Implementation of the Cahn-Ingold-Prelog System for Stereochemical Perception in the LHASA Program

Paulina Mata* and Ana M. Lobo

Departamento de Química and SINTOR-UNINOVA, *Campus* of Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Quinta da Torre, 2825 Monte da Caparica, Portugal

Chris Marshall and A. Peter Johnson

School of Chemistry, University of Leeds, Leeds LS2 9JT, U.K.

Received August 2, 1993®

The CIP system for identification and specification of stereogenic units has been implemented in the LHASA program. This new implementation of the CIP system, according to its 1982 revision, is described. The main improvements introduced are an extension of the capabilities for handling the representations of stereobonds, for the identification of stereogenic centers (chiral and pseudoasymmetric), prochiral centers, and double bonds, and mainly in the implementation of the CIP sequence rules for comparison of the ligands.

1. INTRODUCTION

Stereochemistry has a fundamental role in modern organic chemistry, particularly in the synthesis of drugs and natural products. For a computer program designed to assist organic synthesis planning to be useful, generating sophisticated synthetic schemes, it is important that is can recognize different stereoisomers, has in its database a range of stereoselective transforms, and has a deeply implemented stereochemical strategy. This requires the capability to deal in an efficient way with the stereochemistry of a wide variety of molecules, including the identification and specification of the most relevant stereogenic units in the target molecule and each generated precursor.

Due to the complexity of this process, most of the programs designed to assist the synthetic chemist do not deal well with stereochemistry. The more sophisticated, which can deal with it, usually do not use the same conventions for the identification and specification of stereogenic units[1-3] as do chemists (CIP (Cahn-Ingold-Prelog) system). They use other conventions considered efficient and more suitable for computer use. Apart from the speed, another reason for this is the fact that some conventions used by the CIP system[4] suffer from a lack of precision, which makes them difficult to implement for computer use. An attempt to overcome this problem was made in the 1982 revision of the CIP system,[5] which constitutes a very important improvement and makes its implementation more straightforward.

When conventions other than the CIP conventions are used to deal with stereochemistry, the specification of the stereogenic units obtained has no correspondence with the CIP one which is used by the chemists in their daily life. Thus it cannot be used easily in the interface with the chemists. The implementation of the CIP system is of major interest, at least for communication with the chemist, to ensure that the configuration of the stereogenic units has been correctly and unambiguously defined.

The CIP system is implemented in several programs, namely, CHIRON[6] and QUANTA;[7] however, the implementations achieved are not exhaustive. The complete implementation of the system requires a systematic method for the comparison of the ligands attached to the potential stereogenic units, which takes into consideration all the relevant properties according to the CIP sequence rules. In the programs mentioned just some of the rules are considered, and even these are implemented in an incomplete way, leading in several cases to erroneous identification and assignment of the stereochemistry. Such was also the case of the LHASA[8] (logic and heuristics applied to synthetic analysis) program, and a new implementation of the CIP system, according to its 1982 revision,[5] was made.

2. CIP SYSTEM

2.1. Introduction. The CIP system is a set of conventions through which the description of the absolute configuration of molecules containing stereogenic units can be done compactly enough to allow its inclusion in the name of the compound. This system has been widely accepted by chemists due to its compactness and applicability to most of the molecules which they need to describe.

The basis of the system is provided by the conventions proposed in the first publication, in 1951,[9] and several modifications and extensions have been made since then to achieve generality.[4,5,10,11] The 1982 revision[5] resulted in a new version of the CIP system, whose main success was the introduction of the concept of hierarchical digraphs to represent the ligands and the methodology developed for the analysis of cyclic molecules, which constitute very important improvements to the system and make its implementation for computer use much easier.

In spite of its advances, deficiencies have already been identified in some aspects of this revised version.[12,13] In the course of the implementation of the CIP system in the LHASA program some limitations of the applicability, consistency with the theory, and generality of the 1982 sequence rules for comparison and ranking of the ligands of each stereogenic unit were also encountered by us.[14,15] These are discussed in another paper[15] along with proposals for extensions or modifications to the CIP sequence rules to overcome encountered deficiencies. In the implementation described here we have tried to respect the rules, as they appear in the 1982 revision,[5] as much as possible. However, some modifications were required.

* To whom correspondence should be addressed.
® Abstract published in *Advance ACS Abstracts,* March 1, 1994.

0095-2338/94/1634-0491$04.50/0

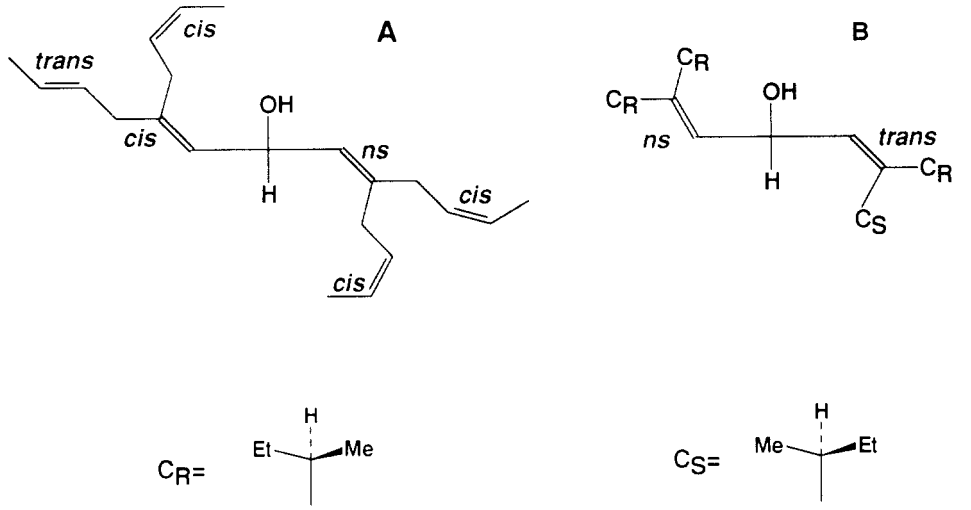


Figure 1. Stereogenic units whose ligands cannot be ordered by the comparison of the highest ranked double bonds that they contain in spite of the differences between them. ns = nonstereogenic.

2.2. CIP Sequence Rules. 2.2.1. Generalities. The procedure to derive a CIP descriptor can be summarized in the following three steps: (i) factorization of the stereomodel assigned to the molecule, into stereogenic units; (ii) determination of the ranking of the ligands around each stereogenic unit; (iii) determination of a descriptor for each stereogenic unit.

The ordering of the ligands (step ii) has a fundamental role in the procedure and can be considered the most complex step. Although simple molecules are easily treated, for the more complex cases difficulties arise and the general case is extremely intricate.

In the 1982 revision of the CIP system[5] the meaning of the term "ligand" was clarified, even in complex cyclic molecules, by the introduction of hierarchical digraphs. These are equivalent acyclic structures into which monodentate, polydentate, and cyclic ligands must be converted for analysis and comparison.

For the determination of the rank of the ligands around a center, these are represented by the hierarchical digraphs and their properties are compared. The CIP sequence rules are a set of conventions defining the hierarchy and the methodology for this comparison.

Sequence Rules (1982)[5]

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*-position and the other in a *trans*-position to the core of the 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 or double bonds.)
4. When two ligands have different descriptor pairs, then the one with the first-chosen *like* descriptor pair has priority over one with a corresponding *unlike* descriptor pair. *Like* descriptor pairs are *RR*, *SS*, *RRe*, *SSi*, *ReRe*, *SiSi*, *MM*, *PP*, *RM*, *SP*, *ReM*, and *SiP*. *Unlike* descriptor pairs are *RS*, *ReSi*, *SRe*, *RSi*, *MP*, *PR*, *SM*, *ReP*, and *SiM*.

