Pergamon

Tetrahedron: Asymmetry Vol. 5, No. 5, pp. 835-861, 1994

0957-4166/94 $7.00+0.00

0957-4166(94)E0114-P

Stereochemistry and Sequence Rules
A Proposal for Modification of Cahn-Ingold-Prelog System

Marko Perdih* and Marko Razinger
National Institute of Chemistry, Hajdrihova 19, 61115 Ljubljana, Slovenia

Abstract: A modification of the CIP system is proposed to combine the advantages of the CIP system and the methods for computerised treatment of stereogenicity. The proposal is based strictly on the topology. The basis for ranking of ligands in the computer application is a method for the canonical numbering of the molecular graphs, therefore no transformations to the hierarchical digraphs are necessary. For the manual application of the proposed system, a transformation of rings and multiple bonds to chains which is equivalent to the canonical numbering, can be used. New terms *constitutional stereogenic centre* and *configurational stereogenic centre* are introduced to specify whether a stereogenic centre has the ligands with identical constitution or not. The term *configurational stereogenic centre* is a topological alternative to the geometrical term *pseudoasymmetric stereogenic centre*. *Set of interdependent stereogenic centres* is proposed as a basis for the treatment of the stereochemical differences between the ligands. The topological definition of the interdependence of the stereogenic centres is also presented. The modified system of the sequence rules consists of four rules.

I. INTRODUCTION

The Cahn-Ingold-Prelog (CIP) system of sequence rules[1-4] was developed to enable description of absolute configurations of molecules containing stereogenic units. Stereogenic units introduced by the CIP system are stereogenic centres, stereogenic axes and stereogenic planes[4].

The foundations of the CIP system were laid down in 1951[1]. Later some modifications and extensions have been presented[2], and the work of the authors of the CIP system was achieved in 1966[3].

Since then, the work on the CIP system was continued in three directions. The first direction was the discussion about classification of the molecular model to stereogenic units, which were introduced but not defined by Cahn et al[3]. Definitions of stereogenic units were proposed by Hirschmann and Hanson[5]. The second direction was the discussion about the completeness of the sequence rules and their consistency. Some extensions to the CIP rules were proposed[6]. The third direction of activities was the computer application of the CIP system. Some of the attempts were based on canonical numbering of molecular graphs[7-9], while Meyer[10] implemented the comparison of ligands according to the first two CIP rules.

Prelog and Helmchen[4] published a paper about basic principles of the CIP system and proposed its revision. They wished to put the CIP system on a solid theoretical ground and to enable its computer implementation.

Soon after the revision some new critiques appeared. Mislow and Siegel[11] criticised the CIP classification of stereogenic units. Dodziuk and Mirowicz[12] proposed a modification of the classification of molecular model to stereogenic units. Instead of three types of stereogenic units they proposed only two types: stereogenic centre and helical axis.

In the meantime Custer[13] found out that the sequence rules were incomplete. Recently Mata et al also published a proposal[14] for revision of the CIP sequence rules to overcome some limitations in applicability, consistency with the theory and generality of the CIP system.

Stereogenic units

In the beginning of the development of the CIP system stereogenic centres were used as the only cause of stereoisomerism[1]. Later Cahn et al.[2] introduced also the concept of axial and planar asymmetry. In the third paper about the CIP system[3], the stereogenic units were classified into three types: stereogenic centre, stereogenic axis and stereogenic plane. Some examples of each type of stereogenic units were shown, but they were not strictly defined. Stereogenic units were defined later by Hirschmann and Hanson[5]. They also noticed that in the case when molecules contain at most tetraligant atoms, stereogenic axes and planes can be replaced with only one type of stereogenic unit. The problem of choice of independent stereogenic units was also pointed out.

Mislow and Siegel[11] stressed the difference between stereogenicity and local chirality. While stereogenicity is based on graph theory, local chirality is rooted in theory of symmetry groups. They concluded that stereogenic units are based on stereogenicity. They showed that in some cases molecules with the same local symmetry possess different stereogenic units, as is the case in vespirenes and biphenyls.

Dodziuk and Mirowicz[12] proposed the application of only two types of stereogenic units: stereogenic centre and helical axis. In some cases these two stereogenic units overlap like in the case of allenes. The choice of the type of stereogenic unit depends on convention. The authors also pointed out the problem of choosing the independent stereogenic units in connection with treatment of conformational chirality.

Sequence rules

The definition and implementation of sequence rules makes possible a separate and sequential treatment of constitutional, geometrical, and stereochemical properties of molecules[1-4]. The constitution is treated with regard to material differences between the atoms, i. e. between their atomic and mass numbers, and with regard to topology of molecular graphs. Geometrical differences between ligands are taken into account with comparison of configurations of double bonds and comparison of configurations of pairs of stereogenic units. In the last step stereochemical properties of ligands, i.e., configurations of individual stereogenic units, are compared. The 1982 revision by Prelog and Helmchen[4] improved the comparison of constitutional differences between ligands, although a complete selectivity was not obtained[13]. The sequence rules after the revision[4] were as follows:

1. Higher atomic number precedes lower.
2. Higher atomic mass number precedes lower.
3. When two ligands differ only in that one has an atom or atom group of higher rank in a *cis*-, and the other in a *trans*- position to the core of stereogenic unit, then preference is given to the former. (This rule is restricted to ligands which differ in *cis-trans* isomerism of planar tetraligant atoms and double bonds.)
4. When two ligands have different descriptor pairs, then the one with the first-chosen *like* descriptor pair has priority over the one with unlike descriptor pair. *Like* descriptor pairs are: *RR*, *SS*, *RRe*, *SSi*, *ReRe*, *SiSi*, *MM* and *PP*, and also *RM*, *SP*, *ReM*, *SiP*. *Unlike* descriptor pairs are: *RS*, *ReSi*, *SRe*, *RSi*, *MP*, *PR*, *SM*, *ReP* and *SiM*. For each ligand, the descriptor ranked as first is paired with all the remaining descriptors. The hierarchical rank of the descriptor pair is given by the rank of the second descriptor in the pair.

5a. A ligand with descriptor *R* or *M* precedes the ligand with descriptor *S* or *P*.
  b. A ligand with descriptor *r* precedes the ligand with descriptor *s*.

A great improvement was achieved by the introduction of hierarchical digraphs. Hierarchical digraphs, which are acyclic directed graphs, enabled a clear definition of ligands, since all cyclic structures were transformed into corresponding acyclic ones. The treatment of constitution was performed by two rules, i. e., comparison of atomic numbers and of mass numbers. Topological properties were implicitly defined by hierarchical digraphs. Introduction of hierarchical digraphs also simplified comparison of configurations of pairs of stereogenic units. Unfortunately each molecular structure has to be transformed into hierarchical digraphs as many times as there are stereogenic units in the structure. Another problem is the size of hierarchical digraphs which may be much larger compared to the size of molecular graphs.

In spite of the improvement of the CIP system, some new proposals were presented and some old proposals were made again. Hirschmann and Hanson[6] proposed the introduction of descriptors *seqCis* and *seqTrans* for geometrically enantiomorphic double bonds. Geometrically enantiomorphic double bonds should be compared by rules 4 and 5 and not by the rule 3.

Custer[13] demonstrated that the CIP rules were incomplete and that the constitutional differences were not treated properly in all cases. In some cases, the ligands in the hierarchical digraphs are equivalent although they are different in the molecular graph. Custer introduced a subrule to rule 1 to guarantee the constitutional completeness. This addition states that:

> A duplicated atom with its predecessor node having the same label closer to the root ranks higher than a duplicated atom with its predecessor node having the same label farther from the root, which rank higher than any non-duplicated-atom-node.

Besides the subrule mentioned above, Custer proposed some other changes. He proposed subrule 5b[4] to be included in rule 4, as it was the case in the 1966 version[3]. He also proposed formal inclusion of the rule which was first proposed by Prelog and Helmchen[4], but its rank had not been defined, as the first subrule of rule 4:

> Chiral stereogenic units precede pseudoasymmetric stereogenic units and these precede non-stereogenic units.

Recently Mata et al.[14] proposed a revision of the CIP system sequence rules. They collected proposals of Custer[13], and Hirschmann and Hanson[6] and also proposed some novel solutions. They proposed a modification of rule 3, which should be restricted to geometrically diastereomorphic planar atoms and double bonds. Instead of the comparison of relative *cis* and *trans* descriptors, they proposed the comparison of descriptors *seqcis* and *seqtrans* as it had been the case in the 1966 version[3]. According to the proposals of Mata et al.[14], the sequence rules are as follows:

1a. Higher atomic number precedes lower;

b. A duplicated atom, with its predecessor node having the same label closer to the root, ranks higher than a duplicated atom, with its predecessor node having the same label farther from the root, which ranks higher than any non-duplicated-atom-node.

2. Higher atomic mass number precedes lower.

3. *seqcis* stereogenic units precede *seqtrans* stereogenic units and these precede non-stereogenic units.
The domain of application of this rule is restricted to geometrically diastereomorphic planar tetraligant atoms and double bonds.

