figure, the values of C and SIZE have been given so that the reader can see the effect of using the Tanimoto coefficient to obtain a single continuous ranking, as against the series of discrete rankings that are obtained in a type A search.

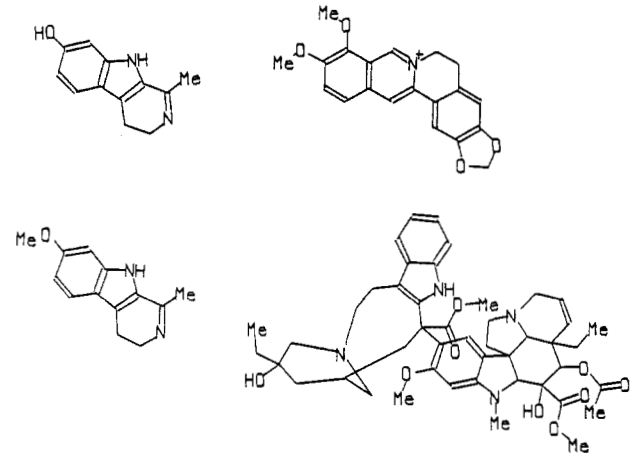
It will be clear that the structures identified in the two types of search are rather different in this case. Specifically, the type A search results in molecules that are more closely related to the target compound than are the molecules identified in the type B search. The latter allows a more exploratory, browsing-like retrieval mechanism, which permits the identification of what are, apparently, more distantly related classes of structures. This is not always the case, of course, as is illustrated by the molecule



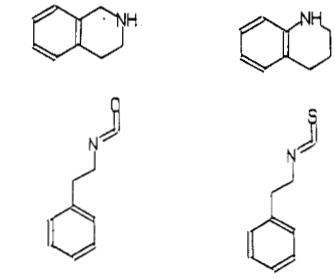
having

$N(P)$:	0	0	4	4	26	1602	4878

where there are very few structures that are closely related to the query compound. In this case, the type A search yielded



as the top four structures, whereas the type B search with $P = 50$ results in the rather less disparate

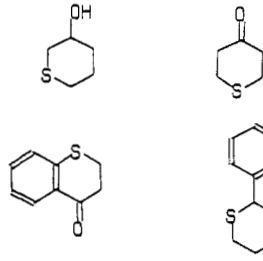


Another example of a structure for which there are very few closely-related molecules is provided by

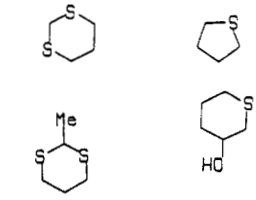
having

$N(P)$:	1	8	10	13	29	180	5188

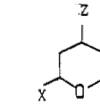
The type A search here yielded



whereas the type B search with $P = 50$ gave the equally heterogeneous



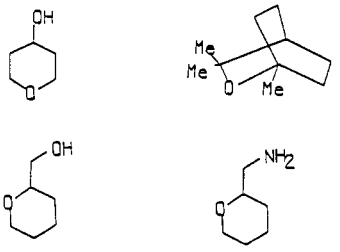
All of the target compounds so far have been fully defined structures. However, there is no reason why substructural queries cannot be searched in this manner as is evidenced by the final example



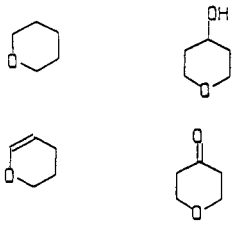
having

$N(P)$:	42	127	177	232	441	2165	6610

where Z is any atom and X is any atom other than hydrogen. The type A search for this query gave



while the type B search with $P = 75$ gave



It will be clear from the examples presented here that the two types of search can provide alternative sets of compounds for consideration by the chemist who is running the search, while additional flexibility may be obtained by varying the threshold used in the type B search. It is, of course, possible to set P to a very low value, such as $P = 25$ or even $P = 0$, but this is unlikely to throw up any interesting structures while decreasing the speed of response owing to the need to sort large numbers of similarity coefficients.