Methodology for Pairing Descriptors: For each ligand the descriptor chosen as first (highest ranked descriptor) is paired with all the remaining descriptors, and the hierarchical rank of the descriptor pair is given by the rank of the second descriptor in the pair.

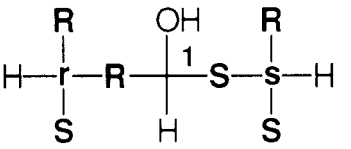


Figure 2. Hierarchical digraph corresponding to a molecule in which the specification of center 1 can be ambiguous using the 1982 version of the sequence rules.[5] How is center 1 specified? Using subrule *R* > *S* or subrule *r* > *s* (the correct way)?

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

5b. A ligand with descriptor *r* has preference over one with descriptor *s*.

A rule was proposed[5] whose rank was not defined: "Chiral stereogenic units precede pseudoasymmetric stereogenic units, and these precede nonstereogenic units."

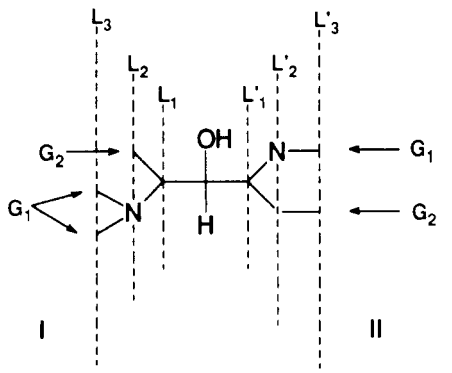
Rule 2 was not considered, as LHASA does not yet use isotopes, and some modifications to rules 3, 4, and 5 were introduced in the present implementation.

2.2.2. Rule 3. It is possible to conceive molecules, such as the ones in Figure 1, in which a double bond that is stereogenic (presenting *cis–trans* isomerism) should be compared with a corresponding nonstereogenic double bond. These cases are not considered by the 1982 version of the CIP rules, and the ordering of the ligands has to be made using other differences. However, it was considered (similarly to the case of tetrahedral stereogenic centers) that stereogenic units have preference over nonstereogenic units, so rule 3 was implemented in the following form: *cis* > *trans* > nonstereogenic

2.2.3. Rules 4 and 5. In the 1982 revision of the CIP system[5] a rule was introduced concerned with the ranking of chiral, pseudoasymmetric, and prochiral units whose hierarchy was not defined. As with other authors,[13] it was considered[15] that this should constitute rule 4a.

It was considered by ourselves[15] and other authors[13] that subrule "*r* precedes *s*" should constitute rule 4c, as it was in the 1966 version of the rules,[4] and not rule 5b.[5] The introduction of this subrule in rule 5 introduces ambiguity in the specification of stereogenic units (Figure 2), and such a modification is not necessary.

Pairs of descriptors containing the descriptors *Re* and *Si* were not included in the present implementation, as it is considered that they are dispensable for the comparison of



LHASA ranking of atoms:
Old implementation

Ligand I			Ligand II		
Level	Group	Atoms	Level	Group	Atoms
1	1	1C	1	1	1C
2	1	1N	2	1	1N
2	2	1C	2	2	1C
3	1	2C	3	1	1C
			3	2	1C

Figure 3. Ranking of atoms as made by LHASA (old implementation of the CIP rules). L = level (distance from the core of the stereogenic unit); G = hierarchical group.

ligands.[14,15] Also the pairs of descriptors including *P* and *M* descriptors were not included, as these should be assigned to stereogenic units that are not recognized by the LHASA stereochemical perception module. For the same reason in rule 5 these descriptors are not compared.

The methodology proposed by Prelog and Helmchen[5] to rank the pairs of descriptors for comparison according to rule 4 does not allow the ordering of two pairs of descriptors in which the first descriptors have the same priority as each other and likewise for the second descriptors. A methodology for the ranking of these cases was proposed[13] using the relationship between the nodes in the digraph. This is included in our implementation of the CIP sequence rules.

Sequence Rules as Implemented in LHASA:

1. Higher atomic number precedes lower.

3. When two ligands differ only in that one has an atom or atom-group of higher rank in a *cis*-position and the other in a *trans*-position to the core of the stereogenic unit, then preference is given to the former. *Cis* and *trans* double bonds rank higher than nonstereogenic double bonds (*cis* > *trans* > nonstereogenic). (This rule is restricted to ligands which differ in *cis–trans* isomerism of planar tetraligant atoms or double bonds.)

4a. Chiral stereogenic units precede pseudoasymmetric stereogenic units, and these precede nonstereogenic units.

4b. When two ligands have different descriptor pairs, then the one with the first-chosen *like* descriptor pair has priority over one with a corresponding *unlike* descriptor pair. *Like* descriptor pairs are *RR*, *SS*. *Unlike* pairs are *RS*, *SR*.

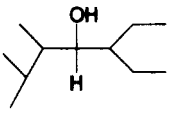
Methodology for Pairing Descriptors: For each ligand the descriptor chosen as first (highest ranked descriptor) is paired with all the remaining descriptors. The following characteristics determine the hierarchical rank of the pairs of descriptors: (i) higher rank of the second descriptor in the pair; (ii) lower rank of the least common ancestor in the graph.

4c. A ligand with descriptor *r* has preference over one with descriptor *s*.

5. A ligand with descriptor *R* has priority over its enantiomorph with descriptor *S*.

3. OLD IMPLEMENTATION

There has been an implementation of the CIP system in LHASA since 1971; however, this implementation was quite limited, particularly in ligand comparison. The sequence rules were implemented in their 1966 version,[4] which is difficult to implement in a computer program. Also, just rules 1, 3, and 5 were selected for implementation.



LHASA ranking of atoms:
Old implementation

Ligand I			Ligand II		
Level	Group	Atoms	Level	Group	Atoms
1	1	1C	1	1	1C
2	1	2C	2	1	2C
3	1	2C	3	1	2C

I II

LHASA comparison: I = II

CIP comparison: I > II

Figure 4. Constitutionally different ligands that could not be differentiated by the old implementation of the CIP rules in LHASA.

Rules 1 and 3 were implemented in a simplified way that allowed LHASA to deal only with very simple cases, which although the most frequent, are not considered sufficient in a program with the degree of sophistication that LHASA has attained.

Without the implementation of rule 4, rule 5 could only be used to identify some stereogenic units, but not to classify and specify them in a correct way.

Rule 2 was not implemented. This was not a problem as LHASA does not at present deal with isotopes.

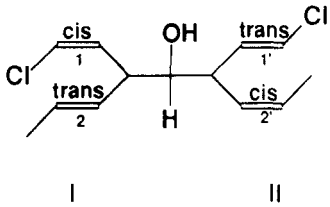
In this simplified implementation of the CIP rules the analysis of the ligands was not made according to the CIP methodology. For comparison according to rule 1, the different atoms in each ligand were divided into groups, and atoms in each group were considered hierarchical equivalents. Atoms were then compared by decreasing order of priority. The factors determining the introduction of an atom in a given group were first the distance from the core of the stereogenic units (level), then the atomic number of the atom to which it was bonded and that was in the previous level, and finally its atomic number, as illustrated in Figure 3.

