Exhaustive Generation of Organic Isomers. 5. Unsaturated Optical and Geometrical Stereoisomers and a New CIP Subrule

M. L. Contreras,*,† G. M. Trevisiol,‡ J. Alvarez,‡ 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 July 24, 1998

This work, based on graph theory, presents a US-CAMGEC software developed for analysis, generation, and counting organic unsaturated stereoisomers with isolated or cumulated double bonds in structures that may contain chiral carbon and heteroatoms with variable valences within the molecule. A new extension to *N*_tuple notation was done for describing accordingly either *Z*, *E*, *r*, or *s* double bond and *R*, *S* stereoisomerism. Algorithms for transforming both constitutional isomer *N*_tuples to their corresponding graphs and graphs to the corresponding configurational *N*_tuples were developed. Computational implementation of Cahn−Ingold−Prelog (CIP) rules was developed including a new subrule for discriminating priorities to rank complex ligands already containing chiral stereocenters and having no different atomic number partial ranks. Test results for families of symmetric and nonsymmetric hydrocarbons and their monohalogenated, mono- and dioxygenated, and mono- and dinitrogenated derivatives are presented.

1. INTRODUCTION

Stereochemistry has a fundamental role in modern organic chemistry, particularly in the field of drugs and natural products. Computational organic chemistry has greatly stimulated the interest in developing software able to recognize and manage 2D and 3D molecular structures to assist in solving problems of synthesis,[1] structure elucidation,[2] database managing[3,4] similarity,[5−7] drug design,[8,9] and others. Active research has been done in the organic isomer enumeration[10−15] and generation[16,17] field particularly in computer aided stereoisomer generation.[18,19]

Numerous computer systems of very different global strategies have been reported for generating and counting constitutional isomer molecular structures.[20−31] Among them, CAMGEC,[22] generates and counts, in an exhaustive and irredundant way, all of the constitutional or topological isomers with different combinations of double bonds and cycles starting just from a given molecular formula. Many of these systems are related to structure elucidation and can integrate some spectrum interpretation capabilities with constrained structure generation;[28,31] other systems are oriented to molecular design[32−35] and combinatorial chemistry.[10]

Development of stereoisomer modules was considered as a natural further step following CAMGEC.[22] For instance, S-CAMGEC, a very specific system for stereoisomers of saturated acyclic chiral compounds was developed.[23] Also specific was a system that enumerates alkane isomers making a special reference to decane stereoisomers[36] and another for fluoroalkane stereoisomers.[37] More general programs for generation of stereoisomers were informed[38−40] and have also been a further contribution to pioneer works on stereoisomer generation systems.[41−43]

On the other hand, application of Cahn−Ingold−Prelog (CIP) rules to computer programs has shown to be a nontrivial task[44−46] sometimes appearing unresolved priority assignments which must be solved. Proposals for a revision of CIP rules have been reported.[44]

In particular, the known sequence rules for properly ranking ligands and in this way for assigning corresponding stereocenter configurations briefly consist of a comparison of ligands based on atomic numbers, followed by a comparison of atomic masses (not relevant in our case). Then, the procedure is applied recursively comparing total summation of atomic numbers by levels until a difference arises. If it does not, there is no topological difference among ligands. If it does, priority is assigned to the ligand with larger atomic number. In the case of cycles and multiple bonds, they must be decomposed in terms of single bonded groups, and then previous rules are applied. Strategies such as hierarchical digraphs[45] and canonical numbering methods[44] have been used. Proposals for ranking ligands around a center were done for ligands already containing stereogenic units,[45,47] for instance

- chiral > pseudoasymmetric > nonstereogenic
- cis > trans > nonstereogenic
- *alike* descriptors pairs (such as *RR* or *SS*) > *unlike* descriptor pairs (such as *RS* or *SR*)
- dv with higher description number > dv with lower description number for a set of interdependent stereogenic centers characterized by description vectors (dv).[44]

The following sections show our efforts on this subject focused to generation and counting of organic compounds which may contain heteroatoms, chiral carbon atoms, and several isolated or cumulated double bonds. Software resulting in this way was called US-CAMGEC (for **U**nsaturated **S**tereoisomers in **C**omputer **A**ssisted **M**olecular **GE**neration and **C**ounting).

* E-mail address: lcontrer@lauca.usach.cl.
† Faculty of Chemistry and Biology.
‡ Faculty of Engineering.

10.1021/ci9700586 CCC: $18.00
Published on Web 04/28/1999

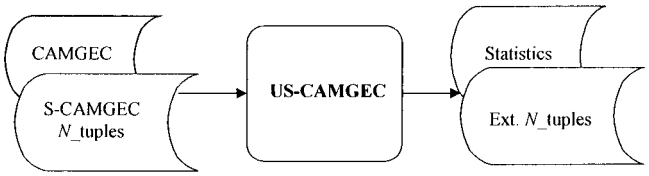


Figure 1. US-CAMGEC global data flow.

