Computerized Stereochemistry: Coding and Naming Configurational Stereoisomers

Marko Razinger* and Marko Perdih

National Institute of Chemistry, Hajdrihova 19, 61115 Ljubljana, Slovenia

Received May 18, 1993®

Several methods for computer generation of binary codes representing all possible configurational isomers of a given structure have been developed in past years. The problem of finding descriptors for naming the generated stereoisomers is not yet completely solved. Established systems for obtaining stereoisomer names are not very well adapted to computer use. An attempt to find descriptors, easily obtainable from the computer-generated parity vectors and as similar to the well established Cahn–Ingold–Prelog descriptors as possibe, is described.

INTRODUCTION

Computerized structure elucidation systems generate a set of constitutional isomers compatible with the input physicochemical data. Some of them generate also possible stereoisomers.[1] In all the implementations, described in the literature, the generated stereoisomers were represented by binary parity codes. Structural interpretation of these codes, i.e. establishing which configuration in three-space is represented by a given parity code, stopped in all the described systems with the establishment of local chirality; it was never pursued to its logical end, which is not only determining the spatial distribution of ligands around a stereocenter but also giving the whole structure its stereochemical name. Such name is obtained by merging the information about the nature and number of constituents of a chemical structure (constitution) and the information about the spatial arrangement of these constituents (configuration) into a single expression. We are presenting here our efforts to find such stereochemical names for configurational isomers; we are not addressing at this stage the question of conformational stereoisomerism nor that of stereogenic units with coordination number greater than 4.

There is no generally accepted and unique scheme for computer-coding and subsequent naming of stereoisomers. Without doubt the Cahn–Ingold–Prelog (CIP) system is by far the most known, established, and accepted system for naming stereochemical features of molecules. The basis of the system was laid down in 1951 by Cahn and Ingold, who were joined early in the 30-year development of the system by Prelog, and the last author's revision was published in 1982 by Prelog and Helmchen.[2] Unfortunately, the CIP system is not very well adapted to computer use. Meyer was the first among the few who attempted to implement some of the CIP priority rules in a computer program; he recognized that, "*...the (CIP) rules were not conceived with regard to the potential computer logic...*", as he had put it.[3] Later suggestions for the revision of the CIP system, demonstrating some of its deficiencies and suggesting supplementary rules, were published by Custer in 1986,[4a] Dodziuk et al. in 1990,[4b] and Mata et al. in 1993.[4c]

Simultaneously to the development of the CIP system, several approaches and algorithms were developed to enable computer manipulation of molecular graphs. All these approaches had in common the goal to number canonically the nodes of molecular graphs, i.e. the atoms in the molecule. To obtain the canonical numbering, some sort of priority rules had to be defined. The well-known Morgan algorithm,[5] based on the extended connectivities of graph vertices, is important as being the first algorithm for canonical numbering of molecular graphs. Extended connectivity of each atom is an *ad hoc* graph invariant, obtainable by summing the connectivities (degrees) of all its neighbors. Years after its first use the graph-theoretical basis of extended connectivity, i.e. its equivalence to the number of walks in the graph, was established[6a] and later theoretically proven.[6b,c] The iterative nature of the calculation of extended connectivity broadens the description of each atom in a way to contain information about the whole graph. The recursive procedure serves the same purpose as the method for ordering groups bound to a stereocenter in CIP approach, in which priority order is induced on atoms, equidistant from the stereocenter, repeating the procedure in ever widening concentric circles.

Approximately at the same time, in the mid-1960s, Dubois began to develop the DARC system.[7] The DARC linear code for chemical structures has some characteristics of both CIP and canonical numbering approach. Its main features are the unique description of structure by limited, concentric environments,[8] a set of rules for canonical labeling, separate description of the basic molecular graph, of the nature of atoms and bonds, explicit description of cycles, and a separate stereochemical descriptor. Some of these features were incorporated later in the CIP system and in the extensions of the Morgan algorithm approach.

Wipke was the first to introduce the idea that canonical numbering can be used for stereodescription:[9] in the Stereochemical Extension of the Morgan Algorithm (SEMA) approach he used canonical numbers of the first neighbors of a stereocenter for the determination of priority-rank order of ligands. Despite the similarity of CIP naming rules and Wipke's parity description of stereocenters there is no simple correspondence between the two systems. CIP system was considered also by Wipke as inconvenient for computer application.

In the SEMA approach the nature of graph vertices (atoms) is considered after their connectivity, while in the Shelley–Munk algorithm for finding topological symmetry[10,11] the order is inverse. In this algorithm, the vertices of the molecular graph are first described with a selected set of graph invariants and then partitioned into automorphism orbits. If atomic number would be chosen as hierarchically the first graph invariant, the Shelley–Munk algorithm would be closer to the logic of the CIP system than the algorithms based on or similar to the Morgan approach.