4a. Chiral stereogenic units precede pseudoasymmetric stereogenic units and these precede non-stereogenic units. Geometrically enantiomorphic two-dimensional units precede two-dimensional non-stereogenic units.

b. When two ligands have different descriptor pairs, then the one with the first-chosen *like* descriptor-pair has priority over the one with a corresponding *unlike* descriptor-pair. *Like* descriptor-pairs are: *RR*, *SS*, *MM*, *PP*, *seqCisseqCis*, *seqTransseqTrans*, *RseqCis*, *SseqTrans*, *MseqCis*, *PseqTrans*, *RM* and *SP*. *Unlike* pairs are: *RS*, *MP*, *RP*, *SM*, *seqCisseqTrans*, *RseqTrans*, *SseqCis*, *PseqCis* and *MseqTrans*.
Methodology for pairing descriptors:
The descriptor assigned to geometrically enantiomorphic double bonds should be associated in the digraph with the first node corresponding to the atoms involved in the double bond.
For each ligand the descriptor chosen as the first (highest ranked descriptor) is paired with all the remaining descriptors. The following characteristics determine the hierarchical rank of the pairs of descriptors:
i) Higher rank of the second descriptor in the pair;
ii) Lower rank of the least common ancestor in the graph.

c. *r* precedes *s* and *p* precedes *m*.

5. A ligand with descriptor *R*, *M* or *seqCis* has priority over its enantiomorph with descriptor *S*, *P* or *seqTrans*.

Computer applications

Morgan's algorithm for the canonical numbering of molecular graphs[7] can be considered as the starting point of attempts for computerised coding of chemical structures. Some years later Dubois and Viellard[8] showed that the iterative canonical numbering algorithms were equivalent to the method for comparison of ligands in the CIP system, since all the methods compare atoms in concentric circles around the centre. Contrary to the CIP system, where ligands of stereogenic centres have to be priority ranked separately for each stereogenic centre, the canonical numbering methods rank the ligands of all stereogenic centres simultaneously. Wipke and Dyott[9] used sequence numbers of atoms, obtained by canonical numbering, as a measure of priority rank of ligands of stereogenic centres. Unfortunately they found out that their method and the CIP system are not compatible. The reason for incompatibility of treatment of differences in constitution was in the different treatment of multiple bonds. The methods for canonical numbering treat double bonds as two-membered rings and triple bonds as bridged rings. In the CIP system multiple bonds are transformed into single bonds by the introduction of duplicated atoms.

In 1980 Meyer tried to implement the CIP system in a computer program[10]. He took into account only the first two rules, i. e., the treatment of differences in the constitution of the ligands. He found the CIP system unsuitable for computer use and suggested that the system should be based on graph theory to enable easier application.

Instead of obtaining equivalence between the canonical numbering methods and the CIP system, the differences between these methods were increased by the introduction of hierarchical digraphs; multiple bonds and also rings are in general treated quite differently with hierarchical digraphs than with canonical numbering methods. It is important to note that canonical numbering methods treat molecular graphs and therefore no transformations are needed. Hierarchical digraphs are in general larger than corresponding molecular graphs, especially in the case of bridged systems. The size of hierarchical digraphs may cause serious problems in computerisation of the CIP system.

In the present contribution, the possibilities of integration of canonical numbering methods with the CIP system will be discussed. The integration will enable assignment of descriptors to stereogenic units based directly on molecular graphs instead of on hierarchical digraphs, which could be discarded. The Shelley-Munk method for canonical numbering[15] will be used for canonical numbering, equivalent to comparison of ligands in the CIP system. Instead of graph invariants, originally proposed by Shelley and Munk, sequence rules of the modified system will be applied. Since the version of Shelley-Munk method from 1977 incompletely treats bridged ring systems[16], the completeness will be achieved by applying the ideas used in the HOC procedures[17,18]. The CIP system will be discussed in the light of these proposals of modifications and some new concepts will be proposed to simplify the rules and their manual and computer application.

II. A PROPOSAL FOR MODIFICATIONS

To describe the configuration of a molecule, the molecule has to be modelled as a graph first. The procedure is then continued in three steps[4,14]:

a. Factorisation of the graph into stereogenic units,
b. Determination of rank order of the ligands around each stereogenic unit and
c. Determination of descriptor for each stereogenic unit.

Stereogenic centres as the most frequently met type of stereogenic units will be discussed first. Then sequence rules will be proposed and discussed.

Definition of stereogenic centres

According to Dodziuk and Mirowicz[12] two types of stereogenic units can be distinguished: stereogenic centre and helical axis. We propose to keep their concept of two types of stereogenic units. When stereogenic centres and helical axes could be applied alternatively, like in the case of allenes and biphenyls[12], preference should be given to stereogenic centres.

As pointed out by Mislow and Siegel[11], stereogenic units are rooted in graph theory, therefore they have to be defined in terms of graph theory. The definition of the stereogenic centres is proposed since this definition is also the basis for the novel concepts introduced later in this paper.

We propose a definition of the stereogenic centres based on graph symmetry and combinatorics. Graph symmetry is mathematically expressed by those vertex permutations which are members of the group of automorphic permutations of the graph.

> *A **potential stereogenic centre** is every atom with three or four ligands like carbon, nitrogen, silicium etc. For some of these atoms, it is known that they usually do not appear as stereogenic centres since the barrier for inversion of configuration is very low. The barrier height cannot be predicted by application of graph theory, therefore such cases have to be considered separately.*
>
> *A potential stereogenic centre* **really is** *a **stereogenic centre** if there exist no automorphic permutation causing an inversion of the configuration of only the potential stereogenic centre under consideration. Inversion of configuration occurs when the automorphic permutation causes an exchange of two equivalent ligands.*
>
> *If there exists at least one automorphic permutation causing the inversion of the configuration of only the stereogenic centre under consideration, then the potential stereogenic centre* **can be** *a **stereogenic centre** if the number of topologically equivalent neighbours (ligands) of potential stereogenic centre is less than or equal to the number of different configurations of these ligands.*

Potential stereogenic centres that fulfil the last condition are stereogenic centres only in the case when their topologically equivalent ligands differ in configuration. In such a case, the consideration of topology is not sufficient, therefore differences in stereochemistry must be found by application of the sequence rules.

The potential stereogenic centre in Fig. 1a is atom 3. It has four neighbours: atoms 2, 4, 6 and 7. Atoms 6 and 7 are topologically equivalent, therefore an automorphic permutation, which permutes them, must exist. The first condition is not fulfilled. Since topologically equivalent methyl groups have no stereogenic centres, they have only "one configuration". This means that the number of topologically equivalent ligands is greater than the number of different possible configurations and the second condition is also not fulfilled, hence atom 3 is not a stereogenic centre.

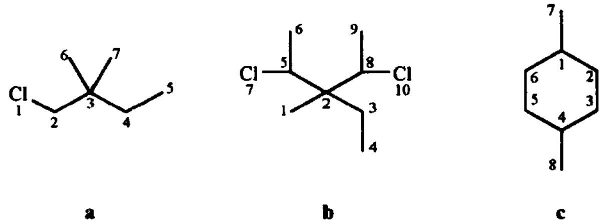


Figure 1: An example of three molecules with potential stereogenic centres.

a: Atom 3 is a potential stereogenic centre.

b: Atoms 2, 5 and 8 are potential stereogenic centres.

c: Potential stereogenic centres are atoms 1 and 4.

In Fig. 1b, the potential stereogenic centres are atoms 2, 5 and 8. The neighbours of atom 5 are all different, i. e., none of them is topologically equivalent to the others. The first condition is therefore fulfilled, and atom 5 is a stereogenic centre. Similarly, atom 8 is also a stereogenic centre. Atom 2 has two topologically equivalent neighbours - atoms 5 and 8 and the first condition is not fulfilled, the same way as for atom 3 in Fig. 1a. Since each of both topologically equivalent ligands of atom 2 has a stereogenic centre (atoms 5 and 8, respectively), each ligand can have two different configurations. The number of different configurations is equal to the number of topologically equivalent ligands, therefore the second condition is fulfilled, so atom 2 is from the graph-theoretical point of view a stereogenic centre. Atom 2 is an actual stereogenic centre only in the case when the stereogenic centres 5 and 8 have different configurations. This has to be established by the application of the sequence rules.

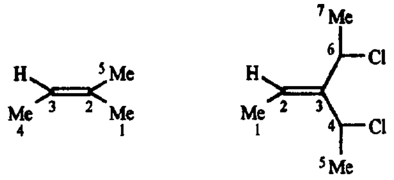
The molecule in Fig. 1c has two potential stereogenic centres - atoms 1 and 4. There exists an automorphic permutation that exchanges atoms 2 and 6, neighbours of atom 1, and so causes the inversion of its configuration. This permutation simultaneously exchanges also atoms 3 and 5, neighbours of atom 4, which is also a potential stereogenic centre. Since the permutation causes more than one inversion of configuration, the first condition is fulfilled, and consequently atoms 1 and 4 are stereogenic centres. There is no automorphic permutation that would cause an exchange of the neighbours of only one of the potential stereogenic centres.