However, in this method the connectivity of the atoms was not fully considered (Figure 4). The same happened with several features of the cyclic molecules, leading in some molecules to incorrect comparison and ordering of the ligands.

Considering rule 3, the simplified version implemented did not allow a correct evaluation of the hierarchy of the double bonds (Figure 5). Also, the particular case of mesomeric systems was not considered, and aromatic double bonds were analyzed as localized double bonds. Additionally, rule 3 was substantially changed in the 1982 revision of the CIP system,[5] thus requiring a completely new implementation.

Other points requiring a revision included particular cases in which the program could not deal properly with the set of stereobonds drawn by the user or when this set was ambiguous.

The specification of double bonds suffered also from all the problems resulting from the simplified implementation of the CIP sequence rules. An extension of the range of double



LHASA analysis (Rule 3)
Old implementation

Ligand I	Ligand II
1 Cis double bond	1 Trans double bond
+	+
1 Trans double bond	1 Cis double bond

LHASA comparison: I = II

CIP analysis (Rule 3)

Ligand I	Ligand II
1 Cis double bond (1)	1 Trans double bond (1')

CIP comparison: I > II

Figure 5. Different ligands which could not be differentiated by the old implementation of the CIP sequence rules in LHASA.

bonds analyzed was also required as only olefins were considered and not other stable double bonds, such as oximes.

Considering the shortcomings of the implementation of the CIP system that resulted in the impossibility of identifying certain stereogenic units and in the erroneous specification of others, and also that an extension was not possible, a complete reformulation of the module was undertaken following a different approach.

4. NEW IMPLEMENTATION

Programs for computer-assisted organic synthesis always start the analysis of a given target molecule, or the evaluation of a generated precursor, by doing the perception of its structural characteristics. This perception, as well as the search for instability and reactive functionality, is controlled in LHASA by PRTARG. The stereochemical perception is controlled by PRSET, and it supervises the perception of double bonds and sp^3 centers of which the configuration is determined by PRSTCT and PRSTER, respectively.

4.1. Analysis of sp^3 Centers. 4.1.1. General Process. The method used, in this implementation of the CIP system in LHASA, for the perception of the absolute configuration of stereogenic centers is the original one,[16] based on a linear representation assigned to the center, considering the molecule drawn and the stereochemical information conveyed by the use of wedge-shaped bonds and dotted bonds. The main improvements introduced are concerned with an extension of the capabilities for handling the representations of stereobonds, an extension of the capabilities for the identification of stereogenic (chiral and pseudoasymmetric) centers and prochiral centers, and mainly in the implementation of the CIP sequence rules for comparison of the ligands. The aims were to achieve a complete identification and specification of the most important stereogenic units in organic chemistry and to produce a solid basis for future extensions of the module.

It was decided to restrict the work to the analysis of stereogenic centers, as it is considered that in most of the cases the stereochemical information relevant in synthesis planning is completely defined by their configuration and relationship in the molecule. From the stereogenic centers, just those having four different ligands are considered, as they are by far the most common in organic chemistry. The present implementation identifies and specifies stereogenic centers located in carbon and noncarbon elements.

The general process is summarized in Scheme 1. In a first stage all stereobonds and their nature are identified. Potential stereogenic centers or prochiral centers are also identified. These include all atoms having four, three, or two explicit ligands that could constitute a stereogenic or prochiral center (for example neutral C and Si atoms; N in quaternary salts or N oxides; N in a bridgehead or in a three-membered ring where it is connected to an atom containing an unshared electron pair; Ge; Sn; P; S in sulfoxides, sulfites, sulfonium salts, etc.).

For the analysis of each stereogenic center LHASA requires four explicit ligands and an unambiguous set of stereobonds (dashed bonds between two potential stereogenic centers are considered as potentially ambiguous). Stereocenters are thus divided into two groups, those that have all the information required and those that do not have it. This information is reevaluated at several stages of the process.

Centers for which all the required information is available are processed first. A linear representation is attributed to each one having an unambiguous set of stereobonds, and a comparison of pairs of ligands is made according to the CIP sequence rules in an exhaustive but nonredundant way.

The comparison of ligands using the CIP sequence rules 4 and 5 must always be done using auxiliary descriptors, but these correspond to the definitive ones in all acyclic molecules and in most cyclic molecules. The exceptions are the cases where the specification of all the centers depends on the specification of the others. Specification using auxiliary descriptors is extremely time consuming. Thus at an initial stage an attempt is made to identify and specify all the units in the molecule without recourse to their use, and only if this is impossible are auxiliary descriptors used.

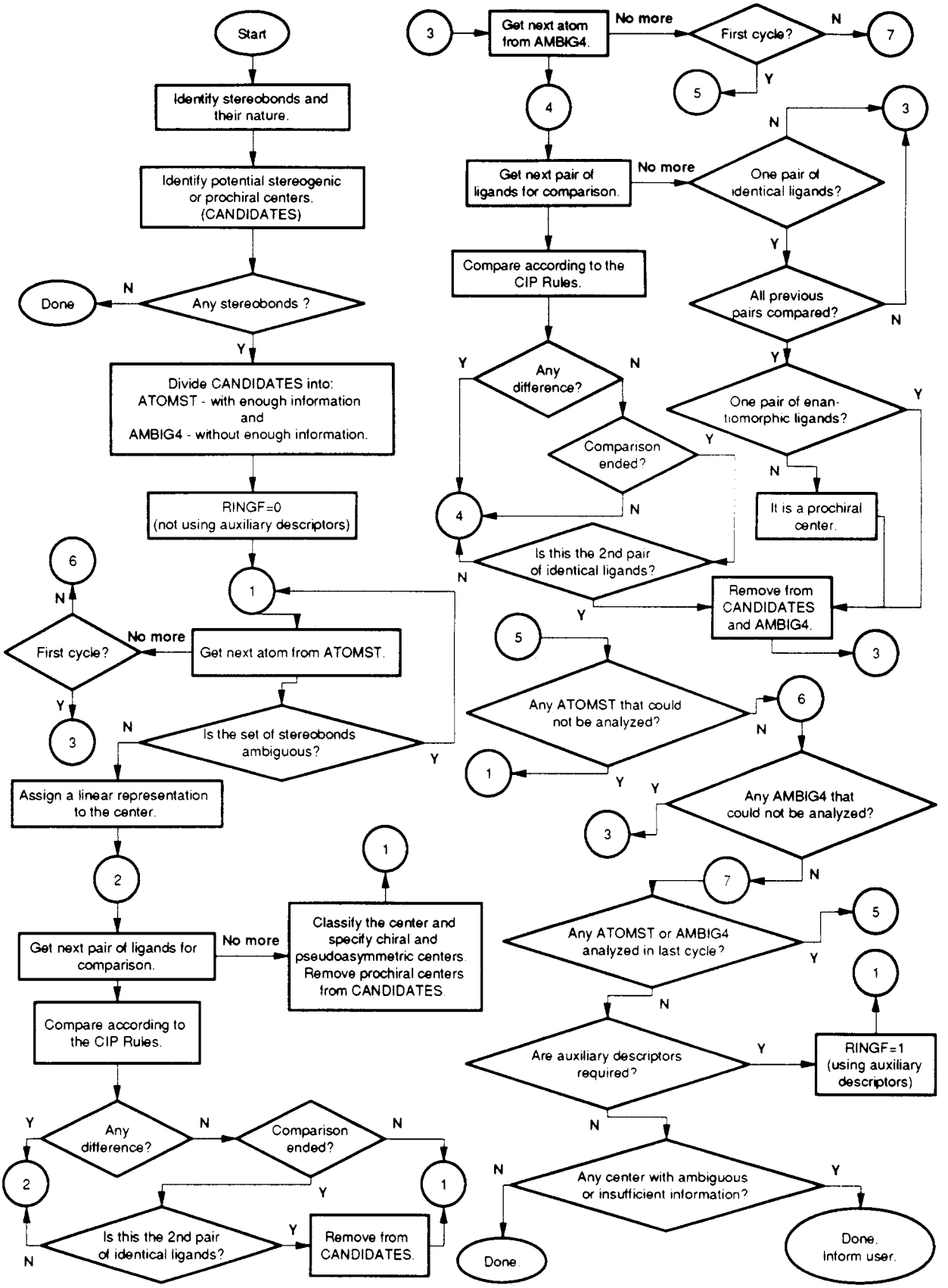
If the comparison of a pair of ligands cannot be concluded because it requires information about centers that have not yet been specified, it is suspended and resumed latter. If the ligands can be compared, a decision must be made as quickly as possible about whether the center is stereogenic or prochiral, and if it is a stereogenic center, it is specified.

