Exhaustive Generation of Organic Isomers. 4. Acyclic Stereoisomers with One or More Chiral Carbon Atoms

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

Chemical Sciences Department, Faculty of Chemistry and Biology, University of Santiago de Chile,
Casilla 5659, Santiago-2, Chile, and Computer and Information Department, Faculty of Engineering,
University of Tarapaca, Casilla 6-D, Arica, Chile

A stereochemical computer assisted molecular generation and counting system, S-CAMGEC, based on graph theory has been created for the exhaustive and nonredundant generation of (R) and (S) carbon stereoisomers of acyclic compounds, which may contain heteroatoms of different valences and multiple bonds. The program takes constitutional N-tuples, searches and labels chiral centers, searches for symmetry, assigns priorities according to Cahn, Ingold, and Prelog rules, and generates stereoisomers in excellent agreement with results coming from reported enumeration procedures. For that a new molecular code was created which defines each stereoisomer in a single and unique way shown through several examples.

INTRODUCTION

A single molecular formula can represent a very big number of isomers. For instance there are 461 366 acyclic isomers with the formula $C_{10}H_{18}N_2$ having three double bonds.[1]

Computational programs for the enumeration and generation of isomers from a molecular formula have to do its work in a reasonable time. For that, the algorithm has to consider at least the following two points:

(a) a method for determining redundancy in the generation process and

(b) a reduction to a minimum in the number of elements to be checked out during such a process.

Actually to accomplish the first point, generator systems use different algorithms but with the same principle: the use of isomer canonical representations. Hendrickson and Parks[2] mention two ways of applying that representation. The first one consists in generating each isomer canonical representation, and, after verifying that the new representation does not exist among the already stored ones, it is kept as a unique representation. Otherwise, it is discarded.

However this alternative is not efficient because the number of structures to be stored could increase considerably and the main memory could become insufficient. In addition, disk access, both for the storage and for the mentioned comparison, influences negatively over the system performance.

The second alternative, widely employed, uses canonical representation for redundancy identification just at the moment of generation. In this alternative, redundancy is detected inmediately by checking the new generated isomer, meanwhile the information is still in the main memory with the advantage of requiring just one disk access for storing the generated representation.

This last method had been used by different systems such as SKEL_GEN in which the canonical representation is given by the maximum superior triangle of adjacency matrix;[2] CAMGEC that employs as a canonical representation the

maximum lexicographic value of a N-tuple[1,3,4] and other pioneering works;[5,6] DENDRAL[7] which uses the topological symmetry; and CHEMICS[8,9] which defines a maximum value of a connectivity stack as the canonical representation.

A generator system that considers spatial factors needs a special notation for distinguishing between one stereoisomer and another one, different from the existing notation[3,10,11] for constitutional isomers. That is proposed in this work.

An approach informed by Tatsuya[12] offers a new way of transforming stereochemical formulas into a graph. For that, a list of adjacency is used together with two typical computational processes: one for transformation and another one for naming each graph in a unique way.

On the other hand, Whyte and Clugston[13] enumerate alkane isomers making a special reference to decane stereoisomers. They describe an algorithm for the stereoisomers obtention from already generated isomers. This general strategy will be used in this work but with a different implementation because the algorithm for generating the isomers is not recursive in our case.

STEREOISOMERS GENERATOR SYSTEM DESCRIPTION

The S-CAMGEC system is an extension of CAMGEC[1,3,4] taking stereoisomerism in consideration. The new stereoisomers generator system employs a modification of the N-tuple type notation proposed by Contreras, Rozas, and Valdivia[4] and considers Cahn, Ingold, and Prelog rules for group priority assignements.

The system uses the second strategy of generation mentioned by Hendrickson and Parks[2] and employs as a canonical representation a configurational vector whose content at a given moment during the generation process indicates which is the configuration corresponding to each of the chiral centers occurring in the molecular compound.

Extended N-tuple Notation. The N-tuple notation here developed contains the following descriptors for each node:[4]

•Tree node number or number of the N-tuple component.

•Symbol of the chemical element representing the node. When the element is represented by a symbol having two characters, two digits are used. Otherwise only one digit is used.

•Number of sons hanging from the node in the tree.

•Type of bond between the node and its father. It can be **r** (if it is root obviously it has not a father), **s** (single bond), **d** (double bond), or **t** (triple bond)).

Canonical N-tuple uses a maximum inverse lexicographical order and the precedence rules for the atoms and for the bonds described before.[4]

Consideration of the spatial factor makes necessary to take into account five additional digits in the new N-tuple for allowing differentiation between one stereoisomer and another one. In this way if the node is a carbon atom, then the following digits are added:

•**q** if the carbon atom is chiral or **n** if it is not.

If the node represents a chiral center, then, Cahn—Ingold—Prelog rules for assigning priorities to each of the four groups joined to it are applied. For that, the following four digits are used as indicated:

•Father's priority: means priority exerted by the father over the chiral carbon atom representing the node.