The primary use of this searching mechanism is to allow the chemist a certain degree of browsing ability, allowing him or her to become familiar with the full range of structural types that are present in the file and that are related to the chosen target compound. Perhaps the greatest advantage of this system, albeit an unquantifiable one, is that it enables the retrieval of unexpected, but structurally similar, molecules in response to a query that may act as a powerful spur to the chemist's imagination, suggesting new structural approaches to the problem in hand.

CONCLUSIONS

In this paper we have described the use of ranking methods to enhance the structure and substructure search components of chemical information systems. The methods appeal to chemists on several accounts. The ranking mechanism reduces the need for queries that have been finely honed so as to produce an acceptable volume of output, thus making end-user chemical retrieval more feasible than with conventional substructure searching systems; chemists can access the company's internal data banks using a fully specified compound of known interest, rather than having to express the query in terms of variable substituent patterns, alternative heteroatomic types, etc.; the somewhat imprecise nature of the search mechanism allows a form of automated browsing that can throw up compounds of interest that might not have otherwise been considered. Apart from the search facilities provided, the methods described here may be implemented with little difficulty: in all, they required about three man-months of development time and are to be made available on a routine basis within Pfizer in the near future.

ACKNOWLEDGMENT

We thank Trevor Devon and Michael Lynch for helpful comments on the manuscript and Pfizer (U.K.) and the Science and Engineering Research Council for financial support.

REFERENCES AND NOTES

(1) Ash, J. E.; Chubb, P. A.; Ward, S. E.; Welford, S. M.; Willett, P. "Communication, Storage and Retrieval of Chemical Information"; Horwood: Chichester, England, 1985.
(2) Willett, P. "An Algorithm for Chemical Superstructure Searching". *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 114–116.
(3) Adamson, G. W.; Bush, J. A. "A Comparison of the Performance of Some Similarity and Dissimilarity Measures in the Automatic Classification of Chemical Structures". *J. Chem. Inf. Comput. Sci.* **1975**, *15*, 55–58.
(4) Wilkins, C. L.; Randic, M. "A Graph Theoretic Approach to Structure-Property and Structure-Activity Correlations". *Theor. Chim. Acta* **1980**, *58*, 45–68.
(5) Gabanyi, Z.; Surjan, P.; Naray-Szabo, G. "Application of Topological Molecular Transforms to Rational Drug Design". *Eur. J. Med. Chem.* **1982**, *17*, 307–311.
(6) Willett, P. "A Comparison of Some Hierarchal Agglomerative Clustering Algorithms for Structure-Property Correlation". *Anal. Chim. Acta* **1982**, *136*, 29–37.
(7) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. "Atom Pairs as Molecular Features in Structure–Activity Studies: Definition and Application". *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 64–73.
(8) Bawden, D.; Catlow, J. T.; Devon, T. K.; Dalton, J. M.; Lynch, M. F.; Willett, P. "Evaluation and Implementation of Topological Codes for Online Compound Search and Registration". *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 83–86.
(9) Willett, P. "Use of Similarity and Clustering Methods in Chemical Information Systems"; Research Studies Press: Letchworth, England; in press.
(10) Willett, P.; Winterman, V. "A Comparison of Some Measures for the Determination of Inter-Molecular Structural Similarity". *Quant. Struct.-Act. Relat.*, in press.
(11) Salton, G.; McGill, M. J. "Introduction to Modern Information Retrieval"; McGraw-Hill: Englewood Cliffs, NJ, 1983.
(12) Adamson, G. W.; Cowell, J.; Lynch, M. F.; McLure, A. H. W.; Town, W. G.; Yapp, A. M. "Strategic Considerations in the Design of a Screening System for Substructure Searches of Chemical Structure Files". *J. Chem. Doc.* **1973**, *13*, 153–157.
(13) Winterman, V. Ph.D. Thesis, in preparation.
(14) Bawden, D. "Computerized Chemical Structure-Handling Techniques in Structure–Activity Studies and Molecular Property Prediction". *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 14–22.
(15) Willett, P. "The Calculation of Intermolecular Similarity Coefficients Using an Inverted File Algorithm". *Anal. Chim. Acta* **1982**, *138*, 339–342.
(16) Noreault, T.; Koll, M.; McGill, M. J. "Automatic Ranked Output from Boolean Searches in SIRE". *J. Am. Soc. Inf. Sci.* **1977**, *28*, 333–339.