For centers that do not have all the information required, the process is similar except that the linear representation is not attributed to the center, as there is not information available for it.

This cycle is repeated as many times as required until all possible centers are analyzed. If, after an exhaustive analysis, it is proved that the use of auxiliary descriptors is required, then this is made.

In the end of the process the results of the identification and specification obtained are stored and centers whose classification or specification was not possible, due to lack of information, are pointed out to the user. If there is any ambiguity in the set of stereobonds, the user will also be informed and a correction will be requested.

Scheme 1. Analysis of sp^3 Centers by LHASA: General Process



4.1.2. Linear Representation. PRSTER calls FILL_ROT to generate the linear representations.[16] These have the form **A-B-C-D** where **A**, **B**, **C**, **D** are the atoms at the vertices of the tetrahedron defined by the attachments to the central atom. **A** is considered to be above the plane defined by the other three atoms, which appear in clockwise order. FILL_ROT looks for the atoms and bonds adjacent to the center, checks if the adjacent set of stereobonds is unambiguous to the program, chooses a reference stereobond, and assigns a linear representation to the center. This information is stored in "linked list" format for rapid access and use by other program modules and is also used to assign descriptors to stereogenic

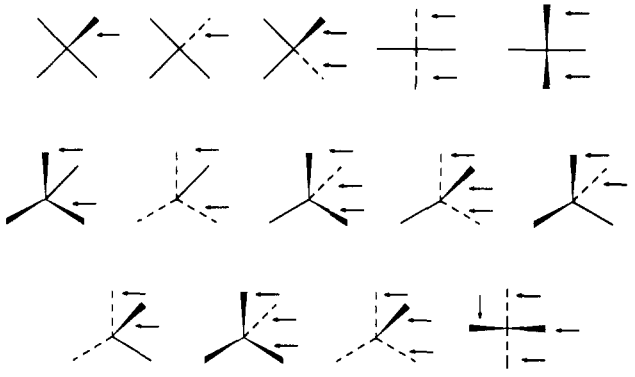


Figure 6. Stereo representations that can be accepted by the program. Arrows point to stereobonds that can be used as reference bonds.

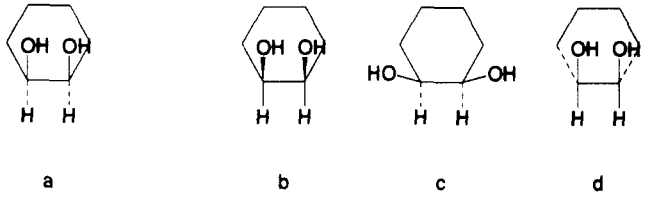


Figure 7. Ambiguity in the perception of stereo representations drawn for cyclic molecules. **a** is ambiguous, while **b**, **c**, and **d** are not ambiguous.

units by comparison of it with the ranking of the ligands according to the CIP sequence rules.

Using the linear representation, LHASA does not perceive stereochemistry in exactly the same way as chemists do. The necessarily two-dimensional method does not always guarantee a good perception of all the sets of stereobonds that can be unambiguous to the chemist, and from each set of stereobonds just some of them can be used as reference bonds. In the present implementation the set of stereochemical representations accepted by the program was considerably extended, and presently they include all those that are not ambiguous for the chemist (Figure 6), with some limitations on the stereobonds that can be accepted as reference stereobonds.

However, there are still cases for which the stereochemical representation is accepted by the program, but the two-dimensional diagram drawn by the user is ambiguous due to the way the perception is done. This may happen in some cyclic molecules in which the program assigns one descriptor or the other to the same center according to the way the molecule is drawn (Figure 7). This aspect must be considered in the future by the implementation of a module which analyzes ambiguous cyclic systems. Now the problem can be overcome by informing the users that they must use a visually realistic representation for each center.

4.1.3. Ligand Comparison. General Process. The ligand comparison is controlled by PRSTER. Every relevant pair of ligands attached to a potential stereogenic unit is successively compared (Scheme 2) according to the rules described in section 2.2 until the first difference is encountered, an exhaustive comparison is made and no difference is encountered, or it is concluded that there is not enough information for the comparison.

The process starts by the comparison of the constitutional properties of the ligands (CIP rule 1). While this comparison is made, the hierarchical digraph is created and stored and basic information about double bonds in the ligands is also stored. This information will be used in later stages of the comparison process if required.

It was already proved that rule 1 as stated by Cahn, Ingold, and Prelog does not allow the ranking of all constitutionally distinct ligands, and a supplement to it was proposed[13] to overcome the deficiencies. However, this supplement was not implemented, as the cases covered by it are extremely rare, but an extension of the implemented rules to consider this supplement can be made if required.

Rule 2 is not implemented but can be easily added when isotope information is available in LHASA.

If ligands have similar constitutional properties and if they contain any double bonds, the next stage is their comparison according to CIP rule 3, plus an extension (cf. 2.2). For this purpose the previously stored information about double bonds is reorganized to rank them and a comparison is made.

If corresponding double bonds are similar in both ligands, or if there are no double bonds, stereogenic centers in the ligands are compared. The first stage in this process is to verify if the ligands contain any identified or potential stereogenic centers. Simultaneously, information about stereogenic centers is stored, and it is checked if corresponding centers have the same nature (rule 4a). If some potential stereogenic centers whose definitive identification has not yet been made are detected, comparison is suspended.

If there are stereogenic centers in the ligands and no difference was yet detected, the next stage is a comparison of the relationship between the chiral center descriptors (rule 4b). If necessary, a comparison of the descriptors attributed to the pseudoasymmetric units in the ligands is also made (rule 4c).

Finally, if the information previously compared did not allow the ranking of the ligands, a check is performed to see if they are enantiomorphic (rule 5).

During the entire process described above the hierarchical rank of the nodes in the digraph is reordered whenever necessary, according to the characteristics of the information under comparison.

