Available online at www.sciencedirect.com
SCIENCE DIRECT®
Tetrahedron: Asymmetry
Tetrahedron: *Asymmetry* 16 (2005) 2215–2223

The Cahn, Ingold and Prelog System: eliminating ambiguity in the comparison of diastereomorphic and enantiomorphic ligands

Paulina Mata* and Ana M. Lobo

Secção de Química Orgânica Aplicada, Departamento de Química, CQFB-REQUIMTE and SINTOR-UNINOVA, Campus FCT–UNL, Quinta da Torre, 2829-516 Monte de Caparica, Portugal

Received 27 February 2005; accepted 26 May 2005

Abstract—The Cahn, Ingold and Prelog (CIP) system is an essential tool in organic chemistry nomenclature for the specification of stereogenic units. However, over the years, examples of molecules for which such specification was found impossible, ambiguous or inconsistent, have steadily accumulated. Herein, a simple methodology for the comparison of combinations of chiral units, namely to evaluate if two ligands are diastereomorphic, enantiomorphic or identical, is proposed, based on a reference descriptor not linked to a digraph node.

1. Introduction

The Cahn, Ingold and Prelog (CIP) system was originally proposed in 1951[1] for the description of the relative configuration of chiral molecules. Soon after this proposal, methods became available to determine absolute configurations,[2] an adaptation of the system to this new situation was proposed in 1956.[3] The system was soon widely adopted by chemists, and the experience acquired with its use, coupled with new developments in chemistry, were the causes for its two revisions in 1966[4] and 1982.[5] Each of these contributed to improve its logic, consistency, scope and applicability. In fact, the 1982 version allowed the specification of the great majority of the stereogenic units commonly encountered in organic molecules. Several authors have, however, reported examples of structures for which specification is impossible, ambiguous or inconsistent by using the present CIP system.[6–11]

For ease of discussion an abbreviated version of the CIP Sequence Rules[5,8] is summarized as follows:

Rule 1—Higher atomic number precedes lower.
Rule 2—Higher atomic mass precedes lower.
Rule 3—When considering double bonds and planar tetraligand atoms, *seqcis* stereogenic units precede *seqtrans* stereogenic units.[12]
Rule 4—

(**a**) Chiral stereogenic units precede pseudoasymmetric stereogenic units and these precede nonstereogenic units.
(**b**) When considering chiral stereogenic units, if 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.
(**c**) *r* precedes *s*.

Rule 5—*R* precedes *S*.

The shortcomings of **Rule 4**, formulated to compare chiral and pseudoasymmetric stereogenic units in ligands, have been noted since 1986.[6,8] In fact, the methodology proposed in this rule to compare combinations of chiral units can still lead to ambiguity in the evaluation of pairs of ligands. Herein a new simple methodology that solves such problem is proposed.

2. General methodology

The construction of hierarchical digraphs to represent stereogenic units and its ligands for specification by the CIP system was proposed in 1982[5] to unify the methodology used for cyclic and noncyclic molecules

* Corresponding author. Tel.: +351 21 2948354; fax: +351 21 2948550; e-mail: mpm@dq.fct.unl.pt

0957-4166/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tetasy.2005.05.037

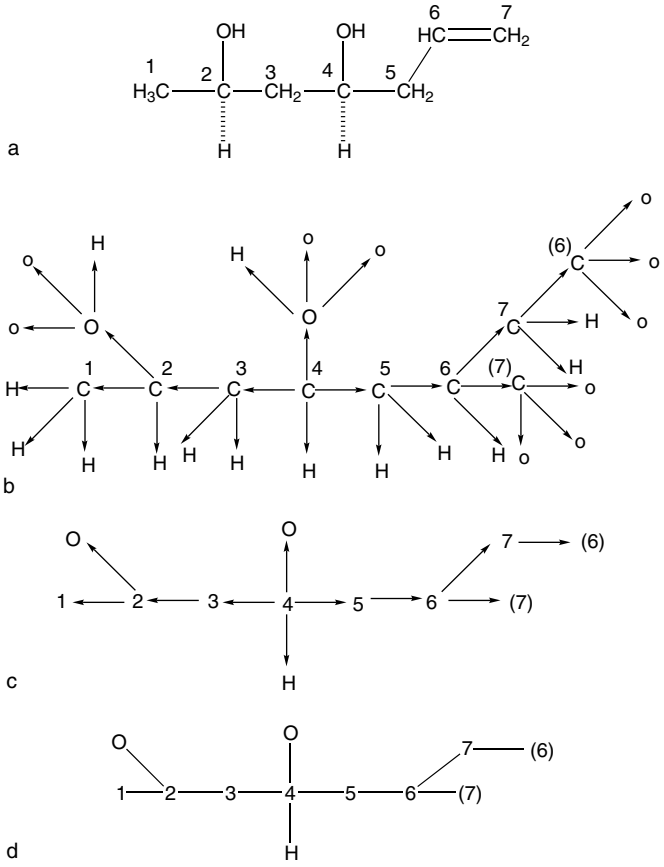

Figure 1. Digraphs of stereogenic units: (a) chemical formula; (b) complete digraph for stereogenic centre 4; (c) simplified digraph for stereogenic centre 4; (d) further simplified digraph for stereogenic centre 4. Note: Multiple bonds and rings require phantom atoms, represented by o, and duplicated atoms, represented by numbers within brackets.

and clarify the concept of a ligand. A set of conventions allow the unambiguous construction of the digraphs. Figure 1 illustrates this process in which a stereogenic unit and its ligands must be converted into a tree-graph representing the connectedness (topology) and make-up of atoms.

