

Exhaustive Generation of Organic Isomers. 3. Acyclic, Cyclic, and Mixed Compounds

M. L. Contreras,*† R. Rozas,† and R. Valdivia‡

Chemistry Department, Faculty of Science, University of Santiago de Chile, Casilla 5659, Santiago-2, Chile,
and Faculty of Engineering, Computer and Information Department, University of Tarapacá,
Casilla 6-D, Arica, Chile

Received September 5, 1993*

A program has been designed which allows for the exhaustive and nonredundant generation and enumeration of organic compounds with cyclic, acyclic, and mixed structures which may contain heteroatoms of variable valence and multiple bonds. Filtering of redundant structures is achieved by use of a disconnection-connection algorithm based on a new tree representation concept of a cyclic graph and on a special N -tuple code for these compounds derived from graph theory. Specific algorithms based on a depth first search (DFS) strategy are developed and applied for making the generation process of mixed isomers selective. The molecular formula is the only data required for making the program to work. One or more structural patterns could be used for that, having multiple bonds, cycles, and heteroatoms with different valencies in the same pattern. These facilities make the program to be a useful tool for molecular design, organic synthesis, structure elucidation, structure-activity relationships, and other applications. The program, CAMGEC, has been written in C under Unix in a very friendly menu driven way.

INTRODUCTION

Enumeration of organic isomers has been considered a very interesting subject,¹⁻⁹ and its development continues. One of the principal tools for that aim is Polya theorem,¹⁰ nowadays called Redfield-Polya theorem,¹¹ taking in consideration the work by Redfield published in 1927.¹² Much of the work in this area is principally treated from the mathematical point of view, and no generation of molecular structures is done.

On the other hand, generation of organic isomers has been accomplished under different approaches and for different scopes such as organic synthesis^{13,14} and structure related problems.¹⁵⁻¹⁸

Similarly important and related topics are molecular codification and computer readable notation of organic isomers. The first molecular boundary code was defined specifically for benzenoid compounds.¹⁹ A similar code was proposed later⁴ also for benzenoid compounds, which have been the matter of much topological and graph theory studies.²⁰⁻²⁵ One more general approach is based in N -tuple notation,³ which has been extended to benzenoid compounds,²⁶ and to acyclic and cyclic compounds having heteroatoms and multiple bonds.^{29,30} However that was not applicable to mixed molecules, i.e., molecules that could have a part cyclic and another part acyclic with several branches on it.

This work presents a way of defining mixed structures based on graph theory.³¹⁻³³ For that a new description of a rooted tree is offered which defines different kinds of leaf nodes. Some of them are treated as vertex node descriptors and others like cycle node descriptors. Also, a new definition is proposed for the maximum number of nodes a mixed isomer tree can have. Additionally, a procedure for achieving selectivity in the generation process is proposed and explained in detail. Examples are given to show reliability and usefulness of the program.

PROCESS OF MIXED STRUCTURES GENERATION

The system here described is an extension of the program CAMGEC^{29,30} (for computer assisted molecular generation

and counting). For making to work properly a program like the one proposed, three basic conditions are needed: (a) provision of a combinatory structure generation process, (b) detection of chemical redundant structures, and (c) selection of a canonical structure.

(a) In the first condition, the combinatory process for structure generation must not produce the same combination more than one time. For that, trees are generated based on the N -tuple notation,²⁹ in such a way that the size of the tree is adapted to the number of nodes obtained from the following formula:

$$\text{no. of nodes of the tree} = \text{no. of graph vertex} + 2(\text{no. of cycles of the graph}) \quad (1)$$

As it can be deduced for acyclic molecules, the number of nodes of the tree is equal to the number of vertices of the graph. For cyclic molecules that do not contain any acyclic substituent or branching, the only trees that should be considered have to accomplish the following rule:

$$\text{no. of leaf nodes} = 2(\text{no. of cycles}) \quad (2)$$

Here each leaf node is associated with another one. By considering the distance between both of the associated leaf nodes of the tree, it is possible to infer the size of the cycle they are defining.³⁰

For mixed molecules eq 2 remains like

$$\text{no. of leaf nodes} \geq 2(\text{no. of cycles}) \quad (3)$$

This is evident because some of the leaf nodes will remain free without being associated with another one for defining a cycle just because they are now originating an acyclic branching. Anyway, for mixed molecules, association between leaf nodes really occurs to an extension that depends only on the molecular structure requirements.

In that way, a tree will give several structures that could accomplish with the applied conditions. For example, if someone wants to build structures with 14 atoms and 3 cycles, expression 1 will give

$$\text{no. of nodes of the tree} = 14 + 2 \times 3 = 20$$

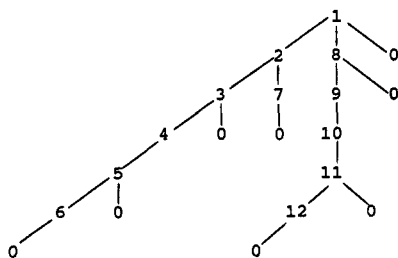
* University of Santiago de Chile.

† University of Tarapacá.

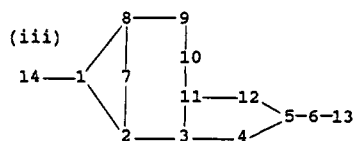
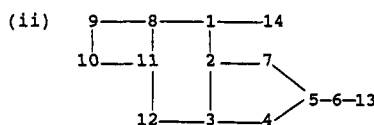
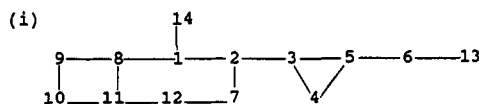
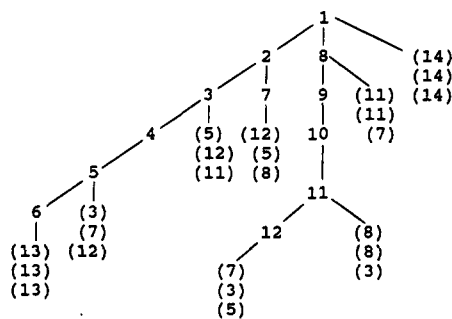
• Abstract published in *Advance ACS Abstracts*, March 1, 1994.