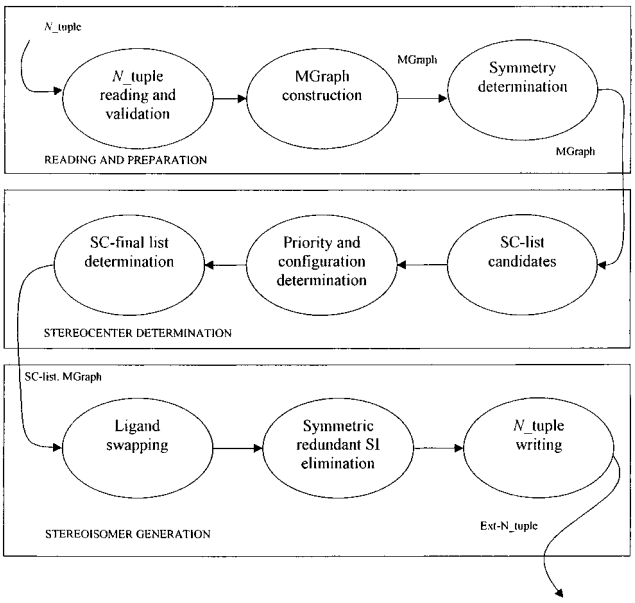


Figure 2. US-CAMGEC main processes.

2. SYSTEM DESCRIPTION

The main purpose of the US-CAMGEC system is to analyze organic molecules containing double bonds (db) to verify if one or more true db-stereocenters or constitutional stereogenic centers are occurring and, in that case, exhaustively and irredundantly generate correct stereoisomers. Two types of db-stereoisomers will be considered:

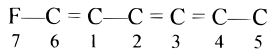
(a) Geometrical isomers, described normally with descriptors *Z* and *E*. The stereogenic unit in this case contains one or any odd number of adjacent (cumulated) db. The simplest case is the cis−trans isomerism of ethylene hydrocarbons.

(b) Optical isomers, described here with the descriptors *r* and *s* to differentiate them from chiral isomers known as *R* and *S*. The stereogenic unit in this case contains an even number of cumulated db. The most thoroughly studied cumulenes of this type are allenes.

Input and output of the US-CAMGEC system (see Figures 1 and 2) are described as follows:

Input. It could be a file with one or more constitutional isomers, each one represented in linear *N*__tuple notation. It also could include stereoisomer molecules with chiral centers. Specifically, it could be the output of a system generator of constitutional isomers like CAMGEC,[22] or the output of a stereoisomer generator system as S-CAMGEC,[23] or even it could be manually generated.

Output. It will be a file containing the corresponding extended *N*__tuples representing *Z*, *E*, *r*, or *s* stereoisomers generated over db stereocenters. The extension procedure of the original *N*__tuple notation for incorporating chiral atoms and db stereoisomer characteristics will be explained below through respective data structures. Global data flow



(a) 1c2r 2c1s 3c1d 4c1d 5c0s 6c1d 7f0s

(b) 1c2rE1z 2c1sE2r 3c1dE2 4c1dE2r 5c0s 6c1dE1z 7f0s

(c) 1c2rE1z 2c1sE2s 3c1dE2 4c1dE2s 5c0s 6c1dE1z 7f0s

(d) 1c2rE1e 2c1sE2r 3c1dE2 4c1dE2r 5c0s 6c1dE1e 7f0s

(e) 1c2rE1e 2c1sE2s 3c1dE2 4c1dE2s 5c0s 6c1dE1e 7f0s

Figure 3. Molecule I. Its graph and *N*__tuples: (a) canonical *N*__tuple; (b) extended *N*__tuple for the *Z*,*r*-stereoisomer; (c) extended *N*__tuple for the *Z*,*s*-stereoisomer; (d) extended *N*__tuple for the *E*,*r*-stereoisomer; and (e) extended *N*__tuple for the *E*,*s*-stereoisomer.

of US-CAMGEC in relation to input and output is given in Figure 1.

Key data structures of the systems are represented by

- *N*__tuple external representation of a molecule in the input text file.
- Internal graph representation of a molecule for processing at the computer memory (MGraph).
- Stereocenter list of molecular graph at the computer memory (SC-list).
- Extended *N*__tuple for external representation of the generated molecule into an output text file.

*N*__tuple external notation for constitutional isomers at the input is a canonical tree representation of a graph which depth-first search produces a linear sequence as explained before.[22]

Internal memory representation of molecular graphs was chosen as a graph representation; i.e., nodes and pointers linked together in a convenient way.

Main processes are shown on the data flow diagram in Figure 2 and are listed below:

- *N*__tuple input and validation.
- Graph construction.
- Symmetry determination.
- List of potential stereocenters or stereogenic centers (SC) construction.
- Potential SC__list analysis and SC determination.
- Stereoisomer generation, elimination of symmetric redundant stereoisomers, final extended *N*__tuple creation, and its output.

In Figure 3 graph and canonical *N*__tuple (a) and some extended *N*__tuples (b−e) for stereoisomers generated for a molecule with two SC are presented. Stereocenter 1, constituted by atoms 1 and 6, is designated as E1, while stereocenter 2, constituted by atoms 2, 3, and 4, is designated as E2. *N*__tuple notation used in this work is similar to one previously proposed.[23] For atoms that are terminal atoms, a character following the stereocenter specification (E1, E2, or other) indicates stereochemical attribute. For instance character z indicates *Z* configuration. Atoms constituting a nonterminal part of a stereocenter, as atom 3 in Figure 3, are not provided with that character. This notation facilitates interaction with graphical interfaces and with topological indices and other calculation programs.