•Priority of the son 1: indicates priority of the first son hanging from the node in the tree, over the chiral atom.

•Priority of son 2: indicates priority of the second son.

•Priority of son 3: indicates priority of the third son.

A special case is presented when the node in consideration corresponds to the first node of the extended N-tuple. In this case, the four values represent sequentially assigned priorities of the four sons over the root node.

For a particular node, order of priorities will indicate configuration of the represented atom. An increasing order will represent configuration (R) and a decreasing order of priorities an (S) configuration. For instance, the canonical N-tuple for 2-butanol is (1c3r 2c1s 3c0s 4c0s 5o0s). Extended N-tuple for (R)-2-butanol is (1c3rq1230 2o0s 3c1sn 4c0sn 5c0sn), and extended N-tuple for (S)-2-butanol is (1c3rq3210 2c0sn 3c1sn 4c0sn 5o0s). Priorities for each group are the same (always oxygen > ethyl > methyl). Priorities order code however is different for each stereoisomer (1230 for (R) configuration vs 3210 for the (S) one). In Figure 1 are depicted the corresponding trees.
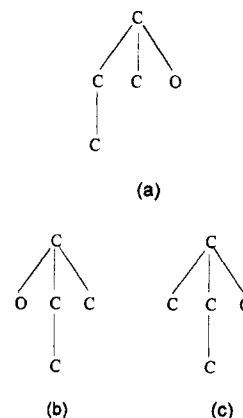
Generator System. The stereoisomers computer assisted molecular generation and counting system, S-CAMGEC, consists basically of a main module, whose aim is to accomplish the generation process starting from a previous process, CAMGEC,[4] that creates an N-tuple for each register of the constitutional isomers archive. S-CAMGEC involves four additional subprocesses:

1. labeling of chiral carbon atoms,
2. symmetry determination,
3. priority assignment, and
4. rearrangement of the extended N-tuples.

These subprocesses are integrated in the way shown in Figure 2. The first three subprocesses were integrated in a unique module, because they share common data for the obtention of the extended N-tuple.

Subprocess 4, also named generation, was implemented as a separated module. This module is in charge of the obtention of the configurational extended N-tuples that is possible to generate from a single N-tuple. The corresponding global diagram is illustrated in Figure 3.

Labeling of Chiral Carbon Atoms. The N-tuple notation properties and chiral center characteristics offer two ways of doing the job: labeling as chiral for those carbon atoms having four different joined groups or labeling as no-chiral

Figure 1. Representation of 2-butanol stereoisomers: (a) canonical representation, (b) (R)-2-butanol, and (c) (S)-2-butanol.
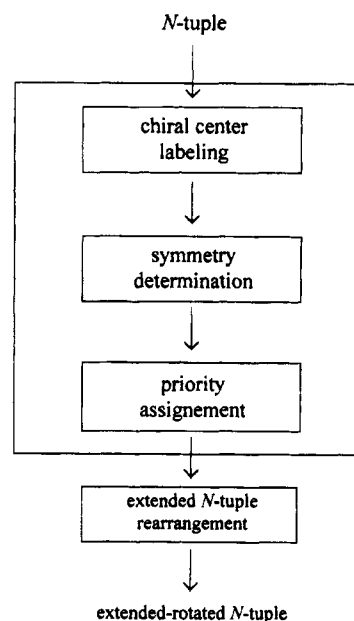
Figure 2. Subprocesses of the stereoisomers generator system.

for those atoms having two equal groups joined to them. Both approaches require the N-tuple being completely analyzed. Each potential chiral atom must obey the following conditions:

(a) It has to be joined to three or four different groups (if it is joined to one or two groups it is assumed that the rest of the joined groups are hydrogen atoms; in this case it is labeled as a non-chiral one.).

(b) It has to not be joined by any double or triple bond.

If the atom accomplishes both preceding conditions, it is labeled with a **q** for chiral; otherwise with an **n** for nonchiral.

To simplify comparison between the four branches joined to a carbon atom, a process of rotation[1] of the N-tuple is done with the aim of keeping the atom which is in course of being labeled as the root node. The resulting N-tuple is called rotated N-tuple. This is the tuple used for the verification of the preceding conditions. The result of this rotation process is illustrated in Figure 4. Original $C_9H_{20}O$ isomer tree is represented in Figure 4a, while Figure 4b represents a rotated tree that allows one to analyze node number 2 (now the root position).