The nodes of the digraph represent atoms and the edges reflect their connectedness. The edges of these graphs are directed: they are considered as originating from the root of the tree-graph (the core of the stereogenic unit, a chiral centre, e.g.) and pointing towards the branch ends, thus the name of directed graph or digraph. Frequently it is not only sufficient, but also advantageous, especially when large molecules are considered, to use simplified digraphs in which the nonrelevant information is omitted.

The nodes of the digraph are hierarchically ranked during the comparison process. This hierarchy depends on the distance from the core of the stereogenic unit and the application of the Sequence Rules. The hierarchy is continually refined, although never inverted, as analysis proceeds. In each sphere of an hierarchical digraph (set of nodes at the same distance from the root) the hierarchy of the nodes is reflected in their position. The convention used is nodes on top precede, or are hierarchically equivalent, to the nodes closer to the bottom.

The Sequence Rules are sequentially and exhaustively applied to the whole digraph until the first difference is encountered. After comparison of the constitutional properties of ligands (**Rules 1** and **2**) and comparison of their double bonds (**Rule 3**), the next step is the comparison of chiral and pseudoasymmetric units (**Rule 4**). This operation will evaluate if ligands are diastereomorphic, enantiomorphic or identical and will permit the ranking of diastereomorphic ligands. This, among other steps, requires a comparison of the interrelations between the descriptors of the chiral units (*R* and *S*). If needs be, **Rule 5** is next applied. For this purpose simplified hierarchical digraphs are used. Figure 2 illustrates how these are obtained.

3. Application of Rule 4-b and its shortcomings

The methodology for the comparison of the combination of chiral units in ligands (**Rule 4-b**) uses pairs of descriptors described as *like* (*l*) or *unlike* (*u*). For this purpose it defines which pairs of descriptors should be considered relevant and then proceeds to classify them as *like* and *unlike*. Note that *like* (*l*) pairs are, for example, *RR* and *SS* and *unlike* (*u*) pairs are *RS* and *SR*. In the following paragraphs only stereogenic centres and their descriptors (*R* or *S*) will be considered.

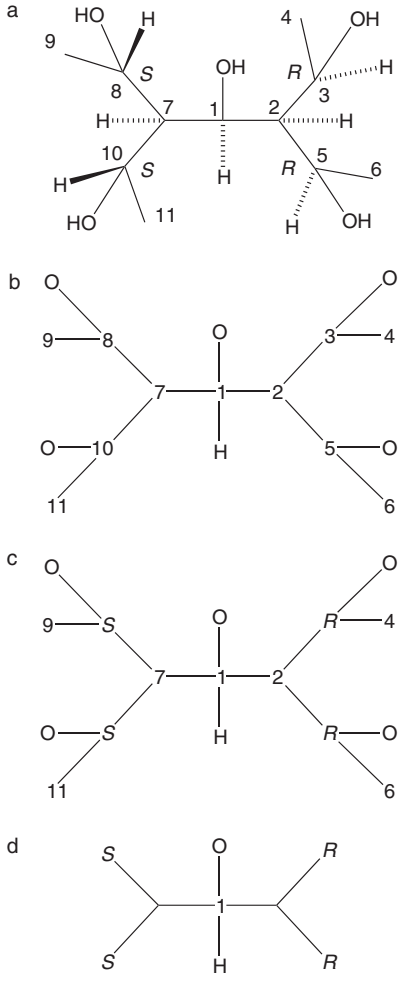

Figure 2. Molecule with constitutionally identical ligands. The process to obtain simplified digraphs for comparison purposes according to **Rules 4** and **5** is illustrated. (a) chemical formula; (b) simplified digraph for stereogenic centre 1; (c) simplified digraph for stereogenic centre 1 with information about descriptors of chiral centres in ligands; (d) further simplified digraph with only the relevant information for analysis by **Rules 4** and **5**.

So the first step in this procedure is the critical choice of the *first descriptor* for each ligand. This descriptor is then paired with the remaining descriptors respecting their connectivity and hierarchy in the digraph. This is followed by comparison of the corresponding pairs in both ligands.

The *first descriptor* ***R*** or ***S*** (depicted in italic bold) of the ligands represented in the digraphs in Figures 3–5 is as follows: (a) the one associated with the highest ranked node corresponding to a chiral centre in the ligand (Fig. 3); (b) the one that occurs the most in the set of equivalent highest ranked nodes (Fig. 4); (c) sequentially both ***R*** and ***S*** if these occur in the same number in the set of equivalent highest ranked nodes (Fig. 5).

As illustrated in Figure 3, the descriptor established as *first descriptor* is then paired with the remaining descriptors in the sequence of their hierarchical order, represented by the numbers close to the nodes in digraph **I**, and the descriptor pairs thus formed are evaluated as *like* or *unlike*. For example, for ligand A of **I** the first pair evaluated is 1,3, that is (***S***,*S*), thus a *like* pair. The first difference encountered, while comparing pairs 1,6, determines the hierarchy of the ligands, that is, B > A.

