Exhaustive Generation of Organic Isomers. 6. Stereoisomers Having Isolated and Spiro Cycles and New Extended *N*__Tuples

M. L. Contreras,*,† J. Alvarez,‡ M. Riveros,‡ G. Arias,† and R. Rozas†

Faculty of Chemistry and Biology, Department of Chemical Sciences, and Faculty of Engineering, Department of Information Technology, University of Santiago de Chile, Casilla 40, Santiago-33, Chile

Received November 17, 2000

A new stereoisomer generation system named CAMGEC2 for generation of stereoisomers containing isolated and spiro cycles with one or more descriptors among *R*, *S*, *Z*, *E*, *M*, and *P* is developed using Graph Theory. It includes new approaches for symmetry analysis, cycle detection processes in molecular graphs in a modular way, and also an extension of the *N*__tuple format for linear representation of molecular graphs that keeps graph topographical information.

1. INTRODUCTION

Organic chemistry development and its industrial application as in many other chemistry areas have found great help in computational tools. Combinatorial chemistry and HTS (high throughput screening) are very interesting examples of that. There are many other important subjects that can benefit from these possibilities.[1] Particularly, organic chemistry structure elucidation,[2−6] molecular design,[7−10] and predictive studies of molecular interactions[11] are areas which have found valuable help in computer tools. For instance, to elucidate a structure either for compounds coming from natural sources or even from synthetic preparations it has been useful to automatically generate all of the possible isomers, then to simulate their spectra and next, to compare them with the real one, first, for finding fragments, and finally to elucidate the entire target structure.

Isomers enumeration and generation programs[12−23] have proved to be useful in solving structure elucidation problems. Many of these programs generate principally constitutional or connectivity isomers (also known as topological isomers), under different approaches. For a better efficiency many of them were designed to have some useful restrictions, the use of patterns of generation[20−22] or the use of badlists which contain structural features nor allowed to be generated.[23]

Stereochemistry, a fundamental structure property, cannot be left out, of course, and generation of configurational isomers (topographical ones) has to be considered for isomers containing at least one stereogenic center in the molecule. One of the pioneers in this field was the work[24] of Nourse et al. (1970) where concepts of automorphism groups and construction of configurational symmetry sets were used. These ideas were further developed under a mathematical perspective in the construction of MOLGEN program[17,25,26] which is still under further development.

Structure generator computer programs capable of calculating all possible isomers under given conditions can exhaustively provide a whole variety of structures ensuring that all possibilities have been taken into account. In this way they could be a very useful complement (provide input data) to computational methods for pharmaceutical properties, namely, computer programs that can predict the possible chemical, physical, or biological properties of chemical compounds or structure-based design. These tools together with similarity searching algorithms can help to reduce costs as well as environmental pollution facilitating the rapid and cost-efficient discovery of novel drugs through information based design. Combination of these technologies (computer isomer generation, property prediction and structure analysis) gives rise to virtual high-throughput screening (VHTS) which has enabled biotechnology companies to avoid the bottlenecks in synthesis and testing, leading to an accelerated selection of only the most promising drug candidates with considerable less cost.

In this piece of work, a new approach based on Graph Theory[27] has been developed for generating organic compounds using new strategies for addressing these interesting stereochemical problems. This work is an extension of a previous contribution.[28] Special attention is devoted to configuration isomers containing isolated and spiro cycles. This new approach uses a tree making process, an *N*__tuple notation, and a special algorithm for symmetry detection that avoids redundancy occurring as a molecular symmetry consequence before generation process takes place thus allowing for a better efficiency. A number of examples are presented to show some of the program capabilities.

2. SYSTEM STRUCTURE

A new version of S-CAMGEC system[28] for stereoisomer generation has been developed, and it will be briefly described here from two points of view: a data flow architecture and a procedural architecture. From the first point of view, the CAMGEC2 system accounts now for the following principal modules: (1) *N*__tuple reading and molecular graph construction, (2) symmetry detection, (3) minimum cycle detection, (4) true stereocenter or stereogenic center (SC) detection, and (5) stereoisomer (SI) generation.

* Corresponding author phone: (562) 6812575; e-mail: lcontrer@lauca.usach.cl.
† Faculty of Chemistry and Biology, Department of Chemical Sciences.
‡ Faculty of Engineering, Department of Information Technology.

10.1021/ci0003395 CCC: $20.00
Published on Web 06/26/2001

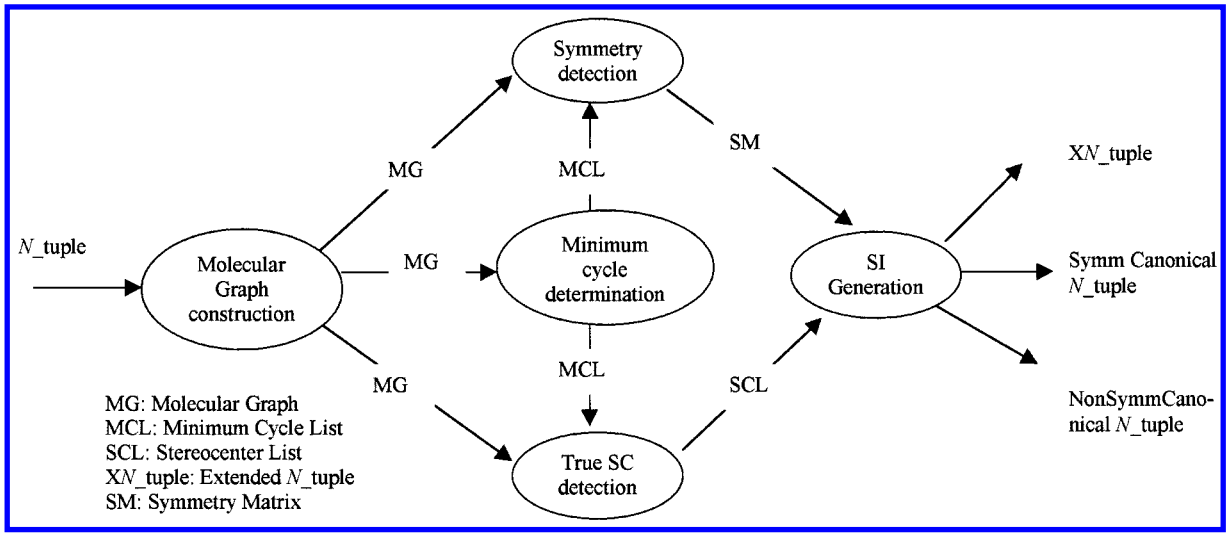


Figure 1. CAMGEC2 system data flow structure.

Relationships between these modules are shown in data flow diagram of Figure 1.

A brief description of module functions, algorithms, and graphical representation is done below:

2.1. *N*_tuple reading and molecular graph construction module takes data input from a string file containing writings of one or more molecular graphs represented as *N*_tuples (a constitutional (topological) linear textual description format).[20−22] This module carries out a reconstruction of the molecular graphs in a way of pointer linked nodes structure.

2.2. Minimum cycle detection module examines a molecular graph and returns a list of minimal cycles (graph cycle factorization). The procedure is shown graphically in Figure 2. For a better understanding it could be useful to consider the following definitions as used here:

Connectivity: node's grade. The number of nodes that are linked to the node in analysis.

Path is a continuous sequence of linked nodes. Path length is the number of elements (nodes) in the sequence.

Cycle is a continuous path (sequence of adjacent nodes) that begins and ends in the same node. Cycle length is the path length of the way that forms the cycle.

Minimum cycle is a cycle of minimal length.

Cyclic node: A node that belongs to any cycle.

Cyclic connectivity (*cc*) of a node: the number of neighbor cyclic nodes.

Opened node: A node whose connectivity is one (does not belong to any cycle).

Closed node: An acyclic node (with cc = 2) that is a part of a kind of bridge between different cycles (a bridge chain).

The algorithm for finding a minimum set of smaller cycles has been adapted from literature[29] and it works modifying molecular graph (Mgraph) as follows:

First step: it recursively eliminates all nodes having connectivity equal to one and updates the connectivity of the remaining nodes. Eliminated nodes are acyclic border or leave nodes. At the end, all acyclic branches are eliminated in this way. Then only cyclic structures and linear chains that are bridge paths between cyclic structures remain (see Figure 3a).

Second step just pretends to eliminate closed nodes (bridge chains) located between isolated cycles or blocks of cycles.

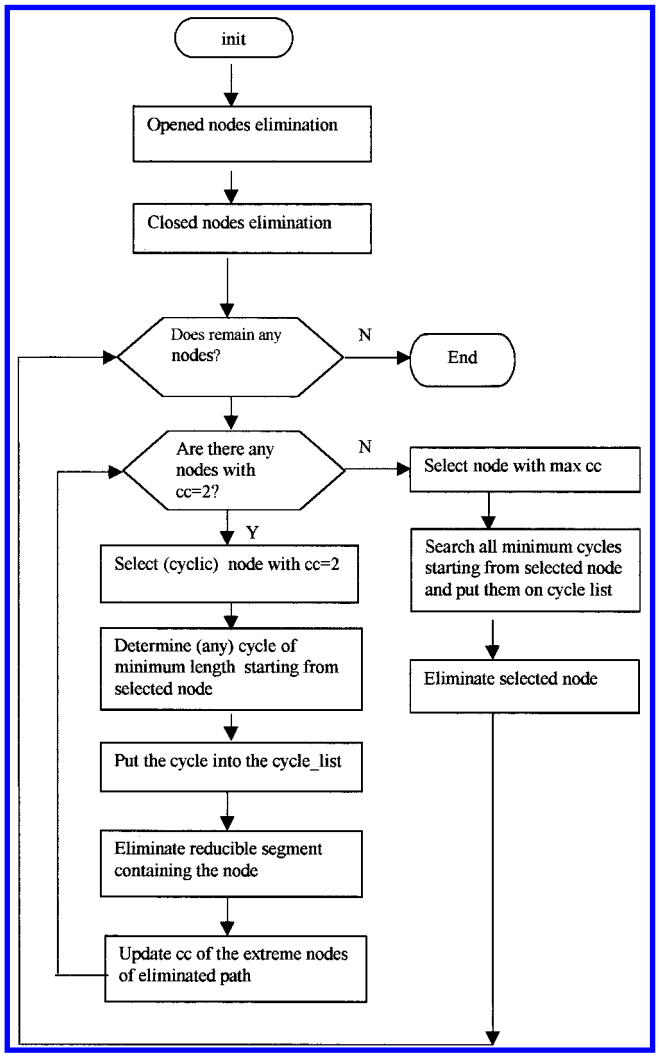


Figure 2. Cycle detection procedure.

