

- had declined further, with TeX rising to 140 (late March). The American Physical Society accepts only their version of TeX (RevTeX): their costs of converting other versions make rekeying the paper their production method of choice (personal communication, March 19, 1990).
- (13) *Chicago Guide to Preparing Electronic Manuscripts for Authors and Publishers*; The University of Chicago Press: Chicago, IL, 1987.
 - (14) On Sept 13, 1989, an informal meeting was held in Miami Beach, FL, attended by representative from the ACS, The Royal Society of Chemistry, John Wiley & Sons, Elsevier Scientific Publishers, and Science Typographers, Inc. The purpose of the meeting was to share experiences in direct author submissions in electronic form. All present agreed it is difficult to support electronic submissions for journals, though it has proven cost effective in some cases for book submissions. Also, all agreed that data content specification as well as the text needed to be captured by the publisher and that authors needed to be enticed to do this. Most agreed that storing (and thus ideally capturing) documents with the content-specific information marked was essential for subsequent use of the data, whether in electronic or in reformatted print product.
 - (15) From 1976 to 1988, 18 papers have been published by using the electronic version of the manuscript, 11 in *Anal. Chem.*, 3 in *Environ. Sci. Technol.*, 2 in *J. Org. Chem.*, 1 in *J. Am. Chem. Soc.*, and 1 in *Inorg. Chem.* Additional papers were available on diskette, but based upon the experience with those published and the complexity of the material, subsequent work was done from the hard copy.
 - (16) Brogan, M. C. Manuscript Submissions in Machine-Readable Form. In *The ACS Style Guide*; Dodd, J. S., Ed.; American Chemical Society: Washington, DC, 1986; Chapter 5, pp 149-157.
 - (17) Brogan, M. Analytical Chemistry—A New Approach. *Anal. Chem.* **1977**, *49*, 557A.
 - (18) Brogan, M. Electronic Manuscripts—One Step Closer. *Anal. Chem.* **1984**, *56*, 784A.
 - (19) Warner, M. Electronic Publishing in Analytical Chemistry. *Anal. Chem.* **1987**, *59*, 1021A.
 - (20) Many publishers do not currently include generic coding in their composition process for data-element identification. In processing non-ACS composition data for file building for Chemical Journals Online (CJO), composition data files must be handled as described in this article. Programs for doing these operations were developed by D. P. Martinsen.
 - (21) *SoftQuad Author/Editor User's Manual*, Version 1.1; SoftQuad, Inc., 1989.
 - (22) Blackmore, J. (Royal Society of Chemistry, Information Services, Thomas Graham House, Science Park, Milton Road, Cambridge, CB4 4WF, U.K.). Personal communication, Aug 11, 1989.
 - (23) It is not apparent how the spreadsheet program Lotus 1-2-3 is used to draw chemical structures, but authors did report relatively high use of this package for that purpose.

Computer-Assisted Structure Generation from a Gross Formula. 3. Alleviation of the Combinatorial Problem[†]

IVAN P. BANGOV

Laboratory of Mathematical Chemistry and Chemical Informatics, Institute of Mathematics, Bulgarian Academy of Sciences, Building 8, Sofia 1113, Bulgaria

Received May 22, 1989

The problem of generation of an exorbitant number of combinatorial operations in the process of isomer enumeration is discussed. The origins of the duplicated structures (isomorphic in the graph-theoretical sense) are examined. A novel approach leading to a substantial reduction of the redundant combinatorial operations is described. Two interrelated schemes: Hierarchical Saturation with Equivalent Saturating Valences (HSESV) and Hierarchical Selection of Saturation Sites (HSSS) are developed and their efficiency is illustrated. Various ways of employment of the available structural and spectral information for alleviation of the combinatorial problem are discussed.

INTRODUCTION

Structure-elucidation systems are designed to produce one or several plausible answers from a limited amount of structural information. This involves the generation of different optional structures, a process which requires structure-generation programs (generators). The most severe problem in the development of such programs is the generation of an enormous number of combinatorial operations for all but the smallest molecules. Most of them result in either chemically inconsistent or redundant structures.

Many approaches to avoid the redundancy have been discussed in literature. Thus, a linear-notation algorithm based on canons of precedence that order the branches of each tree-like acyclic structure was developed by Lederberg et al.² However, a comparison of each newly generated structure with the canonical representations of the structures previously generated is impractical. More advanced is the scheme based on the sequential execution of the "partition" and "labeling" steps exploited in the cyclic structure generator of DENDRAL.³ Here duplication is avoided in the step labeling by taking into account the topological symmetry.⁴

An elegant mechanism using the concept of connectivity stack⁵⁻⁷ was devised in CHEMICS. The formation of the can-

onical (greatest connectivity stack) structure requires an exhaustive generation of all segment permutations. Thus since the segments (primary, secondary, and tertiary components⁵) are of small size, this is a computer-intensive procedure in the cases of larger molecules. Some new developments aiming at reduction of the isomorphism checks have been recently reported,⁷ but this still remains a major problem for the CHEMICS generator. Molecular fragments (pieces of the molecular structure with known connectivity between their atoms) are also employed within this scheme, but they are initially degraded into secondary and tertiary components and then the latter are used in the construction process. The existence or the absence of a given fragment is perceived by a substructure search algorithm⁷ (an additional time-consuming procedure) applied to each generated structure.

A similar approach to the generation of nonisomorphic graphs has been discussed by Faradjiev.⁸ The method implies selection of a graph from a set of isomorphic graphs, which is considered canonical, and the program further generates only canonical representations, using a predicate for canonicity. The maximal adjacency matrix was chosen as the canonicity predicate. Obviously, this predicate corresponds to the connectivity stack in CHEMICS. In the same way, the isomorphism checks are carried out by generating all the permutations, and the fragments can be handled only by a substructure perception algorithm.

[†] For Part 2 of this series see ref 1.

In contrast, our efforts were directed toward the development of a generation scheme which reduces as much as possible the generation of the numerous "trial and error" combinatorial operations to the indispensable ones. We prefer a straightforward handling of the fragments in the same manner as the single atoms instead of their perception after each structure generation.

Recently, a novel approach, structure generation by reduction, (COCOA) was suggested by Munk et al.⁹ They classify all the methods discussed above as well as their own earlier generators ASSEMBLE¹⁰ and COMBINE¹¹ as "structure assembly" methods. In contrast, their new method starts just the other way: a universal set of all possible bondings. The bonds are removed instead of being added as in the assembly methods. Each removal leads to an intermediate structure, called a "hyperstructure", having more bonds than the final one. The advantage of the method, according to the authors, is efficient and flexible handling of various and ambiguous information. It includes overlapping, forbidden, or alternative structural information on one hand and flexible use of symmetry as a constraint on the other. Here however, as in the generators previously described, the presence or the absence of a fragment is tested again by a substructure matching algorithm after each intermediate or complete structure is generated. Hence, the input fragments do not operate as a priori constraints on the number of combinatorial operations.

It is obvious that the most considerable reduction could be achieved with the employment of large fragments referred to as *superatoms*. The concept of superatom was implemented in the CONGEN structure generator.¹² Thus, the fragments are considered superatoms when they possess any number of free valences. The inner structure of the fragment (atoms and their saturated valences) is not taken into account. Whereas the free valences participate in the formation of complete structures, the saturated valences remain "invisible", i.e., they simply do not participate in the combinatorial process. This results in a sharp decrease in computation time when large fragments are employed.

In CONGEN¹² the superatoms participate in the combinatorial process as dummy subunits with no information about their chemical nature. Intermediate structures are first generated by using only their names (partition and labeling steps being executed). Subsequently the "imbedding" technique is applied whereby superatom names are substituted with their identities. At this level the user is able to eliminate a large number of final structures by removing (pruning) a small number of intermediate structures. The user also has the capability to define different constraints under which the generation process is further carried out.

A further development of CONGEN directed toward the processing of the ambiguity is GENOA,¹³ a program which employs overlapping and alternative structural information. The efficiency here is achieved by utilizing comparatively large fragments, although certain parts of them are of ambiguous connectivity.

The handling of the fragments as superatoms in CONGEN and GENOA may be regarded as a very significant development. However, we do not favor this approach because of its complicated algorithmic scheme consisting of numerous levels: partition of the gross formula, generation of vertex graphs by using a catalogue for the cyclic structures (3000 elementary rings are compiled in GENOA), their labeling and further imbedding of the superatoms, etc.