For each chiral carbon atom the corresponding rotated tuple is kept on a chiral centers table. This table is then used for the processes of symmetry detection and priority
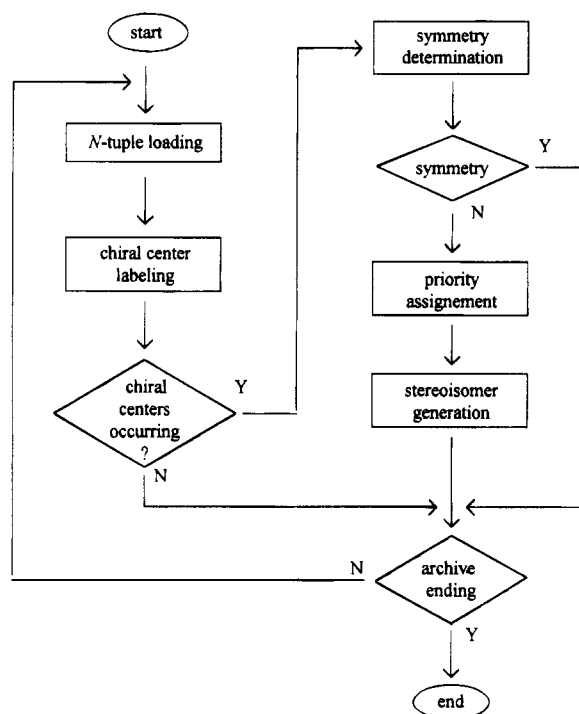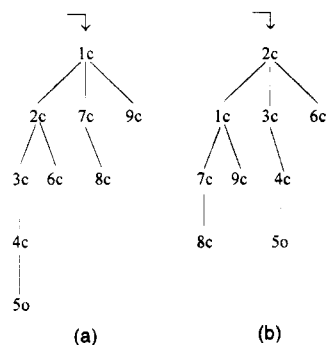
Figure 3. S-CAMGEC global diagram.

Figure 4. Tree representations of a $C_9H_{20}O$ isomer: (a) canonical representation and (b) rotated representation.

assignment. In the case the root node of the original N-tuple is itself a chiral center, then this is the N-tuple stored in the chiral centers table.

Detection of Symmetric Structures. Symmetry is determined with the help of the chiral centers table. The algorithm used is similar to the depth search utilized before.[4] A typical case considers symmetry searching process starting from two joined atoms and making a node to node comparison of their branches. Comparison of nodes occurring in each branch is done taking in consideration the following points:

(i) chemical symbol of the node atom,

(ii) number of vertices that hang from the node, and

(iii) type of bond between the node and its father.

Figure 5 shows the walk followed for symmetry checking in the case where only two chiral centers occur: 1c and 2c. In this figure **chiral[i]** and **chiral[j]** represent pointers to the rotated tuples stored in the chiral centers table. These pointers allow for a descendent walk where comparison is done for nodes pointed by chiral[i] vs chiral[j]. In this case the process finishes at the **end** pointer shown in the figure because in this point node 4c1s defers from node 8c0s in vertices number (sons number).
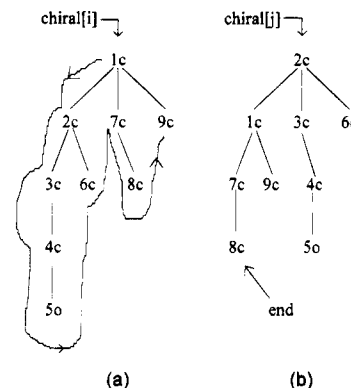
Figure 5. Walking for the search of symmetrical branches.

When two symmetrical branches are found during the walking over the trees, counting of such structures represents a number m of meso forms which will be further discounted from chiral stereoisomers.

Priority Assignment. Priority of the groups linked to the chiral atoms are assigned according to the rules of Cahn—Ingold—Prelog which are based on the atomic number of the elements (given in stablished organic chemistry books[14]).

Numbers from 0 to 4 are assigned to the four groups associated to the chiral center according to their priorities. Biggest priority is assigned number 1 and smallest one, number 4 or 0 (0 is assigned to hydrogen atoms; they are not present in the N-tuple.)

These values are assigned through an algorithm similar to the depth search one and based on the following rules:

1. Groups bonded to the chiral carbon atom are labeled in a descendent order upon the atomic number of the atom directly joined to the carbon (a bigger atomic number implies bigger priority).

2. When two or more atoms joined to the asymmetric atom are equal, then the sum of the atomic number of the following elements in the branch directly joined to each of the previous elements are considered. A bigger value means bigger priority. This rule is recursively applied until different priorities are found.

3. Double and triple bonds are considered as two or three single bonds, respectively.

Branches are ordered upon priority of the atom directly bounded to the chiral center in such a way that finally an extended notation results which takes care of priorities of each of the radicals joined to each chiral center.

An extended notation with an increasing numerical order of group priorities represents an (R) configuration. On the contrary, a decreasing order of priorities represents a (S) configuration (see Figure 1).

Stereoisomers Exhaustive Generation. Number of chiral centers and number of meso forms found according to precedent paragraphs allow for the enumeration of the stereoisomers that can be generated from each N-tuple in analysis. The extended N-tuple obtained by the precedent described procedure is submitted to a rearrangement of the branches joined to each chiral center. As it was explained a rearrangement in increasing numerical order of son priorities will give an (R) configuration (second digit after symbol q is smaller than the third one), and a rearrangement in a decreasing numerical order of priorities will give an (S) configuration.