Bridge chains are ended in nodes whose connectivity is equal to or greater than 3 (see Figure 3b). After this process is finished, it remains only isolated, condensed, spiro, and nested cycles. From these elements, the set of minimum cycles has to be determined. For doing that, a node with cc = 2 (if there is any) is selected. Then determination of a minimum cycle starting from the selected node is done, and

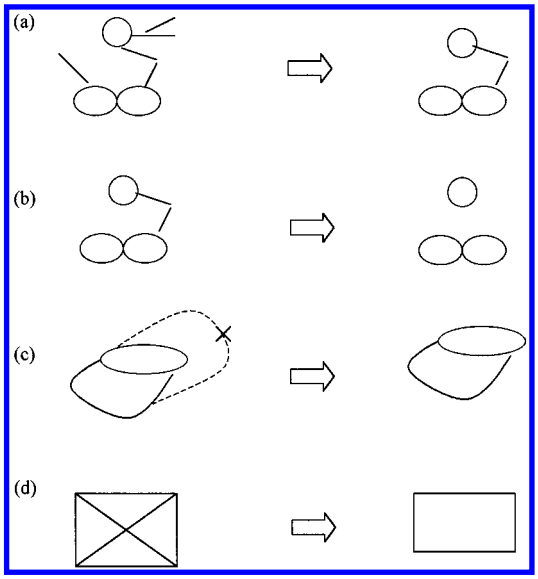


Figure 3. Steps for finding minimum set of smaller cycles: (a) elimination of acyclic branches, (b) elimination of closed nodes (bridge chains), (c) elimination of branches whose nodes have a cc = 2, and (d) elimination of nodes having highest connectivity.

the next elimination of that branch (all nodes with cc = 2) that contains the starting node is accomplished (see Figure 3c). The process is repeated until only cycles having nodes with cc > 2 remain.

Last step: A node with the greatest cc (if any) is chosen, and cycles starting with it are determined and are put on the cycle list and then elimination of that node is done. The procedure continues as explained before for cycles having nodes with cc = 2 (see Figure 3d). At the end of this procedure a complete list of minimum cycles is obtained.

2.3. Symmetry detection is a procedure that obtains a list of symmetry center candidates (SCC) which can be any node or any bond in the molecular graph and examines each of them, determining if such a candidate can be a real topological symmetry center. This search for topological isomorphism is done by building a tree having the SCC as the root and looking for identical branches in the tree. If such isomorphism exists, and if the molecular graph is acyclic, the candidate is considered as a real symmetry center, and a symmetry plane, SP (reflection), passing through the node, or perpendicular to the bond if this is the case, is defined. In case of cyclic molecular graphs, and taking into consideration tetrahedral geometry of saturated carbon atoms, additional nodes could remain in the plane which should be analyzed recursively as a new associated symmetry center candidate. If all associated candidates are symmetry centers, the plane is in effect considered a symmetry plane and an identification number is assigned to it. As a result, a vector (or list) of pairs of symmetry nodes and its corresponding symmetry plane is generated. A graphical representation is shown in Figure 4.

Theoretical symmetry detection methods traditionally have been focused on automorphism groups.[30] The principal problem for that approach is time complexity: it is necessary to examine all *n*! permutations of nodes to check if an adjacency matrix is invariant. That procedure does not take into account any assumption about topographical or spatial considerations. Instead we take advantage of the useful

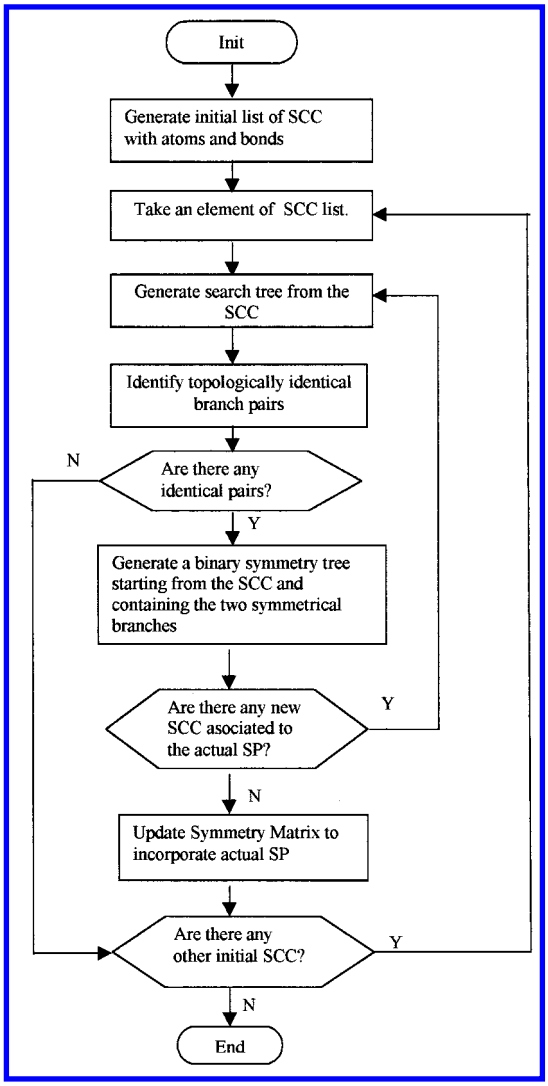


Figure 4. Symmetry detection procedure.

observation that for a simple saturated carbon node located, as known, at the center of a tetrahedron, any pair of its substituents defines a plane that goes through that node or center.

2.4. SC Detection. This module receives information about both the list of SC candidates and the list of minimum cycles, it examines the molecular graph, and it returns a true SC list (SC factorization). SC candidates are tetra- or trisubstituted carbon atom nodes or one or more cumulated carbon−carbon double bonds having one or two substituents at both extremes of the unsaturated bond or bonds as long as they are not a part of a cycle having less than eight members. The procedure works according to the following principal steps (a diagram is shown in Figure 5 for detection of a chiral SC): (1) sequential analysis of all potential SC of the SC candidate list, (2) determination of the relative distances of the rest of the atoms to the SC candidate being analyzed, and (3) construction of a hierarchical digraph or search tree (an acyclic representation of the corresponding molecular graph[21,31,32]) for each potential SC substituent. If the SC being analyzed (which is the root of the search tree) is part of a cycle, this is opened in such a way that the last node for both resulting branches is again the root. For instance, for a cyclobutane ring where root has the number 1 the two resulting branches (starting from the root) should be as

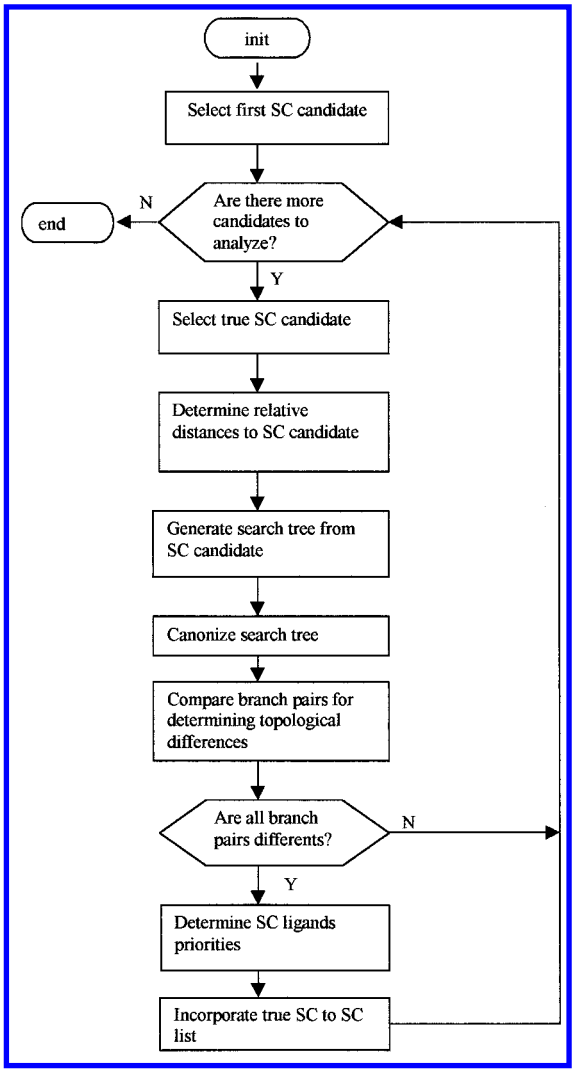


Figure 5. Stereo center (SC) detection procedure.

follows: (a) 1−2−3−4−1 and (b) 1−4−3−2−1 (i) canonization of the search tree: ordering of their branches in such a way that a maximum *N*_tuple can be obtained (resulting a left-loaded tree) and (ii) branches evaluation and their comparison based on constitution, topology, and topography[28] applied to each corresponding substituent. Candidates having identical pairs of branches will be dismissed, otherwise they will be incorporated to the true SC list, and relative CIP priorities[31] for each corresponding SC substituent will be determined and kept for being further included at the final extended *N*_tuple.

Para-stereocenters (including configuration dependent or interdependent SC) are not taken into account yet at this stage.

2.5. Stereoisomer Generation Module. This module is in charge of exhaustively generating stereoisomers (SI) in an orderly manner according to a generation vector (GV), and a set of configuration vectors (CV), taking into consideration symmetry occurrence and avoiding redundancy. Procedure is graphically represented in Figure 6.

Generation involves the permutation of a pair of ligands in each of the stereogenic centers in such a way that two questions have to be considered: where has the exchange to be done (which SC) and will this new molecule be redundant with some other produced before.