In recent papers,^{1,14} we have shown that the same results can be achieved by an appropriate modeling of both the molecular graph representation and the generation process. However, some problems such as graph isomorphism and the capability of handling different constraints were left unre-

Table I. Correspondence between Some Graph-Theoretical and Chemical Notions

graph-theoretical notion	notation	chemical notion	notation
vertex of graph	$v \in V$	atom from chemical structure ^a	$a \in A$
vertex degree	d	atom valency	n
edge of graph	$e \in E \subseteq (V \times V)^b$	chemical bond	$b \in B$
graph ^c	$g \in G = (V, E, f)$ (f is incidence function) ^d	chemical structure	$s \in C^s$
subgraph	$sg \in SG \subseteq G$	chemical group ^e fragment ^f	$cg \in C^g$ $fr \in F$
		segment ^g	$sg \in (A \cup C^g \cup F)$
bonding site	$BS \in BS$	free valence	$f \in FV$

^a Only heavy (non-hydrogen) atoms are included in this definition. ^b $V \times V$ is a Cartesian product of the elements $v \in V$ resulting in all the pairs (v_1, v_2) which define the elements $e \in E$. ^c The chemical structure is usually represented in colored irregular graph. ^d The incidence function f is a function defined on E which assigns to each edge $e \in E$ exactly one pair of nodes $v_1, v_2 \in V$.¹⁵ ^e Chemical group is defined here as a group consisting of a heavy atom with its attached hydrogen atoms. ^f Fragment is defined here as a group of atoms consisting of more than one heavy atoms with or without its attached hydrogen atoms. ^g Segment is defined as a general name encompassing single atoms, fragments, and chemical groups (see ref 6).

solved. Accordingly, we report our efforts toward their solution. They are based on the following presumptions:

First, the generation of duplications must be avoided rather than traced afterwards. Hence, we found that the handling of the fragments within our generation scheme requires new ideas concerning graph isomorphism.

Second, any spectral and/or chemical information should be incorporated in the generation process in such a way that only combinatorial operations and structures consistent with this information are produced. Thus, instead of the usually applied scheme, generate-test-survive or prune, we favor a modeling of the structure representation and the generation process which a priori discards the generation of structures not complying with the input structural and spectral information.

Third, each step of the generation process must be dynamically controlled and directed toward the formation of structures consistent with the input information.

FUNDAMENTALS OF THE METHOD

The basic ideas of our structure-generation approach were reported in recent papers.^{1,14} To aid in clarifying the following discussion, its general features are outlined here.

As is well known, a one-to-one correspondence exists between chemical structure and some classes of topological graphs. The basic notions determining this relation are provided in Table I.

Let us substitute the set E in the graph definition from Table I with its equal $V \times V$. Then the following expression holds:

$$g = (V, E, f) = (V, V \times V, f) \equiv (A, A \times A, f) \quad (1)$$

The product $V \times V$ ($A \times A$) can be further substituted with an operator Γ which maps vertices (atoms) into other vertices (atoms) in the vertex set V (atom set A). Each edge is defined by such a mapping operation.

We consider the fragments superatoms. In order to encompass them in this definition we were forced to use the set

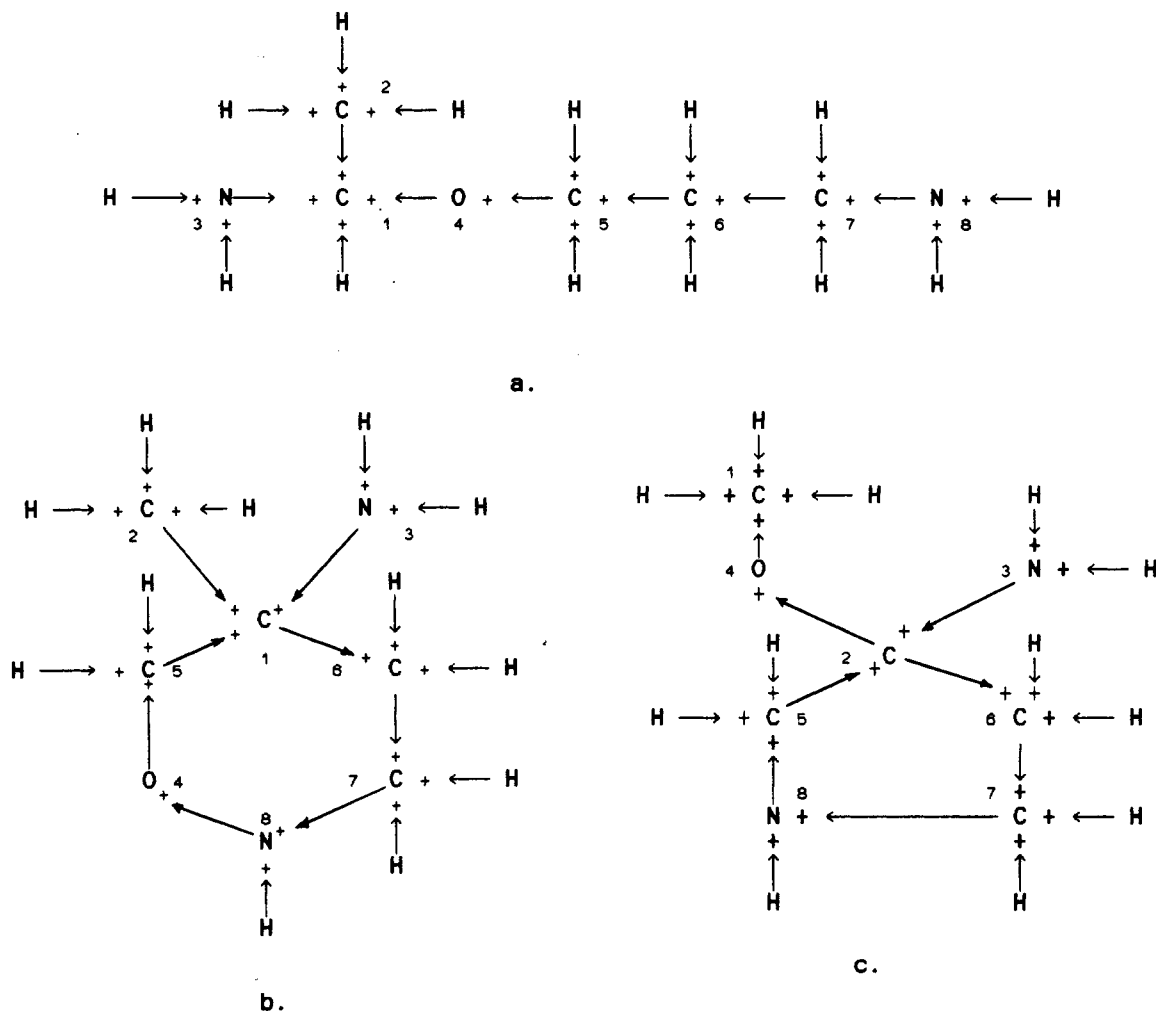


Figure 1. Drawing of directed graphs: (a) acyclic structure representation; (b) and (c) cyclic structure representations.

of bonding sites BS (free valences FV, respectively) instead of the set of vertices (atoms A). Accordingly, the definition (eq 1) can be modified as follows:

$$g = (BS, \Gamma \cdot BS, f) \equiv (FV, \Gamma \cdot FV, f) \quad (2)$$

If $\Gamma = P$ is a permutation generation operator, then from a set of N BSs one can build $N!$ permutations resulting in structures. Most of them have no chemical sense, a great part are disjointed, and a very small part are real nonisomorphic (topologically distinct) structures.

For further reduction of the redundancy we have employed directed graphs.¹ Structures represented by directed graphs can be drawn as shown in Figure 1. One can see that there are two types of bonding sites: \leftarrow and $+$ (in ref 1 they are denoted as * and +). Hereafter we shall name them Saturating Valences (SVs) and Saturation Sites (SSs), respectively. The following rules can be drawn from Figure 1a: (i) all the atoms but the first have $n - 1$ SSs and one SV; (ii) the first atom has n SSs and no SV. Here n is the number of bonding sites (free valences) of a given vertex (atom), i.e., the degree of this vertex.

Cyclic structures have a slightly different representation; e.g., in the case of Figure 1b the cycle closes at the first atom. It is seen that the closure bond BS at this atom is already of the SV type, i.e., the first atom has $n - 1$ SSs and one SV. Hence, the following rule holds: (iii) For every cycle closing atom one BS is transformed from SS to SV. The closure bond SVs were denoted as] in the present version of the program.

This is exemplified additionally in Figure 1c with the second atom selected as a closure atom. Here again one SS is transformed to a closure bond SV. Thus, the second atom already has $n - 2$ SSs and two SVs, the former because of rule

ii and the latter following rule iii as a closure bond SV.

Accordingly, our approach to structure generation is based on the following assumptions: All the bonding sites of the separate segments are partitioned into two sets

$$SS \in S \text{ and } SV \in V$$

according to rules i–iii. A mapping of the SV elements onto the set S produces a set of graphs G . Following eq 2 this can be written symbolically as follows:

$$g = (V, P \cdot S, f) \quad (3)$$

Practically eq 3 can be explained as follows: Let the set V contain m elements and the set S contain L elements. The mapping operator generates mP_L permutations of m elements selected from all L SS elements of the set S without repetition, and maps them to the m SV elements from the set V .

Fragments are considered superatoms, and n is equal to the number of their bonding sites. The same rules are applied to them: (iv) $n - 1$ BSs of each fragment but the first are of the SS type, and one BS is of the SV type; (v) the first fragment has all its n BSs of SS type. In contrast to the single atoms and chemical groups, the BSs of the fragments are no longer equivalent. Hence, each one must be in turn one of the SV type, while the other $n - 1$ are of the SS type.