® Abstract published in *Advance ACS Abstracts,* February 15, 1994.

0095-2338/94/1634-0290$04.50/0

Another important dissimilarity between the CIP and the canonical numbering systems is in their basically different approach to stereodescription. The CIP rules serve to recognize the priority of ligands around each stereogenic unit separately, thus perceiving the local chirality of one unit at a time. This has the adverse consequence that the priority ordering of ligands, which may be quite difficult in highly symmetrical structures, has to be repeated for each stereogenic unit in turn. Also *local order*, imposed on the structure by CIP rules, does not necessarily coincide with the *global order*, used when numbering the atoms and naming the molecule. Canonical numbering of graph vertices on the other hand has to be established only once for the whole graph. The global order thus obtained needs however additional rules when used for the interpretation of local order.

In this context, it is appropriate to cite Mislow whose deep insight into stereochemistry, apparent from his basic textbook[12] and from the great number of papers he wrote on the subject, is leaving a lasting impression on the field. The quote, commenting on the often seen confusion between stereogenicity and chirotopicity, is taken from one of his papers: "*...Molecular segments must be viewed from two separate and distinct aspects: their character as stereogenic units and their local symmetry. The first is dependent on bonding connectivity (constitution) and is rooted in graph and permutation group theory, whereas the second is independent of constitution and is rooted in the theory of symmetry groups. Although these two aspects are in principle distinct and serve different purposes, they happen to overlap in the case of the regular tetrahedral permutation center....*"[13] In the present paper we are describing our endeavor to find a system for coding and naming of configurational stereoisomers which would combine the advantages of both CIP and canonical numbering approaches. The advantage of the CIP system is the fact it enables, at least for small systems, a simple "manual" way of establishing stereochemical descriptors; the advantage of SEMA on the other hand is in the simple computer generation of stereochemical descriptors, regardless of the system size. Therefore, our principal intention was to develop a system that is practical as well for the manual generation of stereochemical descriptors (which SEMA is not) as it is for computer use (which CIP is not).

COMPUTER GENERATION OF STEREOISOMERS

During the past 25 years, a number of research groups developed program packages for the computer manipulation of chemical structure. We will not discuss here the many interesting and important applications of computers in the fields of computer assisted organic synthesis nor in the fields of chemical documentation and information retrieval. We will focus instead on the not so many applications in the field of computer assisted structure elucidation which deal with stereochemical problems.

Without attempting to be exhaustive we may mention the "stereochemical extensions" of some of the more known structural generators. The most known is certainly the DENDRAL project generator whose stereochemical module was implemented by Nourse et al.[14,15] The program module Stereo made possible exhaustive but not redundant generation of stereoisomers of a given constitution. The development of the concept of the so-called *configuration symmetry group* was certainly a breakthrough in the computational treatment of stereoisomerism. Stereochemical interpretation of generated codes in the sense of *R*/*S* and cis/trans designation was limited however to those stereocenters whose configuration depended on constitutional differences according to CIP rules (*local chirality*). Sasaki et al. reported the development of stereochemical modules of their system CHEMICS;[16] they used SEMA notation, developed by Wipke[9] for local chirality designation. We reported the development of a stereochemical program module to be implemented into the SESAMI system.[17] The module enables the exhaustive nonredundant computer generation of stereoisomers of a very broad range of structural classes with emphasis on the so-called *pseudo-asymmetric* cases. Configuration of ligands around a stereocenter is deduced from the parity label of the stereocenter and respective number labels of the ligands defined by the canonical numbering.

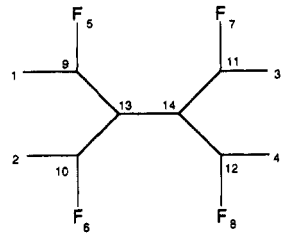
In all of the above mentioned attempts to computerize some aspects of stereochemistry, the computer coding of stereoisomerism was made in a similar manner, i.e. by using the parity vector. This vector consists of a string of binary characters, associating a *binary label to each and every stereogenic unit* in the structure. Such labels, be it 1 and 0 (or +1 and −1, or + and −), serve for the identification and naming of stereoisomers only; since they are of topological origin, they do not indicate the (topographical) symmetry relationships and thus should not be considered as chiral descriptors.[13]

PARITY VECTORS: BINARY STEREOCHEMICAL DESCRIPTORS

The so-called "topological symmetry" of a graph (a molecular structure) can be described and identified by vertex (atom) permutations.[13,18] A structure possessing N stereogenic units, each of which may exist in two different configurations, can have 2^N stereoisomers at most. Checking the permutations we can find which among the stereoisomers are symmetry (i.e., permutationally) equivalent, thus representing really the same isomer. As the stereogenic units may exist in two configurations only, binary descriptors should be sufficient to describe stereoisomers properly.