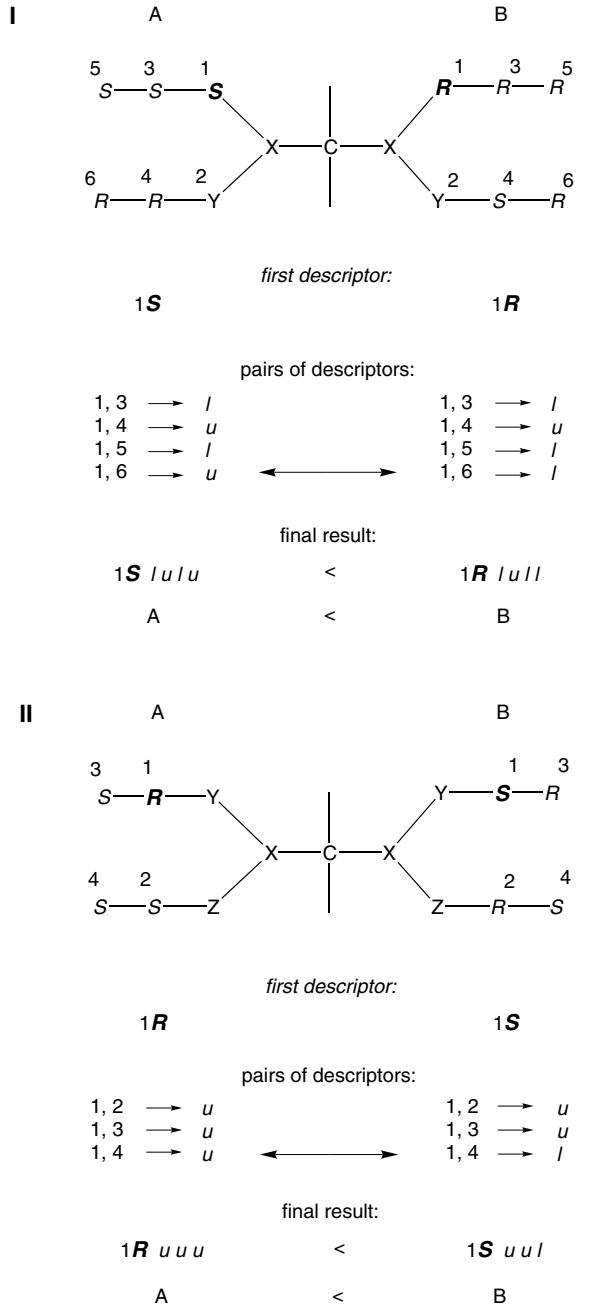

Figure 3. CIP system methodology for formation and comparison of pairs of descriptors. The *first descriptor* is the one associated with the highest ranked node corresponding to a stereogenic centre in the ligand. Note: X, Y and Z can be any atom giving rise to a nonstereogenic centre and allowing the hierarchy of the nodes represented. In **I**, for example, X = N and Y = C and in **II**, X = N, Y = O and Z = C. The numbers in the digraph reflect the hierarchy of the nodes and pairs. The double arrow (↔) highlights the first difference encountered in the comparison.

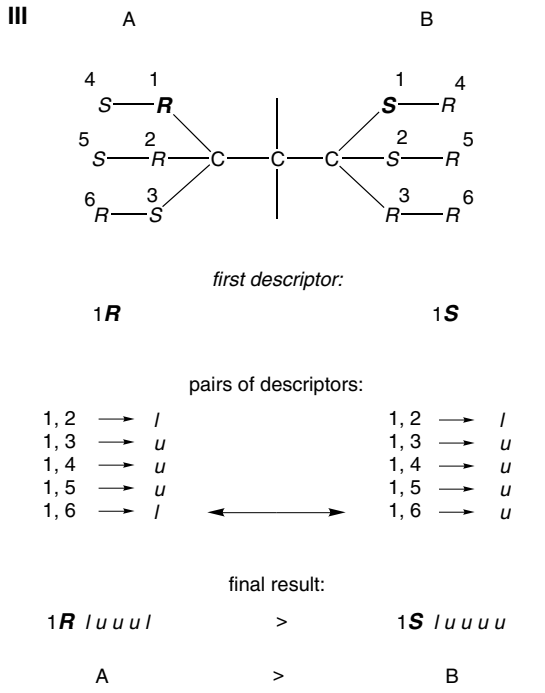

Figure 4. CIP system methodology for the formation and comparison of pairs of descriptors. The *first descriptor* is the one that occurs the most in the set of equivalent highest ranked nodes.

If both descriptors, ***R*** and ***S***, are used as *first descriptors* (see Fig. 5), they should be independently and sequentially used to form pairs of descriptors. Then all pairs situated at the same rank level are compared. The first difference encountered is used to rank ligands. This means that in the examples in this figure, pairs 1,2 and 2,1 (the order is not important) should be simultaneously compared in both ligands and the number of *l* and *u* pairs evaluated. Pairs 1,3 and 2,4 and pairs 1,4 and 2,3 are then compared. If only one of the descriptors, for example, ***R***, was used as *first descriptor* in the analysis of digraph **V**, the ligands would be evaluated as geometrically different ($ull > uuu$) *when in fact they are enantiomorphic ligands*. When both descriptors in **V** are used as *first descriptor* the ligands are perceived as identical according to **Rule 4-b** ($uulu lu = uulul$). In fact, being *enantiomorphic ligands*, they can only be ranked by **Rule 5**.

In the CIP procedure, the hierarchical rank of the descriptor pairs is given by the rank of the second descriptor in the pair. Custer[6] considers that this does not allow the ordering of two pairs of descriptors in which the second descriptors have the same priority. He thus proposed a rule, based on the relationship in the digraph between the nodes in the pair, to overcome this incompleteness of the methodology. This rule states that in such cases, the higher ranking pair is the one with the lower ranking of the least common ancestor in the digraph.[6,†]

[†] Being the ancestor of *x*, any node belonging to the chain of nodes leading from the root to *x*. The least common-ancestor of both *x* and *y* is the ancestor of both *x* and *y*, which has the smallest distance to *x* of all ancestors of *x* and *y*.[6]