Symmetry determination process involves construction of graph automorphism groups[48] joined to a structural analysis

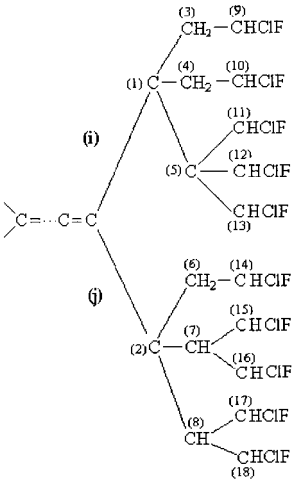


Figure 4. SC ligands with same number and type of configurational centers and different connectivities which cannot be ranked using pre-established rules.[44,45] Configuration *R* for carbon atoms (9) to (18) is considered.

of constitutional *N*_tuples and an algorithm to eliminate redundancy.[49]

One of the most complex and important subprocesses is potential SC_list analysis to get the real stereogenic centers list. Potential SC_list contains no terminal db, i.e., no H-ends, which are discarded right from the beginning. Analysis implies ranking of ligands of each potential stereocenter and then establishing its configuration assignment which will allow its transformation into a real generating SC. According to general organic chemistry knowledge and some recent specialized studies,[44,45] there is a number of CIP rules proposed for properly ranking ligands and in this way to assign corresponding stereocenter configuration descriptors. However, this is a problem not completely solved, and these rules do not allow for properly ranking SC containing ligands such as those described in Figure 4. Let us define one ligand in Figure 4 as ligand (***i***) and the other as ligand (***j***). Analysis of them by levels will be as follows:

	Ligand (***i***)	Ligand (***j***)
level 1:	C1	C2
level 2:	C3; C4; C5	C6; C7; C8
level 3:	C9, H, H; C10, H, H; C11, C12, C13	C14, H, H; C15, C16, H; C17, C18, H
level 4:	5 × HClF	5 × HClF

As it is shown both paths are equivalent from the point of view of atomic numbers, but they have different connectivity: ligand (***i***) presents at level 2 three carbon atoms with connectivities 1, 1, and 3; in this case each of those carbon atoms are bonded to 1, 1, and, respectively, 3 chiral carbon atoms. On the other hand, ligand (***j***) presents also three carbon atoms at level 2 but with connectivities 1, 2, and 2. If both ligands have the same number of identical configuration descriptors at level 3 (5 for each one, all of them *R*), there is no way of applying known rules[44−46] for properly ranking these ligands, and, in that case, a new criterion for elucidating ligand priorities is needed.

At this level then, and as a way of establishing priorities for ranking both ligands, a new concept for determining configuration values is proposed in this work. For that we have given a numerical value 1 to the *S* configuration and a

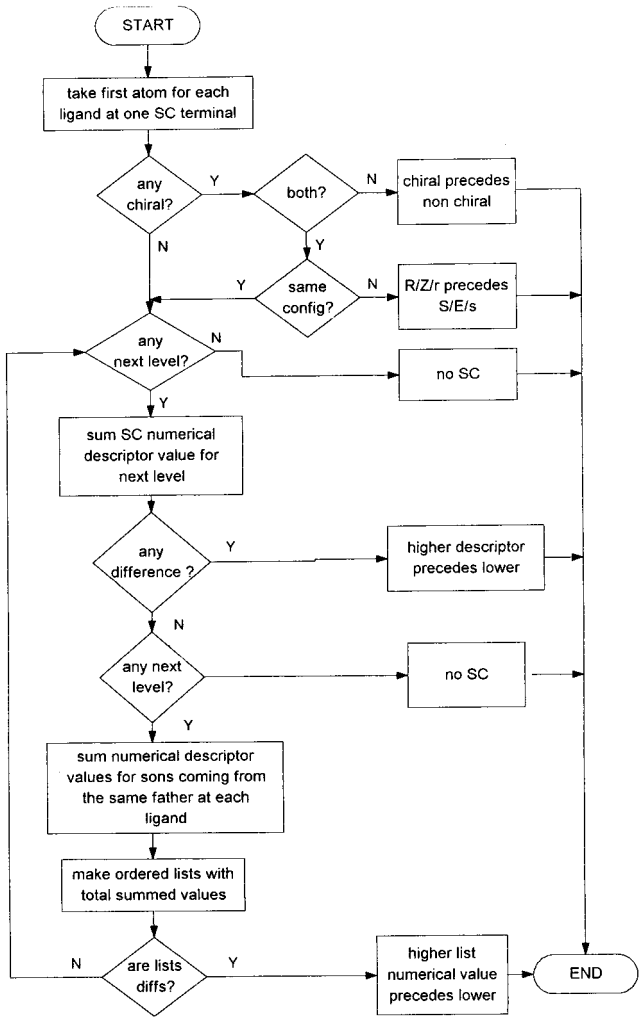


Figure 5. US_CAMGEC flow diagram procedure for ranking ligands. It can be applied to ligands with identical atomic numbers at each level and identical number and type of descriptors as those of Figure 4.

value 2 to the *R* configuration. Then, a subtotal value is calculated for each carbon atom of level 2. This subtotal value is defined as the sum of configuration values for all the chiral carbon bonded to that particular atom. In other words and as shown in Figure 5 the sum of numerical descriptors values for sons coming from the same father are considered.

path ***i***	path ***j***
C3: 2	C6: 2
C4: 2	C7: 2 + 2 = 4
C5: 2 + 2 + 2 = 6	C8: 2 + 2 = 4

As shown, for ligand (***i***) we have three subtotals: the first two with a value of 2 and the third one with a value 6. These subtotals are increasingly ordered and numerically considered. Finally they can form the sublevel chiral value of 226. For ligand (***j***) subtotals are 2, 4, and 4, and a final sublevel chiral value of 244 is obtained. So, ligand (***j***) ranks higher than ligand (***i***) because 244 > 226. The same is valid if the configuration of all chiral carbons occurring at both ligands are *S* type because in that case final counting is 122 > 113. On the other hand, if the five chiral carbon atoms of ligand ***i*** are all *R* and for the ligand ***j*** are all *S*, then our rule will rank ligand ***i*** higher than ligand ***j***, in complete accord with existing rules.