Besides the atoms some bonds or cumulated bonds can also be treated as stereogenic centres. This is the case with biphenyls, alkenes, allenes etc[12]. In these cases the atoms, bound with such bonds, are sp^2 hybridised and have two geminal ligands. In the case of allenes, one has to take into account only both terminal atoms and may neglect the atoms between them. To find out whether a bond or system of cumulated bonds can be interpreted as a stereogenic centre, both atoms that are bound with such a bond (or terminal atoms) have to be considered.

> *A bond can be treated as a stereogenic centre if no automorphic permutation causing an inversion of configuration on the bond under consideration alone exist. This means that the inversion does not occur on the bond at all or it occurs on the bond under consideration and on at least one other bond or stereogenic centre. The inversion of the configuration is caused if an automorphic permutation exchanges one pair of geminal ligands of the bond.*
>
> *The bond can be treated as a stereogenic centre if an automorphic permutation causes an inversion of its configuration only when the number of different configurations of the pair of topologically equivalent geminal ligands, which are exchanged by the permutation, is greater than or equal to two (the number of topologically equivalent geminal ligands).*

Similarly to the atoms that are stereogenic centres, the bonds with topologically equivalent geminal ligands are stereogenic only if the ligands differ in the configuration. This also has to be determined by the sequence rules.

Two molecules with potential stereogenic centres and double bonds that may be interpreted as stereogenic centres are shown in Fig. 2.



a b

Figure 2: Two structures with double bond that can be interpreted as potential stereogenic centres.

In the case in Fig. 2a the double bond cannot be treated as stereogenic centre since atom 2 is bound to two methyl groups and so the first and the second condition are not fulfilled. Situation is different in Fig. 2b, where two topologically equivalent ligands are bound to atom 3, but they contain one stereogenic centre each (atoms 4 and 6, respectively). Each of the two ligands can be found in one of two possible different configurations. The double bond can be interpreted as a stereogenic centre if the configurations of the ligands are different. The potential difference has to be established by sequence rules.

Ranking the ligands

Canonical numbering method is proposed to be used instead of hierarchical digraphs. The comparison of the ligands is performed in concentric circles around the stereogenic centres[8] in both methods, with the hierarchical digraphs and with the canonical numbering method. There are several methods for canonical numbering. We propose the use of the Shelley-Munk method[15] since it is the most similar to the method of the CIP system. The criteria for canonical numbering, i. e., ranking the ligands, originally applied by the Shelley-Munk method, are replaced with the criteria, contained in the proposed sequence rules. There was found a lack of the criteria for ranking some ligands that are parts of bridged ring systems[16]. It was a general failure of the canonical numbering methods, which can be corrected in several ways. The most appropriate way seems to be the solution of the HOC procedure[17,18] since it is not suitable only for the computer use, but can quite easily be performed manually. The HOC procedure introduces additional criteria for ranking and therefore enables the basic method to remain unchanged. Consequently, the graphs do not have to be modified as it is the case in the CIP method.

Sequence rules

We propose a modification of the sequence rules to enable the assignment of descriptors to the stereogenic centres on the basis of molecular graphs. The hierarchical digraphs will not be used in our approach. Since the modified sequence rules enable the treatment of the molecular graphs with the canonical numbering method[15], canonical sequence numbers will become a measure of the priority rank order. Wipke and Dyott[9] found out that the canonical sequence numbers of the atoms express the priority rank order of the ligands of the stereogenic centres. There are two possibilities for the interpretation of the priority rank order according to the canonical sequence numbers of the atoms. The first possibility is that an atom with a lower sequence number ranks higher than the atom with a higher sequence number. The alternative is that an atom with a higher sequence number ranks higher than the atom with a lower sequence number. We chose the former possibility.

Descriptors derived by the application of the modified sequence rules will be the same as the old ones in as many cases as possible. The rules have to be applied sequentially. The first rule remains the same:

1. *Higher atomic number precedes lower.*

This rule is not sufficient to find all topological differences between ligands in some bridged ring systems[16,17]. The information about the ring sizes is missing and has to be added explicitly as a rule, as it is proposed in the HOC procedures[17,18]:

2a. *An atom, member of a larger ring, precedes an atom, member of a smaller ring.*

For the theoretical completeness the second part of the rule is proposed although usually will not have to be used:

b. A bridgehead atom that has the two neighbours, members of the same ring, of higher rank than the neighbours of another bridgehead atom, members of a ring of the same size are, ranks higher than the latter.

With the third rule the differences in atomic mass numbers are taken into account. The rule is the same as before:

3. *Higher atomic mass number precedes lower.*

The first three rules are necessary to take into account all constitutional differences. For a great part of molecules these three rules are sufficient, since the ligands of the majority of stereogenic centres are priority ranked, therefore their descriptors can be determined. From now on only the differences between topologically equivalent ligands have to be considered. Non-topological differences will be dealt with the rule 4, which is presented below.

Atoms 1 and 4 in 1,4-dimethyl cyclohexane (see Fig. 1c) are stereogenic centres. Atoms 2 and 6, the neighbours of the stereogenic centre 1, are topologically equivalent. They cannot be distinguished by the application of rules 1 to 3. The same problem occurs with the stereogenic centre 4. Since the molecular graphs are applied instead of the hierarchical digraphs, auxiliary descriptors cannot be determined to find the differences between the ligands. On the other hand, if one of the ligands of the stereogenic centre 1 preceded the other, than the topologically equivalent ligands of the stereogenic centre 4 would also be ranked. If atom 2 preceded atom 6, for example, than atom 3 would precede atom 5. This is the consequence of the application of the canonical numbering method for priority ranking of ligands. In other words, the priority rank of the neighbours (ligands) of the stereogenic centre 1 depends on the priority rank of the neighbours (ligands) of the stereogenic centre 4 and vice versa. The priority rank of the ligands of the stereogenic centres is **interdependent**, therefore also the descriptors of the configuration of the stereogenic centres are interdependent.

We propose the introduction of a novel concept: the **set of interdependent stereogenic centres**. Members of a set of interdependent stereogenic centres can be found by inspecting the automorphic permutations, therefore the concept is based in graph theory. In order to find the stereogenic centres, members of a set of interdependent stereogenic centres, one has to select from the group of automorphic permutations those which cause the inversion of the configuration of a selected stereogenic centre, member of the set of interdependent stereogenic centres. These permutations exchange the topologically equivalent neighbours of the stereogenic centre, while the stereogenic centre is permuted to itself. Among the chosen permutations are further selected those which cause the minimal number of inversions. All stereogenic centres, on which the inversion of configuration is caused by the application of those permutations, are members of the set of interdependent stereogenic centres.

Stereogenic centres 1 and 4 (Fig. 1c) are interdependent due to the existence of an automorphic permutation that exchanges atoms 2 and 6, and atoms 3 and 5, respectively. They are members of a two membered set of interdependent stereogenic centres.

A set of interdependent stereogenic centres is the basic independent stereogenic unit. A set of interdependent stereogenic centres may have only one member. This is the case of the stereogenic centres with four topologically different ligands. An example of the one membered sets is the molecule in Fig. 1b, where the stereogenic centre 5 is the only member of a set of interdependent stereogenic centres. The stereogenic centre 8 is also the only member of another set of interdependent stereogenic centres.

The configuration of each set of interdependent stereogenic centres is described by a vector. This vector is called a **description vector**. The components of each description vector are the descriptors of the configurations of the stereogenic centres. The first component of the description vector corresponds to the stereogenic centre with the smallest sequence number, while the last component to the stereogenic centre with the highest sequence number. Since there are two possible descriptors for each stereogenic centre, each description vector can be read as a binary number. Due to the symmetry, the configuration of a set of interdependent stereogenic centres can be described with more than one vector. This happens when the set of interdependent stereogenic centres contains more than one member. Among these vectors, the description vector is the one representing the greatest binary number. This binary number can be converted to a decimal number, which is called a **description number**. The description numbers are the descriptors of the configuration of the sets of stereogenic centres and they can be used for the priority ranking. The details of the procedure are given in the Examples section.

Since the description vectors are of different dimensionality, their pairing cannot be performed like pairing of the descriptors by rule 4 in the CIP system. The new rule for discrimination of the ligands on the basis of the stereochemical differences is:

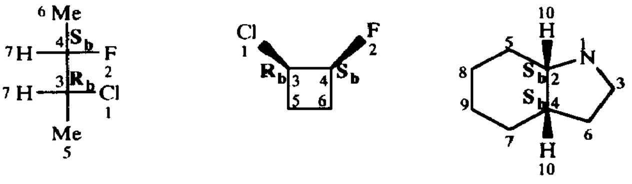
4. *A description vector with higher description number precedes the description vector with lower description number. Description vectors with the smallest number of components are compared first, while description vectors with the highest number of components are compared last.*

Descriptors

In the CIP system[4], the configuration of stereogenic centres is described with the descriptors *R* and *S*, that of pseudoasymmetric stereogenic centres with *r* and *s*, and that of double bonds with *seqcis* and *seqtrans* (*Z* and *E*, respectively). Each descriptor in the CIP system describes the space orientation of the ligands and simultaneously the presence or absence of mirror plane. Since each descriptor of stereogenic centres contains the information on two distinct structural features there are four different descriptors needed to describe all possibilities.