The structure in Figure 1 has 14 nonhydrogen atoms: 6 of them are potential stereocenters. From topological data (connection table) one can find out that 4 of these 6 atoms (atoms 9–12) are bound to 4 constitutionally different neighbors and are thus chiral stereocenters; 2 of the 6 atoms in question (atoms 13 and 14) have each a pair of constitutionally identical ligands which however contain a stereocenter. Thus atoms 13 and 14 can either be chiral or achiral, depending on the configuration of constitutionally equivalent ligands. The details of the computational procedure used to obtain all possible stereoisomers of the input structure are described elsewhere;[17] Figure 1 shows an example of how topological symmetry and generated stereoisomers can be expressed in mathematical terms, i.e. in the form of vertex permutations and parity vectors. The group of automorphism reduces the number of stereoisomers from the theoretically possible 64 to the 10 symmetry non-equivalent (3 pairs of enantiomers and 4 meso forms).

Parity vectors can be interpreted either by graphs or by descriptors. In both cases the priority order of all ligands of each stereogenic unit must be established. A pragmatic way to do this is to take the number labels of individual atoms, determined by canonical numbering, for priority ordering. The space orientation of ligands around the stereogenic unit may be determined in a way parallel to the CIP system. We can arbitrarily interpret a +1 label in parity vector as meaning the same configuration as *R* does in the CIP system (and −1 the same as *S*, of course). Also the diastereoisomerism in



GROUP OF AUTOMORPHISMS

```
p1 : (1)(2)(3)(4)(5)(6)(7)(8)(9)(10)(11)(12)(13)(14)
p2 : (1  2)(3)(4)(5  6)(7)(8)(9  10)(11)(12)(13)(14)
p3 : (1  2)(3  4)(5  6)(7  8)(9  10)(11  12)(13)(14)
p4 : (1)(2)(3  4)(5)(6)(7  8)(9)(10)(11  12)(13)(14)
p5 : (1  3)(2  4)(5  7)(6  8)(9  11)(10  12)(13  14)
p6 : (1  4  2  3)(5  8  6  7)(9  12  10  11)(13  14)
p7 : (1  4)(2  3)(5  8)(6  7)(9  12)(10  11)(13  14)
p8 : (1  3  2  4)(5  7  6  8)(9  11  10  12)(13  14)
```

PARITY VECTORS

	9	10	11	12	13	14
e1	1	1	1	1	1	1
	-1	-1	-1	-1	-1	-1
e2	1	1	1	-1	1	1
	-1	-1	-1	1	-1	-1
e3	1	1	1	-1	1	-1
	-1	-1	-1	1	-1	1
m1	1	1	-1	-1	1	1
m2	1	-1	1	-1	1	1
m3	1	-1	1	-1	1	-1
m4	1	-1	1	-1	-1	-1

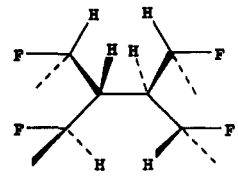
Figure 1. Computational generation of the stereoisomer codes: graph representing the molecular structure, its automorphism group with eight elements, and the 10 symmetry nonequivalent parity vectors representing 10 nonredundant stereoisomers out of the theoretically possible 64.

substituted double-bond systems can be interpreted in this way, taking the +1 label as *Z* and −1 as *E*.

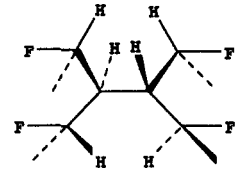
The graphical interpretation of some of the parity vectors in Figure 1 are shown in Figure 2, where we tried to use wedged lines to make visible the space orientation. The use of wedged lines is familiar to chemists to interpret graphically a three-planes model; in Figure 2 this use is somewhat stretched beyond its abilities since there are four planes in the figure.

Such simple interpretation of parity vectors in the frame of the CIP system is possible only for chiral stereogenic units with four constitutionally different ligands (chiral stereocenters, called also "asymmetric atoms", are ligand-constitution dependent stereocenters; the examples of such centers are shown in Figure 3 and also in atoms 9–12 in Figure 1). In the case of stereogenic units whose chirality depends on the configuration of their ligands (the so called "pseudo-asymmetric atoms" are ligand-configuration dependent stereocenters, atoms 13 and 14 in Figure 1) such interpretation is not possible in all cases because the CIP system and the generation of parity vectors are not always compatible.