The atoms forming multiple bonds are considered varieties of the basic atoms. Thus, carbon, oxygen, and nitrogen atoms forming double bonds are represented as the newly defined $=C$, $=O$, and $=N$ atoms, having $n = 3$, 1, and 2, respectively. They will produce again $n - 1$ SSs and one SV in the structure formation process. The same is true for the triple-bond-forming atoms coded as $\#C$ and $\#N$ and having $n = 2$ and 1, respectively.

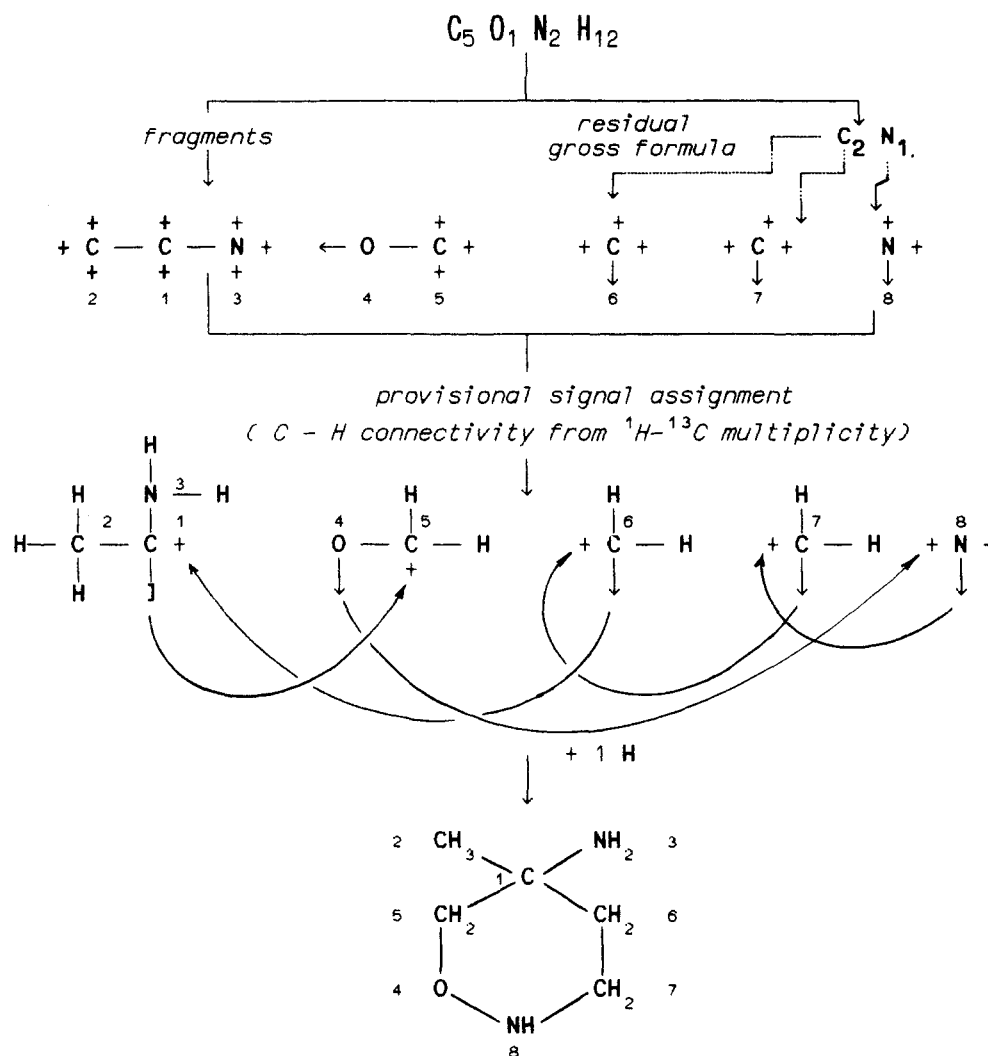


Figure 2. Depiction of the structure generation presented in Figure 3.

Our structure generation method is depicted in Figure 2 with the generation of the isomer b from Figure 1. Its program implementation is illustrated in Figure 3. Initial and basic information is provided by the gross (molecular) formula C₅ON₂H₁₂. The input of fragments provides a substantial constraint on the generation process. As shown in Figures 2 and 3, the use of the fragments C—C—NH₂ and O—C partitions the gross formula into two sets of atoms: the former of known fixed mutual connectivity and the latter of unknown connectivity, i.e., a reduced gross formula—C₂NH₁₀. This pattern mimics in some way the knowledge that a chemist has in the majority of cases about an unknown compound. Usually he has some knowledge about the connectivity within a part of the structure and no precise knowledge about the connectivity within the rest of it. As shown in Figure 3, often in an overlapping area between the "fixed connectivity knowledge" and the "no connectivity knowledge" parts lies the "ambiguous (alternative) connectivity knowledge". At this stage of development of our structure-elucidation approach, we do not deal directly with the ambiguous connectivity knowledge. Some elements of its processing only will be outlined in this paper.

The available structural information is further transformed into the mathematical representation of the molecular structure (Figure 3). The SVs form a vector array SUBS, and the SSs, a two-row array GRAPH. Following rule v formulated above, in the general case no atom from the first fragment should provide a SV. However as long as the first atom is a closure bond atom (rule iii), one BS is transformed into SV (denoted as J). Following rule iv the second fragment O—C provides

one SV (that of the atom O). All other atoms (6,7,8) from the reduced gross formula provide one SV and $n - 1$ SSs, as was outlined above. The atom C of the second fragment is not equivalent to the atom O. Hence, after a complete generation of all the isomers, a new set of SVs is formed with the second fragment SV taken from atom C instead of from atom O, and all the isomers are again generated.

The ± elements of the second GRAPH row indicate free SSs, i.e., parts of the molecular structure of unknown connectivity, while the occupied second-row elements indicate SSs saturated with SVs in the parts of known fixed connectivity within the structure. The juxtaposing of a GRAPH(1) to a entry different from + in GRAPH(2) represents a bonding of a SS with a SV. In this way the GRAPH array represents both the fixed fragment connectivity knowledge and the lack of knowledge about the connectivity within some parts of the query structure.

As the GRAPH and SUBS arrays are formed by taking into account the proper number of free valences, a permutation (5,1,3,4,2 in Figure 3) produces a new chemical structure with no further matching of valences required.

The hydrogen atoms possess only SV-type BSs. Since they are equivalent, no permutations among them are carried out. They simply fill the residual unsaturated SS elements after the complete formation of the all-heavy-atom structure skeleton. We call this approach **Hydrogen Atom Saturation of the Residual Bonding Sites (HASRBS)**. It produces in a natural way various alternative attachments of the hydrogen atoms forming different groups, e.g., a nitrogen atom may appear as both NH and NH₂ groups in different structures.