We propose the use of the descriptors that contain the information on one feature only. Each of our descriptors describes only the space orientation of the ligands and therefore our descriptors are binary. Due to tradition we propose the same letters for descriptors. Subscript *b* is added to each descriptor to distinguish our **binary** descriptors from the **quaternary** CIP descriptors. From the theoretical point of view, it would be sufficient to use only descriptors R_b and S_b for stereogenic centres. Written communication is however easier if one can distinguish at first sight the descriptors of stereogenic centres with topologically equivalent ligands from those with topologically different ligands. We propose to use the small letter descriptors r_b and s_b for the stereogenic centres with topologically equivalent ligands. The configuration of each such stereogenic centre depends on the configurations of other stereogenic centres. The use of this type of descriptors is broader in comparison to the CIP system from 1982[4], since these descriptors are applied to all stereogenic centres with topologically equivalent ligands. The configurations of the stereogenic centres with four topologically different ligands do not depend on the configurations of the other stereogenic centres. For such stereogenic centres we therefore propose the term **constitutional stereogenic centres**, while for the other stereogenic centres we propose the term **configurational stereogenic centres**. For constitutional stereogenic centres, we propose the application of descriptors R_b and S_b to distinguish them from the configurational stereogenic centres. The descriptors of constitutional stereogenic centres obtained by the modified system are in most cases the same as those obtained using to the CIP system. We propose a similar solution for double bonds. For double bonds with topologically different geminal ligands the descriptors Z_b and E_b should be applied, while for double bonds with topologically equivalent geminal ligands we propose the introduction of descriptors z_b and e_b. Space orientation of the ligands is the same for capital letter and small letter descriptors. The descriptors R_b, Z_b, r_b and z_b in description vectors are interpreted as 1 (one), while S_b, E_b, s_b and e_b are interpreted as 0 (zero). The use of four distinct descriptor-pairs (R_b-S_b, r_b-s_b, Z_b-E_b and z_b-e_b) might seem to be in contradiction with the statement that the descriptors are binary. The fact that each descriptor is mapped either to 1 or 0 proves that they are binary indeed.

III. EXAMPLES



a **b** **c**

Figure 3: Three structures where the same stereochemical descriptors are obtained with both methods, i. e., with the CIP system and with the proposed modified system. The numbers are obtained by the canonical numbering method and they correspond to the priority rank of the atoms. Notice that hydrogens in the case **a** have the same number, since they are by tradition omitted from the graphs, and therefore they obtain the highest sequence numbers, i. e., the lowest priority. The situation is similar with the case **c**.

Some examples where the descriptors remain unchanged by the modifications proposed are shown in Fig. 3. In all the cases in Fig. 3, the molecules possess only constitutional stereogenic centres. The ligands of each stereogenic centre were ranked by the application of rule 1. An example of a molecular graph, where rule 2 also has to be applied for ranking of the ligands, is shown in Fig. 4.

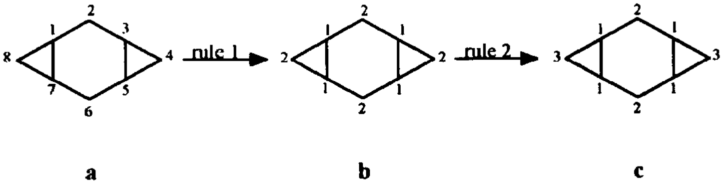


Figure 4: An example of a molecular graph where rule 2 has to be applied.

a: A molecular graph with four stereogenic centres (atoms 1, 3, 5 and 7). Numbering is arbitrary.

b: The stereogenic centres are topologically equivalent. The neighbours of each of the stereogenic centres are one tertiary carbon atom and two secondary carbon atoms, which cannot be ranked by rule 1. The sequence numbers of the atoms after the application of rule 1 are shown.

c: To rank the secondary carbon atoms, the neighbours of each stereogenic centre, rule 2 has to be applied. The secondary carbon atoms, members of the six-membered rings precede the secondary carbon atoms, members of the three-membered rings. Each stereogenic centre (sequence number 1) has three nonhydrogen neighbours, which are priority ranked, therefore the configuration of each stereogenic centre can be specified.

Stereogenic centres 1, 3, 5 and 7 (Fig. 4a) are constitutional stereogenic centres. Their ligands are ranked in the same way as with the CIP system[4], therefore the descriptors of the configuration remain unchanged.

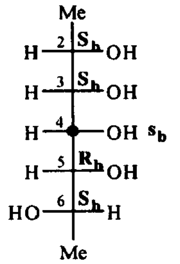


Figure 5: One of the stereoisomers of 2,3,4,5,6-heptanepentol, where constitutional stereogenic centres 2, 3, 5 and 6 have the same descriptor as according to the CIP system. The descriptor of the configurational stereogenic centre 4 however is changed with regard to the CIP system. According to the CIP system, the descriptor of the pseudoasymmetric stereogenic centre 4 would be *R*.

The new descriptors are often different from the CIP descriptors in the case of the configurational stereogenic centres. An example of a configurational stereogenic centre with a different descriptor compared to that in the CIP system is shown in Fig. 5. The difference is a consequence of the revised rule 4, since there is no comparison of pairs of descriptors.

An example of a molecule containing a set of interdependent stereogenic centres is one of the two different stereoisomers of 1,4-dimethyl cyclohexane, which is shown in Fig. 6.

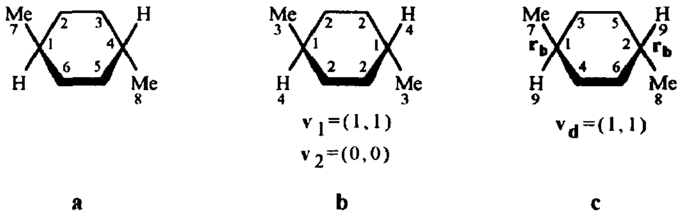


Figure 6: One of the stereoisomers of 1,4-dimethyl cyclohexane.

a: The stereogenic centres are atoms 1 and 4. The starting atom numbering is arbitrary.

b: The atoms and consequently the ligands are priority ranked by rule 1. The labels correspond to priority ranks of the atoms. Rules 2 and 3 provide no further ranking of the ligands. The stereogenic centres form one set of interdependent stereogenic centres. The configuration of the set can be described with two vectors.

c: The atoms are further ranked by the application of rule 4 and the final canonical numbering is obtained. The description vector and the descriptors of both stereogenic centres are also presented.

The atoms of the stereoisomer of 1,4-dimethyl cyclohexane (Fig. 6a) are ranked according to rule 1 (Fig. 6b). Atoms 2 and 6, the neighbours of the stereogenic centre 1, are topologically equivalent, therefore they cannot be ranked by the first three rules. The same is valid for atoms 3 and 5, the neighbours of the stereogenic centre 4. The stereogenic centres 1 and 4 are also topologically equivalent, i. e., they are symmetry related, therefore the sequence order of the stereogenic centres in the vectors, representing the configuration of the set of interdependent stereogenic centres, has to be set arbitrarily. We can choose that the first component of the vector represent the configuration of the stereogenic centre 1, while the second component represents the configuration of the stereogenic centre 4. If we choose that atom 2 precedes atom 6 (see also Fig. 6a), the canonical numbering shown in Fig. 6c is obtained and the configuration is described by vector $\mathbf{v}_1$. If we chose however that atom 6 precedes atom 2, then the configuration would be described by vector $\mathbf{v}_2$. The description vector of the set of interdependent stereogenic centres is that vector to which corresponds the highest description number. Number 3 corresponds to vector $\mathbf{v}_1$ and number 0 corresponds to vector $\mathbf{v}_2$. Therefore, vector $\mathbf{v}_1$ is the description vector ($\mathbf{v}_d$). Consequently, both stereogenic centres obtain descriptors r_b (Fig. 6c).

In the example described above (Fig. 6), the stereogenic centres were the members of only one set of interdependent stereogenic centres. In the next example, a stereoisomer with three sets of interdependent stereogenic centres will be presented. Stereogenic centres of the stereoisomer, shown in Fig. 7, belong to three sets of interdependent stereogenic centres.

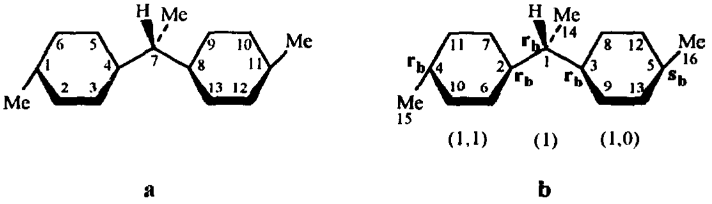


Figure 7: An example of a stereoisomer with three sets of interdependent stereogenic centres.
a: The stereogenic centres are atoms 1, 4, 7, 8 and 11. The starting atom numbering is arbitrary.
b: The description vectors of two sets have to be compared to determine the configuration of the third. The graph is canonically numbered.