So, the program will produce trees having 20 nodes each and should discard all of the trees that do not follow expression 3, which is equivalent in the present case to discard any tree with a number of leaf nodes lower than 6.

Later the process continues by pairing 6 of the leaf nodes leaving the rest (if any) free. For instance, the following tree



has 12 internal nodes and 8 leaf nodes from which 6 leaf nodes should match to account for the three cycles and the other two leaf nodes will derive in acyclic branching. Some of the possibilities are



(b) The second condition for making the program work properly is to have a transformation process able to determine which "combinations" are similar, i.e., represent the same figure. In this way all of the redundant structures will be detected. For acyclic structures the method of rotation around an imaginary *Y*-axis and around an imaginary *Z*-axis²⁹ was proved to be effective. Cyclic and mixed molecules instead are submitted to a process of disconnection-connection.³⁰ When this process is applied to acyclic molecules, satisfactory results are obtained; however, a bigger response time is required.

(c) The third condition consists of the application of a convention for establishing which of the similar structures detected before corresponds indeed to the representative one (canonical figure). The convention used is the maximum inverse lexicographical order of the *N*-tuple.

In this way the skeleton of the graph is built. This is the first of the three general features of GEMAC. For alkanes

that will be the principal part of the whole generation process. If there is a pattern to be present in the generated structures, this is the moment in which its search is done according to a strategy that will be explained later. The term "pattern" will be used from now on, for substructures that have to be present in the final generated structures.

The other two general features of GEMAC refer to heteroatoms incorporation and to multiple bonds incorporation which is done in that order^{29,30} and correspond to condition c mentioned above. For that, some precedence rules have been defined³⁰ for the atoms and for the bonds between node sons and their node fathers in the tree:



where *r* is assigned to the root node of the tree; *s*, *d*, and *t* refer to single, double, and triple bonds, respectively.

Each selection rule is applied at the end of each step (skeleton generation, heteroatoms incorporation, and multiple bonds incorporation), and the whole process is done in dynamic memory. Also, when there are patterns (substructures) to be considered, the search of the pattern in the structure that is being processed in dynamic memory is done at the end of each of these mentioned steps, as will be explained next.

SELECTIVE ISOMER GENERATION AND ENUMERATION

This process allows for the generation of isomers having a defined pattern into their structures. So, only a selective group of isomers that contain a particular user defined substructure can be generated. The pattern is input using the *N*-tuple notation, although not necessarily in their canonical form.

The graph under processing will be called base graph just to make a difference with the user defined pattern graph used as a constraint for the generation process and which will be simply called pattern graph. During the pattern search process, a comparison between the pattern graph and the base graph is made by the use of the depth first search (DFS) strategy³⁴ adapted to the extended *N*-tuple code.

The search of the pattern is done at the end of each of the three principal generation steps, namely, (i) skeleton generation, (ii) heteroatoms incorporation, and (iii) multiple bonds incorporation.

The way of searching along the pattern graph is always the same. But, in the case of the base graph the walking can be changed to find the matching.

The DFS strategy functions in the following way: the vertex most recently visited will be called *x* and any vertex joint to it will be called *y*. The depth search will select a not explored *x*-*y* bond emerging from *x* (*x* → *y*). If the vertex *y* has already been visited, the procedure will search another not explored bond emerging from *x*. If *y* has not been visited, the procedure will mark *y* as visited and it will initiate a new search over *y*. After completing this search through all of the ways emerging from *y*, the search will return to *x*. The selection process for nonexplored arcs emerging from *x* will continue until all of the *x* arcs had been explored.

The application of this strategy to the *N*-tuple of an acyclic graph is easy and natural because the *N*-tuple describes the graph as a rooted tree and that preestablishes a walking based on the maximum lexicographical order.²⁹

For cyclic graphs things are more complicated because the extended *N*-tuple notation in this case leads to some variations in the walking due to the use of leaf nodes for the description

of the cycles. Here, it is compulsory to understand that those leaf nodes employed for cycle description are "special" nodes of the tree that are used just as cycle descriptors.

A tree is essentially acyclic. The analysis of the tree should then consider two kinds of nodes: the cycle descriptor nodes and the vertex descriptor nodes. As will be shown, the position of the corresponding vertex descriptor node will be recovered when the search arrives to a cycle descriptor node. Also a mark should be used to avoid passing twice for the same place during the walking. Additionally this mark will be helpful for cycle analysis, as will be seen later.

The process works recursively. It makes a comparison point to point for each call. The walking is established by applying the DFS strategy to the pattern graph. Instead, the walking over the base graph is adapted to the DFS strategy used for the walking over the pattern graph. Comparison starts at the root node of both graphs. If no matching is found starting from this point, a new search is initiated. Now the point in the tree that follows in the lexicographical order for the base graph is taken as the first one to be compared with the pattern. According to the level of processing at which the comparison takes place, a particular verification step is done as follows:

(i) If the level of working is the structure or skeleton of the graph, then a verification should be done to ensure that the degree of the vertex of the pattern graph is lower or equal to the degree of the base graph vertex. The degree is considered as the number of associated vertices joined by a bond to the vertex in study.

(ii) If the level of working is at the step of processing the kind of vertex (at the end of the heteroatoms incorporation step), it should be verified that the type of atom be identical at both graphs.