Table 1. Binary Matrix Used for Controlling Stereoisomers Generation Process for a Molecule Having Three SC

SC1	SC2	SC3	swapping indication
0	0	0	original molecule
0	0	1	ligand swapping of SC3
0	1	0	ligand swapping of SC2
0	1	1	ligand swapping of SC2 and SC3
1	0	0	ligand swapping of SC1
1	0	1	ligand swapping of SC1 and SC3
1	1	0	ligand swapping of SC1 and SC2
1	1	1	ligand swapping of SC1, SC2, and SC3

In that way, a new subrule was implemented for solving this situation, and in Figure 5 a general ranking procedure implemented here for ligands having the same number and type of SC but different connectivity is given.

The new subrule proposed in this work to be added to the previously reported rules should be described as follows:

*If two ligands contain identical number of stereocenters of the same configuration (same descriptors) and have different connectivities, a sublevel chiral value is calculated at the level of different connectivities. For that, to configuration descriptor R (also Z and r) a numerical value of 2 is assigned and to configuration descriptor S (also E and s) a numerical value 1 is assigned, and a sublevel value is calculated for each of the nodes of this level. These obtained values are arranged in increasing numerical order to get the chiral sublevel value of the ligands. Then the subrule: "**Higher sublevel chiral value precedes lower one"** is applied.*

Once stereogenic centers are validated, the stereoisomers generation process has to be accomplished. This consists basically of swapping a pair of ligands at one of the SC terminal places. Different molecules (stereoisomers) are generated when different combinations of SC swapping are carried out. This point involves two problems: (i) decision of which SC to swap and (ii) swapping of SC ligands. Fortunately ligand swapping is a single pointer swapping computer operation. For the first problem, it is considered that each of the *n* occurring SC can be found in one of two configurations. For the entire molecule, that could be visualized as a binary vector of length *n* with values 0 or 1 corresponding to the configurations *R* or *S*, *Z* or *E*, and *r* or *s*, as required for each of the actual *n* SC, where a value of 1 means "swap" and 0 means "do not swap".

Initial values for the binary vector are chosen for all the SC having configuration 0. The next value and the following ones are obtained by adding 1 to the immediate precedent vector. In that way, only two molecular graphs are kept in memory each time: the actual molecular graph and the next one to be generated. Table 1 shows a generation matrix for a molecule having three SC. Each row corresponds to a binary control vector state for the generation of one of the 2^n possible stereoisomers by swapping a pair of ligands at the SC specified with number 1 in the matrix. As explained before and as is shown in Figure 2 redundancy effect occurring for symmetric molecules is properly managed. US-CAMGEC is not addressed to aromatic compounds. It can work with molecules that contain cycles; however, it does not consider at the moment SC constituted only by atoms that belong to a cyclic structure.

Table 2. Acyclic Hydrocarbon Stereoisomers Generated by US-CAMGEC[a]

formula	no. C	db	tb	CI	no. SC	with SC	IEC	2EC	3EC	uncorr SI	real isomers
C_nH_{2n}	5	1		5	4	1	1			6	6
	6	1		13	9	4	4			17	17
	7	1		27	18	9	9			36	36
C_nH_{2n-2}	5	2		6	4	2	2			8	8
	6	2		16	10	6	5	1		24	23
	7	2		44	23	21	18	3		71	70
C_nH_{2n-4}	5	3		2	2					2	2
	6	3		10	6	4	4			14	14
	7	3		32	17	15	13	2		51	51
C_nH_{2n-4}	5	1	1	4	3	1	1			5	5
	6	1	1	12	8	4	4			16	16
	7	1	1	34	21	13	13			47	47
C_nH_{2n-6}	5	4		1	1					1	1
	6	4		3	3					3	3
	7	4		15	9	6	6			21	21
C_nH_{2n-6}	5	2	1	1	1					1	1
	6	2	1	7	5	2	2			9	9
	7	2	1	29	15	14	13	1		45	45

[a] Abbreviations: no. C: number of carbon atoms; db: number of double bonds; tb: number of triple bonds; CI: number of constitutional isomers generated by CAMGEC; no. SC: number of isomers without any stereocenter; with SC: number of isomers that contain one or more stereocenters; 1EC, 2EC, and 3EC: number of isomers that contain, one, two, or three stereocenters, respectively; uncorr SI: uncorrected stereoisomers including isomers without any stereocenter; and real isomers: number of irredundant total isomers.