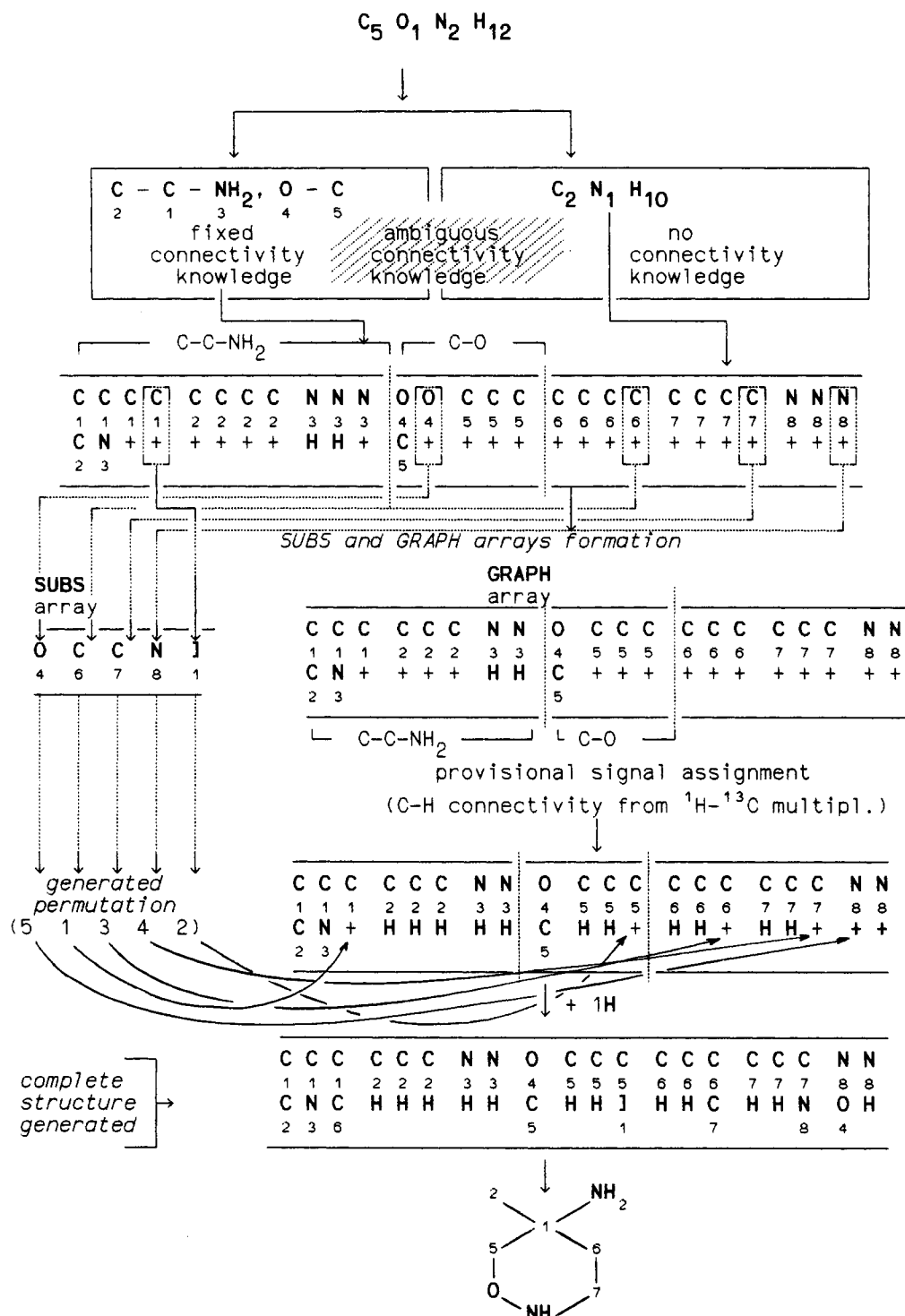


Figure 3. Partitioning of the gross formula into atoms of known (fixed and ambiguous) and unknown connectivity; formation of the GRAPH and SUBS matrix representation of the molecular structure and structure generation.

However to reduce the number of the SSs, we favor a provisional assignment of the carbon atoms with the ^{13}C NMR signals of the query structure spectrum. Each signal is fed into the computer with its chemical shift and 1H - ^{13}C direct multiplicity. While the former is of importance for the structure elucidation, the latter directly participates in the structure-formation process specifying the number of hydrogen atoms adjacent to each carbon atom. The known connectivity part of the structure (the input fragments) is assigned by the user, while the remaining signals are automatically assigned to the carbon atoms of the unknown connectivity part by the program (practically this procedure follows the order of their transformation from the reduced gross formula to the GRAPH and SUBS array representation). Every experienced chemist

familiar with the fragment structure is able to carry out such a provisional assignment. Moreover, we assume the user-driven assignment to be correct only in its multiplicity part. Thus, an interchange of the chemical shifts of two signals of the same multiplicity will produce no error since the program carries out further automated assignment of the chemical shifts.¹⁶

The ^{13}C - 1H direct multiplicity assignment decreases the number of SSs (see Figure 3), and accordingly the number of the combinatorial operations by saturating $M - 1$ SSs with H atoms (here M is the signal multiplicity). Thus, while ${}^5P_{15} = 360\,360$ permutations are necessary without provisional assignment for the example provided in Figure 1a, only ${}^5P_6 = 720$ permutations are generated after the provisional assignment is carried out. Consequently, the HASRBS approach

is applied here only with respect to the heteroatoms if no information about their bonding with hydrogen atoms is provided.

It should be emphasized here that the present method might be developed without any user-driven signal assignment. However, this will require more alternatives to be considered which leads to a substantial increase in the number of combinatorial operations and generated structures. At this stage in development of the program, we find that a compromise between the automated and human-driven inference is more adequate.

Generally speaking, the employment of directed graphs within this approach leads to a substantial decrease of the number of combinatorial operations. For instance, there are 11 BSs in the case of the gross formula $C_5O_2H_{14}$ obeying the following constraints: NH_2-C-CH_3 , $-O-$, $-CH_2-$, $-CH_2-$, $-CH_2-$, NH_2- . If one carries out all the $11! = 39\,916\,800$ permutations the problem becomes practically unfeasible. The use of directed graphs according to this representation lowers their number to ${}^5P_6 = 720$. However, although very spectacular, this reduction is not sufficient for most real-world problems.

ALLEVIATION OF THE COMBINATORIAL PROBLEM

The expression ${}^mP_L = L!/(L-m)!$ shows that any reduction of the combinatorial operations can be achieved by reducing both the number m of the SV elements and the number L of the SS elements. Such a reduction can be affected also by finding a way to substitute the explosion-making permutations with less numerous combinatorial operations, e.g., combinations.

The restrictions on SSs and SVs can be imposed either on purely graph-theoretical and/or on chemical grounds. Further, we shall discuss their practical implementation to the method described above.

It is apparent that the duplications (isomorphic graphs) are due to permutations of equivalent elements. Hence, the following proposition will be proved:

Let G be an irregular graph having overall m bonding sites. A set G^m of $m!$ graphs is generated after carrying out $m!$ permutations of the m bonding sites. If two of the bonding sites are equivalent then $m!/4$ nonisomorphic graphs result.

Proof: Consider the graph g followed by the following two-row array of size m :

a b c c d q u w v
d q u e f c c a b

Here the elements q,b,c...w,v are the bonding sites. All the permutations of the m elements from the second row form a symmetric group S^m which is represented by the set of graphs G^m . Let two elements (c,c) be equivalent. We partition the set of elements E^m into two subsets: E^2 containing the two equivalent elements and the set E^{m-2} formed from the remaining $m-2$ elements.

The E^2 elements produce 2P_m permutations of two elements selected from the m different BS positions, and for each permutation the E^{m-2} elements produce $(m-2)!$ permutations, resulting in the subset G^{m-2} of subgraphs. Accordingly the following equation holds:

$$m! = {}^2P_m(m-2)!$$

As the $(m-2)!$ permutations trace all the branchings within the G^{m-2} subset, two equivalent subsets G^{m-2} of one-to-one isomorphic graphs for each pair of the 2P_m permutations result.¹⁷ Each subgraph of the former and the latter subsets forms a full graph by bonding with either the initial or the

second permutation of the (c,c) elements. Inasmuch as the two permutations are equivalent, two sets of one-two-one isomorphic graphs result. Hence the number of nonisomorphic graphs equals half of the whole number of graphs:

$$m!/2 = {}^2P_m(m-2)!/2 = {}^2C_m(m-2)! = {}^2C_m \prod_{i=1,n} {}^1C_{(m-2)-i} \quad (4)$$

2C_m is the number of combinations of two elements selected from m elements.

The two equivalent elements (c,c) in the first row are the second source of isomorphic graphs. Thus, permuting two nonequivalent elements bonded to them is equivalent to a permutation of the latter. Here we have ${}^2P_{m-2}$ different permutations of two elements selected from the remaining $(m-2)$ elements. In the same way as above we obtain

$$(m-2)!/2 = {}^2P_{m-2}(m-4)!/2 = {}^2C_{m-2}(m-4)!$$

Hence the total number of nonisomorphic graphs is

$$m!/4 = {}^2C_m {}^2C_{m-2}(m-4)!$$

Thus, any permutation of the two equivalent (c,c) second row elements leads to a repeated bonding of the q,u first row elements to the element c. In the same way any permutation of the u,e elements of the second row leads to a repeated bonding with a c element from the first row.

Two approaches aiming at excluding as much as possible the duplications were developed: the Hierarchical Saturation with Equivalent Saturating Valences (HSESV), dealing with the first type of isomorphism, and the Hierarchical Selection of the Saturation Sites (HSSS), dealing with the second type.

The HSESV Approach. This approach implies a partitioning of the SVs into equivalence (automorphism) classes and carrying out only the permutations which are between elements of different classes. The number of combinatorial operations in this case is equal to the number of combinations given by the following generalization of eq 4:¹⁸

$${}^{m_1}C_L {}^{m_2}C_{L-m_1} {}^{m_3}C_{L-m_1-m_2} \dots {}^{m_n}C_{L-\sum_i m_i} \quad (5)$$

This expression assumes a hierarchical ranking of the SVs from the separate equivalence classes into different levels of generation. Here ${}^{m_r}C_{L-\sum_{i<r} m_i}$ is the number of combinations generated at the r th level of m_r SSs selected from the $L - \sum_{i<r} m_i$ SSs left unsaturated from the lower level. For each combination at a given level the m_r SSs are saturated with the m_r equivalent SVs, and all the combinations of the nested higher levels are consequently generated (a depth-first procedure).

The automorphism partitioning of the SVs is carried out as follows. Initially, they are partitioned into the following basic levels:

h levels, consisting of all the segment SVs originating from heteroatoms.

c levels, consisting of SVs originating from carbon atoms of segments having valency greater or equal to 2. Here the molecular skeleton is formed.

] levels, consisting of SVs forming ring-closure bonds. Here the cyclic structures from the molecular skeleton (spanning tree) are formed.

1 levels consisting of SVs from univalent segments. Here these segments are attached to the remaining SSs on the molecular skeleton.