(iii) If the level of working is the type of bond (at the end of the multiple bond incorporation step), it should be verified that the kind of bond from which the vertices in comparison come, according to the walking, be identical.

Each time a matching occurs, a new recursive call is produced until a full matching is found. Many partial comparisons are done depending on the number of vertices of the base graph. On the other hand, each time a matching fails, the search continues with another vertex associated with the last one that did match with the pattern graph. When none of these vertices is coincident, then the vertex from which they come is discarded. The situation is repeated for the vertex from which this last vertex did come. In this way the process fails if the discarding continues until arriving to the starting point. In this case a new point of the base graph should be considered as the starting point, as explained before.

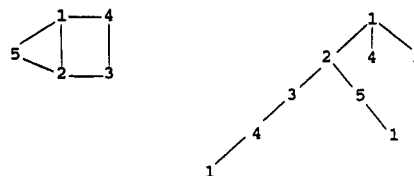
The pattern search process for cyclic molecules needs to use a special mark related to the number of the graph vertex which is the one being compared at that moment. Three rules are applied:

(1) If that mark does not exist, neither in the x node of the tree that describes the pattern graph nor in the y node of the tree that describes the base graph (x and y are the vertices in comparison), then the characteristic number of node x becomes the mark for node y and the characteristic number of node y becomes the mark for node x .

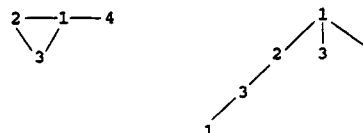
(2) If that mark does exist both in node x and in node y which are being compared, then it is concluded that both cycles are of the same size if the characteristic number of node x is equal to the mark of node y and the characteristic number of node y is equal to the mark of node x .

(3) Any other situation constitutes a cycle discarding result.

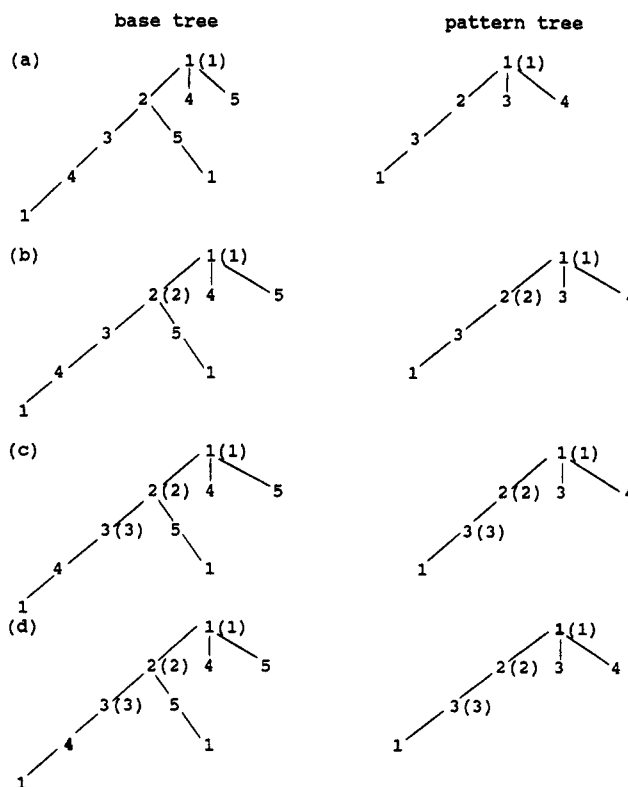
For instance, considering the following base graph and its corresponding tree,



and the following pattern graph with its corresponding tree,



the step-by-step comparison would be (numbers in parentheses are the marks)



As can be seen in cases a-c, selection rule 1 was applied. In the last case, part d, comparison arrives to the first leaf node of the pattern which has the number 1, and it is a cycle descriptor node. Then, the corresponding vertex descriptor node mark has to be recovered. In this example that is the root of the tree, and its mark is (1). Meanwhile in the base tree the descriptor node 4 has no mark yet. So, a cycle discarding result is obtained because of rule 3 expressed before.

A different way of walking through the base graph from node 2 on, just for asserting the search, is as follows:

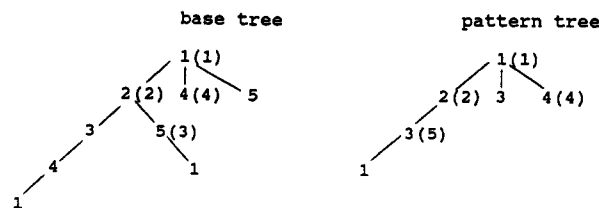


Table 1. Output Archive for the Mixed Isomers C_7H_{14} of Degree 4, Having One Cycle in Their Structure Generated by GEMAC

| isomer | molecular code for mixed structures | | | | | | | | |
|--------|-------------------------------------|------|------|------|------|------|------|------|------|
| 1 | 1c4r | 2c3s | 3c1s | 1c0s | 4c0s | 5c0s | 3c0s | 6c0s | 7c0s |
| 2 | 1c4r | 2c2s | 3c2s | 1c0s | 4c0s | 5c0s | 3c0s | 6c0s | 7c0s |
| 3 | 1c4r | 2c2s | 3c1s | 4c1s | 2c0s | 4c0s | 5c0s | 6c0s | 7c0s |
| 4 | 1c4r | 2c2s | 3c1s | 4c1s | 1c0s | 5c0s | 4c0s | 6c0s | 7c0s |
| 5 | 1c4r | 2c2s | 3c1s | 1c0s | 4c1s | 5c0s | 3c0s | 6c0s | 7c0s |
| 6 | 1c4r | 2c2s | 3c1s | 1c0s | 5c0s | 4c1s | 6c0s | 3c0s | 7c0s |
| 7 | 1c4r | 2c2s | 5c0s | 6c0s | 3c1s | 4c1s | 1c0s | 4c0s | 7c0s |
| 8 | 1c4r | 2c2s | 3c2s | 4c1s | 1c0s | 5c0s | 4c0s | 6c0s | 7c0s |
| 9 | 1c4r | 2c1s | 3c1s | 4c1s | 5c1s | 1c0s | 5c0s | 6c0s | 7c0s |
| 10 | 1c4r | 2c1s | 3c1s | 4c1s | 1c0s | 5c1s | 6c0s | 4c0s | 7c0s |
| 11 | 1c4r | 2c1s | 3c1s | 1c0s | 4c1s | 5c1s | 6c0s | 3c0s | 7c0s |
| 12 | 1c4r | 2c1s | 3c1s | 1c0s | 4c1s | 6c0s | 5c1s | 7c0s | 3c0s |