Table 3. Acyclic Monohalogenated Stereoisomers Generated by US-CAMGEC[a]

formula	no. C	db	tb	CI	no. SC	with SC	IEC	2EC	3EC	uncorr SI	real isomers
$C_nH_{2n-1}X$	5	1		21	12	9	9			30	30
	6	1		56	30	26	26			82	82
	7	1		149	76	73	73			222	222
$C_nH_{2n-3}X$	5	2		20	9	11	10	1		33	33
	6	2		69	26	43	36	7		126	126
	7	2		228	75	153	122	31		443	442
$C_nH_{2n-5}X$	5	3		7	4	3	3			10	10
	6	3		37	14	23	20	3		66	66
	7	3		165	47	118	88	29	1	347	347
$C_nH_{2n-5}X$	5	1	1	14	7	7	7			21	21
	6	1	1	50	24	26	26			76	76
	7	1	1	166	77	89	89			255	255
$C_nH_{2n-7}X$	5	4		1	1					1	1
	6	4		9	5	4	4			13	13
	7	4		60	20	40	34	6		112	112
$C_nH_{2n-7}X$	5	2	1	3	2	1	1			4	4
	6	2	1	25	10	15	14	1		42	42
	7	2	1	134	42	92	77	15		256	256

[a] Same as in Table 2.

3. RESULTS AND DISCUSSION

Results for exhaustive and irredundant double bond stereoisomers generation process developed in this work are shown in Tables 2−8. Hydrocarbons and their monohalogenated, mono- and dioxygenated, and mono- and dinitrogenated derivatives were used for isomer generation. Input files coming from CAMGEC[22] were created for sets of acyclic constitutional isomers under different constraints over bond type. The unsaturated stereoisomer generation process was successfully checked by decodifying final extended *N*_tuples for studied compounds with less than 500 total stereoisomers. From analysis of results shown in Tables 2−8 the following points can be made.

Table 4. Acyclic Monooxygenated Stereoisomers Generated by US-CAMGEC[a]

formula	no. C	db	tb	CI	no. SC	with SC	IEC	2EC	3EC	uncorr SI	real isomers
$C_nH_{2n-6}O$	5	4		5	5					5	5
	6	4		34	24	10	10			44	44
	7	4		182	85	97	87	10		299	299
$C_nH_{2n-6}O$	5	2	1	11	9	2	2			13	13
	6	2	1	67	39	28	27	1		97	97
	7	2	1	322	148	174	156	18		532	532
$C_nH_{2n-4}O$	5	3		22	16	6	6			28	28
	6	3		97	50	47	43	4		152	152
	7	3		405	163	242	197	44	1	741	740
$C_nH_{2n-4}O$	5	1	1	29	20	9	9			38	38
	6	1	1	103	64	39	39			142	142
	7	1	1	342	99	143	143			485	485
$C_nH_{2n-2}O$	5	2		44	26	18	17	1		64	64
	6	2		151	76	75	66	9		244	243
	7	2		485	211	274	232	42		843	842
$C_nH_{2n}O$	5	1		41	28	13	13			54	54
	6	1		109	69	40	40			149	149
	7	1		294	176	118	118			412	412

[a] Same as in Table 2.

Table 5. Acyclic Dioxygenated Stereoisomers Generated by US-CAMGEC[a]

formula	no. C	db	tb	CI	no. SC	with SC	IEC	2EC	3EC	uncorr SI	real isomers
$C_nH_{2n-6}O_2$	5	4		30	24	6	6			36	36
	6	4		216	123	93	87	6		321	319
	7	4		1341	523	818	682	133	3	2443	2438
$C_nH_{2n-6}O_2$	5	2	1	52	39	13	13			65	65
	6	2	1	362	200	162	153	9		542	541
	7	2	1	1966	876	1090	963	127		3310	3309
$C_nH_{2n-4}O_2$	5	3		115	70	45	43	2		164	164
	6	3		614	273	341	296	44	1	1049	1046
	7	3		2840	1013	1827	1429	384	14	5519	5517
$C_nH_{2n-4}O_2$	5	1	1	122	84	38	38			160	160
	6	1	1	519	322	197	197			716	716
	7	1	1	2009	1165	844	844			2853	2853
$C_nH_{2n-2}O_2$	5	2		234	126	108	99	9		360	359
	6	2		907	421	486	419	67		1527	1519
	7	2		3328	1358	1970	1639	331		5960	5949
$C_nH_{2n}O_2$	5	1		204	133	71	71			275	275
	6	1		641	389	252	252			893	893
	7	1		1946	1132	814	814			2760	2760

[a] Same as in Table 2.

Table 6. Acyclic Mononitrogenated Stereoisomers Generated by US-CAMGEC[a]

formula	no. C	db	tb	CI	no. SC	with SC	IEC	2EC	3EC	uncorr SI	real isomers
$C_nH_{2n-5}N$	5	4		9	9					9	9
	6	4		60	44	16	16			76	76
	7	4		326	172	154	140	14		508	508
$C_nH_{2n-5}N$	5	2	1	25	20	5	5			30	30
	6	2	1	134	84	50	48	2		188	188
	7	2	1	587	308	279	254	25		916	916
$C_nH_{2n-3}N$	5	3		37	28	9	9			46	46
	6	3		165	96	69	64	5		244	244
	7	3		664	309	355	300	54	1	1133	1132
$C_nH_{2n-3}N$	5	1	1	50	38	12	12			62	62
	6	1	1	166	116	50	50			216	216
	7	1	1	531	348	183	183			714	714
$C_nH_{2n-1}N$	5	2		69	46	23	22	1		94	94
	6	2		228	131	97	88	9		343	342
	7	2		725	368	357	313	44		1170	1168
$C_nH_{2n+1}N$	5	1		56	42	14	14			70	70
	6	1		149	105	44	44			193	193
	7	1		398	265	133	133			531	531

[a] Same as in Table 2.