Each one of these levels are additionally split into sublevels according to an index assigned to the atoms that the SVs are originating from. Recent papers^{16,19} introduce the Atom-in-Structure Invariant Index (ASII):

$$ASII = ASII_0 - N_H + Q_{at} \quad (6)$$

Here N_H is the number of hydrogen atoms attached to a given

Table II. Initial Atom-in-Structure Invariant Index (ASII₀) Values

atom (hybridization state)	ASII ₀	atom (hybridization state)	ASII ₀
C		O	
sp ³	4	sp ³	23
sp	7	sp ²	25
sp ² (olefinic)	11	F	32
sp ² (aromatic)	13	S	28
N		Cl	33
sp ³	15	Br	34
sp ²	18	I	35
sp	20		

heavy atom, Q_{at} is the charge density calculated through a fast method (we used the IPEOE charge-computational scheme of Gasteiger et al.²⁰), and ASII₀ is an initial value characterizing the atom in a hybridization state (the corresponding values are provided in Table II). In fact the Q_{at} term accounts for the α , β , γ , ..., etc. environments of the atom. A similarity between ASII and Morgan's Extended Connectivity Index²¹ or other local invariants^{22,23} may be found.

As discussed in ref 16 the ASIIs are equal for equivalent free atoms, chemical groups (e.g., two or more CH groups), and topologically symmetric atoms and groups in fragments; different for nonequivalent atoms, groups, and topologically nonsymmetric atoms in fragments; and nonoverlapping for atoms of different types and hybridization states.

Consequently, the SV elements occupy various levels according to the following hierarchical rules:

- the larger the ASII of a given SV, the lower level it occupies
- equivalent SVs (having the same ASIIs) occupy the same levels (with the exception of j-type SVs)
- j-type SVs always occupy different levels (This is because closing of a cycle depends on the atom numbering. Hence, all the ways of closing a cycle must be traced, i.e., permutations instead of combinations for the j-type SVs to be generated.)

Further, the program selects some specific levels:

HiLev is the highest level. A combination at this level produces a complete structure.

SkLev is the highest level of the c-type. The all-heavy-atom skeleton of the structure is formed here. After each combination at this level, the program checks whether a complete skeleton is generated or a disjointed skeleton results. If the skeleton test fails (an empty circle in Figure 4), the combination generation proceeds at the current level, otherwise it proceeds to the higher level.

CyLev is the lowest level of the j-type. A spanning tree is formed after the generation of a combination at the SkLev level. Then the rings are closed by saturating the remaining SSs with the closure-bond SVs. Starting from this level a size-of-cycle test is carried out at all higher j-type levels. This test will be discussed below.

The HSESV approach can be visualized by the combination tree presented in Figure 4. For clarity in the drawing, only one level for each level type (h, c, j, l) is presented. Every node of this tree represents a combination generated at the corresponding level. The tree-like structure of the generation process shows the way to further reduction of the number of combinations. It is clear that most of the combinations at each level are redundant. They must be recognized at the lowest level and subsequently pruned. Thus, the various combinations are divided into "successful" (denoted as filled circles) and "unsuccessful" (denoted as empty circles). Our approach to this problem is discussed in the next sections.

The HSESV approach to structure generation is exemplified in Figure 5 with the case of the gross formula $C_{10}OH_{18}$. Here ${}^8P_{26} = 6.2990928 \times 10^{10}$ permutations of eight SSs selected

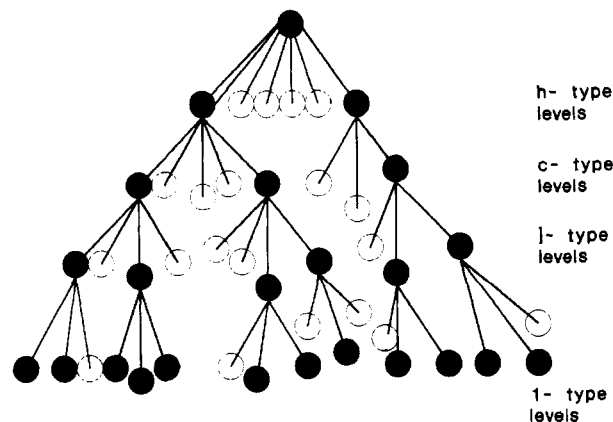


Figure 4. Combination tree representation of the structure generation process. Only one of each of the level types: h, c, j, l is presented. The filled circles depict successful combinations; the empty circles, unsuccessful combinations.

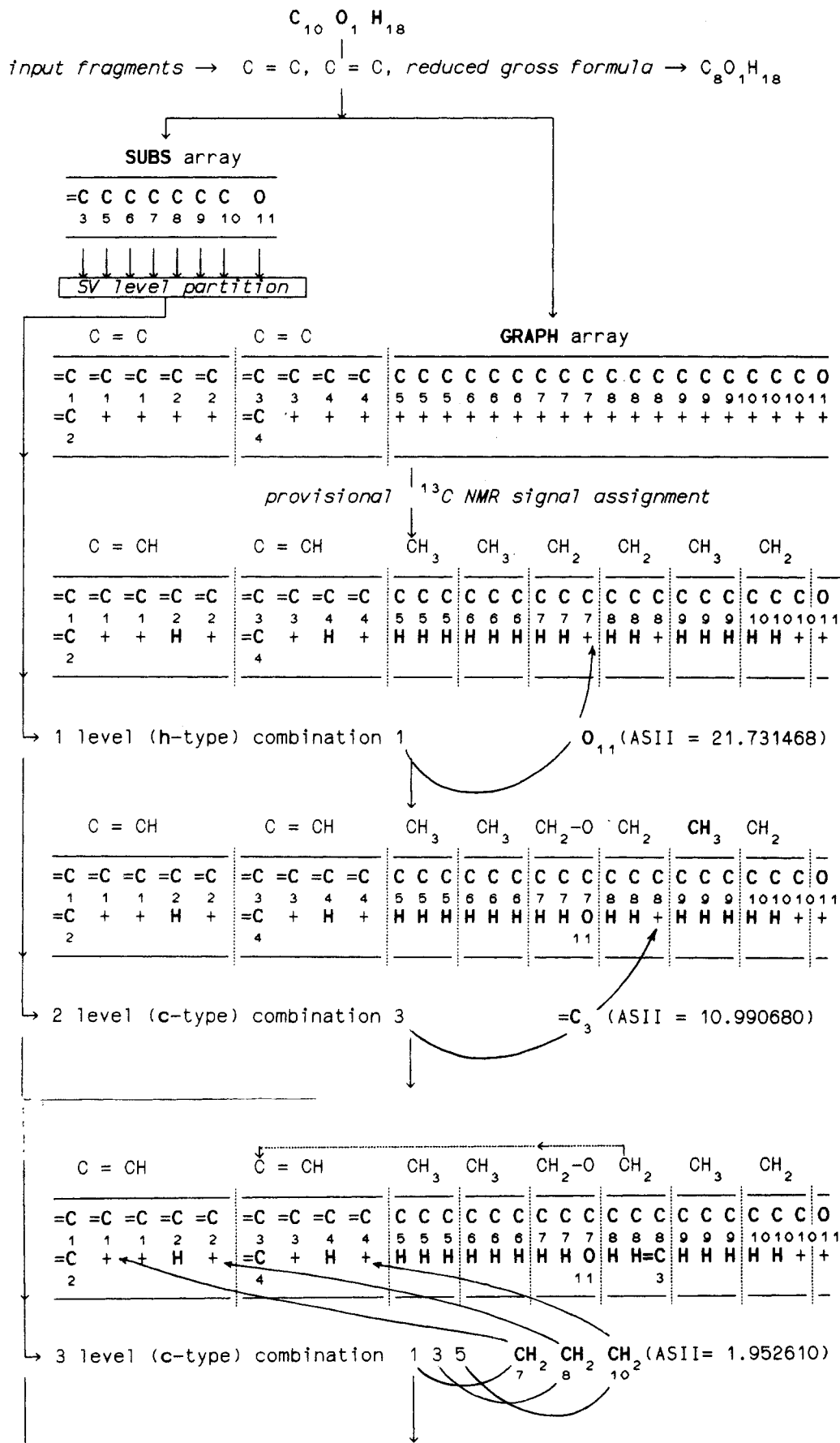
from the 26 SSs without repetition and saturated with eight SVs would have been generated by the standard version of our method. The employment of the 1H - ${}^{13}C$ multiplicities within the provisional ${}^{13}C$ NMR signal assignment reduces their number to ${}^8P_9 = 362880$ (eight SVs versus nine SSs). By applying the HSESV approach, the SV elements are partitioned into four equivalence levels according to their ASIIs (see the ASII values in Figure 4). The first is a h-type level (one oxygen atom SV). ${}^1C_9 = 9$ combinations are generated here. The next two levels are of the c-type with ${}^1C_8 = 8$ combinations for the second, and ${}^3C_7 = 35$ for the third. The fourth level is a l-type level involving ${}^3C_4 = 4$ combinations. Thus, the total number of combinations is ${}^1C_9 \cdot {}^1C_8 \cdot {}^3C_7 \cdot {}^3C_4 = 10080$. Evidently this number is still rather large and its further reduction is carried out through the HSSS approach discussed below.