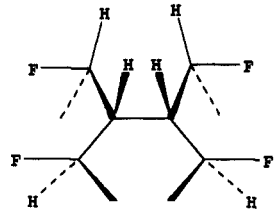
The source of their incompatibility lies in the fact that parity vectors are *binary* for every stereoisomer, while CIP descriptors are *binary* for some configurations and *quaternary* for some



v = (1,1,1,1,1,1)

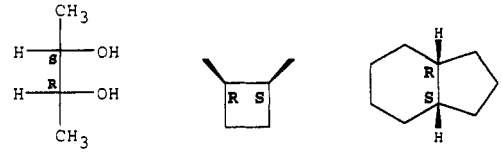


v = (-1,-1,-1,-1,-1,-1)



v = (1,-1,1,-1,1,-1)

Figure 2. Graphical interpretation of some of the parity vectors in Figure 1, a pair of enantiomers and a meso form.



Mapping: +1 ----> R ; -1 ----> S

Figure 3. Example of three structures for which straightforward mapping of parity labels onto CIP descriptors poses no problems.

others. Whenever a stereocenter lies in a symmetry plane (called also a σ-plane or mirror plane), the usual upper-case CIP descriptors **R** and **S** are no longer sufficient for the description of the configuration. In such cases, lower-case descriptors **r** and **s** are needed too; they are sometimes referred to as "chiral descriptors", although such labels bear no relation to the local sense of chirality.[13]

The assignment of CIP descriptors of stereocenters in one of the four possible stereoisomers of 1,2,3,4-tetramethylcyclobutane is shown in Figure 4. The rather complicated procedure[2] has to be done for every stereocenter in turn: first the drawing of the hierarchical digraph and the assorted graph, then assignment of temporary descriptors R_0 and S_0, then the comparison and description of temporary descriptor pairs with labels ***l*** and ***u*** (standing for *like* and *unlike*), and then finally the assignment of the descriptor of the central stereocenter. Whenever the graph has a plane of symmetry passing through the central atom (i.e., through the stereocenter whose descriptor one is just assigning), this atom gets a lower-case descriptor. The example in Figure 5 shows that the simultaneous use of both lower- and upper-case descriptors, i.e. four different descriptors, is necessary in the CIP system indeed. In this case, without the use of quaternary descriptors, using only lower- or only upper-case descriptors, we would

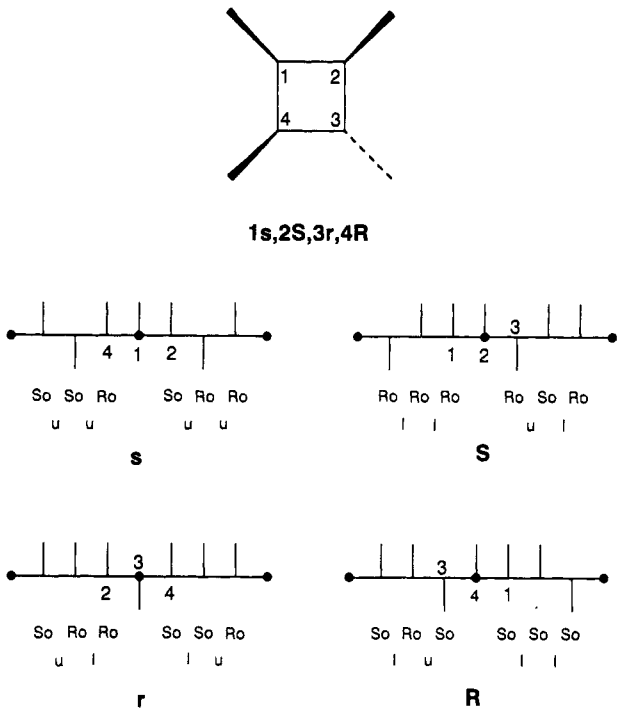


Figure 4. Assignment of CIP descriptors of stereocenters for one of the four possible stereoisomers of 1,2,3,4-tetramethylcyclobutane.

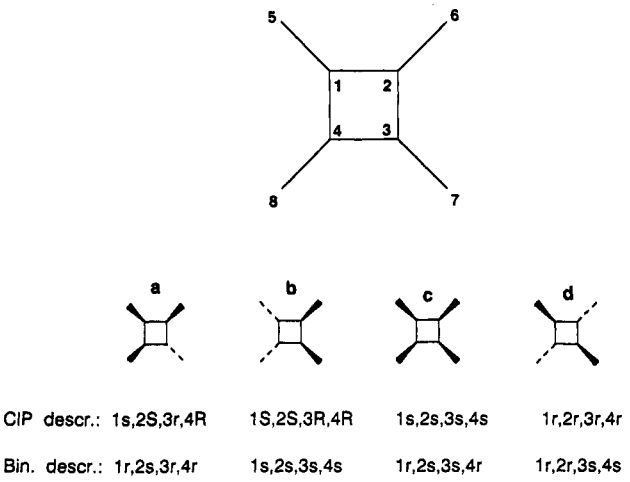


Figure 5. CIP descriptors and descriptors obtained by canonical labeling of all four configurations of 1,2,3,4-tetramethylcyclobutane; the CIP system needs quaternary descriptors while in the canonical labeling system the four stereoisomers are described with binary descriptors.

assign the same four-CIP-descriptor set to two obviously different configurations in Figure 5a,b.