In the pattern tree, the walking is coincident with the maximum lexicographical order. So, after leaf node 1, it corresponds to continue with node 3. But this leaf node is a cycle descriptor node whose vertex descriptor node is the one from which leaf node 1 comes. Then, node 4 of the pattern graph is the next to be compared. In the base tree, after leaf node 1 the search continues with node 4. This is evident because node 2 has a mark and leaf node 5 is a cycle descriptor

(leaf node 1 comes from its vertex descriptor node). On the other hand in the pattern tree, node 4 is a leaf node but is not a cycle descriptor.

In this way a matching was found: the base graph contains as a substructure a ring of three members with a lateral atom.

Finally, the pattern will be contained in the graph in study if the DFS walking over the pattern graph ends successfully at least in one comparison with the base graph for each one of the general steps of the generation process (skeleton, heteroatoms, and multiple bonds).

RESULTS AND DISCUSSION

Generation of isomers has been done for a number of molecular structures. Different patterns were used as constraints, and some of the results were checked with the aid of a graphic interface built specially for that aim.³⁵

Table 1 shows the molecular code for all of the 12 isomers C_7H_{14} having one cycle and at least one tetrasubstituted carbon atom. Figure 1 shows the trees and the corresponding graphs of these isomers. By inspection of the molecular code for mixed structures given in Table 1, and taking into consideration that all of the isomers have the same molecular formula, it

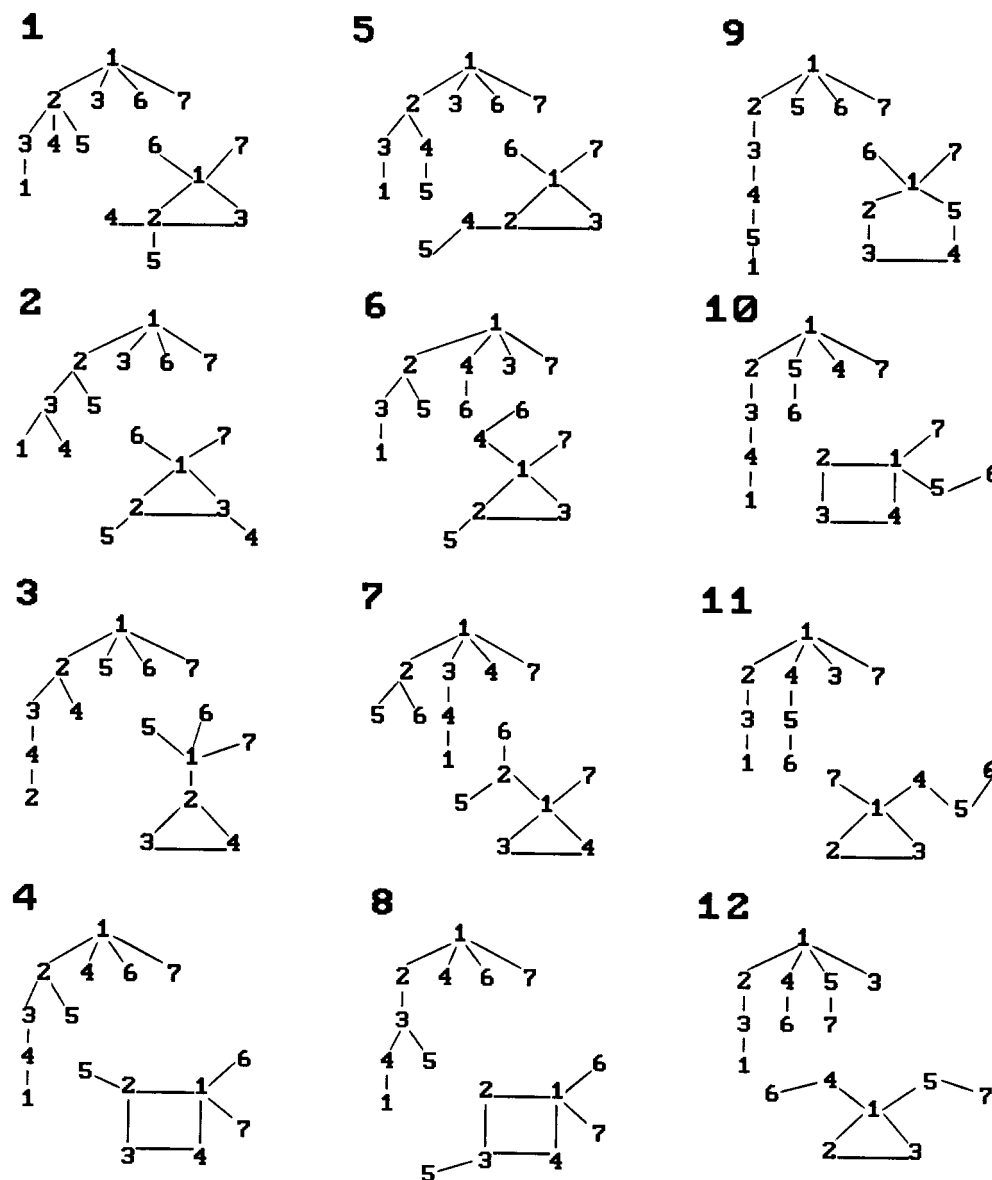
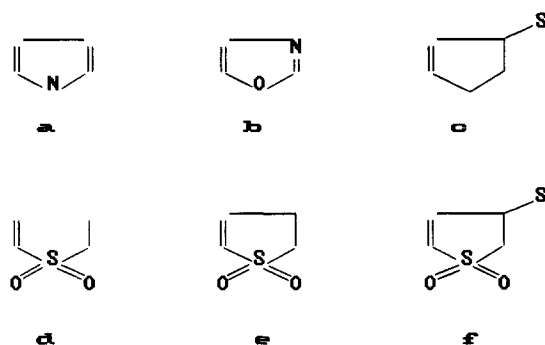