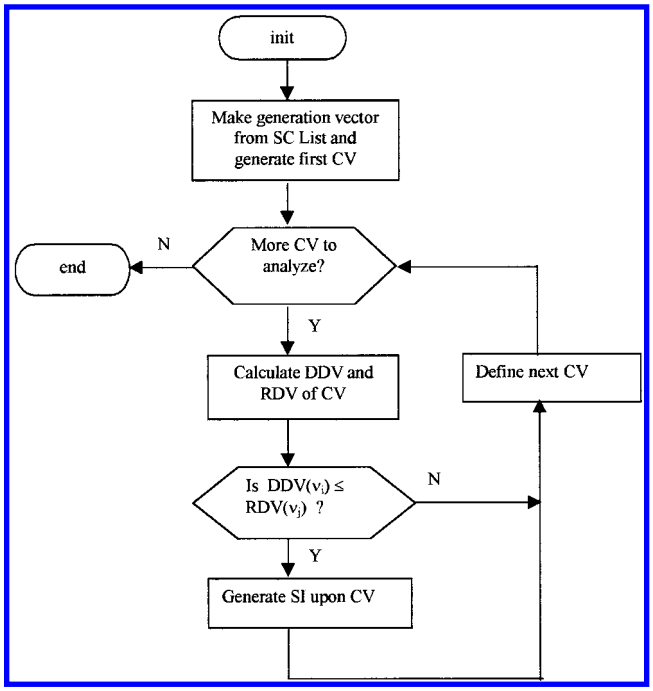


Figure 6. Nonredundant stereoisomer generation procedure.

At the start, a generation vector (GV) is constructed just taking the stereocenter (SC) list (list of the SC identification numbers) and considering it as a vector (i.e. with direct access to each SC).

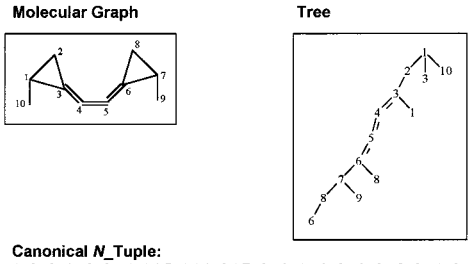
The GV brings information about the number and type of each SC. For chiral SC (*R* and *S* descriptors) the number of the node is indicated in the GV, and for unsaturated SC (*Z*, *E*, *M*, *P* descriptors) node numbers are written in brackets; an even number of nodes in brackets means a descriptor *Z* or *E*; an odd number of nodes in brackets means a descriptor *M* or *P*. An example is given in Figure 7 where extended *N*_tuples are shown also. It is important to note that node numbers are arbitrarily chosen at the beginning, and they are kept during the whole generation process. Their value does not affect any step of that process.

If a symmetry plane exists, the GV is rearranged putting at the center of the vector all the SCs that are contained in a symmetry plane (these SCs are symmetry centers), and then those SCs that constitute symmetry pairs (in relation to a plane) are placed at each corresponding sides (equidistant from center). Symmetry centers and symmetry pairs are indicated by a symmetry matrix in which symmetry centers are located at the diagonal of the matrix and have a value corresponding to the plane that contains them. Symmetry pairs also have the identification number of the plane of symmetry at the corresponding positions in the matrix. In Figure 8 a symmetry matrix for molecule (A) having two symmetry planes is shown.

For each GV with *n* components, a set (with size 2^n) of binary configuration vectors (CV) is defined as a set of vectors v_i representing stereocenter configuration descriptors, i.e.

$$v_i = (c_1, c_2, ..., c_n) \text{ and } \#\{v_i\} = 2^n$$

where for each SC configuration, c_i in the GV, a binary configuration value (represented as 0 for descriptors {*S*, *E*, *M*} and 1 for their corresponding descriptors {*R*, *Z*, *P*}), is



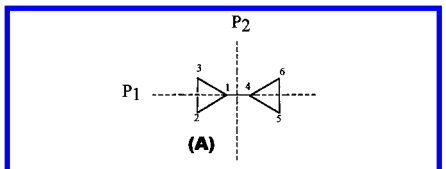
Canonical *N*_Tuple:
1c3r 2c1s 3c2s 4c1d 5c1d 6c2d 7c2s 8c1s 6c0s 9c0s 8c0s 1c0s 3c0s 10c0s

Generation Vector = (1, <3=4=5=6>, 7)

Configuration Representation: 0 = {S,E,M} 1 = {R,Z,P}

Nº	Configuration Vector	Extended *N*_Tuple
1	(0,0,0)	1c3r**S1(3,2,10,0)** 2c1s 3c2s**E2(2,1)** 4c1d**E2** 5c1d **E2** 6c2d**E2(8,7)** 7c2s**S3(8,6,9,0)** 8c1s 6c0s 9c0s 8c0s 1c0s 3c0s 10c0s
2	(0,0,1)	1c3rS1(3,2,10,0) 2c1s 3c2sE2(2,1) 4c1dE2 5c1dE2 6c2dE2(8,7) 7c2sR3(6,8,9,0) 8c1s 6c0s 9c0s 8c0s 1c0s 3c0s 10c0s
3	(0,1,0)	1c3rS1(3,2,10,0) 2c1s 3c2sZ2(1,2) 4c1dZ2 5c1dZ2 6c2dZ2(7,8) 7c2sS3(8,6,9,0) 8c1s 6c0s 9c0s 8c0s 1c0s 3c0s 10c0s
4	(0,1,1)	1c3rS1(3,2,10,0) 2c1s 3c2sZ2(1,2) 4c1dZ2 5c1dZ2 6c2dZ2(7,8) 7c2sR3-(6,8,9,0) 8c1s 6c0s 9c0s 8c0s 1c0s 3c0s 10c0s
5	(1,0,0)	1c3rR1(2,3,10,0) 2c1s 3c2sE2(2,1) 4c1dE2 5c1dE2 6c2dE2(8,7) 7c2sS3(8,6,9,0) 8c1s 6c0s 9c0s 8c0s 1c0s 3c0s 10c0s
6	(1,0,1)	1c3rR1(2,3,10,0) 2c1s 3c2sE2(2,1) 4c1dE2 5c1dE2 6c2dE2(8,7) 7c2sR3(6,8,9,0) 8c1s 6c0s 9c0s 8c0s 1c0s 3c0s 10c0s
7	(1,1,0)	1c3rR1(2,3,10,0) 2c1s 3c2sZ2(1,2) 4c1dZ2 5c1dZ2 6c2dZ2(7,8) 7c2sS3(8,6,9,0) 8c1s 6c0s 9c0s 8c0s 1c0s 3c0s 10c0s
8	(1,1,1)	1c3rR1(2,3,10,0) 2c1s 3c2sZ2(1,2) 4c1dZ2 5c1dZ2 6c2dZ2(7,8) 7c2sR3(6,8,9,0) 8c1s 6c0s 9c0s 8c0s 1c0s 3c0s 10c0s

Figure 7. Molecular graph for a symmetrical molecule with three SC, its tree, canonical *N*_tuple, generation vector, configuration vectors, and extended *N*_tuples. Isomers Nr 5 and 7 are redundant (with corresponding isomers Nr 2 and 4, respectively), and they are not generated.



molecule	Symmetry Planes and symmetry pairs
A	P_1 : (1,1);(2,3);(4,4);(5,6) P_2 : (1,4);(2,5);(3,6)

Symmetry Matrix

	(1)	(2)	(3)	(4)	(5)	(6)
(1)	(1)	0	0	2	0	0
(2)	0	(0)	1	0	2	0
(3)	0	1	(0)	0	0	2
(4)	2	0	0	(1)	0	0
(5)	0	2	0	0	(0)	1
(6)	0	0	2	0	1	(0)

Figure 8. Molecular graph and symmetry elements.

assigned in such a way that a CV v_i is viewed as a binary string, as shown in Figure 7, representing an integer number named direct decimal value (DDV) calculated as

$$\mathrm{DDV} = \sum c_{n-\mathrm{i}}\, 2^i\,;\, i = 0, ..., n-1$$

and taking vector components in a reversed order, a reverse decimal value (RDV) is calculated as

$$\mathrm{RDV} = \sum c_{i+1}\, 2^i\,;\, i = 0, ..., n-1$$

Table 1. Configuration Vector, DDV, RDV, and Redundant Isomer Information for Molecule U of Figure 9

no.	configuration vector (v_i)	direct decimal no. (DDV)	reverse decimal no. (RDV)	does it generate SI?	it is identical with
1	(0,0,0,0)	0	0	yes	
2	(0,0,0,1)	1	8	yes	
3	(0,0,1,0)	2	4	yes	
4	(0,0,1,1)	3	12	yes	
5	(0,1,0,0)	4	2	no	3
6	(0,1,0,1)	5	10	yes	
7	(0,1,1,0)	6	6	yes	
8	(0,1,1,1)	7	14	yes	
9	(1,0,0,0)	8	1	no	2
10	(1,0,0,1)	9	9	yes	
11	(1,0,1,0)	10	5	no	6
12	(1,0,1,1)	11	13	yes	
13	(1,1,0,0)	12	3	no	4
14	(1,1,0,1)	13	11	no	12
15	(1,1,1,0)	14	7	no	8
16	(1,1,1,1)	15	15	yes	

There are two types of CV: vectors having symmetry related elements in relation to the vector's center ("palindromes"; i.e., numbers 1, 7, 10, and 16 in Table 1) and the remainders. For a palindrome vector v_i in a CV, DDV, and RDV values will be equal, i.e. $\mathrm{DDV}(v_i) = \mathrm{RDV}(v_i)$. However palindromes do not represent redundant cases, and the amount of palindromes in the CV is

$$\begin{cases} 2^{(n/2)} & \text{if } n = 2k \qquad ; k = 1, 2, 3, ... \\ 2^{(n-1)/2} & \text{if } n = 2k+1 \end{cases}$$

For nonpalindrome vectors, in case of symmetrical graphs, redundancy appears which can be visualized by a 180° rotation of the graph. This situation corresponds to a vector v_j that is the inverse of another vector v_i and satisfies the condition $\mathrm{DDV}(v_i) = \mathrm{RDV}(v_j)$. Under this redundancy condition the corresponding v_i and v_j will generate the same stereoisomer (see Table 1). The amount of these vectors in the CV set is

$$\begin{cases} (1/2)\,(2^n\text{-}2^{(n/2)}) & \text{if } n = 2k \qquad ; k = 1, 2, 3, ... \\ (1/2)\,(2^n\text{-}2^{(n-1)/2}) & \text{if } n = 2k+1 \end{cases}$$

For instance, in the case of molecular graph **U** shown in Figure 9, $n = 4$; then the number of redundant SI is (1/2) (16−4)→6 (see Table 1).

