(11) V C3 C3 C2 C2 C2 C1 C1 1 4 7 2 0 1 5 8 3 0 6 7 0 6 8 is uniquely presented as the canonical notation of 3,5-dimethylcyclohexanone.

APPLICATION OF THE CANOST NOTATION SYSTEM

The present notation system is an undoubtedly convenient tool for a variety of data bases in which manipulation of structural formulas is required. Actually, the system has been applied to the system SPIRES (SPectral Information REtrieval System), in which a ^{13}C NMR data base system is contained. The data base system, an interactive retrieval system of structure (substructure)–spectral information, has already been reported briefly,[4] and more detail will be presented in the following paper.[5]

A function to represent stereochemical structure has not yet been included with the system. The problem, however, will be solved without much difficulty by adding new symbols indicating stereochemistry such as, for example, E and Z for geometrical and R and S for configurational isomers and by slightly modifying the program.

REFERENCES AND NOTES

(1) Zupan, J.; Heller, S. R.; Milne, G. W. A.; Miller, J. A. *Anal. Chim. Acta* **1978**, *103*, 141. Bremser, W. *Anal. Chim. Acta* **1978**, *103*, 355. Wipke, W. T.; Heller, S. R.; Feldmann, R. J.; Hyde, E., Ed. In "Computer Representation and Manipulation of Chemical Information"; Wiley-Interscience: New York, 1974.
(2) Lynch, M. F.; Harrison, J. M.; Town, W. G.; Ash, J. E. "Computer Handling of Chemical Structure Information"; Macdonald: London, 1971. Davis, C. H.; Rush, J. E. "Information Retrieval and Documentation in Chemistry"; Greenwood Press: Westport, CT, 1974.
(3) (a) Kudo, Y.; Yamasaki, T.; Sasaki, S. *J. Chem. Doc.* **1973**, *13*, 225. (b) Kudo, Y.; Sasaki, S. *J. Chem. Doc.* **1974**, *14*, 200. (c) Kudo, Y.; Sasaki, S. *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 43. (d) Sasaki, S.; Abe, H.; Hirota, Y.; Ishida, Y.; Kudo, Y.; Ochiai, S.; Saito, K.; Yamasaki, T. *J. Chem. Inf. Comput. Sci.* **1978**, *18*, 211.
(4) Abe, H.; Sasaki, S.; Tanaka, K.; Osada, H. *CODATA Bull.* **1981**, *No. 40*, 31.
(5) Abe, H.; Hayasaka, H.; Miyashita, Y.; Sasaki, S. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 216-219.

Generation of Stereoisomeric Structures Using Topological Information Alone

HIDETSUGU ABE, HIROSHI HAYASAKA, YOSHIKATSU MIYASHITA, and SHIN-ICHI SASAKI*

Laboratory for Chemical Information Science, Toyohashi University of Technology, Toyohashi, 440 Japan

Received August 16, 1983

An algorithm for enumeration of stereoisomers due to asymmetric carbon, C=C double bond, and so on has been developed. By use of this algorithm, all the possible stereochemical structures for a molecule may be generated on the basis of its topological representation. The identification of each distinct stereoisomeric structure is performed by SEMA notation.

An algorithm has been developed to enumerate all the possible stereoisomers due to asymmetric carbons, C=C double bonds, and so on on the basis of topological data of chemical structure, for instance, the connection table, or the connectivity matrix, which is without stereochemical information.

Problems on computer-assisted enumeration of all the possible isomeric structures consistent with the molecular formula and/or spectral data of a certain compound have been studied, coupled with the studies of automated structure elucidation of organic compounds. The basic concept of the structure elucidation system is to infer the chemical structures that are not contradictory to such structural information of a sample compound. Up to now, such computer program systems as DENDRAL,[1] CHEMICS,[2] and CASE[3] have been developed for that purpose. The method of structure generation in these systems is the method that utilizes some of the distinctive features of computer, high-speed calculation and exhaustive enumeration and provides the key technique in the systems of structure elucidation as well as automated analyses of spectral data. Nourse and his co-workers[4] extended the isomer generation task from ordinary constitutional isomer level to stereoisomer level. Their work suggests that consideration for stereoisomers should be prerequisite in construction of the structure elucidation system in the future.