Figure 1. Tree and graph representation for the mixed isomers C_7H_{14} of degree 4 having one cycle. Principal numbers refer to the isomer number and molecular code described in Table 1.

Table 2. Acyclic, Cyclic, and Mixed Isomers with Heteroatoms and Multiple Bonds Generated by GEMAC. Selective Generation of Mixed Isomers

| run | formula | generated isomers | | | | pattern ^a |
|-----|--|-------------------|--------|----------|---------------|----------------------|
| | | linear | cyclic | mixed | mixed/pattern | |
| 1 | C ₇ H ₁₁ N | 1 195 | 18 | 4 188 | 22 | a |
| 2 | C ₆ H ₉ NO | 3 721 | 66 | 12 892 | 13 | b |
| 3 | C ₆ H ₁₀ S ₂ O ₂ | 560 639 | 2625 | 3126 361 | 2289 | c |
| 4 | C ₆ H ₁₀ S ₂ O ₂ | 560 639 | 2625 | 3126 361 | 417 | d |
| 5 | C ₆ H ₁₀ S ₂ O ₂ | 560 639 | 2625 | 3126 361 | 283 | e |
| 6 | C ₆ H ₁₀ S ₂ O ₂ | 560 639 | 2625 | 3126 361 | 14 | f |

^a Patterns are described in Figure 2. Molecular formula for runs 3–6 contains one sulfur atom with valence 2 and the other sulfur atom with valence 6. Cyclic generated isomers refer to monocyclic ones.

**Figure 2.** Chemical substructures corresponding to the patterns used in examples of Table 2.

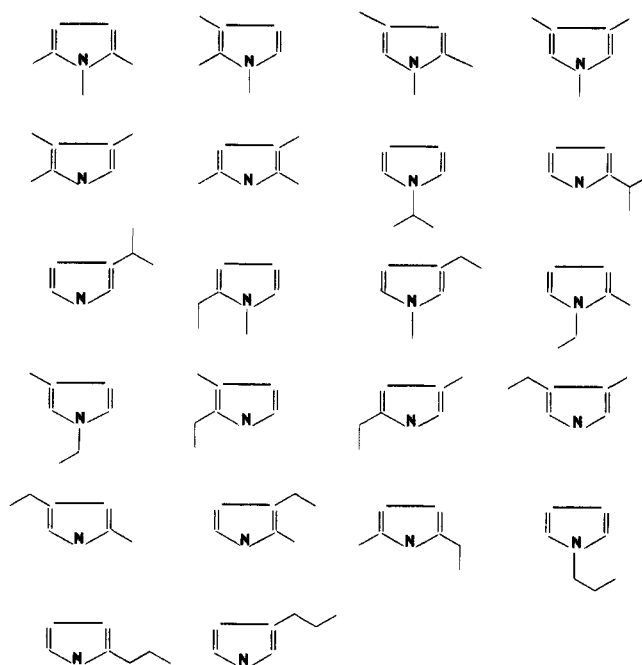
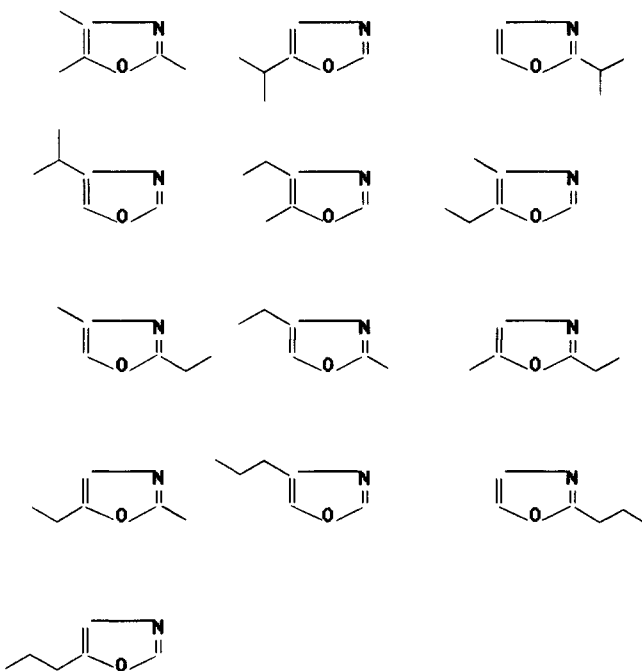
is evident that some of them will have a cycle and 2, 3, or 4 branches because they have 4, 5, or 6 leaf nodes, respectively. This means that knowing the number of cycles and the corresponding molecular code, it is easy to deduce the number of branches the isomers have. This fact constitutes an advantage when compared with other codification systems.^{4,19,26} For instance, in Table 1 all of the isomers have one cycle, so isomers 1 and 2, with 6 leaf nodes, have 4 branches; isomers 3 to 8, with 5 leaf nodes each, have 3 branches; isomers 9–12, with 4 leaf nodes each, have 2 branches. That is also observed from Figure 1.