To sum up, for avoiding generation of redundant vectors it is necessary to calculate DDV and RDV values for each v_i and compare them to verify if they satisfy the redundancy condition and in that case to avoid generation. Thus generation proceeds only for vectors where $\mathrm{DDV}(v_i) \leq \mathrm{RDV}(v_j)$. As an example, this nonredundant generation criteria is verified for molecule **U** in Figure 9 as shown in Table 1, and it is also used in generation procedure as shown in Figure 6.

For chiral stereogenic centers (atoms), stereoisomer generation is done by exchanging any pair of substituents once. In extended *N*_tuple representation, we have chosen to exchange the first two substituents having the greatest priorities according to Cahn, Ingold, and Prelog (CIP) se-

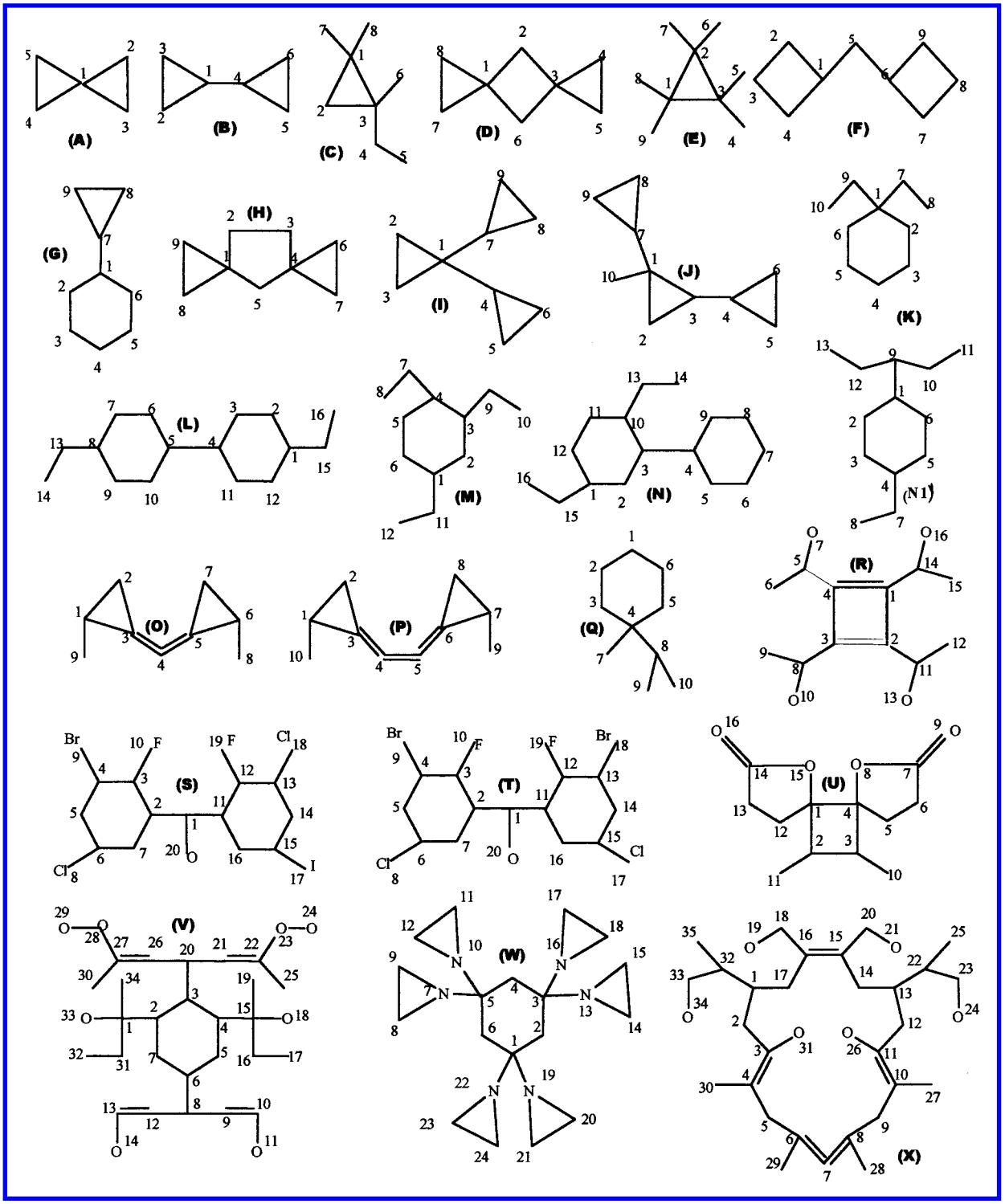


Figure 9. Some generation examples[33] solved by CAMGEC2.

quence rules.[31] If priorities of the first two substituents are written in a decreasing order, the configuration of that node is *R*, otherwise it is *S*.

For unsaturated stereocenters exchange is done just for the two ligands of the first end node of the SC (see Figure 7).

The output of the stereoisomer generation module is a file containing the corresponding extended *N*_tuples for each stereoisomer. This code represents the molecular graph topographical information in a compact codified way easily understandable by computers as it is explained below.

3. NEW LINEAR REPRESENTATION OF A MOLECULAR GRAPH

An *N*_tuple as was defined before[20−22] is a tree representation of a molecular graph. If the molecular graph contains cycles, these are opened to build a hierarchical digraph.[21,31,32] A linear representation (list or string) of this tree is obtained searching in a deep first way, starting from the root node.

A canonical *N*_tuple considers only constitutional and topological information in such a way that each node contains the following attributes: (1) node identification (integer), (2)

Table 2. Hydrocarbons Selectively Generated by CAMGEC2 Classified upon Type of Cyclicity and Symmetry[a]

		restrictions			cyclicity type			I/S				
run	formula	db	tb	*c*	CI	I/S	*C*	SC	sym	nosym	ratio	GI
1	C_5H_{10}			1	5	5		1	5		os	7
2	C_5H_8	1		1	12	12		3	9	3	3.00	16
3	C_5H_8			2	5	1	4		1		os	1
4	C_5H_6		1	1	5	5		1	5		os	6
5	C_5H_6	2		1	14	14			9	5	1.80	14
6	C_5H_6	1		2	11	1	10		1		os	1
7	C_5H_4	3		1	6	6			4	2	2.00	6
8	C_5H_4	1	1	1	7	7			5	2	2.50	7
9	C_5H_4	2		2	13	1	12		1		os	1
10	C_6H_{12}			1	12	12		3	11	1	11.00	18
11	C_6H_{10}	1		1	37	37		13	23	14	1.60	57
12	C_6H_{10}			2	17	3	14	1	3		os	4
13	C_6H_8	2		1	53	53		13	25	28	0.89	66
14	C_6H_8		1	1	14	14		4	12	2	6.00	21
15	C_6H_8	1		2	52	8	44	1	7	1	7.00	9
16	C_6H_6	3		1	31	31		1	15	16	0.94	32
17	C_6H_6	1	1	1	30	30		9	15	15	1.00	39
18	C_6H_6			2	71	9	62		7	2	3.50	9
19	C_7H_{14}			1	29	29		13	23	6	3.80	56
20	C_7H_{12}	1		1	108	108		58	53	55	0.96	205
21	C_7H_{12}			2	56	13	43	6	10	3	3.33	24
22	C_7H_{10}		1	1	40	40		19	29	11	2.34	74
23	C_7H_{10}	2		1	181	181		67	67	114	0.59	286
24	C_7H_{10}	1		2	209	42	167	17	26	16	1.63	68
25	C_7H_8	1	1	1	107	107		52	42	65	0.65	174
26	C_7H_8	3		1	140	140		42	49	91	0.54	193
27	C_7H_8	2		2	341	62	279	18	32	30	1.07	82
28	C_9H_{18}			1	185	185		121	99	86	1.15	530
29	C_9H_{16}			2	573	163	410	117	89	74	1.20	557
30	C_9H_{14}		1	1	334	334		234	146	188	0.78	941
31	C_9H_{14}	1		2	2932	782	2146	565	286	504	0.57	2320
32	C_9H_{12}	1	1	1	1304	1304		877	288	1016	0.28	3280
33	C_9H_{12}	3		1	2393	2393		1542	410	1983	0.21	4969
34	C_9H_{12}	2		2	6433	1632	4801	1091	431	1201	0.36	3910

[a] db = number of double bonds; tb = number of triple bonds; *c* = number of cycles; CI = number of constitutional isomers; I/S = number of constitutional isomers having isolated or spiro cycles; *C* = number of constitutional isomers containing condensed rings; SC = number of stereogenic centers; sym = number of constitutional symmetric isomers; nosym = number of constitutional nonsymmetric isomers; ratio = sym/nosym ratio; os = only symmetric isomers; GI = total generated isomers including stereoisomers.

atom type (symbol), (3) number of sons (integer), and (4) type of connection with father (*s*, *d*, *t*).

An extension of this linear coding of molecular graph incorporating stereo information was done adding the following attributes to nodes which constitute a stereogenic center or are a part of it: (1) stereochemical descriptor (i.e., *R*, *S*, *Z*, *E*, *M*, *P*), (2) stereocenter identification (integer), (3) if the node is a chiral atom: array of four elements each one corresponding to the neighbor nodes identification number (integers) ordered in a priority decreasing way wherein priorities are determined upon CIP sequence rules,[31] and (4) if node is an extreme of an unsaturated stereocenter: two identification numbers of the corresponding ligands ordered by priorities as before.