Table 1. Matrix of Configurations for Isomers Having Four Chiral Carbon Atoms Such as Glucose Isomers

| config	chiral atoms			
	1	2	3	4
1	R	R	R	R
2	S	R	R	R
3	R	S	R	R
4	S	S	R	R
5	R	R	S	R
6	S	R	S	R
7	R	S	S	R
8	S	S	S	R
9	R	R	R	S
10	S	R	R	S
11	R	S	R	S
12	S	S	R	S
13	R	R	S	S
14	S	R	S	S
15	R	S	S	S
16	S	S	S	S

In that way for each chiral center only two trees will be generated: one where its branches will hang in an increasing numerical order of priorities and the other where the branch priority numerical order will be a decreasing one.

On the other hand, a matrix will be generated with the different configurations. The matrix will have a special form that will allow for the generation of a jth row from the previous ith row. This will be valid for j values between 2 and 2^q, where q is the present number of chiral centers. For the first column of this matrix, there is an alternating order between configurations (R) and (S). For the second column, the alternating sequence varies two by two, i.e., (R)—(R) and (S)—(S). In general, alternating sequences vary from $2^i/2$ in $2^i/2$ (with i between 0 and q). This indicates that in the qth column half of the configurations will be (R) and half will be (S) as can be seen in Table 1 for glucose isomers. These isomers have four chiral carbon atoms for each one, and they are 16 different compounds.

Properties of this matrix allow for the use of a configuration arrangement that keeps only the relevant information for the configuration of each of the chiral centers of the compound being generated. That means information relative to the jth row. The idea is to keep in a single vector three essential data: the actual configuration being created, the maximum number of actual configurations to be generated, and the number of configurations (R) or (S) already generated. Length of this vector will be given for the number of labeled chiral centers (q). In this way each component of the jth row will have the following structure:

- ith column,
- actual configuration,
- maximum number of these configurations to be generated, and
- number of these configurations already generated.

This structure is initialized with the following: an (R) configuration for every chiral center, the maximum number of (R) configurations for each of the vector elements, and the rest of the configurations values on 1. The structure is properly updated each time a new rearranged tuple is generated. In other words, the number of configurations under generation is increased and when it is equal to the maximum number of configurations of that type, then the next configuration changes to the other configuration (if it was (R) it changes to (S) and vice versa). When finishing
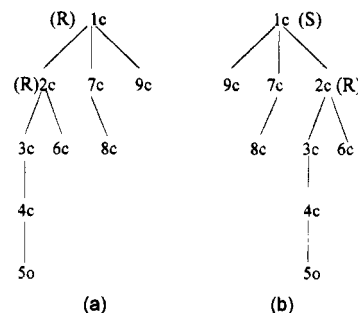
Figure 6. Diastereoisomers of 3,4-dimethylheptane: (a) $(3R,4R)$ isomer and (b) $(3S,4R)$ isomer.
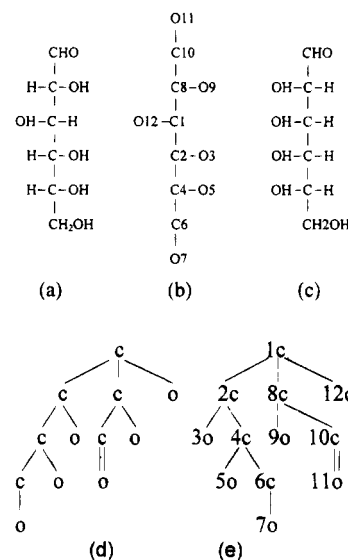
Figure 7. Sugars representation: (a) D-(+)-glucose, (b) D-(+)-glucose graph, (c) L-(−)-talose, (d) tree for hexoses canonical representation, and (e) rotated tree for D-(+)-glucose.

the generation process, the last generated tuple includes only configurations of the (S) type.

As an example in Figure 6 trees of two diastereoisomers of 3,4-dimethylheptane are shown. Figure 6a shows the tree for the $(3R,4R)$ isomer, and Figure 6b shows the tree for the $(3S,4R)$ isomer. Carbon 3 is represented by node 1c, and carbon 4 by node 2c. Also, in Figure 7, the graph and tree for D-(+)-glucose, both numbered accordingly, are depicted.

Note that node number 1 in Figure 7 (parts b and e) represents the glucose carbon atom 3 with an (S) configuration. The rest of the chiral carbon atoms have an (R) configuration. The extended N-tuple for D-(+)-glucose is as follows (see Figure 7e): (1c3rq3210 2c2sq2130 3o0s 4c2sq2130 5o0s 6c1sn 7o0s 8c2sq3120 9o0s 10c1sn 11o0d 12o0s). For comparison, the canonical N-tuple is (see Figure 7d) (1c3r 2c2s 3c2s 4c1s 5o0s 6o0s 7o0s 8c2s 9c1s 10o0d 11o0s 12o0s). When this tuple is extended, without any rotation, it represents L-(−)-talose. This sugar has the four chiral atoms with an (S) configuration, and the corresponding extended N-tuple is (1c3rq3210 2c2sq2310 3c2sq2310 4c1sn 5o0s 6o0s 7o0s 8c2sq3210 9c1sn 10o0d 11o0s 12o0s).