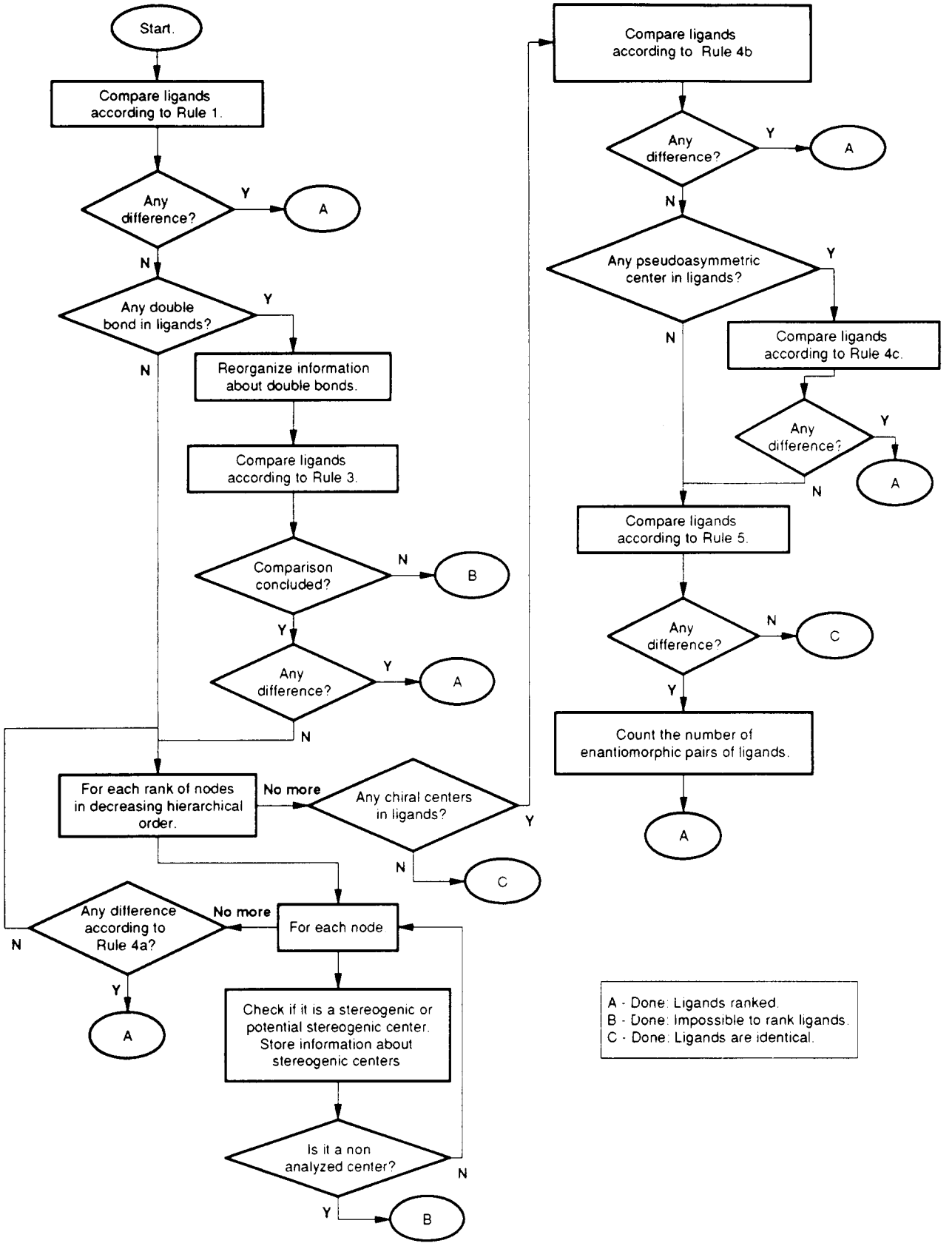
4.1.4. Hierarchical Digraph. A fundamental step in ligand comparison is the generation of the hierarchical digraph, i.e. the equivalent acyclic structure into which monodentate, polydentate, or cyclic ligands must be converted for analysis. There is a set of rules which are used to convert a stereocenter and its ligands into a hierarchical digraph without redundancy and ambiguity.[4,5] The valence complementation and valence bond conventions are fundamental to this process.

According to the conventions, each atom in the molecule is considered tetraligant. Thus each node is connected to one node in the previous level of the digraph and to three other nodes in the following level of the digraph. The digraphs stored are simplified ones in which the nodes corresponding to hydrogen atoms, electronic pairs, and phantom atoms are not represented. However, this information is important for the comparison of each level of the digraph according to rule 1 and can be decisive in the ranking of the ligands.

Some double bonds are treated as special cases as they should be considered as single bonds according to the CIP system. This is the case for double bonds for which there is a contribution of a d orbital, e.g. the double bond involving the sulfur atom in a sulfoxide. In the implementation of this convention all double bonds to heteroatoms adjacent to sulfur or phosphorus are considered single bonds, and if necessary to complete tetravalence of the other atoms involved in the double bond (generally oxygen), an electron pair is added. This allows the program to correctly deal with the great majority of cases.

The hierarchical digraph contains information about characteristics and the ranking of all nodes and is the basis for the comparison of the ligands according to all relevant properties

Scheme 2. Ligand Comparison: General Process



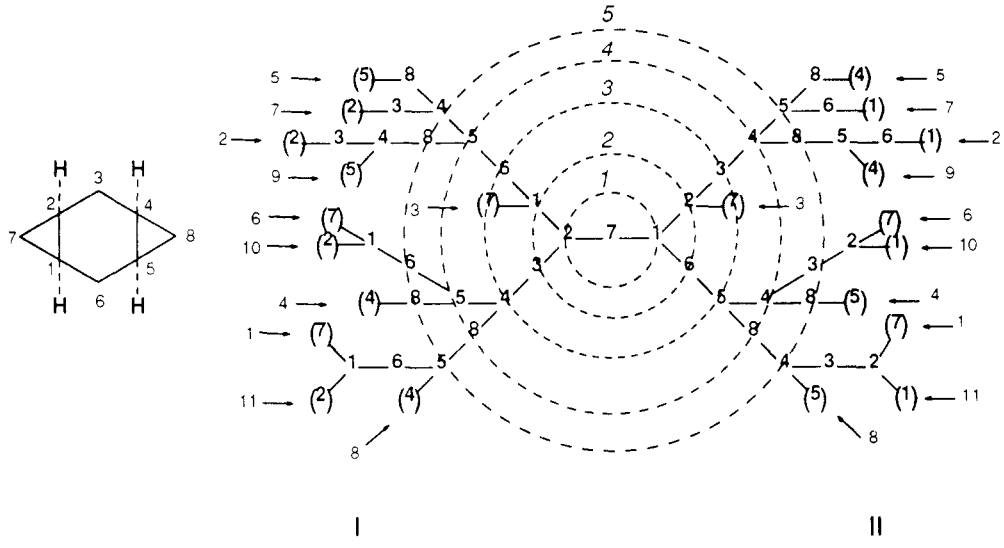
(rules 1–5). Information about it is stored in two arrays (WAY and TWAY) (Figure 8).

WAY has information about the nodes in all paths that is possible to follow from the root of the digraph (potential stereogenic center) to each one of the outermost nodes (leaves). The dimensions of WAY (BR, WN, LE) have the following meaning:

BR identifies each ligand in the pair being compared

WN identifies each path in the digraph leading from the root to the leaves

LE identifies each one of the levels (spheres) of nodes in the digraph. For practical reasons LE = level number + 1.