This study is concerned with a method to make exhaustive and unoverlapped enumeration of stereoisomers caused by asymmetric carbon atoms, C=C double bonds, etc., only from topological structural information such as connection table or connectivity matrix, which includes no stereochemical information. In practice, SEMA (Stereochemically Extended Morgan Algorithm) notation[5] is employed for identification of individual stereoisomeric structures. SEMA is an extension of Morgan method, which Wipke and Dyott have developed for recognition of stereochemical configurations in the organic synthetic design system (SECS).[6]

CONCEPT OF GENERATION OF STEREOISOMERIC STRUCTURES

Stereocenter. The most important things to be considered in generation of stereoisomeric structures are how to recognize the atoms and bonds responsible for stereoisomerism and how to represent the information of their configurations. In accordance with the proposal of Wipke and Dyott,[5] atoms and bonds responsible for the stereoisomerism are called here "stereocenters", which will be defined as follows.

An atom i is called a stereocenter if positional change of two attachments (substituents, atomic groups) bonded to the atom i will result in an alternative three-dimensional structure. As shown in Figure 1, exchange of groups between l and k in formula 1a changes the R configuration into an S configuration (1b). Thus, carbon atom i can be called a stereocenter. As usual, one stereocenter in a molecule results in two stereoisomers, and therefore, n stereocenters may produce 2^n stereoisomers. Of course, there may be cases where identical stereostructures are found in 2^n isomers, because of symmetry in the structure of a compound like tartaric acid. In this study, we will deal with tertiary carbons, quaternary carbons, and C=C double bonds as potential stereocenters but will not consider heteroatoms and double bonds connecting heteroatoms.

Configuration Mold. As one of the characteristics of the present method is that topological information alone is used for stereoisomeric structure generation, configuration mold

0095-2338/84/1624-0216$01.50/0

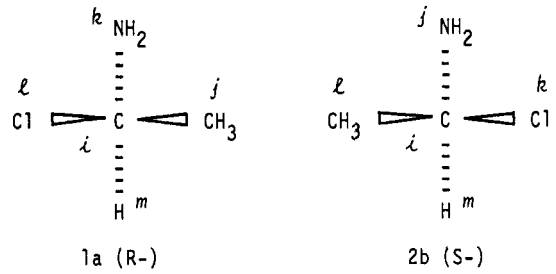
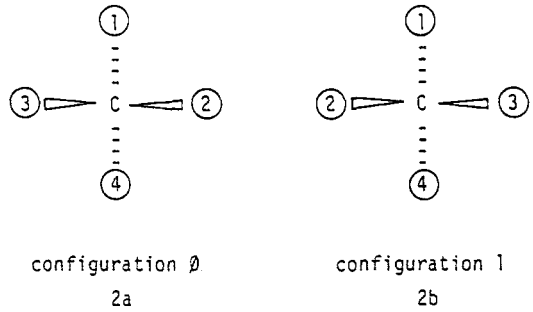


Figure 1. Example of stereocenters.



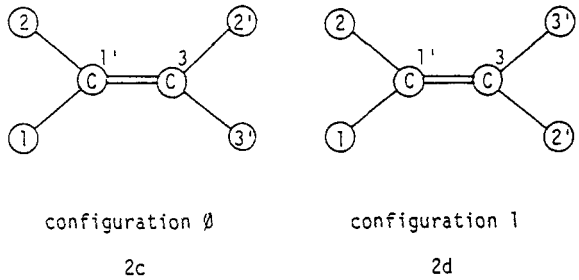


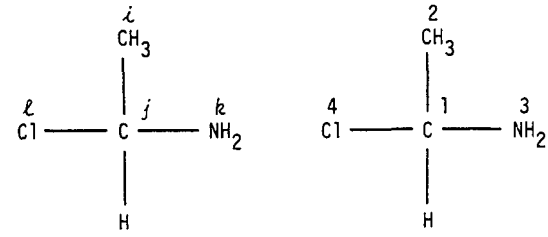
Figure 2. Configuration molds. Numerals indicate the initial labeling orders of the nodes. Structures 2a and 2c are used to determine parities.

is defined by us with reference to the arrangements of attachments around tertiary and quaternary carbons and C=C double bonds. As shown in Figure 2, configurations 2a and 2c are defined as configuration 0 and 2b and 2d as configuration 1. This figure means that numbering of the attachments like in 2a and 2c expresses the three-dimensional mold configurations around a quaternary carbon and a C=C double bond, respectively. In case of C=C double bond, a trigonal carbon atom is regarded as an attachment of the other trigonal carbon. Therefore, as shown at the bottom in Figure 2, carbon *i* and carbon 3 have the attachment groups 1–3 and 1′–3′, respectively. Although the numbering of each atom can be made arbitrarily, the initial numbering must be kept until all the procedures are completely finished.