The stereogenic centres 1 and 4 in Fig 7a belong to one set of interdependent stereogenic centres, the stereogenic centres 8 and 11 correspond to the other, and the stereogenic centre 7 is the only member of the third set of interdependent stereogenic centres. Since both ring ligands of the stereogenic centre 7 have the same constitution, they cannot be ranked by the first three rules. Each of the ring ligands of the stereogenic centre 7 contains two interdependent stereogenic centres. In one ligand the interdependent stereogenic centres are the stereogenic centres 1 and 4, while in the other ligand the interdependent stereogenic centres are 8 and 11. The description vectors and the description numbers of both sets of interdependent stereogenic centres have to be determined to apply rule 4 to rank both ligands. In the same way as in the previous example (see Fig. 6) the description vectors of both sets of interdependent stereogenic centres were determined. According to rules 1 to 3 the stereogenic centre 4 ranks higher than the stereogenic centre 1, therefore the first component of the corresponding description vector corresponds to the stereogenic centre 4, while the second component corresponds to the stereogenic centre 1. In the remaining pair of the stereogenic centres, the stereogenic centre 8 ranks higher than the stereogenic centre 11, therefore the first component corresponds to the former and the second to the latter. The stereogenic centres 4 and 8 rank higher than the stereogenic centres 1 and 11 since the neighbours of the former pair are one tertiary carbon atom and two secondary carbon atoms, while the neighbours of the latter pair are two secondary carbon atoms and one primary carbon atom.

The description number corresponding to the stereogenic centres 1 and 4 is equal to 3 (this is the decimal number corresponding to the description vector (1,1) read as a binary number), while the description number corresponding to the stereogenic centres 8 and 11 is equal to 2 (description vector (1,0)), see Fig. 7b. Therefore, the ligand with the former pair of stereogenic centres precedes the ligand with the latter pair of stereogenic centres. The descriptor of the stereogenic centre 7 can now also be determined. The canonically numbered graph of the molecule with the descriptors of the stereogenic centres and the corresponding description vectors are shown in Fig. 7b.

An example of a polycyclic system is one of the stereoisomers of 2,6-adamantanediol, shown in Fig. 8.

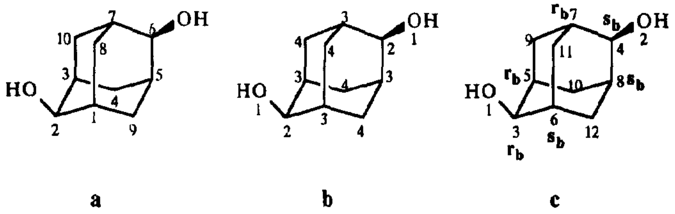


Figure 8: One of the stereoisomers of 2,6-adamantanediol.

a: The starting atom numbering is arbitrary. From the topological symmetry follows that the stereogenic centres 2, 5 and 7 are interdependent. The stereogenic centres 1, 3 and 6 are also interdependent.

b: The priority rank of the atoms after the application of rules 1 to 3.

c: The final canonical numbering and the corresponding descriptors.

In the example presented, the stereogenic centres are atoms 1, 2, 3, 5, 6 and 7. Each of them has two topologically equivalent ligands. Stereogenic centres 2 and 6 rank higher than the other four (Fig. 8b). Stereogenic centres that are interdependent with the stereogenic centre 2 can be found by automorphic permutation that permutes the stereogenic centres 1 and 3, the neighbours of the stereogenic centre 2. This permutation permutes the topologically equivalent neighbours of the stereogenic centres 5 and 7. Since the permutation causes inversion of configuration of the stereogenic centres 2, 5 and 7, they are interdependent. In the same way, it can be found out that the stereogenic centres 1, 3 and 6 are interdependent, too. In the first description vector the first component corresponds to the stereogenic centre 2, while the order of the stereogenic centres 5 and 7 is not determined yet. In the second description vector the first component corresponds to the stereogenic centre 6, while the order of the stereogenic centres 1 and 3 has to be determined later.

The configuration of the stereogenic centre 2 can be determined if the rank of its topologically equivalent neighbours is set. Since the component of the description vector, corresponding to the stereogenic centre 2, is the first, the rank order of the neighbours of the stereogenic centre 2 has to be chosen in such a way as to obtain the descriptor r_b. The descriptor r_b is obtained if the stereogenic centre 3 precedes the stereogenic centre 1. Consequently, atom 10 precedes atom 8 (see Fig. 8a) and atom 4 precedes atom 9, therefore the configuration of the stereogenic centres 5 and 7 can also be determined. When the descriptors of the stereogenic centres 5 and 7 are determined, they can be ranked to obtain the final canonical numbering and to obtain the descriptors of the other stereogenic centres, as shown in Fig. 8c. The first description vector is (1, 1, -1), while the second is (-1, 1, -1). The values of the components of the second description vector are in this case determined on the basis of the first one. This is a consequence of the fact that members of both sets of the interdependent stereogenic centres are members of the same bridged ring system. The name of the isomer is $2r_b,6s_b$-adamantanediol, referring to the initial numbering.

An example of a molecule with a large set of interdependent stereogenic centres is shown in Fig. 9.

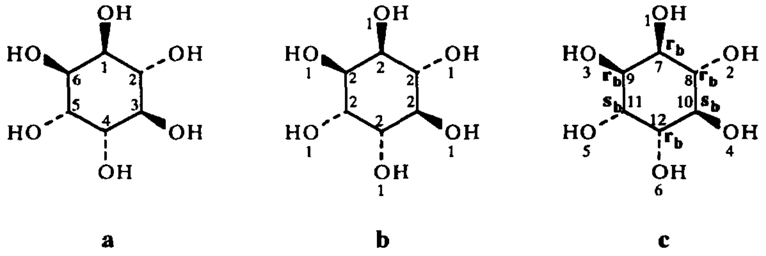


Figure 9: One of the stereoisomers of inositol.

a: The starting atom numbering is arbitrary.

b: The priority rank order of the non-hydrogen atoms after the rule 1 to 3. All six stereogenic centres are topologically equivalent and each has two topologically equivalent neighbours.

c: The atoms are canonically numbered and the configuration of the stereogenic centres is determined.

In the example presented, the stereogenic centres are atoms 1 to 6 (see Fig. 9a). They are topologically equivalent, as shown in Fig. 9b. The determination of the set of interdependent stereogenic centres is in this case not easily surveyed. For example, the stereogenic centres that are interdependent with the stereogenic centre 1 are determined from the permutations that permute its neighbours, while they permute it to itself. There is only one such permutation in this case. This permutation also permutes the topologically equivalent neighbours of the stereogenic centre 4 (see Fig. 9a) and therefore causes an inversion to both stereogenic centres. From this, we can conclude that the stereogenic centres 1 and 4 are interdependent. The stereogenic centres that are interdependent with the stereogenic centre 3 are those, to which an inversion occurs, when the automorphic permutation permutes the neighbours of the stereogenic centre 3 and consequently the neighbours of the stereogenic centre 6. There is again only one such permutation. This permutation causes an inversion of configuration of all stereogenic centres in the molecule, therefore they are all members of the same set of interdependent stereogenic centres.

The reason for this can be seen the best in Fig. 9c, where the atoms are numbered canonically. The stereogenic centre 3 has number 10 in Fig. 9c. The permutation permutes its neighbours 8 and 12 (Fig. 9c), atoms 7 and 11 and the OH groups bound to them. Atom 8 becomes atom 12. The higher ranked neighbour of atom 8 is atom 7 (Fig. 9c). Atom 7 is permuted with atom 11, which is the lower ranked neighbour of atom 12. Therefore the priority rank of the neighbours of atom 8 (and also of atom 12) is reversed by the permutation. In the case of atoms 7 and 11, the situation is the same.

Since carbon atoms are topologically equivalent, any of them is a candidate to be chosen as the highest ranked one. The description number is the largest if atom 1 (see Fig. 9a) is chosen as the highest ranked carbon atom. The atoms are canonically numbered in the way to get descriptor r_b for the stereogenic centre 1, as shown in Fig 9c. Since atom 1 (Fig. 9a) is the highest ranked, it should keep the sequence number 1. Atom 2, the neighbour of atom 1, ranks higher than atom 6, therefore its sequence number remains unchanged. The name of this stereoisomer is therefore $1r_b,2r_b,3s_b,4r_b,5s_b,6r_b$-inositol (initial numbering).

IV. DISCUSSION

Basic stereogenic units are sets of interdependent stereogenic centres. The interdependence of stereogenic centres is defined on the topological basis, and is therefore not the same as Hirschmann and Hanson's interdependence[5]. An example of the interdependent stereogenic centres according to Hirschmann and Hanson is shown in Fig. 10.