RESULTS

Stereoisomers generation program developed for acyclic compounds is of general use. Results obtained for the exhaustive generation of stereoisomers of acyclic alkanes, mono- and dihalogenated alkanes, alcohols, ethers, and amines are presented in Tables 2−8. In Table 2 a compari-

756 *J. Chem. Inf. Comput. Sci., Vol. 35, No. 4, 1995*

CONTRERAS ET AL.

Table 2. Number of Stereoisomers for Alkanes and Some Families of Halogen, Oxygen, and Nitrogen Derivatives

carbon atoms	number of configurations				
	C_nH_{2n+2}	$C_nH_{2n+1}X$	$C_nH_{2n+2}O$	$C_nH_{2n}XY$	$C_nH_{2n+4}N_2$
1[a]	1	1	1	1	2
b	1	1	1	1	c
2[a]	1	1	2	3	6
b	1	1	2	3	c
3[a]	1	2	3	8	16
b	1	2	3	8	c
4[a]	2	5	8	23	49
b	2	5	8	23	c
5[a]	3	11	18	69	143
b	3	11	18	69	c
6[a]	5	28	47	208	436
b	5	28	47	208	c
7[a]	11	74	123	636	1343
b	11	74	123	636	c
8[a]	24	199	338	1963	4194
b	24	199	338	1963	c
9[a]	55	550	934	6098	13136
b	55	551	935	6099	c
10[a]	135	1551	2654	19056	41501
b	136	1553	2657	19059	c

[a] This work. [b] Reference 15. [c] Not calculated.

Table 3. Alkane Stereoisomers Generated by S-CAMGEC

formula	nc[a]	total cca	meso comp	nonster isomers	chiral stereo	no. of config
C_1H_4	1	0	0	1	0	1
C_2H_6	1	0	0	1	0	1
C_3H_8	1	0	0	1	0	1
C_4H_{10}	2	0	0	2	0	2
C_5H_{12}	3	0	0	3	0	3
C_6H_{14}	5	0	0	5	0	5
C_7H_{16}	9	2	0	7	4	11
C_8H_{18}	18	6	1	13	10	24
C_9H_{20}	35	18	1	20	34	55
$C_{10}H_{22}$	75	52	4	35	96	135
$C_{10}H_{22}$	75		5[b]	35[c]	96[d]	136[e]

[a] nc: number of constitutional isomers upon CAMGEC;[4] cca: chiral centers analyzed; nonster: nonstereo (without any asymmetric carbon atom). [b] Forty achiral isomers have been reported[17,19] (we separated them as five meso and 35 nonstereoisomers). [c] Reference 16. [d] Reference 19. [e] Reference 16, 17, and 19.

Table 4. Monohalogenated Alkane Stereoisomers Generated by S-CAMGEC

formula	nc[a]	total cca	meso comp	nonster isomers	chiral stereo	no. of config
C_1H_3X	1	0	0	1	0	1
C_2H_5X	1	0	0	1	0	1
C_3H_7X	2	0	0	2	0	2
C_4H_9X	4	1	0	3	2	5
$C_5H_{11}X$	8	3	0	5	6	11
$C_6H_{13}X$	17	10	0	8	20	28
$C_7H_{15}X$	39	30	0	14	60	74
$C_8H_{17}X$	89	87	0	23	176	199
$C_9H_{19}X$	211	249	1	39	510	550
$C_{10}H_{21}X$	507	703	2	65	1484	1551
$C_{10}H_{21}X^b$			69		1484	1553

[a] Same as specified in Table 3. [b] Reference 19.

son with literature values[15] is done. Number of configurations considers chiral and achiral isomers together.

In Tables 3—8 a number of constitutional isomers (nc) upon CAMGEC,[4] a number of total chiral centers analyzed (cca) for each formula, a number of meso structures recognized by S-CAMGEC together with number of nonstereoisomers (containing no chiral carbon atoms), and chiral generated stereoisomers are presented. Meso compounds, as known, have chiral atoms and symmetry elements in their structures.

There is a direct relationship between data in Table 2 and data in Tables 3—8. For instance, the 19 056 configurational $C_{10}H_{20}XY$ isomers in Table 2 are constituted by three meso compounds, 65 nonstereoisomers, and 18 988 chiral stereoisomers (see Table 6). Results have been presented in both ways just for making a comparison with available literature values.[13,15-19]

Table 9 shows CPU times in the format minutes:seconds.

DISCUSSION

A comparison with values calculated by Read[15] considering Polya theorem,[20] nowadays known as Redfield—Polya theorem[21] (by considering the work by Redfield[22] published in 1927), is presented in Table 2 for particular compound families. Results present an excellent level of reproducibility for every kind of isomers with eight or less carbon atoms. However meanwhile the number of generated configurations is coincident with reported ones; caution should be taken here because there are a number of structures that do not contain chiral atoms at all. Effective chiral stereoisomers are those detailed in Tables 3—8.