Here we shall try to make the meaning of eq 5 and the HSESV method clearer. As discussed above, the most straightforward method to deal with the isomorphism is to determine one out of all isomorphic structures as canonical (that of maximal adjacency matrix in the method of Faradjiev or with greatest connectivity stack in CHEMICS). In other words, at each step of the graph growth we select one out of all permutations producing isomorphic subgraphs. Consider the generation process at level 4 in Figure 5. The CH_3 groups are numbered in an ascending order (5,6,9). Most of the permutations leading to isomorphism violate this arrangement; e.g., the permutation 3,2,1 which is isomorphic to permutation 1,2,3 (permutes equivalent groups) would lead to decreasing the adjacency matrix since it corresponds to the arrangement 9,5,6 of the CH_3 group numbering, hence to noncanonical structure. This is shown in Figure 6 with the corresponding sections of the adjacency matrices and their characteristic vectors.²⁴ It is seen that the vector from the arrangement 9,5,6 is smaller than the one formed from the initial arrangement 5,6,9. The same is true for the three equivalent CH_2 groups at level 3. However, here the isomorphism is not so transparent as in the previous case.

In contrast, the combinations within the HSESV approach preserve the segment arrangement. Hence, whereas in CHEMICS and in the method of Faradjiev the permutations are first generated and subsequently tested for being canonical, most of them are simply skipped by the HSESV approach. This leads to a sharp decrease of the number of combinatorial operations.

Further some practical problems concerning the development of the method are considered.

As stated above, the set V of SVs partitioned on several levels is formed from the segment BSs (one SV from each segment) and from the cycle-closure BSs. While the SV



the same level, only one combination (${}^9C_9 = 1$), i.e., one isomer will be generated. In contrast, by partitioning the SVs into two levels: five equivalent SVs on a c-type and four non-equivalent SVs on a j-type, the permutations between the two types are no more prohibited which leads to generation of different isomers.

(B) Selection of Fragment SVs. As discussed above the fragment BSs are generally not equivalent. They are equivalent only when they are topologically symmetric, e.g., the two CH BSs in the fragment ${}^+CH-N_+-CH^+$. The generation process implies that each nonequivalent BS of every fragment but the first must be in turn once SV, when the others are SSs. This is necessitated by the fact that all the BSs must be mutually bonded. Since all the first segment BSs are of the SS type, if a BS from another fragment has not been converted either into cycle-closure or into fragment SV the two BSs, hence the two fragments, will never form a bond. Accordingly, the following rule applies:

All the BSs of each fragment, but the first, are selected in turn as SVs under the following conditions:

- (i) the current BS is not equivalent to any other previously selected
- (ii) the current BS has not been selected as a cycle-closure SV

Those conditions might be violated if at least one BS is not converted into fragment SV (each fragment but the first must provide one SV in the combinatorial process). For example, there are three SSs at the second fragment $C=CH$ in Figure 5. Following the rules stated above they form ${}^1C_2 = 2$ combinations, one of them (from atom 3) being already converted into SV. The second combination is formed by converting one of the two SSs at atom 8 into SV.

The HSSS Approach. After the generation of a successful combination of the m_r SVs at a level r , the SSs for the higher level are selected. Such a selection allows both the permutations between SVs, bonded to equivalent SSs, to be avoided and any additional topological, chemical, and/or spectral information to be incorporated within the generation process. Basically, these tasks can be approached in two ways:

- (i) by avoiding the selection of SSs which are a priori known as leading to unsuccessful combinations
- (ii) by applying a consistency test after every generated combination.

Whereas ii is applied to one or another form to some generators, e.g., the possibility of imposing constraints on the intermediate structures in CONGEN and COCOA, i shows a new perspective for an a priori avoidance of the generation of unwanted bondings. These two approaches are discussed below.

First, to account for the symmetry of the SSs the following rule is applied:

Consider the level r having m_r equivalent VSs. If the atoms providing SSs for this level are partitioned to n_1 atoms of equivalence class A^1 , n_2 atoms of equivalence class A^2 , ..., etc., then the SSs of each class A^p must be taken from only $p_p \leq n_p$ atoms satisfying the condition $p_p \leq m_r$.

For example, consider the second level in Figure 5. It has $m_2 = 1$ (one $=C$ SV).

Two of the SSs are equivalent (at atoms 8 and 10). Following this rule only one of them (at atom 8) will be selected. This is quite natural since the two CH_2 differ only by their numbering, and the selection of the second SS will produce a set of duplicated structures, as discussed above. Thus, the set of SSs selected for the second level in Figure 5 will be formed of one SS from atom 1, one from atom 2 (the SSs of atoms 3 and 4 are skipped for reasons which will be clarified below), one from atom 8, and one from atom 11, or totally four SSs.

The HSSS approach provides many other ways to reduce the number of SSs, hence the number of combinatorial operations. Further we shall discuss some of them.

(A) Bonding of Atoms to Themselves Is Prohibited. The application of the HSSS in the case of $m_r = 1$ is accompanied by a chemical consistency test. It inhibits selection of any SS originating from the same atom (having the same number) as the sole SV (no atom can be bonded to itself). Thus, in Figure 5 no SS from atom 3 is selected for the sole second level $=C$ SV originating from atom 3.

In the case of $m_r > 1$ such a restriction is obviously impractical. Therefore, the consistency test is carried out afterwards, according to ii. If such a bonding is not detected (a successful combination denoted as filled circle in Figure 4) the process proceeds to the higher level combination generation, otherwise the next combination at the current level is generated and the consistency test is carried out again.

(B) Fragment Atom SS Selection. It is assumed that no input fragment should be bonded to itself, otherwise it must be entered as a cyclic substructure. Hence, all the SSs emanating from the same fragment as the given SV are ignored. For this reason the SSs at atoms 3 and 4 in Figure 5 are not selected for the second level. In the case of more than one SV at a given level, the consistency test is carried afterwards.

(C) Multiple-Bond SS Selection. The multiple bonds are usually formed by pairs of multiple-bond atoms. Therefore, we favor dealing with multiple-bond fragments such as $C=C$, $C=O$, $C=N$, etc. rather than with single atoms ($=C$, $=N$, etc.). Hence, the selection of the SS is subject to the same rules as the processing of the fragment SSs.

Additional tests for the presence or absence of conjugation within the HSSS approach are developed. If the option "no conjugation" is assumed, all the SSs from multiple-bond atoms are ignored during the SS selection for a multiple-bond SV level. In this way any bonding between multiple-bond fragments is a priori avoided, and no conjugation will occur. In contrast, if a conjugation is assumed, then after each combination generation at the highest multiple-bond level the program checks in the GRAPH matrix for at least one bonding between multiple-bond fragments, e.g., a juxtaposing of $=C$ SVs with $=C$ SSs originating from different segments. If such a bonding is detected the combination is considered successful and the process proceeds to a higher level, otherwise it continues with the generation of the next combination at the current level. In this way all branches of the combination tree emanating from unsuccessful combinations are pruned. In cases of ambiguity no test is carried out, and both conjugated and nonconjugated structures are generated.

Thus, for the example presented in Figure 5 and the no conjugation option the number of the second-level combinations is reduced to ${}^1C_2 = 2$ versus 4 in the previous case. As a result only 23 structures are generated versus the total 74 in the case of ambiguous conjugation.

(D) Heteroatom Bonded Carbon Atom SS Selection. It is known that sp^3 carbon atoms attached to heteroatoms resonate at the low-field spectrum regions. Hence, the SVs originating from the heteroatoms are assumed to saturate SSs of sp^3 carbon atoms having chemical shifts in the range 39–100 ppm. Thus for each h-level the program selects only carbon atom SSs having assigned such chemical shifts. Obviously this decreases substantially the number of the SSs at the low-lying h-levels which is conducive to sharp reduction of the number-generated combinations. For example, in the case of Figure 5 the SV from the oxygen atom, according this rule, saturates only the two CH_2 (7 and 10) SSs with chemical shifts equal to 39.70 and 58.70 ppm. Since the two CH_2 are equivalent, only one of them (atom 7) is selected for this level.

Hence, only one instead of the previous nine combinations is generated. In contrast, we allow SVs of carbon atoms with arbitrary chemical shifts to saturate the heteroatom SSs. For instance the SS at atom 11 may be selected for bonding with any carbon atom SV. Thus, structures with abnormally low chemical shifts of carbon atoms bonded to heteroatoms are also generated. In this way some ambiguities in the structural assignment can be treated.

(E) Handling Forbidden Substructures. At each level the generated partial structure can be tested against the presence of forbidden structures. If the selection of a given SS leads to the formation of forbidden substructure, it is simply ignored. Thus, the generation of large classes of unwanted structures is a priori avoided. This procedure can be exemplified with the case of forbidden cycle size. After each j -level combination, the minimum distance matrix is constructed. Hence, any selection of a SS which is to be saturated by a cycle-closure SV at the CyLev equal is tested whether the distance between the SS and SV atoms is equal or greater than required. In this way all the SSs providing smaller cycles are ignored. It must be emphasized, however, that small rings may appear at a higher j -level by intersecting a larger cycle into two smaller cycles. Therefore a procedure for finding the smallest set of smallest rings is also executed after the highest j -type level.