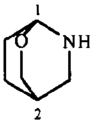


Figure 10: Molecular graph containing stereogenic centres 1 and 2 which are interdependent according to Hirschmann and Hanson[5].

Hirschmann and Hanson interpret stereogenic centre as interdependent because an exchange of two ligands of one stereogenic centre would be in reality followed by exchange of corresponding two ligands of the other stereogenic centre to keep the ring strain as low as possible. This criterion is energetical and geometrical in its basis. We base our definition of interdependence on purely topological arguments. There exists no automorphic permutation that would exchange any two ligands of the stereogenic centres 1 or 2 since their ligands differ in the constitution. Therefore these two stereogenic centres are not interdependent according to our definition and their descriptors can be determined independently. From the combinatorial point of view, four different stereoisomers with constitution presented by the molecular graph in Fig. 10 are possible. Due to the ring strain, only two stereoisomers exist in reality. We already implemented our concept of the interdependence in the method for deriving groups of automorphic permutations of the graphs[19].

We also proposed different naming conventions for the stereogenic centres, based on the topological approach, as an alternative to the terms *stereogenic centres* and *pseudoasymmetric stereogenic centres* according to the CIP system. From our term for the stereogenic centre and our descriptor one would immediately know whether the stereogenic centre has ligands with the same constitution or not. We propose the term *constitutional stereogenic centres* for the stereogenic centres with four ligands that differ in the constitution, because the priority rank order of the ligands depends only on their constitution. The configurations of other stereogenic centres are not important in these cases. Each *constitutional stereogenic centre* is the **only** member of its set of interdependent stereogenic centres, i. e., it is independent by our definition. On the other hand, the *configurational stereogenic centres* have at least two ligands with the same constitution. The topologically equivalent ligands of the configurational stereogenic centres possess stereogenic centres, therefore the descriptor of the configuration of a *configurational stereogenic centre* depends on the constitution of the ligands and on the configuration of other stereogenic centres.

According to the CIP system, the term *pseudoasymmetric stereogenic centre* is used for the stereogenic centres with enantiomorphic ligands. The term *pseudoasymmetric stereogenic centre* is therefore based on the geometrical symmetry approach, while the term *configurational stereogenic centre* is based on the topological symmetry approach.

The introduction of hierarchical digraphs by Prelog and Helmchen[4] was a great step forward in the treatment of the constitutional differences between the ligands. Unfortunately, it caused an insurmountable dichotomy between the CIP system and the computer methods based on canonical numbering. The hierarchical digraphs are suitable for deriving of the descriptors manually, while they are in general not very suitable for computer use for two reasons. The first reason is that the hierarchical digraphs are much larger in the symmetrical bridged ring systems than the molecular graphs are. It is difficult to derive the hierarchical digraphs, either manually or by computers, and to manipulate them. The second reason is that for each stereogenic centre, a separate hierarchical digraph has to be derived to rank the ligands of the stereogenic centre. The consequence is the derivation of a large number of large hierarchical digraphs, which is very time consuming and memory demanding. On the other hand, an advantage of canonical numbering methods is that they rank the ligands of all stereogenic centres in the molecular graph simultaneously. Of course, the canonical numbering is very suitable for the computers, while it is less so for men. Nevertheless, there exists a method, similar to the method applied by the CIP system, which is equivalent to canonical numbering and can be easily performed manually.

The canonical numbering methods "see" the molecular graph the same way as a traveller walking along the graph from one atom to another. The traveller can only walk forward; he can never return from an atom to the previous atom by making a step backward. The process is similar to the building of the hierarchical digraphs. There are, however, two important differences. If the traveller walks in a cycle, he returns to the starting point regardless of the fact that he always goes forward. The traveller of the CIP system stops when he reaches a point where he has already been. He "marks" each atom on his way. Our traveller cannot do this. He walks in circles, therefore the path is periodically repeating. Instead of ending the path the same way as in the CIP system, the molecular graphs are transformed into acyclic graphs with **periodic chains**. Similar to the rings are multiple bonds. In a double bond, there exist two ways to go from one atom to the next. Therefore, the traveller can use one way to go from one atom to the other and return by the other way, without violating of the rule that he must go only forward. The length of the periodic chains that has to be considered depends on the length of the acyclic parts of the original graphs and on the ring sizes. Some examples are presented below.

Two examples of the differences in the treatment of the ligands are shown in Fig. 11. In the first case (Fig. 11a), an acyclic ligand is compared with a ligand containing a double bond. In the hierarchical digraph, the acyclic ligand precedes the ligand with the double bond, while in the graph with periodic chains the latter precedes the former. This difference is a consequence of the treatment of double bonds as rings. In the second example (Fig. 11b), the acyclic ligand is compared with a ligand containing a three membered ring. In this case, the acyclic ligand also precedes the cyclic one according to the CIP system, while the latter precedes the former in the modified system. In both cases, the periodic chains do not have to be compared with the acyclic ligand in their full length. The length of the acyclic ligand is sufficient.

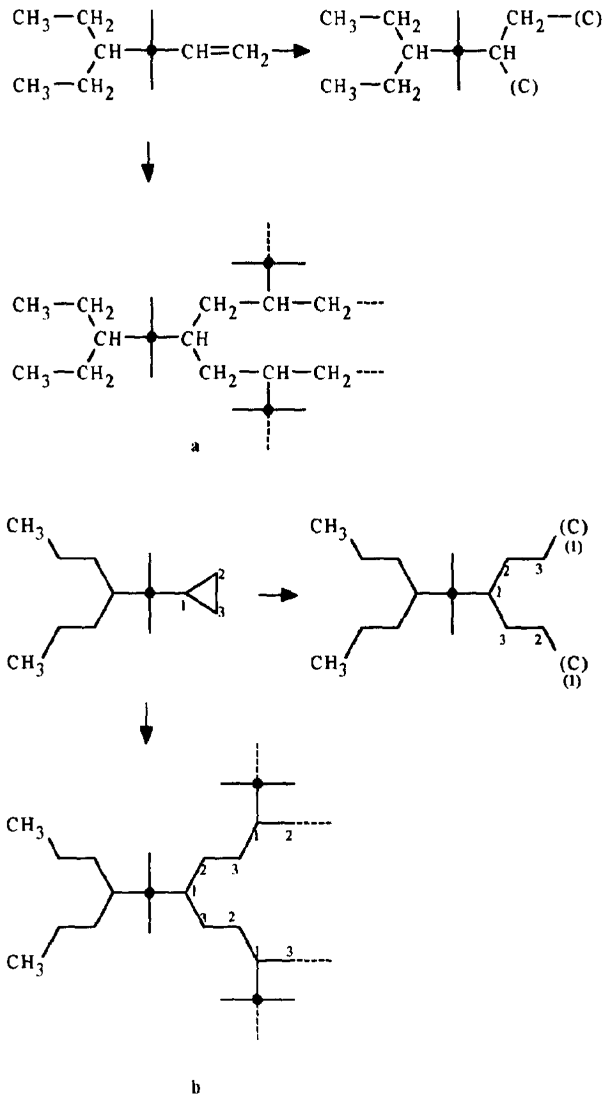


Figure 11: Two molecular graphs where an acyclic ligand has to be compared with a ligand with a double bond (a) and a cyclic ligand (b), respectively. In both cases at the top left is molecular graph, on its right side is the hierarchical digraph, and the graph with periodic chains is below.

When two ligands containing ring systems have to be compared, their periodic chains have to be as long as is the length of the longest period increased by two atoms. In other words, the length of the chains must be equal to the length of the longest chain in the hierarchical digraph, increased by one atom. Additionally, such extension of the hierarchical digraphs to the length of the longest chain, obtained from a ring, and increased by one atom, solves the problem of the constitutional completeness, mentioned by Custer[13].

For manual application of rules 1 and 3 of the modified system of sequence rules, the transformed graphs as presented in Fig. 11 are necessary . Rules 2 and 4 are applied on molecular graph. Transformation of rings and double bonds into periodic chains keeps the immediate neighbourhood of all atoms unaltered. In the hierarchical digraphs, duplicate atoms have different immediate neighbourhood than their predecessors in the original graphs. Transformation of rings in hierarchical digraphs has no chemical background, since the last atom in the chain obtained by the transformation of the ring is a duplicate atom, which would not exist if the ring bond was broken chemically.

The descriptors of the constitutional stereogenic centres remain identical to the CIP descriptors by the modifications proposed when the hierarchical digraph is the same as the molecular graph (acyclic molecules, Fig. 3a) and when the differences in the constitution between the ligands appear before the end of the chain in the hierarchical digraph regardless of the origin of the chain (Fig. 3b, c). In the example shown in Fig. 3c, rule 1 is applied first. With application of the first rule all ligands of both stereogenic centres can be ranked, therefore rule 2 is not needed.