Here it is important to mention that each line of Figure 1 represents one isomer which can be decoded to get an archive as an input to a graphic interface that generates 2D and 3D coordinates.³⁵ The codified archive created by CAMGEC and represented in Table 1 occupies 12 lines (552 characters). Meanwhile each of the 12 decoded archives occupy 42 lines making a total of 24 154 characters for all of the isomers; this shows the efficiency of our molecular code and their utility for the storage and retrieval of the molecular information.

When the program is asked to selectively generate the monocyclic isomers C₇H₁₄ having a cyclopropane as a pattern, 8 structures are generated. They correspond with the isomers numbered 1–3, 5–7, 11, and 12 on Figure 1. For a cyclobutane pattern, 3 isomers are generated (numbers 4, 8, and 10) and only one isomer of this kind has a cyclopentane (number 9).

The program can also make a selective generation by considering only the isomers whose root degree number²⁹ has a particular value. For instance, CAMGEC generates 108 isomers C₇H₁₂ from which 26 isomers are of degree 4 (they have tetrasubstituted atoms), 81 isomers are of degree 3 (they do not have tetrasubstituted atoms but trisubstituted ones), and 1 isomer is of degree 2 (without tetra- or trisubstituted atoms, but disubstituted atoms).

For the acyclic saturated halogenated hydrocarbon with molecular structure C₈H₁₂Cl₃F₃, CAMGEC generates 55 458

**Figure 3.** Organic isomers with molecular formula C₇H₁₁N having the pattern a described on Figure 2.**Figure 4.** Chemical structures of the isomers C₆H₉NO having the pattern b described on Figure 2.

isomers. This coincides with a reported value found by a system based on a predictive counting.⁹ Using the generation process by degrees, CAMGEC produces 50 373 isomers of degree 4, only 5085 isomers of degree 3, and no isomer with degree 2; this last information has not been reported before. As the sum of the number of isomers by degrees equals the total number of isomers, and this total number is a reproduction of predictive reported values, CAMGEC becomes a reliable system.

We have compared our results also with those of more complex published structures,⁷ and there was a coincidence in most of the data. Only two differences were detected here: (i) For 9 carbon atoms and 3 cycles there were 1278 saturated skeletons reported; CAMGEC has found 1307. (ii) For 11 carbon atoms and 1 cycle there were 1231 saturated skeletons

Table 3

(a) Number of Isomer Structures with the Same Molecular Formula Generated by the Systems
CAMGEC, AEGIS, CHEMICS, DENDRAL, ASSEMBLE, GEN, and MOLGRAPH

| mol formula | no. of isomers generated by given method | | | | | |
|---|--|-------|---------|---------|----------|----------|
| | CAMGEC | AEGIS | CHEMICS | DENDRAL | ASSEMBLE | MOLGRAPH |
| C ₂ H ₅ NO ₂ | 84 | 84 | 87 | 84 | 84 | 84 |
| C ₃ H ₇ NO | 84 | 84 | 87 | 84 | 84 | 84 |
| C ₄ H ₇ NO | 767 | 764 | 802 | 764 | 764 | 764 |
| C ₅ H ₈ BrCl | 140 | 140 | 108 | | | 140 |
| C ₆ H ₁₀ O | 748 | 747 | 745 | 747 | | 747 |

(b) Number of Isomer Structures with the Same Molecular Formula and the Same
Structural Constraints on the Generators CAMGEC, GEN, and MOLGRAPH

| mol formula | present pattern | absent pattern | no. of structures generated by given method | | |
|---|-----------------|----------------|---|-----|----------|
| | | | CAMGEC | GEN | MOLGRAPH |
| C ₆ H ₁₁ NO | -CN | | 64 | 64 | 64 |
| C ₆ H ₁₁ NO | -CN, -OH | | 31 | 31 | 31 |
| C ₈ H ₁₆ O ₂ | O=C-O | O=C-OH | 105 | 105 | 105 |

reported; CAMGEC found 1238 instead. If the generation is done using patterns, the following results are obtained with our system: there are 495 isomers having a cyclopropane, or Cy3 = 495, Cy4 = 368, Cy5 = 198, Cy6 = 106, Cy7 = 47, Cy8 = 16, Cy9 = 6, Cy10 = 1, and Cy11 = 1. That makes a total of 1238 isomers.

Working with heteroatoms and multiple bonds, some interesting results are obtained. For instance, in Table 2 some examples are presented for isomers having cyclic structures, multiple bonds, and nitrogen, oxygen, and sulfur atoms. The last four examples contain two sulfur atoms: one of them has valence 2 and the other one valence 6. Acyclic isomers refer to the isomers that do not contain any cycle in their structure. Cyclic isomers are those compounds that do not contain any acyclic branch in their structure. In the examples of Table 2 only monocyclic isomers were considered, so mixed isomers here have one cycle and acyclic branches.

In addition, the number of mixed isomers having a particular pattern is given under mixed/pattern. Patterns used in each example are presented in Figure 2. Mixed/pattern isomers for runs 1 and 2 are shown in Figures 3 and 4, respectively. From the last 4 examples of Table 2 it is easy to observe that the most restricted the pattern used, the smaller the number of isomers is found: mixed monocyclic isomers C₆H₁₀S₂O₂ are 3126 361 from which 2289 have pattern c, 417 have the acyclic pattern d, 283 have the cyclic pattern e, and only 14 isomers have the pattern f which has the structure of pattern e plus the sulfide atom as an additional restriction.