I II

WAY(1,1,1) = 1	WAY(1,4,7) = 400	TWAY(1,1,2,2) = 7	WAY(2,2,4) = 3	WAY(2,7,8) = 100
WAY(1,1,2) = 2	WAY(1,5,1) = 2	TWAY(1,2,1,2) = 2	WAY(2,2,5) ◘ 4	WAY(2,8,1) = 1
WAY(1,1,3) = 3	WAY(1,5,6) = 4	TWAY(1,3,1,2) = 9	WAY(2,2,6) = 8	WAY(2,8,7) = 500
WAY(1,1,4) = 4	WAY(1,5,7) = 8	TWAY(1,4,1,2) = 3	WAY(2,2,7) = 5	WAY(2,9,1) = 2
WAY(1,1,5) = 8	WAY(1,5,8) = 500	TWAY(1,5,1,2) = 6	WAY(2,2,8) = 6	WAY(2,9,8) = 400
WAY(1,1,6) ◘ 5	WAY(1,6,1) = 4	TWAY(1,5,2,2) = 10	WAY(2,2,9) = 100	WAY(2,10,1) = 6
WAY(1,1,7) = 6	WAY(1,6,6) = 6	TWAY(1,6,1,2) = 4	WAY(2,3,1) = 2	WAY(2,10,8) = 100
WAY(1,1,8) = 1	WAY(1,6,7) = 1	TWAY(1,7,1,2) = 1	WAY(2,3,4) = 700	WAY(2,11,1) = 1
WAY(1,1,9) ◘ 700	WAY(1,6,8) = 700	TWAY(1,7,2,2) = 11	WAY(2,4,1) = 1	WAY(2,11,9) = 100
WAY(1,2,1) = 1	WAY(1,7,1) = 5	TWAY(1,8,1,2) = 8	WAY(2,4,5) = 4	
WAY(1,2,3) = 1	WAY(1,7,7) = 3		WAY(2,4,6) = 8	TWAY(2,1,1,2) = 5
WAY(1,2,4) = 6	WAY(1,7,8) = 200	WAY(2,1,1) = 1	WAY(2,4,7) = 500	TWAY(2,1,2,2) = 7
WAY(1,2,5) = 5	WAY(1,8,1) = 1	WAY(2,1,2) = 1	WAY(2,5,1) = 2	TWAY(2,2,1,2) = 2
WAY(1,2,6) ◘ 8	WAY(1,8,7) = 400	WAY(2,1,3) = 6	WAY(2,5,6) = 5	TWAY(2,3,1,2) = 9
WAY(1,2,7) = 4	WAY(1,9,1) = 2	WAY(2,1,4) = 5	WAY(2,5,7) = 8	TWAY(2,4,1,2) = 3
WAY(1,2,8) ◘ 3	WAY(1,9,8) = 500	WAY(2,1,5) = 8	WAY(2,5,8) = 400	TWAY(2,5,1,2) = 6
WAY(1,2,9) = 200	WAY(1,10,1) ◘ 6	WAY(2,1,6) = 4	WAY(2,6,1) = 4	TWAY(2,5,2,2) = 10
WAY(1,3,1) = 2	WAY(1,10,8) = 200	WAY(2,1,7) = 3	WAY(2,6,6) = 3	TWAY(2,6,1,2) = 4
WAY(1,3,4) = 700	WAY(1,11,1) = 1	WAY(2,1,8) = 2	WAY(2,6,7) = 2	TWAY(2,7,1,2) = 1
WAY(1,4,1) = 1	WAY(1,11,9) = 200	WAY(2,1,9) = 700	WAY(2,6,8) = 700	TWAY(2,7,2,2) = 11
WAY(1,4,5) = 5		WAY(2,2,1) = 1	WAY(2,7,1) = 5	TWAY(2,8,1,2) = 8
WAY(1,4,6) = 8	TWAY(1,1,1,2) = 5	WAY(2,2,3) = 2	WAY(2,7,7) = 6	

Figure 8. Tricyclooctane and the hierarchical digraph corresponding to the ligands of C–7: x identifies a node, (x) identifies a node corresponding to a duplicated atom, *x* identifies a sphere of nodes (level), and ← x identifies a path.

If one path X branches from another path Y in level L, the WAY element with LE = 1 is used to store information about the number of the path Y from which X branches out, and apart from this, just the WAY elements with LE > L + 1 are filled for path X.

Each element of WAY contains a number that allows the identification of the atom in the molecule structure corresponding to the node; this is the corresponding atom number in the structure. However, rings and multiple bonds imply the existence of some nodes in the digraph, duplicated nodes, having particular characteristics. In order to allow their identification, these nodes are represented by the value (100 × atom number), the atom number being that of the atom corresponding to the duplicated node in the structure. Since LHASA uses only 64 atoms, this gives a unique number which cannot correspond to an actual atom in the structure.

For a complete knowledge of the hierarchical digraph it is necessary not only to know the nodes constituting it and their connectivity but also their relative hierarchy or, what is equivalent, the hierarchy of the different paths of the digraph. This information is stored in the array TWAY (BR, TW, NW, I) (Figure 8), whose elements contain the number of the paths and whose dimensions have the following meaning:

- BR identifies each ligand in the pair being compared
- TW identifies the hierarchical level of one path or group of paths
- NW identifies each one of the paths belonging to a hierarchical group
- I can have two different values (1 or 2). Elements have I = 1 if the path has not ended in a previous level and 2 if the path has ended in a previous level. When the complete digraph is generated, I is 2 for all the elements.

The rank of the different nodes in the digraph is given first by the level of the digraph to which they belong, and then, in

each level of the digraph, the nodes are ranked on the basis of the highest ranked path to which they belong.

The information contained in WAY and TWAY is not only enough to describe the digraph but also to reconstitute it if necessary. Information in WAY reflects the connectivity of the different nodes in the digraph, and information in TWAY reflects their ranking considering the nature of the atoms to which they correspond, their connectivity, and their distance from the root of the digraph.

Information in TWAY is continuously reordered to reflect the alterations in the ranking of the different nodes due to characteristics of the new levels of the digraph being analyzed according to rule 1 or, in more advanced stages of the process, resulting from the application of other rules. This reordering process is of fundamental importance, and there is a subprogram (ORDER) developed to take care of this job whenever required, which uses all the information available about the properties of the different nodes and their connectivity. This process can be a complex one, as the hierarchical rank of all the nodes can be affected. In fact, the nodes which have been compared change priority according to the sequence rule applied, then all mutually corresponding neighboring nodes of the two compared nodes change priority correspondingly, and then the neighbor's neighbors change priority in the same way, etc., inducing changes of priority in all directions until the whole tree has been altered.

4.2. Implementation of the Sequence Rules. 4.2.1. Comparison of Constitutional Properties (Rule 1). The subprogram PRSTWT is called whenever it is necessary to compare a pair of ligands according to the CIP sequence rule 1.

For the comparison of successive levels of nodes in the digraph according to rule 1 it is necessary to attribute an atomic number to each node. This is a straightforward process, except in the case of duplicated atoms in mesomeric systems.

For comparison purposes the "atomic number" 0 is attributed to electron pairs, as well as to phantom atoms, although no node is assigned to them in the stored digraph. Also, nodes are not assigned to hydrogen atoms, but the atomic number is considered for comparison purposes. A node in the digraph is assigned to each real atom, and its atomic number is determined for comparison purposes.

For duplicated nodes, corresponding to atoms that are not involved in mesomeric systems, the atomic number used for comparison purposes is the one of the corresponding atom in the structure. For duplicated nodes corresponding to atoms involved in mesomeric systems the "atomic number" must be calculated according to the conventions proposed in the CIP system (valence bond conventions).[4]

The present implementation is limited to the aromatic systems recognized by LHASA. The implementation of this convention for such systems is simple; however, its extension to other mesomeric systems would introduce a great complexity, with the unavoidable slowing down of the process. Thus for nonaromatic mesomeric systems double bonds are considered to be localized. Consequences from this are that ligands with identical mesomeric systems which are represented in different ways are considered different by the program and also that some ligands can be incorrectly ranked. These cases are not common, and some of the inconveniences can be avoided if the user introduces the same representation for the same mesomeric system in all of its occurrences in the molecule.