CANONICAL NUMBERING: NAME FROM PARITY CODE

Using canonical numbering as the basic ligand-priority criterion, determined simultaneously and only once for all the stereocenters in the structure, we tried to simplify the complicated procedure of descriptor assignment demonstrated in Figure 4 above. We used the Shelley–Munk algorithm[10,19] earlier to obtain a canonically numbered graph, used further to generate the parity vectors, representing an exhaustive and nonredundant list of stereoisomers.[17] One of the first steps during the determination of the "topological symmetry" of the graph (i.e. of its automorphism group) is just the canonical numbering of graph vertices. Since thus we had an efficient algorithm at hand, we decided to use it also for priority order determination.

We modified however the algorithm to make the priority order it induces as parallel to the CIP order as possible. We took the atomic number as hierarchically the first graph invariant entering the algorithm to mimic the first CIP rule. This parameter takes care of material differences depending on individual graph vertices, i.e. on the nature of atoms represented by the vertices.

Taking ring membership as the second graph invariant, we modified the original Shelley–Munk algorithm and also included explicitly the criterion which is implied in the CIP procedure of breaking rings when drawing hierarchical digraphs.[2] The second parameter, although describing individual vertices, depends also on those graph vertices which are members of the rings. Knowing the smallest set of smallest rings (SSSR)[20] we are able to construct a "ring membership vector" for every vertex. The components of this vector correspond to ring sizes, sorted in the ascending order; the value of each component is equal to the number of rings of the corresponding size the vertex is a member of. In some special cases the above procedure does not differentiate topologically non-equivalent vertices, members of several cycles. However, it is possible to distinguish such bridgehead vertices by comparing the ordered sets of the equivalence-class membership of constituent vertices. This procedure is similar to that used in the HOC-2A algorithm of Balaban et al.[21]

The main advantage of the canonical numbering approach is that original, unchanged molecular graphs can be used for ligand priority determination, in contrast to the transformation of graphs to digraphs (including the ring breaking) as it is done in the CIP approach. The same numbering is valid for the determination of ligand priority of all the stereogenic units in contrast to the CIP method of drawing separate digraphs for every stereocenter in some cyclic cases. What is lost in discarding digraphs is gained with the introduction of topological description implying also structural hierarchy. Thus modified, canonical numbering system makes possible detection of all constitutional differences.

The use of canonical numbering for priority ordering of ligands and for generating by this method the binary descriptors (of all four possible stereoisomers of the structure in Figure 4) is demonstrated in Figure 5. Using canonical labels to establish priority (lower number precedes higher, the unlabeled hydrogen atoms get label $N+1$ by default, N being the number of graph vertices) has two distinct advantages compared with CIP descriptors also shown in the same figure: first, the naming is extremely simple compared to the CIP system naming, shown in Figure 4; second, the resulting descriptors are binary, thus making possible the straightforward mapping of parity labels to descriptors.

CONFIGURATION-DEPENDENT STEREOGENIC UNITS

Dealing with configurational differences calls for a novel approach in computational classification and treatment of stereogenic units. The atoms, traditionally referred to as "pseudo-asymmetric", are separated into two types (see also ref 17). The configuration of the first type of atoms *depends* on the configuration of other stereocenters, present in the structure, as in the case of 2,3,4,5,6-pentahydroxyheptane in Figure 6a; for the second type of atoms, this dependence is mutual, i.e. the configurations of stereocenters of this type are

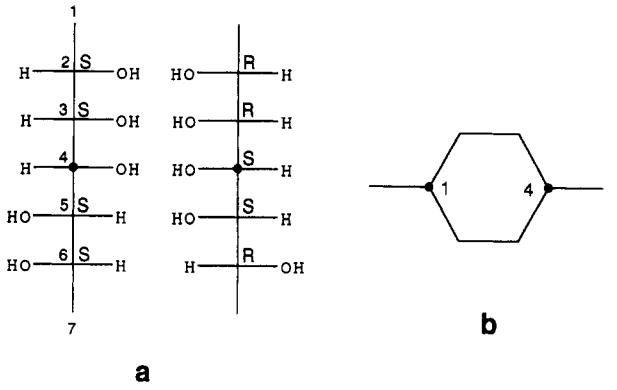


Figure 6. Examples of ligand configuration-*dependent* (a) and ligand configuration-*interdependent* (b) stereocenters (dot-marked atoms).

interdependent, and the example of such structure is 1,4-dimethylcyclohexane in Figure 6b.