The basic idea of stereoisomeric structure generation is that all the possible combinations of configurations for potential stereocenters in a topologically represented structure are generated by using these configuration molds. Then, each generated structure is named by SEMA, which is a stereochemically unique naming algorithm. Finally, the structures that have unique SEMA names are chosen as the true stereoisomeric structures.

EXAMPLES FOR STEREOISOMERIC STRUCTURE GENERATION

By use of toplogical structure 3a in Figure 3 as an example, how all the possible stereoisomeric structures are generated is shown below. The letters *i*, *j*, *k*, and *l* in formula 3a are arbitrarily assigned to CH_3, C, NH_2, and Cl, respectively, as



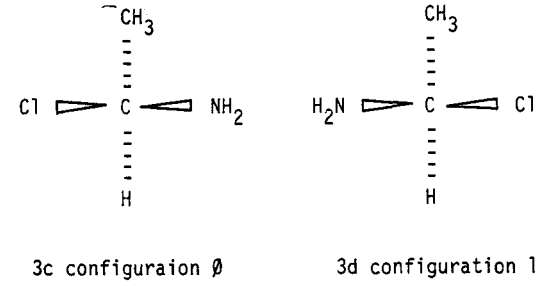


Figure 3. Examples of generation of stereoisomeric structures.

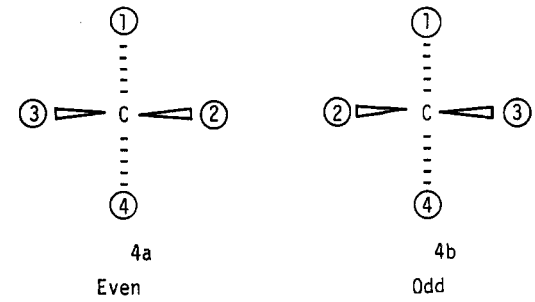


Figure 4. Definition of parities in SEMA name. Numerals show the orders determined by the Morgan algorithm.

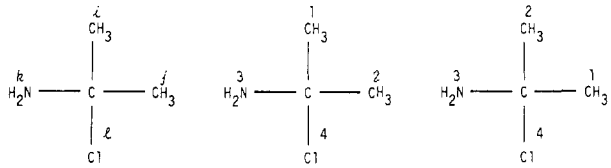
the initial labeling. In this formula, node (atom) *j* is a candidate for a stereocenter since it is a tertiary carbon. The topological information indicates only that the nodes *i*, *k*, and *l* are bonded to node *j*. On the other hand, all the nodes in the structure are labeled by the Morgan method[7,8] as shown in formula 3b.

According to Wipke and Dyott,[5] if the attachments around a stereocenter are arranged like formula 4a (Figure 4), the parity of the stereocenter is defined as Even (0), and if they are arranged like formula 4b, it is defined as Odd (1).

The actual procedure to obtain the parity for a potential stereocenter employed in this paper is as follows.

The attachment list for formula 3a is arranged in a natural order as [*i k l*]. Then, the corresponding labeling list for 3b is made up as [2 3 4]. In this case, no permutation is required for arranging the latter numerical list in ascending order, so the number of the permutation for this attachment list is determined as 0. This is easily understood by comparing formula 2a in Figure 2 with formula 4a in Figure 4. The Even parity is given for the attachment list requiring zero or an even number of permutations, and Odd parity is given for those that require an odd number of permutations.

Let us assume an *n*-place binary numeral that represents potential stereoisomers, where *n* is the number of potential stereocenters. This numeral string is called a configuration list. Then, a value of 0 for each element of the figure means that the corresponding stereocenter is given the parity determined originally, and a value of 1 means it is given the complementary parity. Enumeration of all the possible *n*-place binary numerals is the same as enumerating all the possible



5a 5b 5c

Figure 5. Initial and two alternative Morgan labelings for 2-amino-2-chloropropane.