(F) Chemical Group Specification. Any information on the number of chemical groups such as OH or NH (or NH_2) also leads to the reduction of the number of combinatorial operations. Such information can be derived from various spectroscopic sources, e.g., from the intensities of the ^1H NMR spectra. After this number is specified, e.g., N_{OH} is the number of OH groups, the program checks all the GRAPH array elements for any hydrogen atom juxtaposed to an oxygen SS (the OH group may occur as part of a fragment, previously entered). If such a bonding is detected the N_{OH} is reduced by one. Then N_{OH} heteroatom (O atoms in our case) SSs from the first row of the GRAPH array are saturated with H atoms (second row GRAPH elements being filled with H atoms). If $N_{\text{OH}} = 0$ is specified, the program checks that no oxygen SS of the first GRAPH array row is saturated with an H atom in the second GRAPH array row. If such a juxtaposing is detected the program ignores this structure. In this way only ethers are generated. In the case of ambiguous information no test is carried out and the program generates both the OH group and ether structures by the application of the HASRBS procedure at the HiLev level.

The same procedures are applied to the other chemical groups such as NH (NH_2). In particular, we do not distinguish between the latter two groups as it is rather difficult to infer their discrimination from routine spectroscopic and chemical data. Both are fed to the computer and processed as NH groups, and the NH_2 groups are formed as a result of the application of the HASRBS procedure at the highest HiLev level.

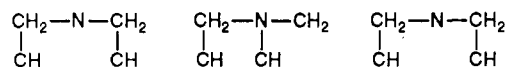
To sum up: the application of the HSSS approach sharply reduces the number of SS elements, hence the number of combinatorial operations. Thus, for the example in Figure 5, their number for the initial set of SVs (second fragment SV from atom 3 taken) is ${}^1C_1 \cdot {}^1C_4 \cdot {}^3C_7 \cdot {}^3C_4 = 560$ (no conjugation and $N_{\text{OH}} = 0$ case) and ${}^1C_1 \cdot {}^1C_3 \cdot {}^3C_6 \cdot {}^3C_3 = 60$ for the $N_{\text{OH}} = 1$ case. By taking into account the ${}^1C_1 \cdot {}^1C_2 \cdot {}^3C_7 \cdot {}^3C_3 = 280$ ($N_{\text{OH}} = 0$) and ${}^1C_1 \cdot {}^1C_2 \cdot {}^3C_6 \cdot {}^3C_3 = 40$ ($N_{\text{OH}} = 1$) combinations generated with the second set of SVs (second fragment SV from atom 4 taken), the total numbers become 840 and 100 combinations, respectively.

However, these numbers might be different because of the tests discussed above and for reasons outlined below.

Dynamical HSESV and HSSS Approaches. Two basic requirements are imposed on every generator: irredundancy and exhaustiveness. Frequently, they contradict. Thus, in our case of an initial determination of the hierarchical levels as constant during the whole generation process, the need for exhaustiveness on a given level requires a full generation of all the combinations at the lower levels. In contrast, the utilization of the topological and chemical information within the HSSS approach restricts the number of SSs. Hence, some irredundant structures might be omitted. For example, if an oxygen SV saturates only one out of several CH_2 SSs (the latter being of higher hierarchical levels), as it is in Figure 5, then a $\text{O}-\text{CH}_2$ fragment results. Therefore the SV from this CH_2 will be no longer equivalent to the other CH_2 SVs. Hence, the permutations between this SV and the other CH_2 SVs are already permissive. In other words the corresponding CH_2 level should be split. We approach this problem by introducing a dynamical determination of the hierarchical levels. After each successful combination at a given level the ASII of the atoms are reestimated, and a new hierarchical ranking of the levels higher than the current one is carried out. For instance whereas three CH_2 SVs at the third level have had initially equal ASII values (see Figure 5), after the first level combination the $\text{O}-\text{CH}_2$ SV has already ASII = 2.077508, the other two mutually equivalent CH_2 ASII remaining equal to 1.952610. Prima facie, the need for calculating the charges at each successful combination makes this approach computationally demanding. However, this is the only way to properly use the HSSS method without omitting some irredundant structures. Obviously, this is the price which one must pay for the sharp reduction of the SSs.

The dynamical estimation of ASII is also required in the process of recognition of duplications which is discussed below.

Additional Controls: Recognition of Duplications. Although the method outlined above avoids a great majority of isomorphic structures, some duplications still appear. They are due to several reasons; e.g., consider the two equivalent CH_2 SSs in the fragment $\text{CH}_2-\text{N}-\text{CH}_2$. They are sequentially substituted with two equivalent, e.g., CH, SVs. Accordingly, ${}^2C_3 = 3$ combinations are generated producing the three possible substructures:



It is obvious that the first and the third substructures are isomorphic.

A second possible source of duplications is the different paths that the cyclic structures close. Each simple cycle has two directions that it can be walked around. Since we have no way to distinguish a priori between them, duplications still appear at the highest j -type level.

It should be admitted that due to the HSESV and HSSS approaches the number of duplicated structures is very small. Nonetheless, to prevent the generation of duplications at an early stage of structure formation we test for isomorphic graphs after each generated combination at a given level. A newly defined global index was used to this end:

$$I = 10 \left(\frac{\sum_{i=j} (\text{ASII}_i \cdot \text{ASII}_j)^{1/2}}{\sum_i (\text{ASII}_i)^2} \right) \quad (8)$$

Here the summation $i = j$ is over mutually connected nodes (i, j) only, and the summation i in the denominator is over all the nodes. For each combination generated at a given level this index is calculated and tested with a list of index values computed from previous combinations. A combination producing partial structure or a set of disjointed partial structures is considered successful, and the generation process proceeds

Table III. Number of Structures, Duplications, and Combinations Generated from a Given Gross Formula and Constraints^a

no.	gross formula	constraints	no. of structures	no. of duplications at HiLev level	no. of combinations
1	C ₁₀ H ₂₂ O	1 OH group, multiplicity of 1-decanol ^b	1	0	17
2	C ₁₀ H ₂₂ O	1 OH group, 1 C=C fragment, multiplicity of 4-decen-1-ol	6	0	57
3	C ₁₀ H ₁₈ O	1 OH group, 2 C=C fragments, multiplicity of linalool, ^b no conjugation	32	0	124
4	C ₁₀ H ₁₈ O	1 OH group, 2 C=C fragments, multiplicity of geraniol, ^b no conjugation	23	0	122
<div style="text-align: center;"> $\begin{array}{c} \text{C} \\ \\ \text{C}-\text{O}-\text{C}-\text{C} \\ \\ \text{C} \end{array}$ fragment </div>					
5	C ₁₀ H ₁₈ O	multiplicity of cineol, ^b 3-membered rings avoided	10	0	94
6	C ₁₀ H ₁₈ O	1 C=C=O, 2 C=C-C, ^c 1 C-C fragments, multiplicity of camphor ^b	14	5	137
7	C ₁₃ H ₂₀ O	1 C=C-C=C-C=O fragment, multiplicity of β -ionone ^b	117	5	782
8	C ₁₁ H ₁₇ ON ₃	1 C-C-C-C-N, 1 NH=C-NH-C=O, and 1 C=C fragments, multiplicity of arenaine ^b	94	0	538

^aAll structures are generated additionally under the heteroatom bonded to carbon atoms of ¹³C NMR chemical shifts in the region 39–100 constraint. ^b¹³C–H multiplicity taken from ref 25. ^cThese fragments are not equivalent due to the different multiplicities of their carbon atoms.

to a higher level if no index from the list coincides with the current index, otherwise the generation continues at the current level. Thus, with the discrimination of the duplicated partial structures at the lowest levels, the combination tree branches from Figure 4 are pruned at an early stage which leads to further reduction of the number of generated combinations. Our experience shows that in the cases of acyclic structures no duplications appear at the highest HiLev level, i.e., there is no need for an isomorphism test of the generated complete structures. However, in the case of cyclic structures some duplications still appear at the HiLev level. Thus, in the particular case of the gross formula C₁₀H₁₆O with the constraints CH₂–C=O, CH–CH₂–CH₂, C–CH₃, and CH₃–C–CH₃ from all the 19 structures generated at the HiLev level 14 are topologically distinct (nonisomorphic), i.e., five pairs of duplicated structures still remain (see also Table III).