An example is presented in Figure 7 where the extended *N*_tuple for different stereoisomers is shown. One can see that this format contains all relevant information and has the advantages of economy and expressivity.

4. RESULTS AND DISCUSSION

Tables 2−5 show a sample of cyclic isomer generation under different restrictions for a number of family compounds such as hydrocarbons, halogenated hydrocarbons, and oxygenated and nitrogenated compounds. Generated constitutional isomers are classified according to their cyclic nature in compounds containing isolated and spiro cycles and on the other hand condensed and nested rings. Further analysis is done over constitutional isomers containing isolated and spiro (I/S) cycles for classifying them in symmetrical and nonsymmetrical isomers. Just for facilitating comparison between different compounds the ratio for symmetrical to nonsymmetrical isomers is also shown. Finally a total number of cyclic isomers containing isolated and spiro cycles which include stereoisomers is given. This quantity represents the exhaustive and nonredundant number of topological and topographical cyclic isomers for a given formula.

From data in Tables 2−5 it is possible to observe that for compounds with two cycles the number of isomers having condensed rings is always more than twice the number of isomers having I/S cycles. This difference increases with the number of double bonds, being higher for smaller compounds. For compounds with one or two oxygen (or nitrogen) atoms the same tendency is observed, nevertheless differences between isomers containing condensed and I/S cycles

Table 3. Halogenated Hydrocarbons Selectively Generated by CAMGEC2 Classified upon Type of Cyclicity and Symmetry[a]

		restrictions			cyclicity type			I/S				
run	formula	db	tb	*c*	CI	I/S	*C*	SC	sym	nosym	ratio	GI
1	C_5H_9F			1	14	14		7	9	5	1.80	30
2	C_5H_7F	1		1	42	42		24	14	28	0.50	76
3	C_5H_7F			2	13	1	12	1	1		os	2
4	C_5H_5F		1	1	12	12		6	8	4	2.00	22
5	C_5H_5F	2		1	41	41		14	10	31	0.32	55
6	C_5H_5F	1		2	34	2	32	1	2		os	3
7	C_5H_3F	3		1	13	13		1	3	10	0.30	14
8	C_5H_3F	1	1	1	18	18		6	7	11	0.64	24
9	C_5H_3F	2		2	28	1	27		1		os	1
10	$C_6H_{11}F$			1	44	44		28	21	23	0.91	121
11	C_6H_9F	1		1	157	157		110	40	117	0.34	366
12	C_6H_9F			2	62	9	53	7	5	4	1.25	26
13	C_6H_7F	2		1	201	201		117	34	167	0.20	356
14	C_6H_7F		1	1	46	46		29	21	25	0.84	111
15	C_6H_7F	1		2	201	26	175	16	13	13	1.00	54
16	C_6H_5F	3		1	100	100		39	16	84	0.19	139
17	C_6H_5F	1	1	1	97	97		61	23	74	0.31	172
18	C_6H_5F	2		2	229	25	204	13	10	15	0.67	40
19	$C_7H_{13}F$			1	135	135		100	49	86	0.57	452
20	$C_7H_{11}F$	1		1	559	559		452	103	456	0.23	1618
21	$C_7H_{11}F$			2	260	51	209	42	22	29	0.76	194
22	C_7H_9F	2		1	878	878		644	114	764	0.15	2008
23	C_7H_9F		1	1	168	168		126	56	112	0.50	505
24	C_7H_9F	1		2	1015	184	831	140	56	128	0.44	539
25	C_7H_7F	3		1	611	611		371	71	540	0.13	1096
26	C_7H_7F	1	1	1	453	453		349	74	379	0.20	1042
27	C_7H_7F	2		2	1466	243	1223	167	57	186	0.31	528
28	$C_9H_{17}F$			1	1190	1190		1038	231	959	0.24	5873
29	$C_9H_{15}F$	1		1	6337	6337		5809	575	5762	0.10	27691
30	$C_9H_{15}F$			2	3735	989	2746	557	221	768	0.29	6311
31	$C_9H_{13}F$		1	1	1973	1973		1742	321	1652	0.19	8866
32	$C_9H_{13}F$	2		1	13691	13691		12220	868	12823	0.01	49893
33	$C_9H_{13}F$	1		2	19350	1642	7358	1526	207	1435	0.14	7774
34	$C_9H_{11}F$	1	1	1	7673	7673		6975	594	7079	0.01	28157
35	$C_9H_{11}F$	3		1	14955	14955		12985	812	14143	0.06	44323
36	$C_9H_{11}F$	2		2	39840	9644	30196	8597	906	8738	0.10	36747

[a] Same as in Table 2.

for compounds having two oxygen (or two nitrogen) atoms are smaller.

On the other hand for hydrocarbons with two I/S cycles a higher number of symmetric related to nonsymmetric isomers is found except for compounds containing heteroatoms. In general for monocyclic compounds the ratio symmetrical/nonsymmetrical isomers decreases for compounds having a triple bond (tb), decreases in a more significant way for compounds having a double bond (db) instead, and still decreases in the presence of both (tb + db) except for C_5 halogenated compounds where this value varies between 0.5 and 2.0. The ratio also decreases for higher numbers of carbon atoms. This is also true for compounds containing one or two oxygen atoms (or nitrogen atoms), differences being smaller for compounds with two heteroatoms. Isomers having two cycles show a higher symmetrical/nonsymmetrical ratio. Interestingly this ratio diminishes to half its value for compounds having two oxygen atoms instead of just one oxygen atom. The same is valid for nitrogenated compounds.

Table 6 shows a small sample of constitutional isomers with two isolated or spiro cycles generated with zero, one, or two double bonds, classified upon symmetry and number of stereogenic centers (SC). A number of redundant cases for symmetrical isomers are also specified. As expected, the number of stereoisomers (SI) increases as the number of double bonds increases. Symmetrical compounds show a higher number of isomers having no SC. Instead nonsymmetrical compounds show a higher number of isomers having one SC and also show a wider distribution of isomers having up to 5 SC specially those dioxygenated or dinitrogenated compounds. This diversity decreases for compounds with two double bonds. Data in Table 6 also show that for compounds having zero or one double bond a higher number of redundant cases occurs, and this number increases for dioxygenated or dinitrogenated compounds.

On the other hand in Table 7 and Figure 9 other generation examples solved by GEMAC2 are presented. All the analyzed examples taken from literature[33] have shown that our results are in excellent agreement with reported values. Interestingly symmetry analysis of molecules such as **O** and **P** as indicated in Table 7 helps one to understand and visualize better molecular geometry of unsaturated compounds.

The system described here can process a single *N*_tuple or a file containing many of them. When the system is processing a single *N*_tuple, optionally it offers information about perception of structural features of the molecule (identification of SC candidates, true SC, planes of symmetry,

Table 4. Isomers Containing One or Two Oxygen Atoms Selectively Generated by CAMGEC2 and Classified upon Type of Cyclicity and Symmetry[a]

		restrictions			cyclicity type			I/S				
run	formula	db	tb	*c*	CI	I/S	*C*	SC	sym	nosym	ratio	GI
1	$C_5H_{10}O$			1	33	33		18	19	14	1.36	67
2	C_5H_8O	1		1	103	103		51	33	70	0.47	174
3	C_5H_8O			2	40	7	33	4	5	2	2.50	13
4	C_5H_6O		1	1	30	30		16	16	14	1.14	52
5	C_5H_6O	2		1	112	112		33	27	85	0.32	145
6	C_5H_6O	1		2	113	15	98	7	10	5	2.00	22
7	C_5H_4O	1	1	1	49	49		14	14	35	0.40	63
8	C_5H_4O	3		1	45	45		1	11	34	0.32	46
9	C_5H_4O	2		2	105	9	96	1	6	3	2.00	10
10	$C_6H_{12}O$			1	102	102		65	50	52	0.96	249
11	$C_6H_{10}O$	1		1	375	375		241	97	278	0.35	789
12	$C_6H_{10}O$			2	174	38	136	27	21	17	1.24	96
13	C_6H_8O	2		1	516	516		258	96	420	0.29	852
14	C_6H_8O		1	1	112	112		71	48	64	0.75	244
15	C_6H_8O	1		2	600	112	488	66	47	65	0.72	224
16	C_6H_6O	1	1	1	250	250		134	53	197	0.27	414
17	C_6H_6O	3		1	307	307		93	54	253	0.21	400
18	C_6H_6O	2		2	745	116	629	58	43	73	0.59	178
19	$C_7H_{14}O$			1	302	302		221	114	188	0.61	889
20	$C_7H_{12}O$	1		1	1297	1297		969	262	1035	0.25	3355
21	$C_7H_{12}O$			2	684	172	512	137	72	100	0.72	581
22	$C_7H_{10}O$		1	1	395	395		290	133	262	0.51	1053
23	$C_7H_{10}O$	2		1	2171	2171		1433	306	1865	0.16	4539
24	$C_7H_{10}O$	1		2	2810	640	2170	471	189	451	0.42	1719
25	C_7H_8O	1	1	1	1127	1127		776	191	936	0.20	2364
26	C_7H_8O	3		1	1702	1702		878	207	1495	0.14	2823
27	C_7H_8O	2		2	4348	889	3459	576	208	681	0.31	1813
28	$C_9H_{18}O$			1	2609	2609		1709	583	2026	0.29	11202
29	$C_9H_{16}O$	1		1	14325	14325		12573	1608	12717	0.13	54939
30	$C_9H_{16}O$			2	9221	2767	6454	2488	663	2104	0.32	15293
31	$C_9H_{14}O$		1	1	4501	4501		3873	862	3639	0.24	17571
32	$C_9H_{14}O$	2		1	32407	32407		27457	2489	29918	0.08	105324
33	$C_9H_{14}O$	1		2	49544	14061	35483	12363	2082	11979	0.17	62765
34	$C_9H_{12}O$	1	1	1	18242	18242		15721	1761	16481	0.11	59351
35	$C_9H_{12}O$	3		1	38114	38114		30056	2408	35706	0.01	102275
36	$C_5H_{10}O_2$			1	196	196		133	81	115	0.70	504
37	$C_5H_8O_2$	1		1	587	587		381	132	455	0.29	1216
38	$C_5H_8O_2$			2	273	67	206	47	31	36	0.86	187
39	$C_5H_6O_2$	2		1	607	607		280	104	503	0.21	957
40	$C_5H_6O_2$		1	1	145	145		93	62	83	0.75	293
41	$C_5H_6O_2$	1		2	720	144	576	90	54	90	0.60	295
42	$C_5H_4O_2$	1	1	1	211	211		90	49	162	0.30	316
43	$C_5H_4O_2$	3		1	240	240		56	41	199	0.21	297
44	$C_5H_4O_2$	2		2	603	97	506	41	33	64	0.52	140
45	$C_6H_{12}O_2$			1	672	672		512	222	450	0.49	2097
46	$C_6H_{10}O_2$	1		1	2438	2438		1843	422	2016	0.21	6296
47	$C_6H_{10}O_2$			2	1292	340	952	215	125	215	0.58	1234
48	$C_6H_8O_2$	2		1	3276	3276		2104	405	2871	0.14	6585
49	$C_6H_8O_2$		1	1	629	629		465	199	430	0.46	1621
50	$C_6H_8O_2$	1		2	4286	1001	3286	741	262	739	0.36	2770
51	$C_6H_6O_2$	3		1	1904	1904		881	208	1696	0.12	2979
52	$C_6H_6O_2$	1	1	1	1333	1333		860	216	1117	0.19	2586
53	$C_6H_6O_2$	2		2	4960	1018	3942	658	231	787	0.29	2042
54	$C_7H_{14}O_2$			1	2231	2231		1851	542	1689	0.32	8489
55	$C_7H_{12}O_2$	1		1	9516	9516		7976	1188	8328	0.14	30404
56	$C_7H_{12}O_2$			2	5607	1601	4006	1402	413	1188	0.35	7544
57	$C_7H_{10}O_2$		1	1	2551	2551		2091	590	1961	0.30	8164
58	$C_7H_{10}O_2$	2		1	15662	15662		12567	1368	14294	0.01	39989
59	$C_7H_{10}O_2$	1		2	22374	5829	16518	4876	1001	4828	0.21	21065
60	$C_7H_8O_2$	3		1	12184	12184		7929	914	11270	0.08	24481
61	$C_7H_8O_2$	1	1	1	7022	7022		5311	813	6209	0.13	17435
62	$C_7H_8O_2$	2		2	33064	7950	25114	6146	1082	6868	0.16	21143
63	$C_9H_{18}O_2$			1	23402	23402		21341	3069	20333	0.15	128267