In bridged ring systems, where the ligands are parts of ring systems as in the example shown in Fig. 4, the first rule is sometimes not sufficient in the modified system. In the CIP system, the data about the ring sizes are implicitly taken into account in the hierarchical digraphs. In the modified system the ring sizes have to be expressed explicitly (rule 2). In the hierarchical digraphs a ligand, which has been obtained from a larger ring, precedes the ligand, which has been obtained from a smaller ring, if the first difference between the ligands appears after the duplicated atom of the shorter chain has been compared with an atom from the longer chain. This rather complicated implicit rule of the CIP system is replaced with the much simpler explicit rule 2 in our proposal: The atoms, members of larger rings, precede the atoms, members of smaller rings. The second part of rule 2 serves more for theoretical completeness than because of its frequent necessity. The graphs that have to be treated by the second part of rule 2 are usually not a part of chemist's everyday work[18].

Graph theory makes the unification of the treatment of double bonds and stereogenic centres possible. From this point of view there is no need to rank the ligands according to the configurations of the double bonds first and later, if necessary, according to the configurations of the stereogenic centres the same way as it is performed by the CIP system (rules 3 to 5). One of the basic principles of the CIP system is that the first difference between the ligands, i. e. the difference that appears closer to the stereogenic centre is decisive. We propose that this principle should be taken into account consistently. A molecule, containing a configurational stereogenic centre, where in the topologically equivalent ligands the stereogenic centre lays closer to the central stereogenic centre than does the double bond, is shown in Fig. 12.

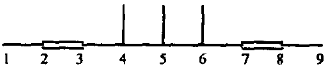


Figure 12: A molecular graph with the constitutional stereogenic centres 4 and 6 lying closer to the configurational stereogenic centre 5 than the double bonds are.

The configurational stereogenic centre 5 has two topologically equivalent ligands. According to the CIP system the configuration of the double bonds is compared first. In the system we propose, the configuration of the double bonds is decisive only if the constitutional stereogenic centres 4 and 6, laying closer to the configurational stereogenic centre, have the same configuration. Rule 4 of the modified system is therefore consistent with the spirit of the CIP method.

Our rule 4 is certainly the most important change proposed. Priority relations are no longer based on the stereogenic centres but on the sets of interdependent stereogenic centres, instead. The comparison of the configurations of the sets of interdependent stereogenic centres degenerates to the comparison of the configurations of individual stereogenic centres only in the case of sets with a single member as in the case shown in Figure 5. The comparison of the configurations of stereogenic centres performed by the CIP system is therefore only a special case of something more general.

The comparison of the pairs of descriptors as performed by rule 4 of the CIP system[4] cannot be performed by the modified system due to the possible differences in the dimensionality of the description vectors. Although the existence of mirror planes is not explicitly expressed by the descriptors proposed, the information is still contained. An advantage of the proposed system is that on the basis of the descriptor one immediately knows whether a stereogenic centre has topologically equivalent ligands or not. In the example shown in Fig. 5, the small letter descriptor in the name of the substance makes evident the fact that the stereogenic centre 4 has two topologically equivalent ligands. The other stereogenic centres have no topologically equivalent ligands.

The derivation of the description vectors is similar to looking for the "best names" introduced by Wipke and Dyott[9]. Looking for the "best names" is typical for the molecular graphs, therefore no derivation of auxiliary descriptors is needed. The difference between this and our method is that all stereogenic centres of the molecules are included by their method, while in our method this has to be performed for each set of interdependent stereogenic centres separately. The problem is thus frequently reduced. For example, in the case of the molecule, shown in Fig. 7, four possible "names" have to be checked to find the "best name", while the "name" is correctly found in the first attempt by our method. Our proposal of rule 4 enables a simpler treatment of the stereochemical differences between the ligands of the stereogenic centres than the CIP system and the method proposed by Wipke and Dyott do. On one hand CIP system is simple for manual use and complicated for computer use for large and symmetric molecules. On the other hand the method proposed by Wipke and Dyott is suitable for computer use and unsuitable for the manual use. We joined the advantages of both methods to enable easy manual and computerised treatment.

In the example shown in Fig. 6, the description vector represents a sort of the "best name". The description vector can be determined by choosing one of the stereogenic centres as the first in the vector. In this case one can choose the order of the stereogenic centres arbitrarily since they are topologically equivalent. In the next step, the topologically equivalent neighbours of the first stereogenic centre have to be ranked in such a way to obtain the descriptor r_b for the stereogenic centre chosen. This step is the consequence of the condition that the description vector is the one represented with the greatest binary number. Therefore the first component of the vector should be 1, which corresponds to descriptor r_b in this case. When the topologically equivalent neighbours of the first stereogenic centre are ranked, the topologically equivalent neighbours of the second stereogenic centres are ranked simultaneously, therefore the descriptor of its configuration can be determined.

The procedure is similar in the case shown in Fig. 7. First, the configurations of those stereogenic centres which are ring members, have to be determined. In both rings, the topologically equivalent neighbours of the higher ranking stereogenic centre are ranked to obtain the descriptor r_b for the higher ranked stereogenic centre. The rank order of the neighbours of the other stereogenic centre, member of the same set of interdependent stereogenic centres, is determined simultaneously and its descriptor can be specified. Then both description vectors or both description numbers, respectively, are compared according to rule 4 to rank both ring ligands. In this case the descriptors of both two-membered sets of interdependent stereogenic centres have to be derived to enable specification of the configuration of the central stereogenic centre. There is no contradiction with the demand of rule 4 that smaller sets have to be compared first, because the configuration of the central stereogenic centre depends on the configurations of the other stereogenic centres, while the configurations of the other stereogenic centres do not depend on the configuration of the central one.

The example of one of the stereoisomers of adamantanediol, shown in Fig. 8, serves as an illustration of the relative simplicity of the modified system. The manual derivation of the descriptors is not complicated. From the symmetry of the molecule it is evident that each stereogenic centre has two topologically equivalent ligands. The ligands of the stereogenic centres 2 and 6 (Fig. 8a) are ranked with regard to their constitution by inspection of the immediate neighbourhood of each stereogenic centres. The ligands of the other stereogenic centres (1, 3, 5 and 7) can be ranked by inspection of the atoms at the topological distance 2 from the stereogenic centres, therefore only a very small part of the graph with the periodic chains has to be inspected. Rules 2 and 3 cause no further differentiation of the ligands, therefore rule 4 has to be used. Rule 4 is applied on the original presentation of the molecule. Stereogenic centres 2 and 6 rank higher than the other stereogenic centres because of the oxygen atoms bound to them. The neighbours of each of the other stereogenic centres are three carbon atoms and one hydrogen atom. To find the stereogenic centres that are interdependent with the stereogenic centre 2, one has to permute the topologically equivalent ligands of the stereogenic centre. The permutation of the neighbours of the stereogenic centre 2 can be performed in two ways, but the number of inversions of configuration of the stereogenic centres is the smallest in the case of the automorphic permutation, which permutes the neighbours of the stereogenic centres 5 and 7, not exchanging the stereogenic centres 5 and 7 themselves. The determination of the stereogenic centres, interdependent with the stereogenic centre 6, is analogous. The result are two sets of interdependent stereogenic centres. In the next step one of the sets of interdependent stereogenic centres is chosen and the neighbours of the highest ranking stereogenic centre are ranked to get descriptor r_b. When the neighbours of the stereogenic centre 2 are ranked, the neighbours of the stereogenic centres 5 and 7 are ranked consequently and their descriptors can therefore be determined. When the descriptors are determined, the final form of the corresponding description vector can be determined. The consequence is that the topologically equivalent ligands of other stereogenic centres (1, 3 and 6) are ranked and their descriptors can be determined. The procedure could be started also with the choice of the stereogenic centre 6 instead of the stereogenic centre 2, but the name of the compound would be the same. The consequence of the choice of the stereogenic centre 6 would be also that the stereogenic centre 6 would get lower sequence number than the stereogenic centre 2.

The inositol molecule, shown in Fig. 9, is a worst case example in the sense that the manual determination of the name is quite complicated. The determination of the interdependence relation between the stereogenic centres is not as easy as it is in most cases. It is no problem to find the automorphic permutations that have to be used to determine the interdependence. The problem is that the permutation that exchanges the neighbours of the stereogenic centres 1 and 4 (see Fig. 9a) causes the inversion of configuration only to the stereogenic centres 1 and 4, while the similar automorphic permutation, which exchanges the neighbours of the stereogenic centres 3 and 6, inverts the configurations of all stereogenic centres. On the basis of the mathematical principle of transitivity we can conclude that all stereogenic centres are interdependent. The procedure for finding the stereogenic centres to which an inversion occurs due to the automorphic permutation selected is described in the methods for computer-generation of the stereoisomers[22]. To apply this method the atoms should be canonically numbered first, without consideration of rule 4. There are 6 different (although from the graph-theoretical point of view equivalent) canonical numberings. One of them is shown in Fig. 9c. In this case, there are as many possibilities as there are stereogenic centres, therefore the amount of work is about the same as when the CIP system is used. In both methods the molecule has to be analysed six times - in the former due to the fact that any of the stereogenic centres could be chosen as the highest ranked one, while in the latter due to the derivation of six hierarchical digraphs.