Table I. Parities for Alternative Labelings[a]

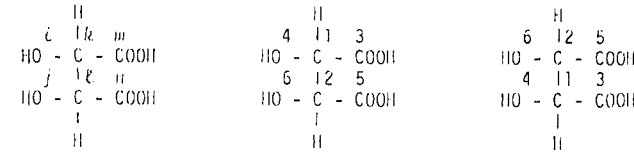
configuration		∅	1
determined parity	5b	E	O
	5c	O	E
selected parity		O	O

[a] The priority order of parity is O > E.

combinations of configurations for structures having *n* potential stereocenters.

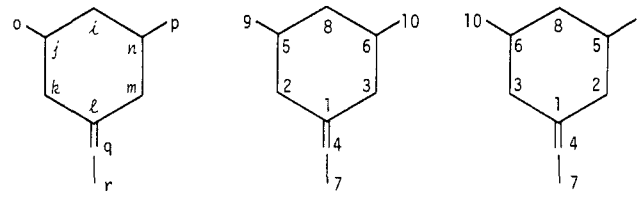
For this example, a 1-place configuration generates lists of 0 and 1 corresponding to two potential stereoisomeric structures. For the element of the configuration lists having a value of 1, the previously determined parity is inverted. Here, we obtained two potential stereoisomeric structures that have Even and Odd parities, respectively. According to definition, each potential stereoisomeric structure that has unique parity is distinct. Thus, two distinct stereoisomeric structures that have 0 and 1 configurations, respectively, are obtained. By using the configuration molds 2a and 2b (Figure 2), two stereoisomeric structures 3c and 3d shown in Figure 3 are finally obtained.

Since every tertiary and quaternary carbon in a structure is regarded as potential stereocenters in the present method, the above-mentioned procedure can be applied to the structure shown in Figure 5. For this example, two alternative Morgan labelings 5b and 5c are possible for 5a. If several labelings are possible from the Morgan algorithm, parities for every labeling are determined by referring to configuration molds 2a and 2b. The corresponding configuration lists and parities for the two alternative labelings are shown in Table I. Then, the one giving the parity combination having the earliest dictionary order is selected. According to the definition of the SEMA-naming algorithm, numerals 1 and 2 are assigned to Odd and Even parities, respectively, for the selection of the earliest one. Since Odd (1) is earlier than Even (2), the Odd parity is selected for the two configurations 0 and 1 of 5a (Figure 2) as shown at the bottom of Table I. This result means that the two configurations are stereochemically identical because their SEMA names are identical; therefore, the quaternary carbon is not a stereocenter. This procedure will be illustrated more clearly in the next example, tartaric acid. Tartaric acid has two stereocenters, so that four ($=2^2$) stereoisomers should exist. Actually, only three stereoisomers exist because of the symmetrical nature of the structure. The following example illustrates the procedure to generate and identify these three structures by the above-mentioned procedures. The initial labeling is shown in formula 6a of Figure 6. For this example, two alternative Morgan labelings 6b and 6c are possible. The parity combinations obtained for every configuration list are shown in Table II. It should be noted that the parity combinations to be compared are not arranged by initial labeling order but by determined Morgan labeling



6a 6b 6c

Figure 6. Initial and two alternative labelings for tartaric acid.



7a 7b 7c

Figure 7. Initial and two alternative Morgan labelings of 1-ethylidene-3,5-dimethylcyclohexane. Carbon 7 (or *r*) has two possible orientations:

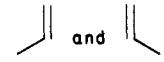


Table II. Generated Parity List for Tartaric Acid

			1	2	3	4
configuration list		*k*	0	0	1	1
		l	0	1	0	1
6b	1	*k*	E	E	O	O
	2	*l*	E	O	E	O
6c	1	*l*	E	O	E	O
	2	*k*	E	E	O	O
selected parities			E	O	O	O
			E	E	E	O

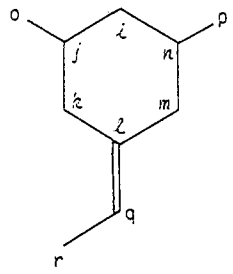
orders. As shown in Table II, the parities selected indicate that configurations [0 1] and [1 0] have the same parity combinations [O E]; therefore, they are identical stereoisomeric structures. Thus, three stereoisomeric structures whose stereocenters have the configurations [0 0], [0 1], and [1 1] are obtained for tartaric acid.