[a] Same as in Table 2.

Table 5. Isomers Containing One or Two Nitrogen Atoms Selectively Generated by CAMGEC2 and Classified upon Type of Cyclicity and Symmetry[a]

		restrictions			cyclicity type			I/S				
run	formula	db	tb	*c*	CI	I/S	*C*	SC	sym	nosym	ratio	GI
1	$C_5H_{11}N$			1	44	44		22	27	17	1.59	83
2	C_5H_9N	1		1	157	157		72	56	101	0.55	254
3	C_5H_9N			2	62	9	53	4	7	2	3.50	15
4	C_5H_7N		1	1	46	46		20	27	19	1.42	74
5	C_5H_7N	2		1	201	201		59	48	153	0.31	262
6	C_5H_7N	1		2	201	26	175	9	17	9	1.89	37
7	C_5H_5N	1	1	1	97	97		29	33	64	0.52	126
8	C_5H_5N	3		1	100	100		8	22	78	0.28	108
9	C_5H_5N	2		2	229	25	204	6	13	12	1.08	31
10	$C_6H_{13}N$			1	135	135		82	67	68	0.99	307
11	$C_6H_{11}N$	1		1	559	559		334	164	395	0.42	1107
12	$C_6H_{11}N$			2	260	51	209	32	29	22	1.32	118
13	C_6H_9N	2		1	878	878		417	178	700	0.25	1413
14	C_6H_9N		1	1	168	168		97	77	91	0.85	338
15	C_6H_9N	1		2	1015	184	831	99	80	104	0.77	348
16	C_6H_7N	3		1	611	611		181	111	500	0.22	801
17	C_6H_7N	1	1	1	453	453		227	119	334	0.36	731
18	C_6H_7N	2		2	1466	243	1223	114	84	159	0.53	383
19	$C_7H_{15}N$			1	403	403		282	158	245	0.65	1103
20	$C_7H_{13}N$	1		1	1907	1907		1343	427	1480	0.29	4585
21	$C_7H_{13}N$			2	1011	237	474	175	106	131	0.81	725
22	$C_7H_{11}N$		1	1	584	584		402	204	380	0.54	1439
23	$C_7H_{11}N$	2		1	3559	3559		2214	556	3003	0.19	7088
24	$C_7H_{11}N$	1		2	4595	1019	3576	729	317	702	0.45	2554
25	C_7H_9N	1	1	1	1924	1924		1243	377	1547	0.24	3863
26	C_7H_9N	3		1	3179	3179		1863	422	2757	0.15	5155
27	C_7H_9N	2		2	8047	1679	6368	1034	395	1284	0.31	3358
28	$C_9H_{19}N$			1	3478	3478		2886	790	2688	0.29	13677
29	$C_9H_{17}N$	1		1	20606	20606		17467	2538	18068	0.14	72610
30	$C_9H_{17}N$			2	13294	3845	9449	3350	968	2877	0.34	19225
31	$C_9H_{15}N$		1	1	6502	6502		5593	1245	5257	0.24	23389
32	$C_9H_{15}N$	2		1	50621	50621		71990	4353	46268	0.09	152940
33	$C_5H_{12}N_2$			1	341	341		218	126	215	0.59	771
34	$C_5H_{10}N_2$	1		1	1341	1341		788	293	1048	0.28	2535
35	$C_5H_{10}N_2$			2	633	117	516	72	49	68	0.72	276
36	$C_5H_8N_2$	2		1	1942	1942		857	313	1629	0.19	2979
37	$C_5H_8N_2$		1	1	326	326		188	114	212	0.54	614
38	$C_5H_8N_2$	1		2	2242	386	1856	226	124	262	0.47	763
39	$C_5H_6N_2$	3		1	1182	1182		218	180	1002	0.18	1489
40	$C_5H_6N_2$	1	1	1	796	796		350	173	623	0.28	1205
41	$C_5H_6N_2$	2		2	2817	468	2349	234	120	348	0.35	763
42	$C_6H_{14}N_2$			1	1176	1176		852	340	836	0.41	3188
43	$C_6H_{12}N_2$	1		1	5354	5354		3732	904	4450	0.20	12319
44	$C_6H_{12}N_2$			2	2868	633	2235	479	210	423	0.50	1971
45	$C_6H_{10}N_2$	2		1	9375	9375		5574	1137	8238	0.14	17634
46	$C_6H_{10}N_2$	1		2	12113	2562	9551	1809	589	1973	0.30	6486
47	$C_6H_8N_2$	3		1	7581	7581		3330	813	6768	0.12	11612
48	$C_6H_8N_2$	1	1	1	4222	4222		2560	649	3573	0.18	7909
49	$C_6H_8N_2$	2		2	19086	3903	15183	2292	699	3204	0.22	7879
50	$C_7H_{16}N_2$			1	3923	3923		3128	831	3092	0.27	12838
51	$C_7H_{14}N_2$	1		1	20302	20302		15911	2481	17821	0.14	56789
52	$C_7H_{14}N_2$			2	12077	3047	9030	2526	714	2333	0.31	12166
53	$C_7H_{12}N_2$		1	1	5462	5462		4262	1014	4448	0.23	15548
54	$C_7H_{10}N_2$	3		1	42167	42167		25368	3224	38943	0.08	79884
55	$C_7H_{10}N_2$	1	1	1	20078	20078		14609	2225	17853	0.13	46425

[a] Same as in Table 2.

and pairs of symmetry; this information is provided using the corresponding node numbers initially chosen) and also detailed information about relative CIP priorities of all the SC attached groups and the different vectors useful for the generation process such as generation vector, symmetry pairs vector, and configuration vectors. CAMGEC2 also includes a statistical archive that contains records of the generation results. This file is being appended each time a new generation process is done. It keeps user name, date and time, file name, and the initial and final number of the *N*__tuples being processed. Also it keeps execution time, number of read *N*__tuples, number of incorrect *N*__tuples if any, number

Table 6. Isomers Containing Two Isolated or Spiro Cycles Selectively Generated by CAMGEC2 and Classified upon Symmetry[a]

				symmetric					nonsymmetric								
run	formula	db	CI	no-SC	1SC	2SC	3SC	4SC	no-SC	1SC	2SC	3SC	4SC	5SC	SI	red	total GI
1	C_6H_{10}	0	3	2	1										4		4
2	C_6H_8	1	8	6	1				1						9		9
3	C_6H_6	2	9	7					2						9		9
4	C_7H_{12}	0	13	7	2	1				1	2				25	1	24
5	C_7H_{10}	1	42	18	7	1			7	5	4				69	1	68
6	C_7H_8	2	62	24	8				20	9	1				82		82
7	C_6H_9F	0	9	2	2	1				2	1	1			26		26
8	C_6H_7F	1	26	7	5	1			3	5	5				54		54
9	C_6H_5F	2	25	5	5				7	7	1				40		40
10	$C_7H_{11}F$	0	51	9	8	5				7	14	6	2		195	1	194
11	C_7H_9F	1	184	28	20	8			16	48	46	18			540	1	539
12	C_7H_7F	2	243	31	20	6			45	90	50	1			528		528
13	$C_6H_{10}O$	0	38	11	8	2				9	5	3			97	1	96
14	C_6H_8O	1	112	31	15	1			15	28	22				224		224
15	C_6H_6O	2	116	29	14				29	42	2				178		178
16	$C_7H_{12}O$	0	172	35	24	13				36	39	22	3		587	6	581
17	$C_7H_{10}O$	1	640	104	69	16			65	181	162	43			1725	6	1719
18	C_7H_8O	2	889	136	64	8			177	340	163	1			1813		1813
19	$C_6H_{10}O_2$	0	340	62	39	22		2		72	82	56	5		1260	26	1234
20	$C_6H_8O_2$	1	1001	158	71	19	3	1	92	282	305	60			2792	22	2770
21	$C_6H_6O_2$	2	1018	163	56	11	1		197	416	174				2052	10	2042
22	$C_7H_{12}O_2$	0	1601	196	125	85	4	3	3	285	464	336	98	2	7615	71	7544
23	$C_7H_{10}O_2$	1	5829	548	312	137	3	1	405	1531	1949	872	71		21135	70	21065
24	$C_7H_8O_2$	2	7950	641	358	82	1		1161	2975	2369	361			21176	33	21143
25	$C_6H_{11}N$	0	51	18	9	2			1	11	7	3			119	1	118
26	C_6H_9N	1	184	56	22	2			29	46	28	1			349	1	348
27	C_6H_7N	2	243	62	22				67	79	13				383		383
28	$C_7H_{13}N$	0	237	59	31	16			3	51	49	25	3		734	9	725
29	$C_7H_{11}N$	1	1019	193	100	24			127	291	266	58			2566	12	2554
30	C_7H_9N	2	1679	270	113	12			375	619	279	11			3361	3	3358
31	$C_6H_{12}N_2$	0	633	125	56	27		2	31	161	154	72	5		2002	31	1971
32	$C_6H_{10}N_2$	1	2562	383	166	34	4	2	370	829	654	119	1		6527	41	6486
33	$C_6H_8N_2$	2	3903	490	179	26	4		921	1618	439	26			7905	26	7879
34	$C_7H_{14}N_2$	0	3047	400	188	119	4	3	121	723	865	514	108	2	12263	97	12166