The difference between *configuration-dependent* and *configuration-interdependent* classes of stereogenic atoms becomes clear when one compares dot-marked atoms in Figure 6. In acyclic structure, the stereochemical behavior of the central C-atom 4 is clearly depending on the configuration on atoms 2, 3, 5, and 6, while the inverse is not true. In the cyclic structure, two neighbors of each stereocenter are constitutionally *and* configurationally equivalent. To solve this stalemate, we choose arbitrarily one of the neighbors of atom 1 (the one with lower canonical number) as having priority over the other neighbor. This choice determines the configuration of atom 1 but at the same time also the configuration of the other stereocenter because its neighbors are no longer equivalent. Because of the interdependence of the priority of the stereocenters' neighbors, we state that both stereocenters are configurationally interdependent. The arbitrariness of the priority choice is in a way analogous to the arbitrariness of the ring-cutting step in the CIP system, where the neighborhood of selected atoms (end atoms in hierarchical digraphs) is effectively altered.

The practical consequence of the interdependence demonstrated above is a significant difference in the interpretation of parity vectors, illustrated in Figure 7. Stereocenters, marked in the figure with dots, belong to three distinct structural units: one unit with the two stereocenters in the first cycle, another unit with two stereocenters in the second cycle, and a third unit with an acyclic stereocenter. The cyclic ligands are 1,4-dimethylcyclohexanes with interdependent stereocenters; the configuration of the acyclic stereocenter can be determined only if we establish a priority relationship between the two constitutionally identical cyclic units. As each of these units has two stereocenters, we must compare simultaneously the two centers in one unit with the two centers in the other. For doing this we describe the stereochemical properties of each unit with a *description vector* which has as many components as there are stereocenters in the unit (two components in the case of the structure in Figure 7, the units being the cyclic parts with atoms 2–4 and 3–5, respectively). The parity codes are transformed into binary numbers, +1 corresponding to 1 and −1 to 0. Topologically equivalent atoms must encode into corresponding components: thus in Figure 7, atoms 2 and 3 encode into the first component of their respective description vectors, and atoms 4 and 5, into the second component. Reading binary description vectors as decimal numbers gives us the *description number* of the corresponding unit (of the group of interdependent stereocenters). A higher description number has priority over the lower; in Figure 7, the unit 2–4 has priority over unit 3–5 in the case of enantiomers (in the third and fourth row of description vectors). The configuration of the central, acyclic stereocenter in its own structural unit is thus established.

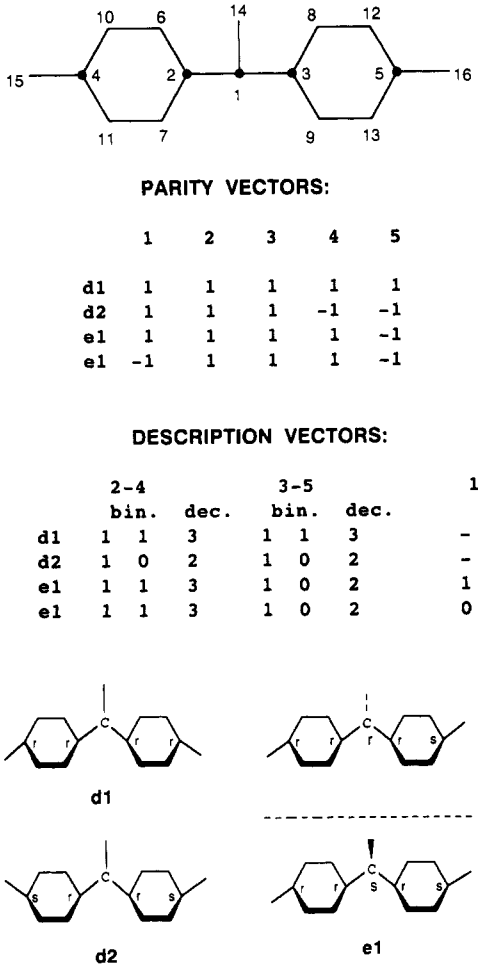


Figure 7. Structure shown has two structural units with interdependent stereocenters (two para-substituted cyclohexane rings). Parity labels are transformed into description vectors, which are sets of binary numbers. Transformation of description vectors to decimal numbers results in description numbers, used for priority determination of groups of interdependent stereocenters.

The concept of description vectors, developed for priority ordering of structural units containing several interdependent stereocenters, may be generalized also to encompass the "asymmetric", constitution-dependent stereocenters. Their description vectors are just a special case having one component only, being degenerated into scalars; however, the priority ordering criteria remain the same.

We may describe the orientation of ligands around configuration-dependent stereocenters with the lower-case descriptors **r** (or **z**) and **s** (or **e**) for **+1** and **0** in description vectors, respectively, bearing in mind that those descriptors will not be equivalent to the CIP **r** and **s** descriptors. However, we may decide to use upper-case letters **R** and **S** for **+1** and **0**, respectively, for the degenerated description vectors characterizing the constitution-dependent stereocenters; these *R* and *S* descriptors will be equivalent to CIP descriptors.