Next, let us consider the structure of 1-ethylidene-3,5-dimethylcyclohexane in Figure 7 as an example that contains two asymmetric carbons and a C=C double bond. Similarly to the previous examples, 7a is considered as a graph, and labeling is made in accordance with the Morgan method, to provide two alternative labelings, 7b and 7c. In the structure, candidates for stereocenters are *j* and *n* nodes and *l*–*q* double bond. Thus, a maximum of eight ($=2^3$) stereoisomers may exist. Attachment lists are [*i k o*] and [*i m p*] for the *j* and *n* nodes and [*k m q*] and [*l r*] for the *l* and *q* nodes of the double bond, respectively. In order to determine the parity of the double bond, the parities for two doubly bonded nodes are independently determined, and the sum of the two is defined as the parity of the corresponding double bond. In this case, [2 3 4] and [1 7] are obtained for the labeling 7b, and the parities are both Even. Therefore, the parity of the double bond corresponding to configuration 0 is Even + Even = Even.

Table III shows the results obtained after all the possibilities are examined. The parity combinations shown in the lowest column indicate that the configurations 1 and 3, 2 and 7, 4 and 5, and 6 and 8 are respectively identical with each other; therefore, four unique stereoisomeric structures that have the configurations [0 0 0], [0 0 1], [0 1 1], and [1 0 1] for stereocenters *j*, *l*–*q*, and *n* are obtained.

Table III. Generated Parity List for 1-Ethylidene-3,5-dimethylcyclohexane

			1	2	3	4	5	6	7	8
Configuration list		j	Ø	Ø	Ø	Ø	1	1	1	1
		$(\ell$-q)	Ø	Ø	1	1	Ø	Ø	1	1
		n	Ø	1	Ø	1	Ø	1	Ø	1
Determined parities	(1-4)	ℓ-q	E	E	O	O	E	E	O	O
	5	j	O	O	O	O	E	E	E	E
	6	n	O	E	O	E	O	E	O	E
Determined parities	(1-4)	ℓ-q	O	O	E	E	O	O	E	E
	5	n	O	E	O	E	O	E	O	E
	6	j	O	O	O	O	E	E	E	E
Selected Parity combinations			O	O	O	O	O	O	O	O
			O	E	O	O	O	E	E	E
			O	O	O	E	E	E	O	E



j	i	k	o
n	i	m	p
ℓ	k	m	q
q	ℓ	r	

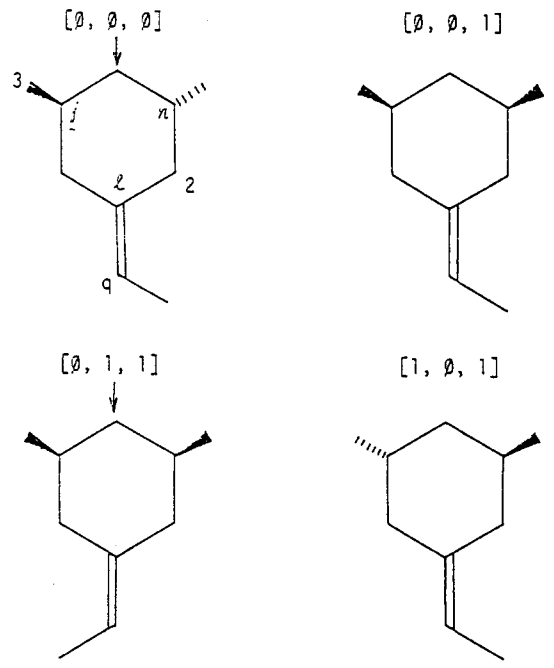


Figure 8. Generated stereoisomers for 1-ethylidene-3,5-dimethylcyclohexane.