Results show in general reliable values for generation and enumeration of mixed organic connectivity isomers. Compared to other systems like AEGIS,³⁶ CHEMICS,^{15,37} DENDRAL,¹³ ASSEMBLE,^{38,39} GEN,⁴⁰ and MOLGRAPH⁴¹ (see Table 3), CAMGEC has been shown to be a useful approach to the structure generation problem due to its exhaustiveness, selectivity, and generality of structures, type of atoms, valencies, and bonds. There are two cases of differences between GEMAC results and the results reported, shown in Table 3 for which there is no reasonable explanation at the moment: for run 3a, 767 isomers C₄H₇NO are obtained instead of the 764 reported, and for run 5a, 748 isomers C₆H₁₀O are found instead of the 747 reported ones.

In the first case, when the program is run for C₄H₇NO using different degrees, the following values are obtained: degree 2 isomers = 117; degree 3 isomers = 540; degree 4 isomers = 110; that makes a total of 767 isomers and confirms the found value. In addition, from the molecular formula the following isomer characteristic options²⁹ are automatically

determined: (a) two double bonds, (b) one double bond and one cycle, (c) one triple bond, and (d) two cycles. CAMGEC found 174 isomers with two double bonds, 387 with a double bond and one cycle, 60 isomers with one triple bond, and 146 isomers with two cycles. That makes again a total of 767 isomers C₄H₇NO.

On the other hand from the 748 isomers C₆H₁₀O, 49 are of degree 2, 498 isomers are of degree 3, and 201 isomers are of degree 4. From the total, 151 isomers have two double bonds, 375 have one double bond and one cycle, 47 isomers have one triple bond, and 175 have two cycles. Again the total makes 748 isomers. Additionally, from the 375 isomers with one double bond and one cycle, 3 are of degree 2, 291 isomers are of degree 3, and 81 isomers are of degree 4. The 375 specified isomers contain 33 isomers with a C=O double bond and 342 isomers with a single C-O bond.

CAMGEC differs from existing programs in many ways; it uses canonically ordered *N*-tuples instead of linear structure lists³⁶ or connectivity matrices.⁴⁰ This point could be of significant value for pedagogic applications. Additionally CAMGEC applies generation algorithms in dynamic memory, optimizing access time and storage capacity. This is because in this way it is not necessary to store all of the temporal generated structures first, before searching for a particular pattern. Even more, it can manipulate molecules having a same heteroatom with different valencies in a single structure or pattern. Also, CAMGEC uses as input data only the molecular formula and one or more patterns or substructures when required. Other generators^{16,36,40} do use some fragments as a starting point or input to the generation process which starts from the substructure represented by the fragment. CAMGEC could do that by defining a special "atom" equivalent to the desired fragment.^{29,30} In this way a different molecular formula, containing the "new" defined atom, has to be described for the input. The last two approaches alleviate combinatorial process. This is a problem difficult to be solved efficiently. The problem of graph isomorphism belongs to NP problems.²⁸ Selectivity possibilities offered by CAMGEC, such as the use alone or in combination, of isomer characteristics, different root degree numbers, and the presence of one or more patterns, help in some measure to get reproducible results in a reasonably efficient way for chemical examples with structures of intermediate size and complexity.

The next natural development of this system will consider stereoisomers generation,⁴² making the system really exhaustive. In summary, a program has been built on the basis of a new conception of a tree description, in *N*-tuple notation,

in depth first strategy, and in a disconnection-connection method which allows for generation and enumeration of organic compounds with cyclic, acyclic, and mixed structures that may contain heteroatoms and multiple bonds. The program has proved selective and reliable. It really generates and enumerates organic isomers that are theoretically predicted through the use of the Redfield-Polya theorem⁹⁻¹² and also it has been demonstrated to be exhaustive and nonredundant by comparison with other isomer generation systems. The use of the program in a selective way allows one to have a better perspective of any structural problem by having an idea of how many different isomers are possible for being considered in a particular work and which structures they have. These facilities make the program a useful tool for organic chemistry teaching, for molecular design, for organic synthesis, for structure elucidation, for structure-activity relationships, and for other related applications.

ACKNOWLEDGMENT

This work was partially financed by the University of Santiago de Chile and the University of Tarapaca.