Data in Table 2 show a small difference at the level of 10 carbon atoms for alkanes and nine carbon atoms for the other families. The reason for that difference is due to the occurrence of pseudoasymmetric atoms which are not yet considered by the program.[23] This aspect can be easily understood by analyzing data in Table 3 as follows.

S-CAMGEC generated values for hydrocarbons, presented in Table 3, have a very good reliability as compared with

informed enumerated values.[16,17,19] The only difference with literature values is the number of meso compounds for decane isomers. A number of 40 achiral isomers (including meso compounds and nonstereoisomers) has been reported by Wang and Gu[17] and by Robinson et al.,[19] meanwhile 35 nonstereoisomers have been published by Henze and Blair[16] in very good agreement with our own results. That fact allows us to conclude that the number of meso compounds reported should be 5; however, S-CAMGEC generates only four meso compounds.

On the other hand, there exists one $C_{10}H_{22}$ isomer that contains a pseudoasymmetric atom, the 3,4,5-trimethylheptane.[13,14] Effectively, carbon atom number 4 in this molecule is achiral when bounded either to two (R)-sec-butyl fragments or two (S)-sec-butyl fragments. But when atom 4 is bound to both an (R)-sec-butyl group and an (S)-sec-butyl group it became chiral. That is why atom 4 is called a pseudoasymmetric atom. However in the last case structures are symmetrical. This situation makes possible the existence of two meso forms of it: the (3R,4R,5S)-3,4,5-trimethylheptane and the (3S,4S,5R)-3,4,5-trimethylheptane. S-CAMGEC considers at the moment just one of them. That is why a total of only four meso compounds are generated. In Table 3, reported values are shown for making clear this point.

The same effect is observed on Tables 4—8. For monohalogenated alkanes results from enumeration procedure[19] report 1484 chiral, 69 achiral, and 1553 steric planted trees for isomers with 10 carbon atoms. Our results (see Table

ACYCLIC STEREOISOMERS WITH ONE OR MORE CHIRAL C ATOMS

J. Chem. Inf. Comput. Sci., Vol. 35, No. 4, 1995 **757**

Table 5. Monooxygenated Hydrocarbon Stereoisomers Generated by S-CAMGEC

formula	nc[a]	total cca	meso comp	nonster isomers	chiral stereo	no. of config
C_1H_4O	1	0	0	1	0	1
C_2H_6O	2	0	0	2	0	2
C_3H_8O	3	0	0	3	0	3
$C_4H_{10}O$	7	1	0	6	2	8
$C_5H_{12}O$	14	4	0	10	8	18
$C_6H_{14}O$	32	14	0	19	28	47
$C_7H_{16}O$	72	45	0	33	90	123
$C_8H_{18}O$	171	138	1	61	276	338
$C_9H_{20}O$	405	406	1	107	826	934
$C_{10}H_{22}O$	989	1183	6	194	2454	2654

[a] Same as specified in Table 3.

Table 6. Dihalogenated Alkane Stereoisomers (Different Halogen Atoms) Generated by S-CAMGEC

formula	nc[a]	total cca	meso comp	nonster isomers	chiral stereo	no. of config
C_1H_2FI	1	0	0	1	0	1
C_2H_4FI	2	1	0	1	2	3
C_3H_6FI	5	3	0	2	6	8
C_4H_8FI	12	10	0	3	20	23
$C_5H_{10}FI$	31	32	0	5	64	69
$C_6H_{12}FI$	80	99	0	8	200	208
$C_7H_{14}FI$	210	301	0	14	622	636
$C_8H_{16}FI$	555	908	0	23	1940	1963
$C_9H_{18}FI$	1479	2714	1	39	6058	6098
$C_{10}H_{20}FI$	3959	8061	3	65	18988	19056

[a] Same as specified in Table 3.

Table 7. Stereoisomers of Diamine Compounds Generated by S-CAMGEC

formula	nc[a]	total cca	meso comp	nonster isomers	chiral stereo	no. of config
$C_1H_6N_2$	2	0	0	2	0	2
$C_2H_8N_2$	6	0	0	6	0	6
$C_3H_{10}N_2$	14	2	0	12	4	16
$C_4H_{12}N_2$	38	10	0	29	20	49
$C_5H_{14}N_2$	97	42	0	59	84	143
$C_6H_{16}N_2$	260	155	3	129	304	436
$C_7H_{18}N_2$	688	538	4	261	1078	1343
$C_8H_{20}N_2$	1856	1791	16	544	3634	4194
$C_9H_{22}N_2$	4994	5774	26	1088	12022	13136
$C_{10}H_{24}N_2$	13550	18251	93	2199	39116	41408

[a] Same as specified in Table 3.