When these configuration lists are fit to the originally labeled formula 7a, the structures shown in Figure 8 are finally generated from the configuration molds shown in Figure 2. For allenes and cumulenes, our method can be applied with minor modification as described below. This idea is originated from Nourse's paper.[4b] By regarding the C=C double bond of the structures in Figure 9 as a four-membered ring, the stereostructures of them can be generated by the use of our algorithm. Actually, we have obtained two stereoisomeric

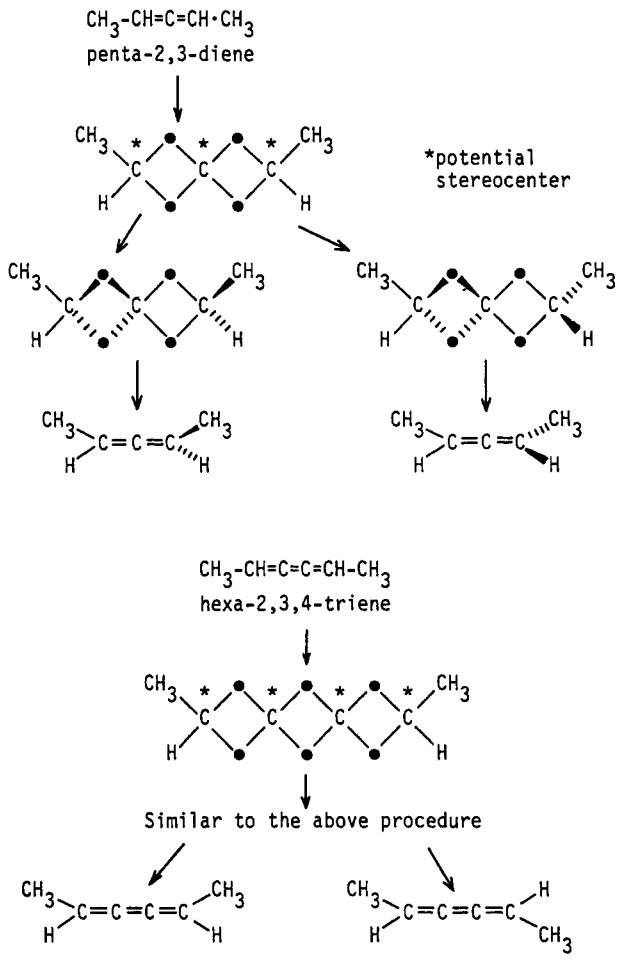


Figure 9. Generation of stereoisomeric structures for allene and cumulene.

structures for penta-2,3-diene and two for hexa-2,3,4-triene as shown in Figure 9.

This method, which is based on a simplified and modified Morgan method, can predict all the possible stereoisomers with their configurations from only the topological information of structural formula, and the method is much simpler and easier for the chemist to handle than the others. In general, generation of stereostructures of highly symmetrical structure has been thought to be difficult. The present algorithm generated four suitable stereostructures with neither omittance nor duplication even for 1,2,3,4-tetramethylcyclobutane. The result was exactly identical with Nourse's. Thus, the method is widely applicable and will undoubtedly become a strong weapon for structure construction in a structure elucidation system.

REFERENCES AND NOTES

(1) Lindsay, R. K.; Buchanan, B. G.; Feigenbaum, E. A.; Lederberg, J. "Application of Artificial Intelligence for Organic Chemistry"; McGraw-Hill: New York, 1980.
(2) Yamasaki, T.; Abe, H.; Kudo, Y.; Sasaki, S. "Computer-Assisted Structure Elucidation". *ACS Symp. Ser.* **1977**, *No. 54*, 108.
(3) Shelley, C. A.; Munk, M. E. "Computer-Assisted Structure Elucidation". *ACS Symp. Ser.* **1977**, *No. 54*, 92.
(4) (a) Nourse, J. F.; Smith, D. H.; Carhart, R. E.; Djerassi, C. *J. Am. Chem. Soc.* **1980**, *102*, 6289. (b) Nourse, J. G. *J. Am. Chem. Soc.* **1979**, *101*, 1210. (c) Nourse, J. G.; Carhart, R. E.; Smith, D. H.; Djerassi, C. *J. Am. Chem. Soc.* **1979**, *101*, 1216.
(5) Wipke, W. T.; Dyott, T. M. *J. Am. Chem. Soc.* **1974**, *96*, 4834.
(6) Wipke, W. T.; Dyott, T. M. *J. Am. Chem. Soc.* **1974**, *96*, 4825.
(7) Morgan, H. L. *J. Chem. Doc.* **1965**, *5*, 107.
(8) Shelley, C. A.; Munk, M. E. *J. Chem. Inf. Comput. Sci.* **1977**, *17*, 110.