[a] db = double bonds; CI = constitutional isomers; no-SC = isomers without any stereogenic center (SC); 1SC = isomers containing one SC; 2SC = isomers containing two SC; similar meaning for 3SC, 4SC, 5SC; SI =number of calculated stereoisomers including compounds without any SC; red = number of redundant isomers; total GI = total number of generated nonredundant isomers.

Table 7. Results for the Examples of Figure 9

graph	cycles	candidate nodes	true SC	priorities	symmetry planes and pairs	generation[a]
A	C1: (1-2-3) C2: (1-4-5)	1	no	no	P1: (1,1);(2,3);(4,4);(5,5) P2: (1,1);(2,2);(3,3);(4,5)	PI: 1 RI: 0 TI: 1
B	C1: (1-2-3) C2: (4-5-6)	1-4	no	no	P1: (1,1);(2,3);(4,4);(5,6) P2: (1,4);(2,5);(3,6)	PI: 1 RI: 0 TI: 1
C	C1: (1-2-3)	1-3	3	1-2-4-6	no	PI: 2 RI: 0 TI: 2
D	C1: (1-2-3-6) C2: (3-4-5) C3: (1-7-8)	1-3	no	no	P1: (1,1);(2,6);(3,3);(4,4);(5,5);(7,7);(8,8)	PI: 1 RI: 0 TI: 1
E	C1: (1-2-3)	1-2-3	no	no	P1: (1,1);(2,3);(6,4);(7,5); (8,8);(9,9) P2: (1,2);(3,3);(4,4);(5,5);(8,6);(9,7) P3: (1,3);(2,2);(6,6);(7,7);(8,4);(9,5)	PI: 1 RI: 0 TI: 1
F	C1: (1-2-3-4) C2: (6-7-8-9)	1-6	no	no	P1: (1,1);(2,4);(3,3);(5,5);(6,6);(7,9);(8,8) P2: (5,5);(1,6);(2,7);(3,8);(4,9)	PI: 1 RI: 0 TI: 1
G	C1: (1-2-3-4-5-6) C2: (7-8-9)	1-7	no	no	P1: (1,1);(2,6);(3,5);(4,4);(7,7);(8,9)	PI: 1 RI: 0 TI: 1
H	C1: (1-2-3-4-5) C2: (4-6-7) C3: (1-8-7)	1-4	no	no	P1: (1,4);(2,3);(8,6);(9,7);(5,5)	PI: 1 RI: 0 TI: 1
I	C1: (1-2-3) C2: (4-5-6) C3: (7-8-9)	1-4-7	no	no	P1: (1,1);(2,3);(4,4);(5,6);(7,7);(8,9)	PI: 1 RI: 0 TI: 1
J	C1: (1-2-3) C2: (4-5-6) C3: (7-8-9)	1-3-4-7	1 3	3-7-2-10 1-4-2-0	no	PI: 4 RI: 0 TI: 4
K	C1: (1-2-3-4-5-6)	1	no	no	P1: (1,1);(2,6);(3,5);(4,4);(7,7);(8,8);(9,9);(10,10)	PI:1 RI:0 TI:1

Table 7 (Continued)

graph	cycles	candidate nodes	true SC	priorities	symmetry planes and pairs	generation[a]
L	C1: (1-2-3-4-11-12)	1-4-5-8	no	no	P1: (1,1);(2,12);(3,11);(4,4);(5,5);(6,10);(7,9); (8,8);(13,13);(14,14);(15,15);(16,16)	PI:1 RI:0
	C2: (5-6-7-8-9-10)				P2: (1,8);(2,7);(3,6);(4,5);(11,10);(12,9);(15,13);(16,14)	TI:1
M	C1: (1-2-3-4-5-6)	1-3-4	1	2-6-11-0	no	PI: 8
			3	4-2-9-0		IR: 0
			4	3-5-7-0		TI: 8
N	C1: (1-2-3-10-11-12)	1-3-4-10	1	2-12-15-0	no	PI: 8
	C2: (4-5-6-7-8-9)		3	4-10-2-0		IR: 0
			10	3-11-13-0		TI: 8
N1	C1: (1-2-3-4-5-6)	1-4-9	no	no	P1: (1,1);(2,6);(3,5);(4,4);(7,7);(8,8);(9,9); (10,12);(11,13)	PI: 1 IR: 0 TI: 1
O	C1: (1-2-3)	1-6	1	3-2-9-0	no	PI: 8
	C2: (5-6-7)	⟨3-4-5⟩	6	5-7-8-0		IR: 0
			⟨3-4-5⟩	3: 1-2 5: 6-7		TI: 8
P	C1: (1-2-3)	1-7	1	3-2-10-0	P1: (1,7);(2,8);(3,6);(4,5);(10,9)	PI: 8
	C2: (6-7-8)	⟨3-4-5-6⟩	7	6-8-9-0		IR: 2
			⟨3-4-5-6⟩	3: 1-2 6: 7-8		TI: 6
Q	C1: (1-2-3-4-5-6)	4-8	no	no	P1: (1,1);(2,6);(3,5);(4,4);(7,7);(8,8);(9,10)	PI: 1 IR: 0 TI: 1
R	C1: (1-2-3-4)	5-8-11-14	5	7-4-6-0	P1: (1,2);(4,3);(5,8);(6,9);(7,10);(14,11);(15,12);(16,13)	PI: 16
		⟨1-4⟩	8	10-3-9-0	P2: (1,4);(2,3);(11,8);(12,9);(13,10);(14,5);(15,6);(16,7)	IR: 6
		⟨2-3⟩	11	13-2-12-0		TI: 10
			14	16-1-15-0		
S	C1: (2-3-4-5-6-7)	1-2-3-4-6-11-12-13-15	1	20-2-11-0	no	PI: 512
	C2: (11-12-13-14-15-16)		2	3-1-7-0		IR: 0
			3	10-4-2-0		TI: 512
			4	9-3-5-0		
			6	8-5-7-0		
			11	12-1-16-0		
			12	19-13-11-0		
			13	18-12-14-0		
			15	17-14-16-0		
T	C1: (2-3-4-5-6-7)	1-2-3-4-6-11-12-13-15	2	3-1-7-0	P1: (1,1);(2,11);(3,12);(4,13);(5,14);(6,15); (7,16);(8,17);(9,18);(10,19);(20,20)	PI: 256
	C2: (11-12-13-14-15-16)		3	10-4-2-0		IR: 120
			4	9-3-5-0		TI: 136
			6	8-5-7-0		
			11	12-1-16-0		
			12	19-13-11-0		
			13	18-12-14-0		
			15	17-14-16-0		
U	C1: (1-2-3-4)	1-2-3-4	1	15-4-2-12	P1: (1,4);(2,3);(11,10);(12,5);(13,6);(14,7); (15,8);(16,9)	PI: 16
	C2: (4-5-6-7-8)		2	1-3-11-0		IR: 6
	C3: (1-12-13-14-15)		3	4-2-10-0		TI: 10
			4	8-1-3-5		
V	C1: (2-3-4-5-6-7)	1-2-3-4-6-8-15-20	1	33-2-31-34	P1: (1,15);(3,3);(2,4);(5,7);(6,6);(8,8);(9,12); (10,13);(11,14);(16,31);(17,32); (18,33);(19,34);(20,20);(21,26); (22,27);(23,28);(24,29);(25,30)	PI: 256
		⟨9-10⟩	2	1-3-7-0		IR: 120
		⟨12-13⟩	4	15-3-5-0		TI: 136
		⟨21-22⟩	15	18-4-16-19		
		⟨26-27⟩	⟨9-10⟩	9: 8-0		
			⟨12-13⟩	10: 11-0		
			⟨21-22⟩	12: 8-0		
			⟨26-27⟩	13: 14-0		
				21: 20-0		
				22: 23-25		
				26: 20-0		
				27: 28-30		
W	C1: (1-2-3-4-5-6)	1-3-5	no	no	P1: (1,1),(2,6);(3,5);(4,4);(7,13);(8,14);(9,15);(10,16); (11,17);(12,18);(19,19);(22,22);(20,21);(23,24)	PI: 1
	C2: (7-8-9)				P2: (1,3),(2,2);(4,6);(5,5);(7,7);(8,9);(10,10);(11,12); (13,19);(14,20);(15,21);(16,22);(17,23);(18,24)	IR: 0
	C3: (10-11-12)				P3: (1,5),(2,4);(3,3);(6,6);(7,22);(8,23);(9,24);(10,19); (11,20);(12,21);(13,13);(14,15);(17,18);(16,16)	TI: 1
	C4: (13-14-15)					
	C5: (16-17-18)					
	C6: (19-20-21)					
	C7: (22-23-24)					
X	C1: (1-2-3-4-5-6-7-8-9-10-11-12-13-14-15-16-17)	1-13-22-32	1	32-2-17-0	no	PI: 256
		⟨3-4⟩	13	22-12-14-0		IR: 0
		⟨6-7-8⟩	22	23-13-25-0		TI: 256
		⟨10-11⟩	32	33-1-35-0		
		⟨15-16⟩	⟨3-4⟩	3: 31-2		
			⟨6-7-8⟩	4: 5-30		
			⟨10-11⟩	6: 5-29		
			⟨15-16⟩	8: 9-28		
				10: 9-27		
				11: 26-12		
				15: 20-14		
				16: 18-17		

[a] PI = potential isomers, RI = redundant isomers, TI = total isomers.

of *N*_tuples having no cycles, having I/S cycles, condensed or nested cycles, symmetrical and nonsymmetrical *N*_tuples, and also the number of *N*_tuples having no SC, one SC, two SC, three SC, four SC, or more SC. Finally the statistical file keeps count of the number of generated stereoisomers and the final total number of isomers containing all the nonredundant topological and topographical isomers.