The convention is fully implemented for aromatic rings having up to nine atoms which are neutral or have a charge of +1 or −1. For all the other systems the convention was simplified and the "atomic number" of the duplicated node is the average of the atomic numbers of the aromatic atoms adjacent to the node being studied. The value attributed in this way can in some rare cases be different from the one attributed by the CIP system.

For combinations of different aromatic mesomeric systems the hierarchy defined by the CIP system to attribute the "atomic number" to the nodes is generally respected.

As the "atomic number" attributed to duplicated nodes in mesomeric systems is not always an integer, to simplify the comparison process, and also to make easier the identification of electronic pairs and phantom atoms, the value that is compared is the integer part of 10 × (atomic number + 1). It was proved that this number is enough to rank ligands in all possible cases.

Simultaneously with the comparison according to rule 1 and the concomitant development of the digraph, fundamental information about structural characteristics of the double bonds in the ligands is gathered and stored in an array (DOUB). Double bonds with stable configuration that can exist in two different isomeric configurations (olefins, oximes, oxime ethers, and haloimines) are considered; double bonds in enols, enamines, and those in rings with less that eight atoms and aromatic double bonds, as well as cumulated double bonds (if an even number, they constitute an axis of chirality that cannot be specified by the present implementation of the rules; if an odd number, they should be compared according to rule 3, and the present implementation does not allow specification of this kind of bond), are excluded.

4.2.2. Comparison of Double Bond Configuration (Rule 3). Prior to the comparison according to rule 3, available information about double bonds is reorganized. Information in DOUB allows recognition of all structural characteristics of the double bonds in the ligands, as well as the paths and levels of the digraph containing the atoms involved. However, it has no information about the hierarchy of double bonds at the same distance from the root, necessary to compare corresponding double bonds in both ligands by decreasing priority order. Combining the information in DOUB with the information in TWAY, a complete set of information can be achieved. This process is performed by subprogram FILL_DOUBLE, and the resulting information is stored in array DOUBLE. This information allows the comparison and the reordering of the digraph (TWAY) and of DOUBLE whenever required.

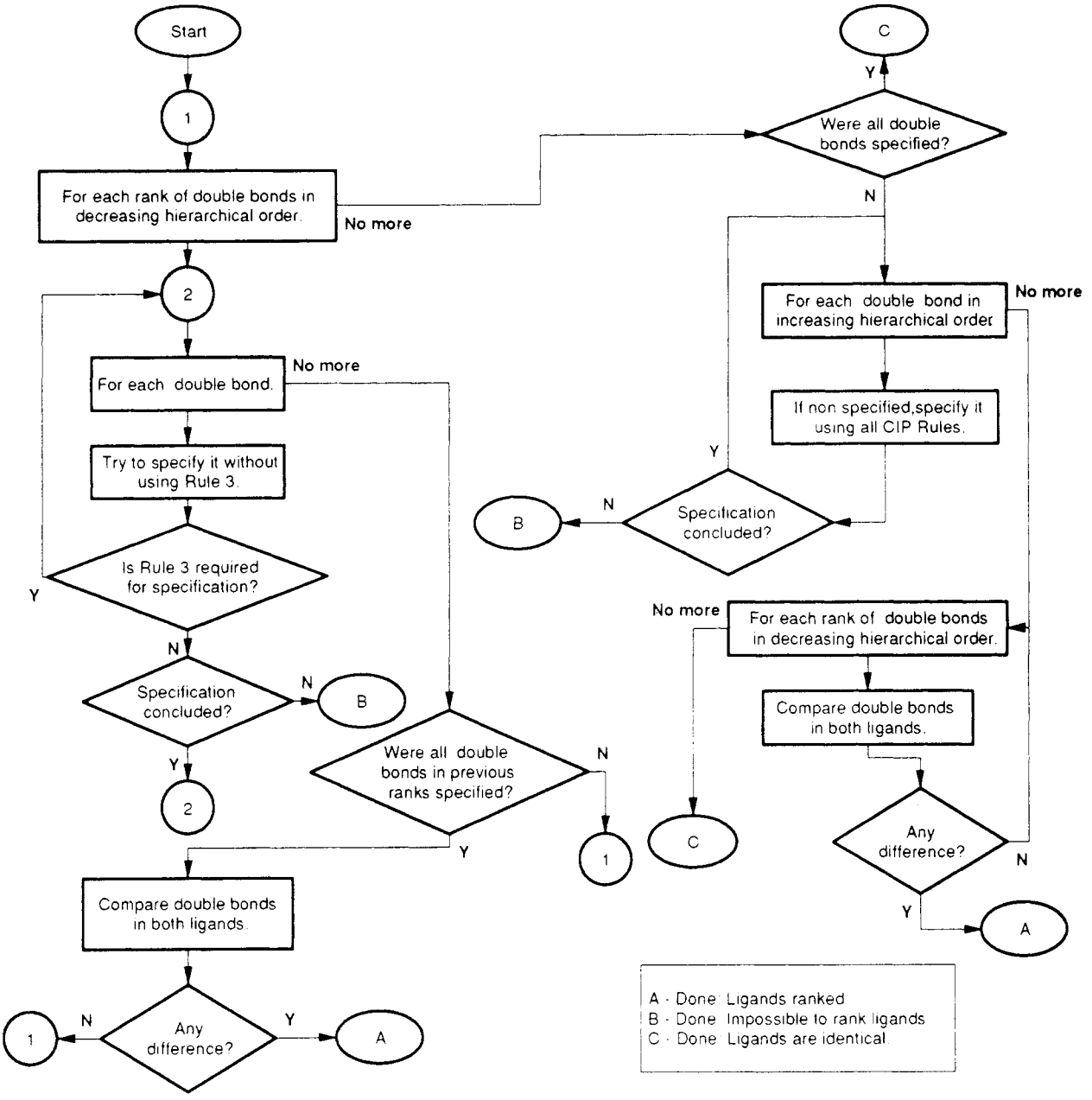
Subprogram RULE3 controls the process of specification and comparison of the double bonds according to CIP sequence rule 3 plus the extension referred to in section 2.2 (Scheme 3).

In the first stage the double bonds are successively specified as *cis* or *trans*, applying these terms to the location of an atom or atom group of higher rank according to the CIP sequence rules in relation to the core of the stereogenic unit, rather than to the configuration of the double bonds themselves. This specification is made by the comparison of the linear representation attributed to each double bond with the ranking of the ligands.[16]

The double bonds whose ligands can be ranked without the recourse to CIP rule 3 are classified and specified in this stage. If it is necessary to make the comparison according to rules 4 or 5 and the required centers are not specified, the analysis of the stereogenic center is suspended.

If specification of a double bond requires the use of rule 3, it is suspended and will be made in another stage. However, the specification process of other double bonds in the ligands

Scheme 3. Comparison According to Rule 3



continues, but the comparison of the ligands is not made, as the information available is not enough.

If specification of all the double bonds in a given hierarchical rank is concluded, and all previously analyzed double bonds could be specified, the next step is the comparison of corresponding double bonds in both ligands. During this process, and whenever required, the hierarchical digraph (TWAY) and DOUBLE are reordered. This is necessary when both ligands have more than one type of double bond (*cis*, *trans*, or nonstereogenic) which previously were in the same hierarchical rank and the higher ranked type is similar in both ligands. The reordering of TWAY is made by the subprogram ORDER and is followed by a reorganization of the array DOUBLE according to the new hierarchical status. This process is repeated as many times as necessary until analysis is complete or a difference is encountered.