It is important to note that the most highly ranked stereogenic centre has to be assigned with the lowest possible sequence number in the traditional numbering method used for naming of the compounds.

In this paper we did not refer to treatment of the mesomeric systems. The possible solution of the 1966 CIP system is suitable. The difference is only that the rank of the phantom atom has to be established in each iteration of the canonical numbering method.

The descriptors proposed by the modification differ in some cases from the CIP descriptors. The descriptors of the constitutional stereogenic centres remain unchanged in most cases, except in the cases specifically mentioned above (see Fig. 11). The descriptors of the configurational stereogenic centres are small letters describing the space orientation of the ligands and the presence of the topological symmetry. That is why the new descriptors z_b and e_b are introduced for double bonds with the topologically equivalent ligands. This is the most important difference with regard to the CIP system[4]: the CIP descriptors describe the space orientation of the ligands and simultaneously the presence of the mirror plane, i. e. the presence of the geometrical symmetry. Each stereogenic centre with topologically equivalent ligands in the CIP system can have one of four possible descriptors. This fact caused serious problems in interpretation of stereochemical codes in the form of the so called parity vectors, which are the result of stereoisomer generators[20-22]. The problems arose because generated stereochemical codes are binary, while CIP descriptors are quaternary due to geometrical symmetry mentioned above. Our proposal enables an elegant solution of this problem.

The most important advantages of the method proposed are the relative ease of manual and computerised implementation of the method, which can be applied on non-modified graphs. Therefore no derivation of some special graphs for each stereogenic centre are necessary. The consequence of the new approach is the fact that the ligands of the stereogenic centres are ranked simultaneously for all stereogenic centres and not separately for each stereogenic centre. The determination of the descriptors is also simplified, since no determination of the auxiliary descriptors is necessary. The determination of the auxiliary descriptors is replaced with the determination of the sets of interdependent stereogenic centres, based on the analysis of the automorphic permutations. The latter task is generally much more easy performed, except in some rare cases like inositol. The computerised treatment of inositol is not a problem, the problem is in this case the manual treatment; nevertheless, the amount of work is about the same as when the CIP system is used.

The other important difference between the method proposed and the CIP system is that the CIP system uses the reflection variant and reflection invariant descriptors. Such differentiation enabled an easy way to find out if two stereoisomers are enantiomers or not. The method proposed does not enable such an elegant way, but the enantiomeric pairs can still be identified. Actually, grouping the stereoisomers into pairs of enantiomers is one of the tasks performed in the computerised generation of the stereoisomers[20-22]. Therefore the grouping, performed by the computers is not a problem. In general, the descriptors of the constitutional stereogenic centres are reflection variant (except in the cases where this descriptors describe the configurations of double bonds). Therefore one has only to find out whether the descriptors of the configurational stereogenic centres change due to the reflection or not, to find the enantiomer of each stereoisomer. The advantage of our approach is, on the other hand, an easy interpretation of the computer-generated parity vectors, as it was mentioned before. Another difference of minor importance is that descriptors *Z* and *E* of the CIP system are in some cases reflection invariant despite the fact that the capital letters in general represent invariant descriptors. In our approach the capital and small letter descriptors are applied consistently, since the topological equivalence of the geminal ligands of double bonds is the criterion the type of the descriptor is based on.

According to Dodziuk and Mirowicz[12] two types of stereogenic units should be applied: stereogenic centres and helical axes. Whenever it is possible the stereogenic centres should be used. The proposed formulation of rule 4 also enables ranking of the ligands with regard to conformations. For example, cyclohexane can be in chair or boat conformation. Each conformation is described by six interdependent helical axes[12]. It is a logical conclusion that two cyclohexane rings, which differ in conformation, should be ranked by comparing the two sets of interdependent helical axes. This example shows the generality of the proposed modifications of the CIP system.

V. CONCLUSION

We propose the term *constitutional stereogenic centre* for the stereogenic centres with four ligands differing in their constitution. We also propose the term *configurational stereogenic centre* for the stereogenic centres with the ligands that have the same constitution. This proposal is based on topology and is an alternative to the term *pseudoasymmetric stereogenic centre* which is based on geometry.

We also propose the definition of the sets of interdependent stereogenic units, which are topologically independent units. The configuration of each set of interdependent stereogenic units is described by the description vector and the corresponding description number. The ligands that do not differ in the constitution can be ranked by comparison of the description numbers of the sets of interdependent stereogenic units.

The method for ranking the ligands is based on the Shelley-Munk canonical numbering method[15]. The sequence rules are sequentially applied instead of graph invariants used previously by the method. For manual application, the graphs with periodic chains can be used. The comparison of the ligands in the concentric circles around the stereogenic centres is equivalent to the canonical numbering. The graphs with the periodic chains should be used for the application of rules 1 and 3, while the original molecular graphs should be used for the application of rules 2 and 4. The sequence rules are as follows:

1. *Higher atomic number precedes lower.*
2a. *An atom, member of a larger ring, precedes an atom, member of a smaller ring.*
  b. *A bridgehead atom that has the two neighbours, members of the same ring, of higher rank than the neighbours of another bridgehead atom, members of a ring of the same size are, ranks higher than the latter.*
3. *Higher atomic mass number precedes lower.*
4. *A description vector with higher description number precedes the description vector with lower description number. Description vectors with the smallest number of components are compared first, while description vectors with the highest number of components are compared last.*

The descriptors applied for the constitutional stereogenic centres are R_b and S_b, while for double bonds that can be treated as the constitutional stereogenic centres, i. e. they have no topologically equivalent geminal ligands, the descriptors used are Z_b and E_b. For the configurational stereogenic centres, the descriptors r_b and s_b are applied, while for the double bonds with the topologically equivalent geminal ligands the descriptors z_b and e_b are used.

The proposed modifications are a synthesis of the advantages of the CIP system[4] and the advantages of the method of Wipke and Dyott[9], proposed for computer applications. The method is suitable both for manual and computer use.

Acknowledgement

We thank the Ministry of Science and Technology of Slovenia for financial support. We also thank the referees for their helpful suggestions.

References

1. Cahn, R.S; Ingold, C.K. *J.Chem.Soc.*, **1951**, 612.
2. Cahn, R.S.; Ingold, C.K.; Prelog, V. *Experientia*, **1956**, 12, 81.
3. Cahn, R.S.; Ingold, C.K.; Prelog, V. *Angew.Chem.Int.Ed.Engl.*, **1966**, 5, 385.
4. Prelog, V.; Helmchen, G. *Angew.Chem.Int.Ed.Engl.*, **1982**, 21, 567.
5. Hirschmann, H.; Hanson, K.R. *J.Org.Chem.*, **1971**, 36, 3293.
6. Hirschmann, H.; Hanson, K.R. *Tetrahedron*, **1974**, 30, 3649.
7. Morgan, H.L. *J.Chem.Doc.*, **1965**, 5, 107.
8. Dubois, J.E.; Viellard, H. *Bull.Soc.Chim.*, **1968**, 913.
9. Wipke, W.T.; Dyott, T.M. *J.Am.Chem.Soc.*, **1974**, 96, 4834.
10. Meyer, E.F. *J.Comput.Chem.*, **1980**, 1, 229.
11. Mislow, K.; Siegel, J. *J.Am.Chem.Soc.*, **1984**, 106, 3319.
12. Dodziuk, H.; Mirowicz, M. *Tetrahedron Asymmetry*, **1990**, 1, 171.
13. Custer, R.H. *MATCH*, **1986**, 21, 3.
14. Mata, P.; Lobo, A.M.; Marshall, C.; Johnson, A.P. *Tetrahedron Asymmetry*, **1993**, 4, 657.
15. Shelley, C.A.; Munk, M.E. *J.Chem.Inf.Comput.Sci.*, **1977**, 17, 110.
16. Carhart, R.E. *J.Chem.Inf.Comput.Sci.*, **1978**, 18, 108.
17. Balaban, A.T.; Mekenyan, O.; Bonchev, D. *J.Comput.Chem.*, **1985**, 6, 538.
18. Balaban, A.T.; Mekenyan, O.; Bonchev, D. *J.Comput.Chem.*, **1985**, 6, 552.
19. Bohanec, S.; Perdih, M. *J.Chem.Inf.Comput.Sci.*, **1993**, 33, 719.
20. Nourse, J.G.; Carhart, R.E.; Smith, D.H.; Djerassi, C. *J.Am.Chem.Soc.*, **1979**, 101, 1216.
21. Zlatina, L.A.; Elyashberg, M.E. *MATCH*, **1992**, 27, 191.
22. Razinger, M.; Balasubramanian, K.; Perdih, M.; Munk, M.E. *J.Chem.Inf.Comput.Sci.*, **1993**, 33, 812.

(*Received in UK* 8 *February* 1994; *accepted* 11 *April* 1994)