The reported tests were performed on a Pentium II dual processor, 233 MHz speed, 128 MB of RAM, Red Hat Linux (V. 6.0), using the compiled software written in C.

5. CONCLUSIONS

A generation program, CAMGEC2, has been developed for the exhaustive and nonredundant generation of topological and topographical isomers with isolated and spiro cycles including heteroatoms and unsaturated bonds. The program is able to analyze and perceive structural features such as SC candidates, true SC, relative groups of CIP priorities, symmetry planes, and pairs of symmetry. The program has shown to be useful for working with one compound or with families of compounds using elements of the Graph Theory, with ad hoc tree making algorithms.

New features developed include (1) generation algorithms based on generation vectors, whose structure is dependent on symmetry occurrence, (2) a symmetry searching algorithm and a symmetry matrix defining corresponding pairs of symmetry in relation with a particular plane of reflection, (3) a configuration vector analysis algorithm that eliminates redundant configuration vectors allowing for an efficient procedure that eliminates possibilities of redundant isomers before the generation process takes place, and (4) a linear textual molecular representation format that consists of an extended *N*_tuple notation which allows for a good man-machine communication where the number, position, and type of SC present in the molecular graph are clearly specified as well as the relative CIP priorities of the SC bonding groups which is extremely useful for a good molecular graphic representation. Also, *N*_tuple notation is a convenient way of computer managing molecular information for further calculation processes, for instance, of graph invariants and their correlation studies with organic compound properties useful in molecular design and in VHTS.

As shown, simple algorithms can be used for getting good solutions in molecular generation of isomers in a modular and detailed way which is useful for studying specific families of compounds.

CAMGEC2 is a computer tool that could be used alone or as a complement to other computer programs in education and in research oriented to molecular structure studies and their relationships with physical, chemical, or physicochemical properties.

ACKNOWLEDGMENT

The authors are pleased to thank the Dirección de Investigación Científica y Tecnológica at the University of Santiago de Chile for financial support.

REFERENCES AND NOTES

(1) See for example, a collection of contributions in *Rev. Comput. Chem.*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 1990−1999; Vol. 1−13.
(2) Carhart, R. E.; Smith, D. H.; Brown, H.; Djerassi, C. An Approach to Computer-Assisted Elucidation of Molecular Structure. *J. Am. Chem. Soc*. **1975**, *97*, 5755−5762.
(3) Carhart, R. E.; Smith, D. H.; Gray, N. A.; Nourse, J. G.; Djerassi, C. GENOA: A Computer Program for Structure Elucidation Utilizing Overlapping and Alternative Substructures. *J. Org. Chem*. **1981**, *46*, 1708−1718.
(4) Gray, N. A. B. *Computer-Assisted Structure Elucidation*; Wiley-Interscience: New York, 1986.
(5) Funatsu, K.; Miyabayaski, N.; Sasaki, S. Further Development of Structure Generation in the Automated Structure Elucidation System CHEMICS. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 18−28.
(6) Christie, 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.
(7) Nilakantan, R.; Bauman, N.; Venkataraghavan, R. A Method for Automatic Generation of Novel Chemical Structures and Its Potential Applications to Drug Discovery. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 527−530.
(8) Pivina, T. S.; Molchanova, M. S.; Shcherbukhin, V. V.; Zefirov, N. S. Computer Generation of Caged Frameworks Which Can be Used as Synthons for Creating High-Energetic Materials. *Propell., Expl., Pyrotech.* **1994**, *19*, 286−289.
(9) Pivina, T. S.; Shcherbukhin, V. V.; Molchanova, M. S.; Zefirov, N. S. Computer-Assisted Prediction of Novel Target High-Energy Compounds. *Propell., Expl., Pyrotech.* **1995**, *20*, 144−146.
(10) Clark, D. E.; Frenkel, D.; Levy, S. A.; Li, J.; Murray, C. W.; Robson, B.; Waszkowycs, B.; Westhead, D. R. PRO_LIGAND: An Approach to *de novo* Molecular Design. 1. Application to the Design of Organic Molecules. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 13−32.
(11) See, for example: *Chem. Rev.* **1993**, *93*(7) dedicated to Molecular Mechanics and Modeling.
(12) Bone, R. G. A.; Villar, H. O. Exhaustive Enumeration of Molecular Substructures. *J. Comput. Chem.* **1997**, *18*(1), 86−107.
(13) Cyvin, S. J.; Wang, J.; Brunvoll, J.; Cao, S.; Li, Y.; Cyvin, B. N.; Wang, Y. Staggered Conformers of Alkanes: Complete Solution of the Enumeration Problem. *J. Mol. Struct.* **1997**, 413−414, 227−239.
(14) Brunvoll, J.; Cyvin, S. J.; Cyvin, B. N. Enumeration of Tree-like Octogonal Systems. *J. Math. Chem.* **1997**, *21*(2), 193−196.
(15) Caporossi, G.; Hansen, P. Enumeration of Polyhex Hydrocarbons to $h = 21$. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 610−619.
(16) Fujita, S. Pseudo-Point Groups and Subsymmetry-Itemized Enumeration for Characterizing the Symmetries of 1,4-Dioxane and 1,4-Oxathiane Derivatives. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 876−884.
(17) Wieland, T.; Kerber, A.; Laue, R. Principles of the Generation of Constitutional and Configurational Isomers. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 413−419.
(18) Molchanova, M.; Zefirov, N. S. Irredundant Generation of Isomeric Molecular Structures with Some Known Fragments. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 8−22.
(19) Agarwal, K. K. An Algorithm for Computing the Automorphism Group of Organic Structures with Stereochemistry and a Measure of its Efficiency. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 402−404.
(20) Contreras, M. L.; Valdivia, R.; Rozas, R. Exhaustive Generation of Organic Isomers. 1. Acyclic Structures. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 323−330.
(21) 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.
(22) Contreras, M. L.; Rozas, R.; Valdivia, R. Exhaustive Generation of Organic Isomers. 3. Acyclic, Cyclic, and Mixed Compounds. *J. Chem. Inf. Comput. Sci*. **1994**, *34*, 610−616.
(23) Luinge, H. J.; Van der Maas, J. H. AEGIS, an Algorithm for the Exhaustive Generation of Irredundant Structures. *Chemom. Intell. Lab. Sys.* **1990**, *8*, 157−165.
(24) Nourse, J. G.; Smith, D. H.; Carhart, R. E.; Djerassi, C. Exhaustive Generation of Stereoisomers for Structure Elucidation. *J. Am. Chem. Soc.* **1970**, *101*, 1216−1223.
(25) Benecke, C.; Grund, R.; Hohberger, R.; Kerber, A.; Laue, R.; Wieland, Th. MOLGEN, A Generator of Connectivity Isomers and Stereoisomers for Molecular Structure Elucidation. *Anal. Chim. Acta* **1995**, *314*, 141−147.
(26) Wieland, T. Enumeration, Generation, and Construction of Stereoisomers of High-Valence Stereocenters. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 220−225.
(27) Read, R. C. In *Chemical Applications of Graph Theory*; Balaban, A. T., Ed.; Academic Press: New York, 1976; pp 11−60.
(28) Contreras, M. L.; Trevisiol, G. M.; Alvarez, J.; Arias, M. G.; Rozas, R. Exhaustive Generation of Organic Isomers. 5. Unsaturated Optical and Geometrical Stereoisomers, and a New CIP Subrule. *J. Chem. Inf. Comput. Sci*. **1999**, *39*, 475−482.

(29) Fan, B. T.; Panaye, A.; Doucet, J. P.; Barbu, A. Ring Perception. A new Algorithm for Directly Finding the Smallest Set of Smaller Rings from a Connection Table. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 657−62.
(30) Balasubramaniam, K. Computational Techniques for the Automorphism Groups. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 621−626.
(31) Prelog, V.; Helmchen, G. Basic Principles of the CIP−System and Proposals for a Revision. *Angew. Chem., Int. Ed. Engl.* **1982**, *21*, 567−583.
(32) Cusler, R. H. Mathematical Statements about the Revised CIP-System. *Match* **1986**, *21*, 3−31.
(33) Figure 10 (examples A−K) from Rücker and Rücker;[34] Figure 10 (examples L−N1) from Balasubramanian and Basak;[35] Figure 10 (example O) Razinger et al.;[36] Figure 10 (example S) Prelog and Helmchen.[31] These examples have been reproduced correctly by CAMGEC2.
(34) Rücker, G.; Rücker, C. On Topological Indices, Boiling Points and Cycloalkanes. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 788−802.
(35) Balasubramanian, K.; Basak, S. C. Caracterization of Isospectral Graphs Using Graphs Invariants and Derived Orthogonal Parameters. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 367−373.
(36) Razinger, M.; Balasubramanian, K.; Perdih, M.; Munk, M. E. Stereoisomer Generation in Computer-Enhanced Structure Elucidation. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 812−825.

CI0003395