The few isomorphic cyclic structures formed at the HiLev level are recognized in our program by employing an agreement factor between the structure and the ¹³C NMR spectrum (AF).¹⁶ The standard approximation error was chosen as appropriate to this end:

$$AF = \left(\frac{\sum (\delta_{C_{13}} - Ag_C - B)^2}{N - 1} \right)^{1/2} \quad (9)$$

This factor provides values equal up to the 11th decimal place for isomorphic structures.¹⁶

¹H–¹³C Direct Multiplicities/C–H Adjacent Connectivity Check. After each successful combination at the last HiLev level the HASRBS procedure is executed. As stated above it fills all the remaining unsaturated SSs with H atoms. In some cases it can attach an extra H atom to a carbon atom with already specified signal multiplicity (C–H adjacency, respectively). The program carries out a ¹³C–¹H multiplicity test by matching the *N_H* of the H atoms adjacent to each carbon atom with its ¹H–¹³C multiplicity exploiting the well known rule *N_H* = *M_s* – 1 (here *M_s* is the signal multiplicity). If all the carbon atoms conform to this rule the structure is considered correct, otherwise it is rejected and a new combination is generated at the HiLev level.

RESULTS

The work of our generator is exemplified with some practical results provided in Table III. The number of generated structures is compared with the number of duplications at the

HiLev level and with the total number of combinations.

Although the examples from 1 to 6 have the same number of carbon (10) and oxygen (1) atoms, they provide a different number of generated structures. Evidently, this difference depends on the symmetry, i.e., on the number of equivalence (automorphism) classes that the vertices (atoms) are partitioned. This partitioning is carried out in our case on the basis of the input structural and spectral information (number of double-bond fragments, ¹³C–¹H multiplicity, etc.). It is seen from Table III that the number of combinations correlates with the number of generated structures. It is smallest (17) in the most symmetric case of example 1. Here the number of automorphism classes (generation levels) is also the smallest (3). In contrast, in spite of the use of fragments, it is greatest (137) in the case of example 6 where we have no equivalent structural segments.

The efficiency of this approach in dealing with duplications at the highest HiLev level is best illustrated with the example 8 where only five pairs of structures out of 117 occur duplicated.

FURTHER DEVELOPMENTS

Evidently, the HSSS approach provides an extremely flexible and powerful tool for the processing of any chemical, structural, and spectral information. Its work resembles the "heuristic process" of taking analytical decisions during the structure formation. Any new information can be incorporated within its scheme.

As seen from Figure 5 each combination generated at a level lower than HiLev produces a set of substructures which grow up with every transition to a higher level. Their conformity to the spectral information can be controlled at each level, and combinations producing substructures inconsistent with this information might be ignored. By doing so the generation of all the higher level combinations related to an inconsistent combination will be avoided. However, this procedure calls for the building up of a knowledge base of rules for matching the substructures with the spectral information. In addition, the connectivity information from 2D NMR techniques can easily be utilized within this approach by reducing it to our 2D structural representation of SV and SS bonding sites.

Any ambiguity such as the presence of more than one possible connectivity of a given SV can easily be treated by selecting more than one optional SSs. In this way the HSSS approach leads both to reducing the number of SSs (hence alleviation of the combinatorial problem) and to the use of more than one SSs in cases of ambiguous information.

The development of these new features are in progress, and the details of the algorithm will be reported in forthcoming papers.

THE PROGRAM

This new version of our generator was initiated on FORTRAN 77¹ as a part of the structure-elucidation system ASECI3. However, in the course of its development it was found that Pascal is more appropriate both for the design of the generator and for the structure-inference and structure-presentation parts of the system. Now the program is written in TURBO-Pascal-5.5 for IBM XT/AT compatible microcomputers.

ACKNOWLEDGMENT

I am indebted to Prof. S. Dodunekov from the Institute of Mathematics at the Bulgarian Academy of Sciences for the enlightening discussions on the combinatorial problems in this work. Thanks are also due to Dr. S. Karabunarliev for providing the first version of a permutation-generation procedure.

REFERENCES AND NOTES

- (1) Bangov, I. P.; Kanev, K. Computer-Assisted Structure Generation from a Gross Formula. II. Multiple Bond Unsaturated and Cyclic Compounds. Employment of Fragments. *J. Math. Chem.* **1988**, *2*, 31–48.
- (2) Lederberg, J. Topology of Molecules. In *The Mathematical Science*; The MIT Press: Cambridge, MA, 1969; p 37.
- (3) Masinter, L. M.; Sridharan, J.; Lederberg, J.; Smith, D. H. Application of Artificial Intelligence for Chemical Inference. XII. Exhaustive Generation of Cyclic and Acyclic Isomers. *J. Am. Chem. Soc.* **1974**, *96*, 7702–7714.
- (4) Masinter, L. M.; Sridharan, N. S.; Carhart, R. E.; Smith, D. H. Application of Artificial Intelligence for Chemical Inference. XIII. Labeling of Objects Having Symmetry. *J. Am. Chem. Soc.* **1974**, *96*, 7714–7723.
- (5) Kudo, Y.; Sasaki, S. The Connectivity Stack, a New Format for Representation of Organic Chemical Structures. *J. Chem. Doc.* **1974**, *14*, 200–202.
- (6) Kudo, Y.; Sasaki, S. Principle for Exhaustive Enumeration of Unique Structures Consistent with Structural Information. *J. Chem. Inf. Comput. Sci.* **1975**, *16*, 43–49.
- (7) Funatsu, K.; Miyabaiyashi, N.; Sasaki, S. Further Development of Structure Generation in the Automated Structure Elucidation System CHEMICS. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 18–28.
- (8) Faradjiev, I. A. Constructive computation of combinatorial objects. In *Algorithmic Investigations in Combinatorics*; Faradjiev, I. A., Ed.; Nauka: Moscow, 1978; p 11 (in Russian).
- (9) Cristie, B. D.; Munk, M. E. Structure Generation by Reduction: A New Strategy for Computer-Assisted Structure Elucidation. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 87–93.
- (10) Shelly, C. A.; Hays, T. R.; Roman, R. V.; Munk, M. E. An Approach to Automated Partial Structure Expansion. *Anal. Chim. Acta* **1978**, *103*, 121–132.
- (11) Lipkus, A. H.; Munk, M. E. Automated Classification of Candidate Structures for Computer-Assisted Structure Elucidation. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 9–18.
- (12) Carhart, R. A.; Smith, D. H.; Brown, H.; Djerassi, C. Applications of Artificial Intelligence for Chemical Inference. XVII. An approach to Computer-Assisted Elucidation of Molecular Structure. *J. Chem. Inf. Comput. Sci.* **1975**, *97*, 5755–5762.
- (13) Carhart, R. E.; Smith, D. H.; Gray, N. A. B.; Nourse, J. G.; Djerassi, C. GENOA: A Computer Program for Structure Elucidation Utilizing Overlapping and Alternative Substructures. *J. Org. Chem.* **1981**, *46*, 1708–1718.
- (14) Bangov, I. P. Computer-Assisted Generation of Molecular Structures from a Gross Formula. I. Acyclic Saturated Compounds. *Commun. Math. Chem.* **1983**, *14*, 235–246.
- (15) *Mathematics at a Glance*: Gillert, W.; Kuestner, H.; Hellwich, M.; Kaestner, H., Eds.; VEB Bibliographisches Institute: Leipzig, 1975; p 688.
- (16) Bangov, I. P. Use of the ¹³C-NMR Chemical Shift/Charge Density Linear Relationship for Recognition and Ranking of Chemical Structures. *Anal. Chim. Acta* **1988**, *209*, 29–43.
- (17) We favor the following definition of isomorphism provided by E. H. Sussenguth, A graph-Theoretical Algorithm for Matching Chemical Structures. *J. Chem. Doc.* **1969**, *5*, 36–43: Two graphs G and G* are isomorphic if there is one-to-one correspondence between the nodes of G and G* which preserves one-to-one correspondence between the branching of the graphs.
- (18) In fact, this product is the well-known expression for the number of the different partitions of n-element set into k blocks, e.g., see Lipski, W. *Combinatorics for programmers*; Mir: Moscow, 1988; p 48 (in Russian).
- (19) Bangov, I. P. Use of ¹³C NMR Chemical Equivalent Signals and Topological Symmetry for Removal of Redundant Structures. *Commun. Dep. Chem. Bulg. Acad. Sci.* **1988**, *21*, 194–199.
- (20) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity—A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219–3288.
- (21) Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures—A Technique Developed at Chemical Abstract Service. *J. Chem. Soc.* **1965**, *5*, 107–113.
- (22) Balaban, A. T.; Mekenyan, O.; Bonchev, D. Unique Description of Chemical Structures on Hierarchically Ordered Extended Connectivities (HOC Procedures). I. Algorithms for Finding Graph Orbits and Canonical Numbering of Atoms. *J. Comput. Chem.* **1985**, *6*, 538–551.
- (23) Filip, P. A.; Balaban, T. S.; Balaban, A. T. A New Approach for Devising Local Graph Invariants: Derived Topological Indices with Low Degeneracy and Good Correlation Ability. *J. Math. Chem.* **1987**, *1*, 61–83.
- (24) This vector is defined in both ref 8 and in Hendrickson, J. B.; Toczko, A. G. Unique Numbering and Cataloguing of Molecular Structures. *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 171–177. It is formed from the rows of the upper triangle of the adjacency matrix concatenated in sequence from the top to the bottom.
- (25) Breitmaier, E.; Voelter, W. ¹³C NMR Spectroscopy, Erbel, H. F., Ed.; Verlag Chemie: Weinheim and New York, 1978.