Table 7. Acyclic Dinitrogenated Stereoisomers Generated by US-CAMGEC[a]

formula	no. C	db	tb	CI	no. SC	with SC	IEC	2EC	3EC	uncorr SI	real isomers
$C_nH_{2n-4}N_2$	5	4		116	96	20	20			136	136
	6	4		782	502	280	265	15		1092	1089
	7	4		4467	2190	2277	1978	294	5	7362	7355
	5	2	1	255	191	64	63	1		321	321
	6	2	1	1362	871	491	467	24		1901	1900
	7	2	1	6324	3538	2786	2555	231		9572	9570
$C_nH_{2n-2}N_2$	5	3		369	255	114	110	4		491	491
	6	3		1783	996	787	712	74	1	2724	2720
	7	3		7730	3635	4095	3460	621	14	13151	13145
	5	1	1	341	272	69	69			410	410
	6	1	1	1298	965	333	333			1631	1631
	7	1	1	4697	3298	1399	1399			6096	6096
$C_nH_{2n}N_2$	5	2		561	374	187	178	9		766	765
	6	2		2071	1228	843	772	71		3056	3047
	7	2		7323	3909	3414	3041	373		11483	11470
$C_nH_{2n+2}N_2$	5	1		375	290	85	85			460	460
	6	1		1162	846	316	316			1478	1478
	7	1		3513	2449	1064	1064			4577	4577

[a] Same as in Table 2.

(1) The total number of stereoisomers (SI) increases with the number of carbon atoms due to the combinatorial nature of generation process. For instance in Table 5, the total number of isomers for molecules $C_nH_{2n-4}O_2$ with three db and having 5, 6, and 7 carbon atoms was 164, 1046, and 5517 isomers, respectively. Also, the total number of isomers increases with heteroatom features, in particular, with their valences and number in the molecule, as was expected due to a higher possibility of combinations. In that way for families of hydrocarbons, monohalogenated, monooxygenated, mononitrogenated, dioxygenated, and dinitrogenated derivatives having seven carbon atoms and three db (see Tables 2−7), total isomer numbers were equal to 51, 347, 740, 1132, 5517, and 13145 for each family, respectively.

(2) For a family of compounds with a constant number of carbon atoms, the total number of isomers shows a different tendency according to the number of db in the molecule: it increases up to a maximum and then decreases. For instance, for hydrocarbon molecules (Table 2) having seven carbon atoms with 1, 2, 3, and 4 db, the total number of isomers was 36, 70, 51, and 21, respectively. When comparing the number of isomers called "with SC" in Table 2, for the same kind of compounds the number of structures with SC are 9, 21, 15, and 6, respectively. The same tendency is observed for the other families studied.

(3) Generation of symmetric stereoisomers eliminates redundancy effects. For instance, stereoisomer generation process applied to 2,5-heptadiene will generate isomers *Z*,*Z*, *Z*,*E*, and *E*,*E* eliminating the *E*,*Z* isomer because it recognizes that *Z*,*E* and *E*,*Z* are the same molecule. Table 8 shows several cases where redundant isomers were eliminated. General cases affected by this redundancy effect are those symmetrical ones with more than one SC. It can be observed a greater number of redundant isomers for molecules containing two N atoms and two O atoms.

US-CAMGEC is able to exhaustively generate db stereoisomers for molecules that may contain heteroatoms, even with different valences (i.e. an S atom with valence 2 and

Table 8. Acyclic Stereoisomers Generated by US-CAMGEC: Symmetry Redundant Cases[a]

formula	no. C	db	tb	CI	sym CI	unsym CI	uncorr SI	red SI	real isomers
C_nH_{2n-2}	6	2		16	4	12	24	1	23
	7	2		44	8	36	71	1	70
$C_nH_{2n-3}X$	7	2		228	5	223	443	1	442
$C_nH_{2n-2}O$	6	2		151	7	144	244	1	243
	7	2		485	6	479	843	1	842
$C_nH_{2n-4}O$	7	3		405	5	400	741	1	740
$C_nH_{2n-2}O_2$	5	2		234	14	220	360	1	359
	6	2		907	29	878	1527	8	1519
	7	2		3328	54	3274	5960	11	5949
$C_nH_{2n-4}O_2$	6	3		614	19	595	1049	3	1046
	7	3		2840	14	2826	5519	2	5517
$C_nH_{2n-6}O_2$	6	4		216	9	207	321	2	319
	7	4		1341	28	1313	2443	5	2438
	6	2	1	362	6	356	542	1	541
	7	2	1	1966	7	1959	3310	1	3309
$C_nH_{2n-1}N$	6	2		228	8	220	343	1	342
	7	2		725	11	714	1170	2	1168
$C_nH_{2n-3}N$	7	3		664	6	658	1133	1	1132
$C_nH_{2n}N_2$	5	2		561	22	539	766	1	765
	6	2		2071	47	2024	3056	9	3047
	7	2		7323	86	7237	11483	13	11470
$C_nH_{2n-2}N_2$	6	3		1783	34	1749	2724	4	2720
	7	3		7730	36	7694	13151	6	13145
$C_nH_{2n-4}N_2$	6	4		782	18	764	1092	3	1089
	7	4		4467	53	4414	7362	7	7355
	6	2	1	1362	11	1351	1901	1	1900
	7	2	1	6324	16	6308	9572	2	9570