REFERENCES AND NOTES

- (1) Henze, H. R.; Blair, C. M. The Number of Structurally Isomeric Alcohols of the Methanol Series. *J. Am. Chem. Soc.* **1931**, *53*, 3042.
- (2) Read, R. C. In *Chemical Applications of Graph Theory*; Balaban, A. T., Ed.; Academic Press: New York, 1976; pp 11-60.
- (3) Knop, J. V.; Müller, W. R.; Jeričević, Z.; Trinajstić, N. Computer Enumeration and Generation of Trees and Rooted Trees. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 91-99.
- (4) Knop, J. V.; Szymanski, K.; Jeričević, Z.; Trinajstić, N. Computer Enumeration and Generation of Benzenoid Hydrocarbons and Identification of Bay Regions. *J. Comput. Chem.* **1983**, *4*, 23-32.
- (5) Knop, J. V.; Szymanski, K.; Klasinc, L.; Trinajstić, N. Computer Enumeration of Substituted Polyhexes. *Comput. Chem.* **1984**, *8*, 107-115.
- (6) Polya, G.; Read, R. C. *Combinatorial Enumeration of Graphs, Groups, and Chemical Compounds*; Springer-Verlag: New York, 1987.
- (7) Hendrickson, J. B.; Parks, C. A. Generation and Enumeration of Carbon Skeletons. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 101-107.
- (8) Wang, J.; Gu, F. Enumeration of Isomers of Polyethers. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 552.
- (9) Gu, F.; Wang, J. The Numbers of Structural Isomers, Stereoisomers, and Chiral and Achiral Stereoisomers of Fluorochloroalkanes. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 407-410.
- (10) Polya, G. Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und Chemische Verbindungen. *Acta Math.* **1937**, *68*, 145.
- (11) Lloyd, D. *Stud. Phys. Theor. Chem.* **1989**, *63*, 85-105.
- (12) Redfield, J. H. The Theory of Group-Reduced Distributions. *Am. J. Math.* **1927**, *49*, 433-455.
- (13) Lindsay, R. K.; Buchanan, B. G.; Feigenbaum, E. A.; Lederberg, J. *Application of Artificial Intelligence for Organic Chemistry*; McGraw-Hill: New York, 1980.
- (14) Trinajstić, N.; Jeričević, Z.; Knop, J. V.; Müller, W. R.; Szymanski, K. Computer Generation of Isomeric Structures. *Pure Appl. Chem.* **1983**, *55*, 379-390.
- (15) Funatsu, K.; Miyabayashi, N.; Sasaki, S. Further Development of Structure Generation in the Automated Structure Elucidation System CHEMICS. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 18-28.
- (16) Bangov, I. P. Computer-Assisted Structure Generation from a Gross Formula. 3. Alleviation of the Combinatorial Problem. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 277-289.
- (17) Akutsu, T. A New Method of Computer Representation of Stereochemistry Transforming a Stereochemical Structure into a Graph. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 414-421.
- (18) Rücker, G.; Rücker, C. Isocodical and Isospectral Points, Edges, and Pairs in Graphs and How to Cope with Them in Computerized Symmetry Recognition. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 422.
- (19) Balaban, A. T. Chemical Graphs. Part 12. Configurations of Annulenes. *Tetrahedron* **1971**, *27*, 6115-6131.
- (20) Dias, J. R. Algebraic Structure Count. *J. Math. Chem.* **1992**, *9*, 253-260.
- (21) Dias, J. R. Benzenoid Isomer Enumeration and a New Topological Paradigm. *Struct. Chem.* **1992**, *3* (6), 389-398.
- (22) Dias, J. R. Enumeration, Theoretical Properties, and Other Aspects of Helical Polycyclic Aromatic Hydrocarbons. *J. Mol. Struct. (THEOCHEM)* **1993**, *284*, 11-22.
- (23) Dias, J. R. Benzenoids to Fullerenes and the Circumscribing Algorithm. *Chem. Phys. Lett.* **1993**, *209* (5, 6), 439-444.
- (24) Dias, J. R. *Handbook of Polycyclic Hydrocarbons*, Part A; Elsevier: Amsterdam, 1987.
- (25) Dias, J. R. *Molecular Orbital Calculations Using Chemical Graph Theory*; Springer-Verlag: Berlin, 1993.
- (26) Randić, M.; Nikolić, S.; Trinajstić, N. Compact Molecular Codes for Polycyclic Systems. *J. Mol. Struct. (THEOCHEM)* **1988**, *165*, 213-228.
- (27) Nikolić, S.; Trinajstić, N. Compact Molecular Codes for Annulenes, Aza-annulenes, Annulenoannulenes, Aza-annulenoannulenes, Cyclazines and Aza-cyclazines. *Croat. Chem. Acta*, **1990**, *63* (2), 155-170.
- (28) Randić, M. Compact Molecular Codes. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 136-148.
- (29) Contreras, M. L.; Valdivia, R.; Rozas, R. Exhaustive Generation of Organic Isomers. 1. Acyclic Structures. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 323-330.
- (30) Contreras, M. L.; Valdivia, R.; Rozas, R. Exhaustive Generation of Organic Isomers. 2. Cyclic Structures: New Compact Molecular Code. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 483-491.
- (31) Harary, F. *Graph Theory*; Addison-Wesley: Reading, MA, 1972.
- (32) Lunnnon, W. F. In *Graph Theory and Computing*; Read, R. C., Ed.; Academic: New York, 1972.
- (33) Toranzo, F. A. *Introducción a la Teoría de Grafos*; Monograph No. 15, Series Mathematics; Organization of American States: Washington, D.C., 1976.
- (34) Aho, A.; Hopcroft, J.; Ullman, J. *Data Structures and Algorithms*; Addison Wesley: Reading, MA, 1983; pp 215-218, 241-242.
- (35) Contreras, M. L.; Lopetegui, P.; Pinto, A.; Rozas, R. Graphic Interface with Automatic 3-D Coordinate Generation. To be published.
- (36) Luinge, H. J.; van der Mass, J. H. AEGIS, an Algorithm for the Exhaustive Generation of Irredundant Structures. *Chemom. Intell. Lab. Syst.* **1990**, *8*, 157-165.
- (37) Abe, H.; Okuyama, T.; Fujiwara, I.; Sasaki, S. I. A Computer Program for Generation of Constitutionally Isomeric Structural Formulas. *J. Chem. Inf. Comput. Sci.*, **1984**, *24*, 220-229.
- (38) Shelley, C. A.; Hays, T. R.; Munk, M. E.; Roman, R. V. An Approach to Automated Partial Structure Expansion. *Anal. Chim. Acta* **1978**, *103*, 121-132.
- (39) Shelley, C. A.; Munk, M. E. CASE, a Computer Model of the Structure Elucidation Process. *Anal. Chim. Acta* **1981**, *133*, 507-516.
- (40) Bohanec, S.; Zupan, J. Structure Generation of Constitutional Isomers from Structural Fragments. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 531-540.
- (41) Kerber, A.; Lane, R.; Moser, D. Ein Strukturgenerator für Molekulare Graphen (A Structure Generator for Molecular Graphs). *Anal. Chim. Acta* **1990**, *235*, 221-228.
- (42) Contreras, M. L.; et al. Manuscript in preparation.