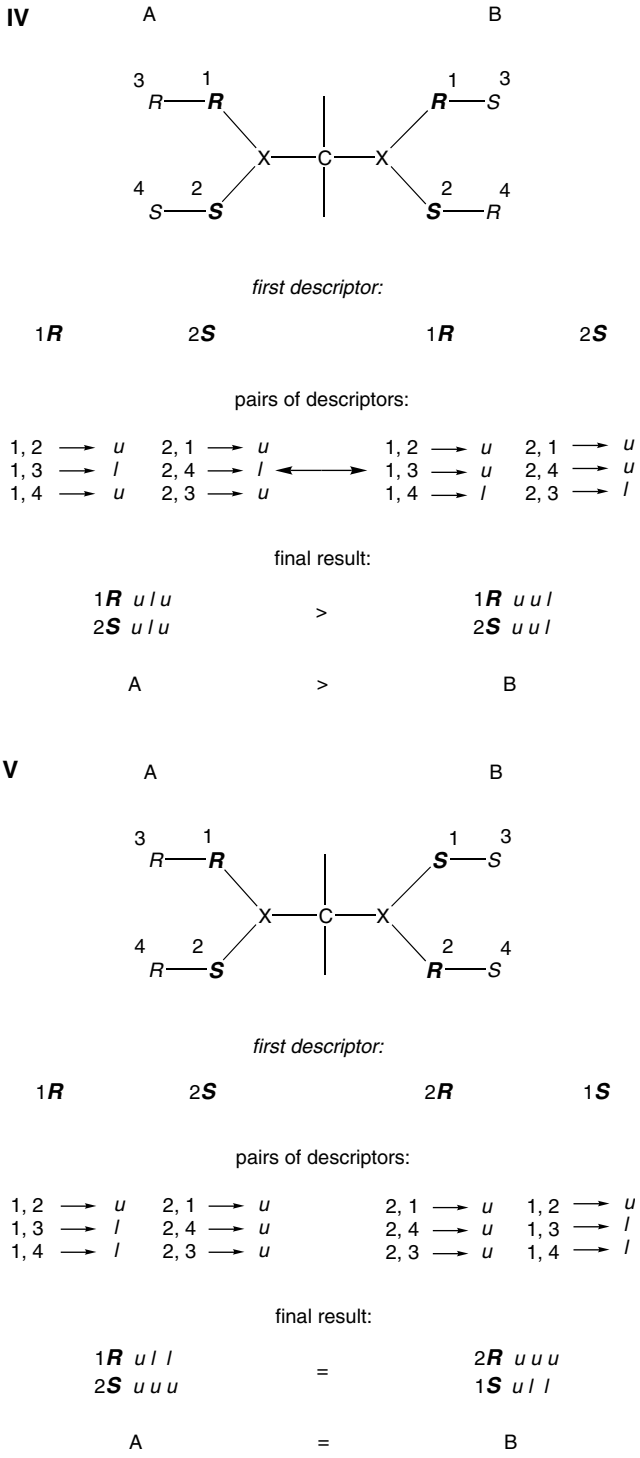

Figure 5. CIP system methodology for formation and comparison of pairs of descriptors. The *first descriptor* is sequentially both *R* and *S* as these occur in the same number in the set of equivalent highest ranked nodes.

This rule deals effectively with molecules, such as those represented by digraphs in Figure 5. In such cases, both descriptors should be sequentially considered as *first descriptors*. However, when in **IV** and **V**, 1***R*** is considered the *first descriptor* in ligand A, pair 1,3 cannot be hierarchically equivalent to pair 1,4. Custer's rule allows the

differentiation of these pairs (in these cases 1,3 > 1,4) and correctly ranks the ligands.

However, this same rule may introduce ambiguity in other situations, such as in Figure 6. If nodes 1, 2 and 3 are hierarchically equivalent when comparison according to **Rule 4** is initiated, both hierarchical digraphs, **VIa** and **VIb** in Figure 6, can be valid for the same molecule. The *first descriptor* in ligand A is ***R*** because it is the descriptor that occurs the most. When ***R*** is associated with node 1, if the digraph considered is **VIa** and Custer's rule used, then pair 1,4 would have precedence over pair 1,5. Thus, ligand B would be considered as having precedence over ligand A, or $A < B$. If alternatively a comparison is made on the basis of digraph **VIb**, then both ligands would be perceived as equivalent according to such rule or A = B.