The detailed and systematic description of the modifications we propose to the CIP system to make it computer-compatible, with the rules, algorithms, and their implementation, will be described in a separate paper.

CONCLUSION

Computer manipulation and processing of chemical structure-related data are increasing constantly. In some areas, like documentation and literature searching, they are indispensable and it would be difficult even to imagine dealing with today's deluge of information without the help of computers. In some others fields, like computer assisted structure elucidation or chemical synthesis for example, the computers may be a great help but should not yet be expected to solve the problems alone; because of their complexity, the problems in these fields are less well defined and are consequently much more difficult to solve algorithmically. Nevertheless, the effects to develop programs helping chemists in these areas continue, and the results are promising.

One of the areas of chemistry where computers' use is not yet fully developed is also stereochemistry. Many efforts, some of them cited earlier in this article, have been made in stereoisomer counting and generation, in CIP rules' implementation, and in coding stereochemical information. In spite of this, the implementation of the CIP system has not gone beyond rules 1 and 2 (constitutional differences), and the treatment of stereocenters in commercial software is mainly limited to establishing that two configurations are possible around a stereogenic unit (local chirality).

In our work on stereoisomer generation[17] we concentrated on highly symmetrical structures containing pseudoasymmetric stereogenic units; in these cases the detection of stereogenicity and its proper analysis are much more difficult than in the case of an ordinary and most frequently met asymmetrically substituted chiral center or double-bond system. We came to the conclusion that in such systems, where the stereogenicity of the central unit depends only on the configuration, and not on the constitution, of its ligands, further differentiation is imperative. As we explained in the former work[17] and mentioned shortly above in this paper, we defined and treated separately *configuration-dependent* stereogenic units and groups of *configuration-interdependent* units. This differentiation made possible the establishment of a priority between ligands, a prerequisite for assignment of stereochemical descriptors to individual configurations. Descriptors compatible with the proposed stereocenter classification have been introduced, thus making possible automatic stereoisomer naming from the computer generated parity codes. The treatment of collective stereogenicity of structural units containing two or more interdependent stereocenters is made possible by the introduction of a new type of description vector. It is used for the priority determination of constitutionally equivalent ligands. The components of this vector are parity descriptors of configurationally interdependent stereocenters. A stereocenter having four non-equivalent ligands represents the degenerate case of the vector with one component only.

The CIP system was revised in 1982 after its incorporation into the IUPAC Nomenclature System. Several revisions have been proposed during the past decade; all of them concerned the assignment of descriptors in special, difficult, or ambiguous systems with multiple cycles, or high symmetry, or pseudo-asymmetric atoms. Rules 1 and 2, defining the role of constitutional differences in priority assignment, remain the rock-solid basis of the system, familiar to and used by every organic chemist. The revision and modifications of later CIP rules, although necessary, are not yet generally accepted nor officially recognized. Therefore our proposal for drastic changes, which might be at first considered to be too extreme, will not change at all the descriptors of the type *R*, *S* and/or *Z*, *E*, attributed to stereogenic units in the most frequently met structures where the stereogenicity depends on the constitution alone.

In contrast to the CIP descriptors, our descriptors are binary. They specify only the configuration of neighbors around a stereocenter while CIP descriptors contain also geometrical information related to molecular symmetry. We decided however to make these binary descriptors quarternary by using upper-case letters to designate the configuration of constitution-dependent stereogenic units and lower-case letters for configuration description of the configuration-dependent stereogenic units. Thus, the upper-case descriptors of our system are equivalent to CIP descriptors, while the lower-case descriptors are different in both systems. In this manner, the upper or lower case of the descriptor clearly indicates the nature of the stereogenic unit. This is also simpler and perhaps more consistent than is the case in the CIP system where pseudo-asymmetric centers can have upper- or lower-case descriptors, depending on the configuration of ligands.

To conclude, we want to stress again that the proposed method of stereochemical-descriptor generation is adapted to both manual and computer implementation. The advantage of the generated descriptors is in the fact that they can be used for stereoisomer naming (in human communication) as well as for computer representation and communication. It is true that the proposed system is dependent of the molecular graph numbering, while CIP descriptors are not; but it is also true that for obtaining CIP descriptors, it is necessary to induce the numbering (called priority order) around each stereogenic center in turn. We succeeded, with the help of a few simple rules, to induce a numbering coinciding with CIP ordering, having to number the molecular graph only once. Additionally, the numbering obtained is not arbitrary but depends on the basic properties of the chemical structure—the nature of the constituent atoms and of the molecular graph. It should be emphasized that the proposed assignment of ligand priority via canonical numbering labels is not limited to the use of the modified Shelley–Munk algorithm: any canonical numbering algorithm will do as long as the hierarchically uppermost criteria for obtaining automorphism partitioning are atomic number and ring membership of atoms. When the starting numbering of the graph (structure) is the conventional IUPAC numbering, the resultant stereochemical descriptors can be mapped again back to the original numbering regardless of which numbering (based on graph invariants!) is used between for the descriptor assignment.