[a] Abbreviations: no. C: number of carbon atoms; db: number of double bonds; tb: number of triple bonds; CI: number of constitutional isomers generated by CAMGEC; sym CI: symmetric CI; unsym CI: unsymmetric CI; uncorr SI: uncorrected stereoisomers including isomers without any stereocenter; red SI: redundant symmetric stereoisomers; and real isomers: number of irredundant total isomers.

another one with valence 6 within the same molecule; in fact they are treated by the system as different atoms[20]), multiple bonds of different types, and cycles. Its results are completely consistent with those in the literature.[30,38] Perception of db stereogenic units in a wide range of cyclic molecular structures works very well. Priority determinations according to CIP rules and assignment of SC configurations are also set correctly.

In relation to cases where it is necessary to rank two ligands each one having a similar number of chiral carbon atoms with the same configuration but with a different level distribution, as in the example of Figure 4, US-CAMGEC offers a new way of solving this problem by assigning different numerical values to configuration descriptors *R* and *S* ($R = 2$; $S = 1$) and at the same time by defining the concept of a chiral vector. This is done by considering all the sons coming from the same parent node in the tree of the molecule (digraph; see system description and Figure 5). As explained before that problem could not be solved by applying known CIP rules which failed for this particular case where no ligand difference is found neither at the atomic number level nor for configuration descriptor pairs.

Even when application of numerical values assigned to configuration descriptors will directly help to rank ligands, care should be taken when applying values so known priority rules[44−47] are not violated. One of the visualized problems refers for instance to the case where one ligand has two atoms with an *S* configuration and another ligand has one atom with an *R* and one atom with an *S* configuration. Known rules will find that the first ligand with the two *S* configurations ranks higher than the other ligand. Our system will find the inverse situation. Its use is recommended only for ranking ligands that have identical configuration descriptors as the case shown in Figure 4. Treatment proposed in this work for *R* and *S* configuration descriptors may be extended to other kinds of configuration descriptors like *Z*-*E* or *r*-*s*.

Finally, the developed representation of extended *N*__tuples for db SI also containing chiral atoms could constitute an invaluable tool for 3-D molecule structure representation, its storage and retrieval, and also the calculation of topological indexes and other calculated variables, like similarity indexes, that depend on molecular structure. In addition, our system has proved to be a good tool specially for studying families of allene derivatives and their optical stereoisomers. All these points undoubtedly make US-CAMGEC a good contribution to molecular design.

4. IMPLEMENTATION ASPECTS

The program was developed on a SUN-IPX computer with SunOS 4.1.3 (28.5 Mips). Inputs to the program were files created under particular structural constraints with the help of CAMGEC[22] and S__CAMGEC.[23] Also, some manually created *N*__tuples, automatically validated by the program, were used as input in order to compare US-CAMGEC generation results with published data. The program is able to work with both canonical *N*__tuples like the ones coming from CAMGEC and also with extended *N*__tuples (noncanonical). Main outputs correspond to the extended *N*__tuple archives where one isomer is stored in one line of the archive. *N*__tuple storage is optional, and the output can be just the number of total isomers generated. Another important output corresponds to the statistical part of the generation results, where the name of the input file, date of processing, empirical formula, number of carbon atoms, number of double bonds, number of triple bonds, number of constitutional isomers generated by CAMGEC, number of isomers without any stereocenter, number of isomers that contain one or more stereocenters, total number of stereoisomers generated by the program, and number of isomers that contain respectively 1, 2, 3, or 4 stereocenters, are considered.

5. CONCLUSIONS

Useful computer software, US-CAMGEC, has been developed for generation, perception, and counting of unsaturated stereoisomers, focused on alkenes and cumulenes, and also including acyclic chiral carbons. It is based on graph theory and provided with the following capabilities:

- It makes perception of double bond stereocenters correctly identifying geometrical and optical stereogenic centers. It makes exhaustive generation of both kinds of stereoisomers.
- It ranks complex ligands that already contain some stereocenters on their structure.
- It contains computer implementation of CIP rules.
- It defines and implements a new subrule for ranking ligands not differentiated by their atomic numbers, containing a given number of stereocenters with identical configuration descriptors and different connectivity, which cannot be ranked by the existing CIP rules and their extensions.

•It develops an extended *N*_tuple notation which could be of great help for studying and calculating molecular properties and graph invariants for many structures.

•Its results are reliable. The program developed in C is easily portable to different hardware architectures.

This work will constitute an important base for educational purposes and for research, especially in organic chemistry, molecular design, structure elucidation, and also combinatorial chemistry. At the moment we are beginning to extend the program to include cyclic chiral carbons for properly treating dependency and interdependency among tetrahedral and planar/helicoidal SC.

ACKNOWLEDGMENTS

Financial support from University of Santiago de Chile is appreciated. Discussions with Mr. J. M. Dorado from University of Salamanca, Spain, and help in symmetry computational aspects by Eng. A. Toro are also appreciated.