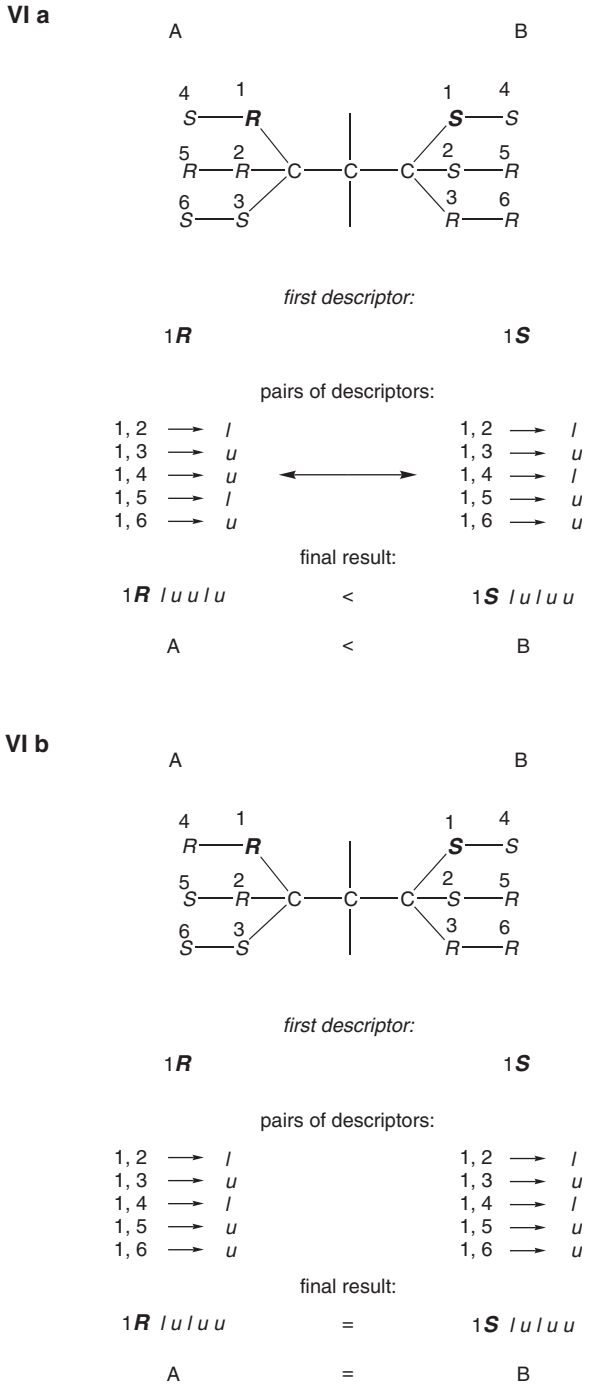

Figure 6. Example of a molecule whose specification, using Custer's rule for ranking pairs of descriptors,[6] is ambiguous.

It can thus be concluded by the examples considered above that the methodology proposed in the 1982 revision of the CIP system is unsatisfactory and that the rule proposed by Custer still introduces ambiguity or inconsistency in some situations.

The 1982 CIP methodology[5] also makes difficult the reordering of the hierarchical digraph, a procedure which can be required during comparison, and an aspect which was never considered by its authors.

4. A simple methodology for unambiguous ranking of ligands

To overcome the problems encountered, a new methodology is proposed next to form and rank pairs of descriptors. For this purpose the concept of a *first descriptor* needs to be enlarged to that of a descriptor called here *reference descriptor*, not linked with any node in the digraph. However, its choice, ***R*** or ***S***, depends on the descriptors of the highest ranked nodes.

(1) The *reference descriptor*, is chosen as proposed for the *first descriptor*,[5] and previously described. This *reference descriptor*, however, *is not associated with any particular node in the digraph*.[‡]
(2) The *reference descriptor* in each ligand is then paired with ***all*** the descriptors associated with nodes corresponding to chiral units in the ligand, respecting their connectivity and hierarchy. Corresponding pairs in both ligands are then compared until the first difference is encountered.
(3) The digraph is reordered, whenever required, to reflect modifications introduced in the hierarchy of the nodes during the analysis process.

An application of this methodology is illustrated in Figure 7. For digraph **I**, previously analyzed (see Fig. 3), all nodes, and consequently all pairs of descriptors, are successively ranked as indicated in the figure. Comparison of pairs of descriptors does not modify their previous hierarchy, and ligands are ranked by the comparison of the lowest ranked pair of ligands, that is, comparison in A between the *reference descriptor* ***S*** and the descriptor *R* of node 6, to give an *unlike* (*u*), and similarly for B to produce a *like* (*l*). Hence, B takes priority to A: B > A.

In the case of digraph **III**, already mentioned (see Fig. 4), nodes 1, 2 and 3 are all equally ranked when analysis

[‡] Note that in their examples Prelog and Helmchen always associate the *first descriptor* with one node of the digraph.

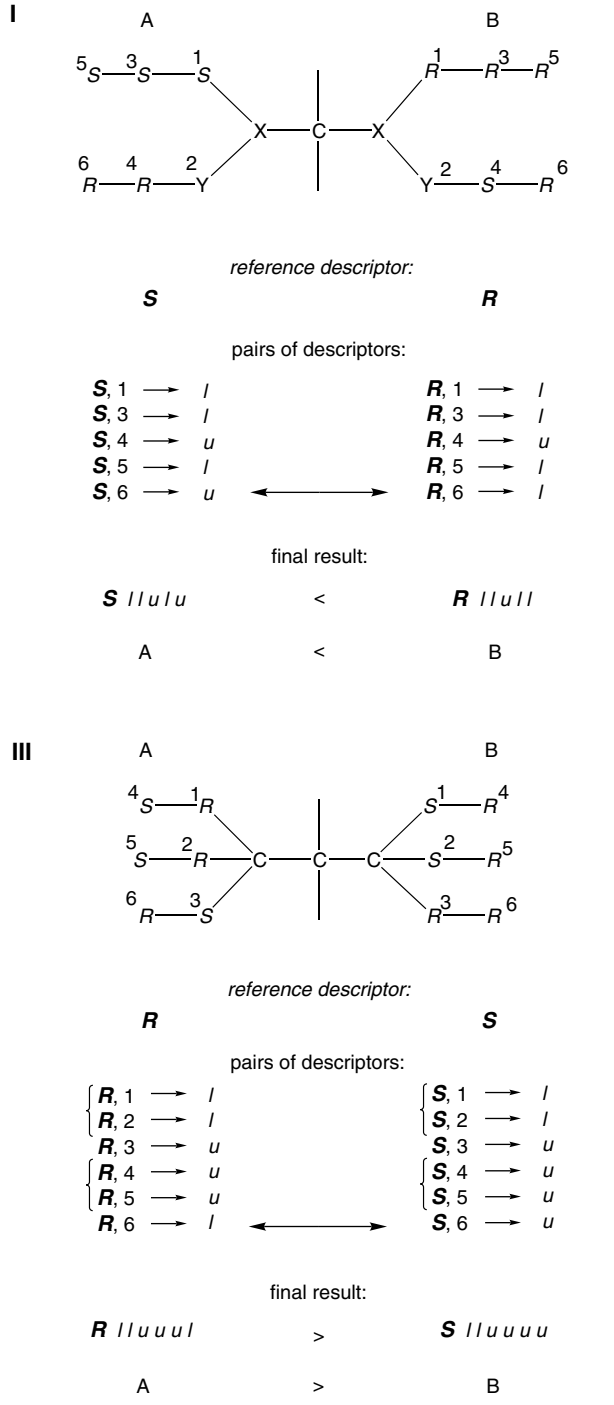

Figure 7. Application of the *reference descriptor* methodology, proposed for the formation and comparison of pairs of descriptors. Note: The sequence of the pairs of descriptors shown reflects its hierarchical rank at the end of the comparison process. Equally ranked pairs are grouped by braces.