ACKNOWLEDGMENT

We acknowledge the financial support of the Ministry of Science and Technology of Slovenia through Research Contract No. P1-0120-104. M.R. thanks Prof. M. E. Munk, Department of Chemistry and Biochemistry, Arizona State University, for many fruitful discussions about the classification of stereogenic units. We also thank Prof. A. P. Johnson, School of Chemistry, University of Leeds, for sending us the draft copy of his manuscript (ref 4c) prior to publication.

REFERENCES AND NOTES

(1) Gray, N. A. B. *Computer Assisted Structure Elucidation*; Wiley: New York, 1986, Chapters 9 and 10.
(2) Prelog, V.; Helmchen, G. Basic Principles of the CIP-System and Proposals for a Revision. *Angew. Chem., Int. Ed. Engl.* **1982**, *21*, 567–583.
(3) Meyer, E. F. The Computational Application of the Cahn-Ingold-Prelog Rules 1 and 2. *J. Comput. Chem.* **1980**, *1*, 229–232.
(4) (a) Custer, R. H. Mathematical Statements about the Revised CIP-System, *MATCH* **1986**, *21*, 3–31. (b) Dodziuk, H.; Mirowicz, M. A

Proposal for a Modification of the Cahn, Ingold, and Prelog Classification of Chirality. *Tetrahedron: Asymmetry* **1990**, *1*, 171–186. (c) Mata, P.; Lobo, A. M.; Marshall, C.; Johnson, A. P. The CIP Sequence Rules: Analysis and Proposal for a Revision. *Tetrahedron: Asymmetry* **1993**, *4*, 657–668.

(5) Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at CAS. *J. Chem. Doc.* **1965**, *5*, 107–113.

(6) (a) Razinger, M. Extended Connectivity in Chemical Graphs. *Theor. Chim. Acta (Berlin)* **1982**, *61*, 581–586. (b) Ruecker, G.; Ruecker, C. Count of All Walks as Atomic and Molecular Descriptors. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 683–695. (c) Figueras, J. Morgan Revisited. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 717–718.

(7) Dubois, J. E. Ordered Chromatic Graph and Limited Environment Concept. In *The Chemical Applications of Graph Theory*; Balaban, A. T., Ed.; Academic Press: London, 1976; pp 330–370.

(8) Dubois J. E.; Viellard, H. Systeme DARC. Theorie de generation-description. *Bull. Soc. Chim.* **1968**, 900–919.

(9) Wipke, W. T.; Dyott, T. M. Stereochemically Unique Naming Algorithm. *J. Am. Chem. Soc.* **1974**, *96*, 4834–4842.

(10) Shelley, C. A.; Munk, M. E. Computer Perception of Topological Symmetry. *J. Chem. Inf. Comput. Sci.* **1977**, *17*, 110–113.

(11) Razinger, M.; Balasubramanian, K.; Munk, M. E. Graph Automorphism Perception Algorithms in Computer-Enhanced Structure Elucidation. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 197–201.

(12) Mislow, K. *Introduction to Stereochemistry*; Benjamin: New York, 1965.

(13) Mislow, K.; Siegel, J. Stereoisomerism and Local Chirality. *J. Am. Chem. Soc.* **1984**, *106*, 3319–3328.

(14) Nourse, J. G. The Configuration Symmetry Group and Its Application to Stereoisomer Generation, Specification, and Enumeration. *J. Am. Chem. Soc.* **1979**, *101*, 1210–1216.

(15) Nourse, J. G.; Smith, D. H.; Carhart, R. E.; Djerassi, C. Computer-Assisted Elucidation of Molecular Structure with Stereochemistry. *J. Am. Chem. Soc.* **1980**, *102*, 6289–6295.

(16) 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.

(17) 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.

(18) Nourse, J. G. *The Permutation Group in Physics and Chemistry*; Lecture Notes in Chemistry; Springer: Berlin, **1979**; pp 19–32.

(19) Shelley, C. A.; Munk, M. E. An Approach to the Assignment of Canonical Connection Tables and Topological Symmetry Perception. *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 247–250.

(20) Gasteiger, J.; Jochum, C. An Algorithm for the Perception of Synthetically Important Rings. *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 43–48.

(21) Balaban, A. T.; Mekenyan, O.; Bonchev, D. Unique Description of Chemical Structures Based on Hierarchically Ordered Extended Connectivities (HOC Procedures). I. Algorithms for Finding Graph Orbits and Canonical Numbering of Atoms. *J. Comput. Chem.* **1985**, *6*, 538–551.