REFERENCES AND NOTES

(1) Barberis, F. Barone, R.; Arbelot, M.; Baldy, A.; Chanon, M. CONAN (CONnectivity ANalysis): A Simple Approach in the Field of Computer-Aided Organic Synthesis. Example of the Taxane Framework. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 467−471.

(2) Balasubramanian, K.; Basak, S. C. Characterization of Isospectral Graphs Using Graph Invariants and Derived Orthogonal Parameters. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 367−373.

(3) Contreras, M. L.; Deliz, M.; Galaz, A.; Rozas, R.; Sepulveda, N. A Microcomputer-based system for chemical information and molecular structure search. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 105−108.

(4) Contreras, M. L.; Deliz, M.; Rozas, R. Personal microcomputer based system of chemical information with topological structure data elaboration. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 163−167.

(5) Sadowski, J.; Wagener, M.; Gasteiger, J. Assessing Similarity and Diversity of Combinatorial Libraries by Spatial Autocorrelation Functions and Neural Networks. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 2674−2677.

(6) Basak, S. C.; Grundwald, G. D. Molecular Similarity and Estimation of Molecular Properties. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 366−372.

(7) Skvortsova, M.; Baskin, I. I.; Stankevich, I. V.; Palyulin, V. A.; Zefirov, N. S. Molecular Similarity. 1. Analytical Description of the Set of Graph Similarity Measures. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 785−790.

(8) Contreras, M. L.; Gonzalez, F. D.; Rozas, R. MM2 Parameterization of the Azido Group with Electron Lone Pairs. Conformational Analysis of 3′-Azidothymidine (AZT) and Structure Activity Relationship for its Analogues as Potential Anti-HIV Drugs. *Bol. Soc. Chil. Quim.* **1995**, *40*(1), 33−40.

(9) Contreras, M. L.; Gonzalez, F. D.; Muñoz, V. C.; Rozas, R. Computer Assisted Molecular Design and Biophore Determination. Application to AIDS Drugs. *Bol. Soc. Chil. Quim.* **1995**, *40*(3), 279−291.

(10) Bone, R. G. A.; Villar, H. O. Exhaustive Enumeration of Molecular Substructures. *J. Comput. Chem.* **1997**, *18*(1), 86−107.

(11) 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.

(12) Brunvoll, J.; Cyvin, S. J.; Cyvin, B. N. Enumeration of Tree-like Octogonal Systems. *J. Math. Chem.* **1997**, *21*(2), 193−196.

(13) Caporossi, G.; Hansen, P. Enumeration of Polyhex Hydrocarbons to $h = 21$. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 610−619.

(14) John, P. E.; Mallion, R. B.; Gutman, I. An Algorithm for Counting Spanning Trees in Labeled Molecular Graphs Homeomorphic to Cata-Condensed Systems. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 108−112.

(15) 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.

(16) Wieland, T.; Kerber, A.; Laue, R. Principles of the Generation of Constitutional and Configurational Isomers. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 413−419.

(17) Molchanova, M.; Zefirov, N. S. Irredundant Generation of Isomeric Molecular Structures with Some Known Fragments. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 8−22.

(18) 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.

(19) Jaritz, R. The Mathematical Modeling of Space Models of Chemical Molecules. *Match* **1998**, *37*, 179−193.

(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) Contreras, M. L.; Rozas, R.; Valdivia, R.; Agüero, R. Exhaustive Generation of Organic Isomers. 4. Acyclic Stereoisomers with One or More Chiral Carbon Atoms. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 752−758.

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

(25) Trinajstic, N.; Jericeviv, Z.; Knop, J. V.; Muller, W. R.; Szymanski, K. Computer Generation of Isomeric Structures. *Pure Appl. Chem.* **1983**, *55*, 379−390.

(26) 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.

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

(28) 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.

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

(30) 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.

(31) Bangov, I. P. Computer-Assisted Structure Generation from a Gross Formula. 3. Alleviation of the Combinatorial Problem. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 277−289.

(32) 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.

(33) 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.

(34) 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.

(35) 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. Comp.-Aided Mol. Des.* **1995**, *9*, 13−32.

(36) 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.

(37) 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.

(38) 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.

(39) 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.

(40) Wieland, T. Enumeration, Generation, and Construction of Stereoisomers of High-Valence Stereocenters. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 220−225.

(41) 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.

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

(43) Abe, H.; Hayasaka, H.; Miyashita, Y.; Sasaki, S. Generation of stereoisomeric structures using topological information alone. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 216−219.

(44) Perdih, M.; Razinger, M. Stereochemistry and Sequence Rules. A proposal for Modification of Cahn−Ingold−Prelog System. *Tetrahedron: Asymmetry* **1994**, *5*(5), 835−861.
(45) Mata, P.; Lobo, A. M. Implementation of the Cahn−Ingold−Prelog System for Stereochemical Perception in the LHASA Program. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 491−504.
(46) Figueras, J. Computer Coding of Configuration. *J. Chem. lnf. Comput. Sci.* **1996**, *36*, 491−496.
(47) Prelog, V.; Helmchen, G. Basic Principles of the CIP-System and Proposals for a Revision. *Angew. Chem., Int. Ed. Engl.* **1982**, *21*, 567−583.
(48) Balasubramanian, K. Computer Perception of Molecular Symmetry. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 761−770.
(49) Contreras, M. L. et al. (to be published).

CI9700586