starts. However, this situation changes upon comparison of the pairs of descriptors. For example, for ligand A, nodes 1 and 2 form *like* pairs with the *reference descriptor* ***R*** and node 3 forms an *unlike* pair. This requires a redefinition of the ranking of these nodes. Thus, the equivalent nodes 1 and 2 take priority over 3. This difference of hierarchy is reflected in the next sphere of the digraph: 4 and 5 are hierarchically equivalent to each other and take priority over 6.

In digraph **VIb** of Figure 8, previously analyzed (see Fig. 6), before starting comparison according to **Rule 4-b**, all nodes in each sphere of the digraph are equally ranked (1 = 2 = 3 > 4 = 5 = 6). Thus, six different hierarchical digraphs are valid to represent each ligand (see lower part of Fig. 8 for ligand A). When analysis of the second sphere of the digraph (nodes 1, 2 and 3) ends, the ranking of the nodes (and pairs of descriptors) is the modified (1 = 2 > 3 > 4 = 5 > 6), and only two hierarchical digraphs can represent each ligand (see lower part of Fig. 8). However, after analysis of the third sphere, node 4 takes priority over 5, taking this priority to 6. Thus, all pairs of descriptors are successively ranked (1 > 2 > 3 > 4 > 5 > 6), as indicated in Figure 8, and only one hierarchical digraph can represent the ligand. Ligands are perceived as identical according to **Rule 4-b**. **Rule 5** is required to finish the ranking process.

Using this methodology, digraph **VIa** in Figure 6 does not represent the hierarchy of nodes after concluding the analysis of the third sphere. Thus, it cannot be used to rank these ligands and therefore they are unambiguously perceived as equivalent.

The comparison in Figure 9 of the ligands of digraphs **IV** and **V** (previously in Fig. 5) by **Rule 4-b** requires the use of both descriptors as *reference descriptors*. While analysing using each one of the descriptors as reference, the hierarchical digraph should be reordered accordingly, thus the pairs are all sequentially ranked as indicated in the figure.

Analysis of ligands proceeds by comparison of hierarchically equivalent sets of pairs formed with both *reference descriptors* (in the same row in Fig. 9) in each ligand. In **IV** the ligands are perceived as different and ranked by the highlighted difference. In **V** the ligands are perceived as identical according to this rule and comparisons should proceed using **Rule 5**. The proposed methodology allows a correct and unambiguous analysis of the ligands, which are enantiomorphic.

It should be noted that when each one of the descriptors is used as reference, the hierarchy of the nodes is necessarily different, leading to two different hierarchical digraphs (Fig. 10). In (a) nodes 1, 2, 3 and 4 are sequentially ranked and, similarly in (b), the sequence of nodes is 2, 1, 4, 3. These reordered hierarchical digraphs are temporary. When the analysis finishes if the ligands are perceived as different and ranked accordingly, the hierarchical digraphs will not be further used. However, if the ligands are not yet ranked, the hierarchical digraph, valid before starting the comparison by **Rule 4-b**, in which nodes 1 and 2 are equally ranked and have priority over the equally ranked nodes 3 and 4, is resumed.

When comparison of ligands requires the use of both descriptors, as *reference descriptors*, a special situation