Table 8. Monooxygenated Hydrocarbon Stereoisomers Generated by S-CAMGEC in an Independent Way as Alcohols and Ethers

formula	nc[a]	total cca	meso comp	nonster isomers	chiral stereo	no. of config
C_1H_4O	t	1	0	1	0	1
	oh/e	1/0		1/0		1/0
C_2H_6O	t	2	0	2	0	2
	oh/e	1/1		1/1		1/1
C_3H_8O	t	3	0	3	0	3
	oh/e	2/1		2/1		2/1
$C_4H_{10}O$	t	7	0	6	2	8
	oh/e	4/3		3/3	2/0	5/3
$C_5H_{12}O$	t	14	0	10	8	18
	oh/e	8/6		5/5	6/2	11/7
$C_6H_{14}O$	t	32	0	19	28	47
	oh/e	17/15		8/11	20/8	28/19
$C_7H_{16}O$	t	72	0	33	90	123
	oh/e	9/33		14/19	60/30	74/49
$C_8H_{18}O$	t	171	1	61	276	338
	oh/e	89/82	0/1	23/38	176/100	199/139
$C_9H_{20}O$	t	405	1	107	826	934
	oh/e	211/194	1/0	39/68	510/316	550/384
$C_{10}H_{22}O$	t	989	6	194	2454	2654
	oh/e	507/482	2/4	65/129	484/970	1551/1103

[a] Same as specified in Table 3; t = total; oh/e = alcohol/ether.

Table 9. CPU Times for the S-CAMGEC Generation of Stereoisomers

carbon atoms	CPU time[a]				
	C_nH_{2n+2}	$C_nH_{2n+1}X$	$C_nH_{2n+2}O$	$C_nH_{2n}XY$	$C_nH_{2n+4}N_2$
1	0:00	0:00	0:00	0:00	0:00
2	0:00	0:00	0:00	0:00	0:01
3	0:00	0:00	0:01	0:00	0:01
4	0:00	0:01	0:01	0:01	0:03
5	0:01	0:01	0:02	0:02	0:06
6	0:01	0:01	0:06	0:05	0:17
7	0:02	0:03	0:13	0:14	0:46
8	0:05	0:06	0:35	0:37	2:08
9	0:12	0:14	1:33	1:41	5:51
10	0:20	0:34	4:12	4:38	12:58

[a] Minutes:seconds. MicroVax-Q2, Model Dec-System 2100, Ultrix-32m (V.4.1), 16 MB RAM, with two disk units (RF72E, 1 GB; RZ24, 209 MB).

4) show 1484 chiral and 67 achiral isomers (65 nonstereoisomers and two meso compounds: 3,4,5-trimethyl-4-halogenheptane and 3,5-dimethyl-4-(halogenmethyl)heptane. Both of these structures have a central pseudoasymmetric carbon atom, and then each one should have two meso forms. S-CAMGEC does not considers pseudoaymmetric atoms yet[23] and generates just one meso form for each structure). On the other hand, for nonanes, 41 achiral planted trees are informed[19] and again there is a difference with our results: 39 nonstereoisomers and one meso compound. The meso isomer, 3,5-dimethyl-4-halogenheptane, contains a pseudoasymmetric carbon atom and that is why S-CAMGEC generates only 550 of the 551 enumerated[19] nonane configurations (see Tables 2 and 4).

Data in Table 5 contain results for $C_nH_{2n+2}O$ isomers and show 934 and 2654 configurations for isomers with nine and 10 carbon atoms, respectively. The corresponding reported values obtained by enumeration[15] are 935 and 2657 (see Table 2). Our results (Table 5) show one and six meso compounds, respectively. For a better analysis, data have been obtained for alcohols and for ethers in a separated way using CAMGEC facilities[1] for that purpose. Results are presented in Table 8 for the total isomers as an ensemble (indicated with a t) and for the two different kind of isomers (indicated by oh/e for the alcohol/ether). Compounds with nine carbon atoms show no meso ether forms and one meso alcohol (see Table 8), the 3,5-dimethyl-4-hidroxyheptane which has a pseudoasymmetric atom. Results for isomers with 10 carbon atoms show two meso forms for alcohols and four meso forms for ether compounds. Analysis for alcohol meso forms is similar to the one for monohalogenated compounds explained before. For ethers, enumeration results[17] show 134 achiral isomers which are one unit bigger

than S-CAMGEC results (four meso + 129 nonstereoisomers in Table 8). In this case the meso compound that has a pseudoasymmetric atom is the 3,5-dimethyl-4-methoxyheptane.

Literature data for dihalogenated compounds obtained by an enumeration process[18] based on Polya—Redfield theorem also showed a very good reliability of our results (see Table 6) with a perfect agreement for achiral and chiral isomers.

758 *J. Chem. Inf. Comput. Sci., Vol. 35, No. 4, 1995*

CONTRERAS ET AL.