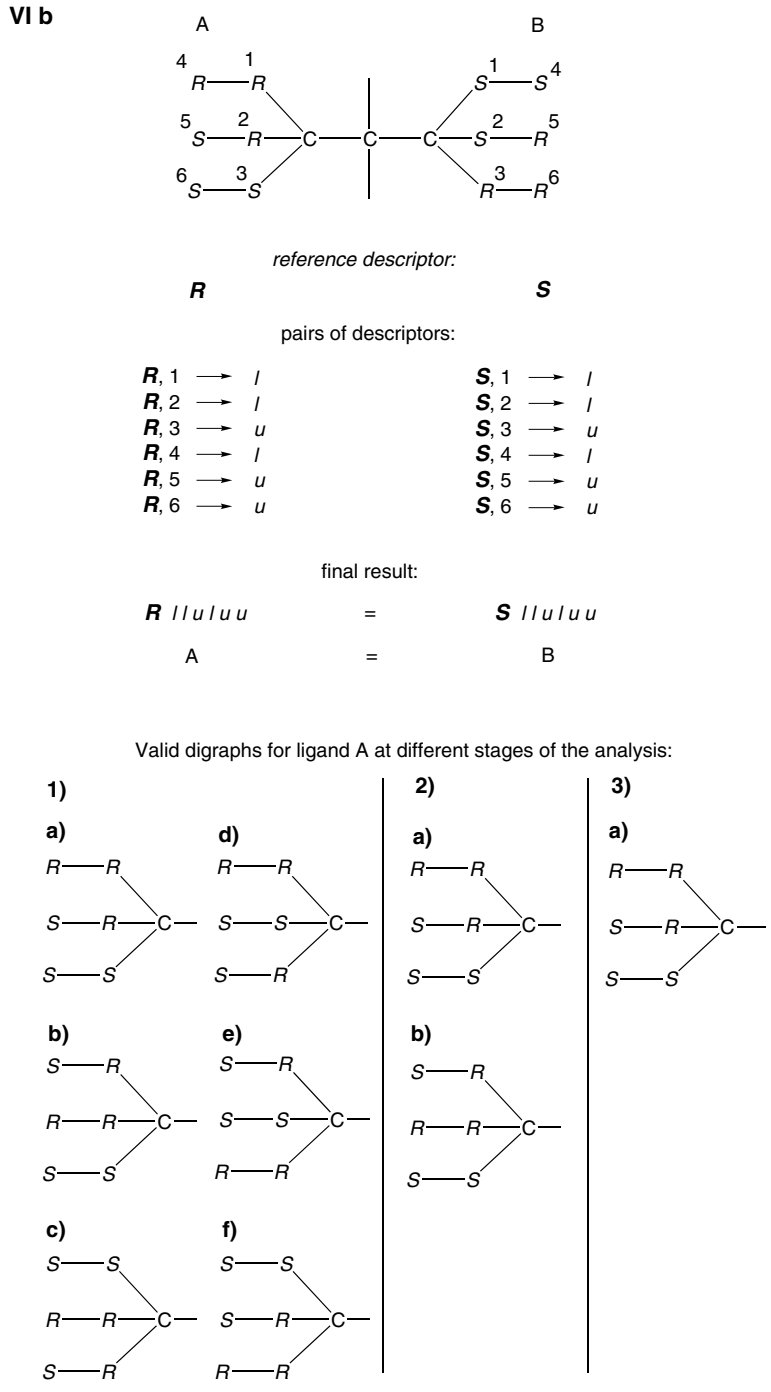

Figure 8. Application of the *reference descriptor* methodology and illustration of the process of the reordering of the digraph. Valid digraphs of ligand A: (1) before starting comparison by **Rule 4-b**; (2) after ending analysis of pairs containing descriptors of nodes 1, 2 and 3; (3) in the end of the analysis process.

can arise, which was not previously considered. This is illustrated in Figure 11.

In ligand A in **VII** there are two different descriptors (*R* and *S*) associated with hierarchically equivalent nodes and both should be considered as *reference descriptors*. However in ligand B there is only one descriptor (*R*) and thus only one *reference descriptor* should be considered. Consequently, in ligand A four pairs of descriptors must be considered for comparison, while in ligand B only two pairs can be formed. Comparison of a different number of pairs of descriptors can be a problem. In this case, no further analysis is required because the first difference is encountered, which is the *number of reference descriptors* in each ligand, and this is enough to rank them. For such situations, it is proposed that ligands with only one *reference descriptor* (***R*** or ***S***) should have priority over ligands with the two *reference descriptors* (***R*** and ***S***).

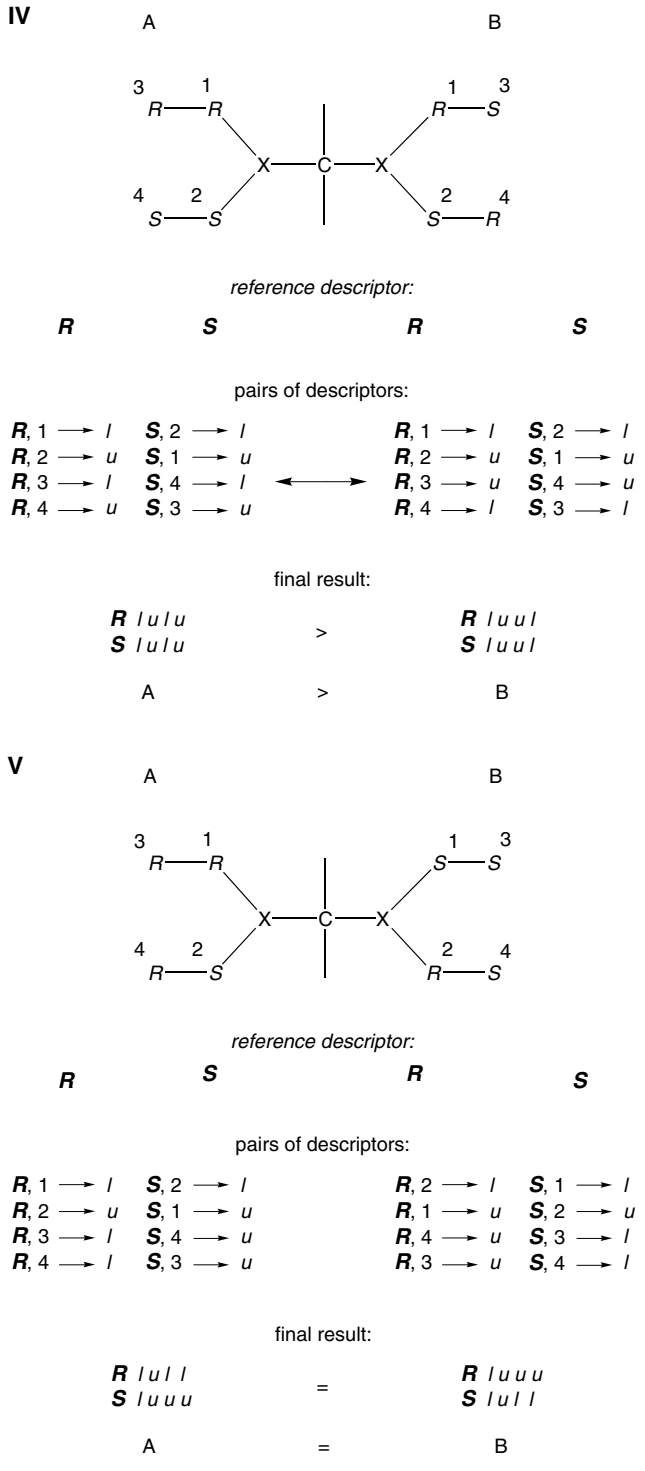

Figure 9. Hierarchical digraphs having ligands whose comparison requires both descriptors to be considered as *reference descriptors*.

The new proposal for **Rule 4-b** is as follows:

A *reference descriptor* (*not associated with any node in the digraph*) is chosen in each ligand and is: (a) the one associated with the highest ranked node corresponding to a chiral unit in the ligand; (b) the one that occurs the most in the set of equivalent highest ranked nodes; (c) sequentially both descriptors (e.g. ***R*** and ***S***), if these occur in the same number in the set of equivalent highest ranked nodes.

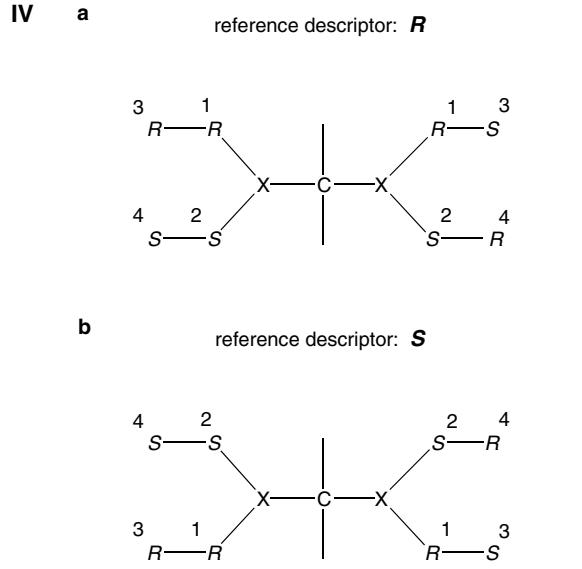

Figure 10. Hierarchical digraphs for molecules represented by digraph **IV**, when each one of the descriptors is used as *reference descriptor*.

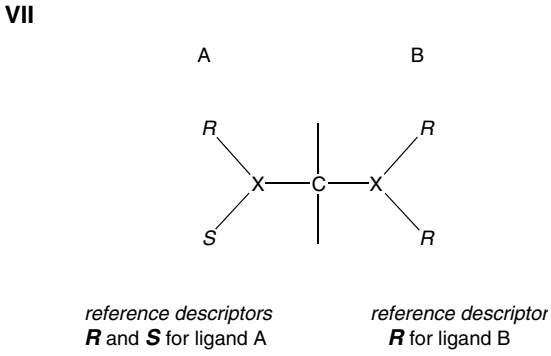

Figure 11. Digraph of a molecule having ligands for which different sets of *reference descriptors* are used.

(i) If the number of *reference descriptors* is different in both ligands then the ligand with *one reference descriptor* has priority over the one with *two reference descriptors*.

(ii) If both ligands have the same number of *reference descriptors*, then the *reference descriptor* is paired with each one of the descriptors associated with nodes corresponding to chiral units, respecting their connectivity and hierarchy in the digraph. Then the one with the first-chosen *like* descriptor-pair has priority over the one with a corresponding *unlike* descriptor-pair. (Note that *like* descriptor-pairs are, e.g., ***RR***, ***SS*** and *unlike* pairs ***RS***, ***SR***).

5. Conclusions

Herein, the CIP methodology for comparing pairs of chiral descriptors is discussed, with examples of situations where ambiguity emerges reported and a proposal to overcome such shortcomings is presented. Its main aspect is related with the introduction of a *reference*

descriptor, which is not directly associated with any node of the digraph, and should be paired with all relevant descriptors for ligand comparison purposes, according to Sequence **Rule 4-b**. The use of the comparison of the *number of reference descriptors* in both ligands to rank them, in some special cases, not considered by the previous rules is also proposed. This methodology is simple, solves the ambiguity encountered and also simplifies the required reordering of the hierarchical digraph.

References

1. Cahn, R. S.; Ingold, C. K. *J. Chem. Soc.* **1951**, 612–622.
2. Bijvoet, J. M.; Peerdeman, A. F.; van Bommel, A. J. *Nature* **1951**, *168*, 271–272.
3. Cahn, R. S.; Ingold, C. K.; Prelog, V. *Experientia* **1956**, *12*, 81–124.
4. Cahn, R. S.; Ingold, C. K.; Prelog, V. *Angew. Chem., Int. Ed. Engl.* **1966**, *5*, 385–415.
5. Prelog, V.; Helmchen, G. *Angew. Chem., Int. Ed. Engl.* **1982**, *21*, 567–583.
6. Custer, R. H. *Match* **1986**, *21*, 3–31.
7. Hirschmann, H.; Hanson, K. R. *Tetrahedron* **1974**, *30*, 3649–3656.
8. Mata, P.; Lobo, A. M.; Marshall, C.; Johnson, A. P. *Tetrahedron: Asymmetry* **1993**, *4*, 657–668.
9. Mata, P.; Nachbar, R. *Tetrahedron: Asymmetry* **1995**, *6*, 693–696.
10. Mata, P.; Lobo, A. M.; Marshall, C.; Johnson, A. P. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 491–504.
11. Nicolaou, K. C.; Boddy, C. N. C.; Siegel, J. S. *Angew. Chem., Int. Ed.* **2001**, *4*, 701–704.
12. (a) Stereogenic unit—unit composed of an achiral core (atom or skeleton of atoms) with distinguishable ligands, whose constitution-preserving interchange leads to a stereomorphic figure; (b) *seqcis* = *Z* and *seqtrans* = *E* in double bond specification, however the descriptors *seqcis* and *seqtrans* are preferred as they satisfy the convention that reflection variant descriptors are marked by capital letters and reflection invariant descriptors by lowercase letters; (c) Pseudoasymmetric centre—stereogenic centre having two ligands, which are enantiomorphic.