There are no detailed reported values for dihalogenated alkanes with nine and 10 carbon atoms, neither for any $C_nH_{2n+4}N_2$ isomer. Results found by S-CAMGEC for these last compounds are presented in Tables 2 and 7. Many meso compounds appear, and an analysis should be done about pseudoasymmetric atoms occurrence in this case.

S-CAMGEC algorithms allow for a relatively easy extension to heteroatom stereoisomerism. A first point should be modification of the utilized *N*-tuple which could be a direct extrapolation of chiral carbon notation.

On the other hand, CPU time increases exponentially with the number of carbon atoms (see Table 9). Relative complexity of the structures considered also influences CPU time due to ramification. A bigger ramification involves bigger probability of finding chiral centers, and this increases the number of processes to be run. Also the occurrence of heteroatoms increases possibilities of finding chiral centers which is observed by comparison of alkane isomers with oxygen derivatives (alcohols, ethers), dihalogenated alkanes, or amine isomers. Effectively, for a fixed number of carbon atoms an increasing number of isomers is obtained when incorporating a halogen, oxygen, two halogens, or two nitrogen atoms in the structure (see Table 2) and that obviously increases the CPU time.

CONCLUSIONS

A system, S-CAMGEC, has been created for the exhaustive enumeration and generation of stereoisomers based on graph theory. The system is able to recognize achiral and chiral isomers. Within achiral structures it is capable of managing in a separated way meso compounds and nonstereoisomers. S-CAMGEC works using CAMGEC output and a new molecular code definition, based on *N*-tuples, for differentiating stereochemical configuration (*R*) or (*S*) according to Cahn, Ingold, and Prelog rules. Results have shown the importance of pseudoasymmetric atoms for the generation of meso compounds and were consistent with values reported by enumeration methods based on Polya—Redfield theorem which make S-CAMGEC of valuable use in several fields like structure elucidation, organic synthesis, stereochemistry, molecular design, and others.

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) Contreras, M. L.; Valdivia, R.; Rozas, R. Exhaustive Generation of Organic Isomers. 1. Acyclic Structures. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 323−330.

(2) Hendrickson, J. B.; Parks, C. A. Generation and Enumeration of Carbon Skeletons. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 101−107.

(3) 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.

(4) 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.

(5) Knop, J. V.; Müller, W. R.; Jericeviv, Z.; Trinajstić, N. Computer Enumeration and Generation of Trees and Rooted Trees. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 91−99.

(6) Knop, J. V.; Szymanski, K; Klasing, L.; Trinajstić, N. Computer Enumeration of Substituted Polyhexes. *Comput. Chem.* **1984**, *8*, 107−115.

(7) Lindsay, R. K.; Buchanan, B. G.; Feigenbaum, E. A.; Lederberg, J. *Application of Artificial Intelligence for Organic Chemistry*; McGraw-Hill: New York, 1980.

(8) Funatsu, K.; Miyabaiyashi, N.; Sasaki, S. Further Development of Structure Generation in the Automated Structure Elucidation System CHEMICS. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 18−28.

(9) 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.

(10) Balaban, A. T. Chemical Graphs. Part 12. Configurations of Annulenes. *Tetrahedron* **1971**, *27*, 6115−6131.

(11) Randić, M.; Nikolic, S.; Trinajstić, N. Compact Molecular Codes for Polycyclic Systems. *J. Mol. Struct. (Theochem)* **1988**, *165*, 213−228.

(12) Tatsuya, A. A New Method of Computer Representation of Stereochemistry. Transforming a Stereochemical Structure into a Graph. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 414−417.

(13) Whyte, J. R. C.; Clugston, M. J. The Enumeration of Isomers-with Special Reference to the Stereoisomers of Decane. *J. Chem. Ed.* **1993**, *70*, 874−876.

(14) March, J. *Advanced Organic Chemistry*; 4th ed.; Wiley: New York, 1992; pp 94−127.

(15) Read, R. C. In *Chemical Applications of Graph Theory*; Balaban, A. T., Ed.; Academic Press: New York, 1976; pp 11−60.

(16) Henze, H. R.; Blair, C. M. The Number of Isomeric Hydrocarbons of the Methane Series. *J. Am. Chem. Soc.* **1931**, *53*, 3077−3085.

(17) Wang, J.; Gu, F. Enumeration of Isomers of Polyethers. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 552−554.

(18) 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.

(19) Robinson, R. W.; Harary, F.; Balaban, A. T. The Numbers of Chiral and Achiral Alkanes and Monosubstituted Alkanes. *Tetrahedron* **1976**, *32*, 355−361.

(20) Polya, G. Kombinatorische Anzahlbestimmungen fur Gruppen, Graphen und Chemische Verbindungen. *Acta Math.* **1937**, *68*, 145.

(21) Lloyd, D. *Stud. Phys. Theor. Chem.* **1989**, *63*, 85−105.

(22) Redfield, J. H. The Theory of Group-Reduced Distribution. *Am. J. Math.* **1927**, *49*, 433−455.

(23) Contreras, M. L. et al. (work in progress).