If there are double bonds in the ligands whose specification was not possible, a second stage in this comparison process is required. In this stage double bonds whose specification was impossible are specified starting from the outermost non-specified double bonds. This can guarantee that comparison of their ligands according to rule 3 can be performed.

When all double bonds in DOUBLE are specified, a comparison of the corresponding ones in both ligands is performed in decreasing hierarchical order. In this process, whenever required, reorganization of TWAY and DOUBLE is performed.

4.2.3. Comparison of Stereogenic Centers (Rules 4 and 5).

4.2.3.1. Rule 4a. Subprogram FILL_BRSTE is mainly concerned with the comparison of the nature of corresponding centers in both ligands, that is, comparison according to rule 4a. It also simultaneously gathers structural and hierarchical information about the chiral centers in the ligands necessary for comparison according to rules 4b and 5, which is stored in array BRSTE.

All the nodes in the digraph are analyzed, by decreasing hierarchical order, to check if they correspond to a real or potential stereogenic center. If any node corresponds to a potential stereogenic center, whose analysis was not previously made or concluded, the comparison of the ligands and the analysis of the center to which they are attached is suspended, as there is not information available to conclude it.

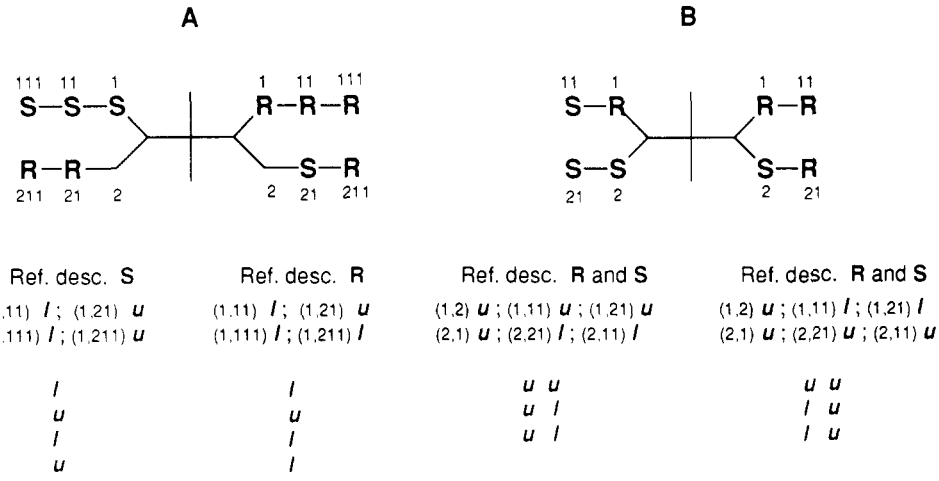


Figure 9. Comparison according to rule 4b.

For each set of hierarchically equivalent nodes, chiral and pseudoasymmetric centers in both ligands are counted, and information about chiral centers is stored in BRSTE. The number of chiral and pseudoasymmetric centers is then compared in an attempt to rank ligands according to rule 4a.

In this process, if the higher ranked nodes are equivalent and there is more than one type of node, a reordering of the digraph is made. Then, if the comparison of the higher ranked nodes allows distinction between the ligands, the comparison is concluded. Otherwise a reordering of BRSTE is made.

4.2.3.2. Rule 4b. Subprogram RULE4 compares the relationship of the descriptors assigned to chiral nodes to identify ligands that are diastereoisomorphic or potentially enantiomorphic.

In this process the first step is the analysis of the higher ranked descriptors to check for any difference or to identify the reference descriptor which is going to be paired with the remaining ones. This is the descriptor assigned to the majority of the higher ranked nodes, or both descriptors if they occur in the same amount[5] (Figure 9).

All relevant, but nonredundant, pairs of descriptors are formed and compared until any difference is encountered or all the pairs are compared.

If both descriptors are used as reference (Figure 9, molecule B), when paired with hierarchically equivalent descriptors, the relationship in the digraph (distance) of the corresponding nodes is calculated and used for the ranking of the pairs (cf. 2.2) for comparison purposes (e.g. pair (1,11) has priority over pair (1,21)).

Hierarchically equivalent pairs are compared to check for any difference. When both descriptors are used as reference, information resulting from the analysis of the pairs formed with both is added before any decision is made (Figure 9, molecule B).

When there is more than one type of pair in both ligands the digraph is reordered prior to the comparison, as described for other rules. If this is the case, the vector BRSTE is filled again considering the new ranking of the nodes and the process is restarted.

Simultaneously, with the comparison according to rule 4b information about the ligand having a higher frequency of *R* descriptors assigned to the highest ranked nodes is stored. This information allows identification of potentially enantiomorphic ligands and, if no difference is encountered according to rules 4b and 4c, allows the ligands to be ranked.

4.2.3.3. Rule 4c. Diastereoisomorphism of ligands can also be due to the existence of pseudoasymmetric centers with different configuration (rule 4c), and this is analyzed by subprogram RULE4C.

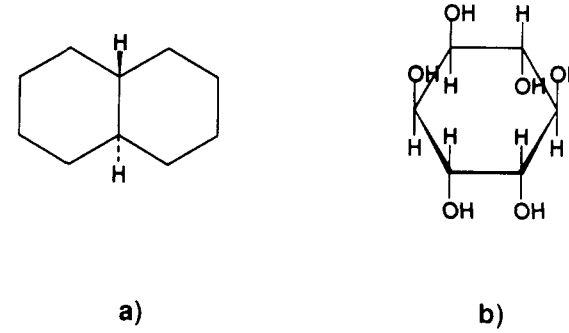


Figure 10. Molecules having stereogenic centers whose identification and specification requires the use of auxiliary descriptors.

All the nodes in the digraph are successively compared, by decreasing order of priority, to check if they correspond to a pseudoasymmetric center. For each hierarchically equivalent group of nodes in this situation the amount of *r* and *s* descriptors is counted and a comparison of corresponding centers in both ligands is performed. As described for the previous rules, the digraph is reordered prior to the comparison whenever required.

4.2.3.4. Rule 5. If the previous rules did not allow the ligands to be ranked, PRSTER checks if they are enantiomorphic or identical. Information stored simultaneously with the comparison according to rule 4b allows identification and ranking of enantiomorphic ligands if just one descriptor was used as a reference descriptor.

However the fact that, for the comparison according to rule 4b, both ligands were considered as reference descriptors also points to the existence of potentially enantiomorphic ligands. In this case both ligands must be analyzed again to determine if they are enantiomorphic or identical.

Information about the number of enantiomorphic pairs of ligands bonded to the center being analyzed is stored, once it is essential to decide whether the center is pseudoasymmetric or chiral.

4.2.4. Auxiliary Descriptors. Usually for acyclic molecules and for most cyclic ones the auxiliary descriptors are equivalent to the definitive ones, and these can be used for specification purposes. However, for the analysis of molecules for which the assignment of descriptors to stereogenic centers depends on descriptors assigned to other centers and vice versa (Figure 10), it is necessary to resort to auxiliary descriptors, a concept introduced in the 1982 revision of the CIP system.[5] These descriptors are temporary and just used for the comparison of ligands according to rules 4 and 5. In this case only the

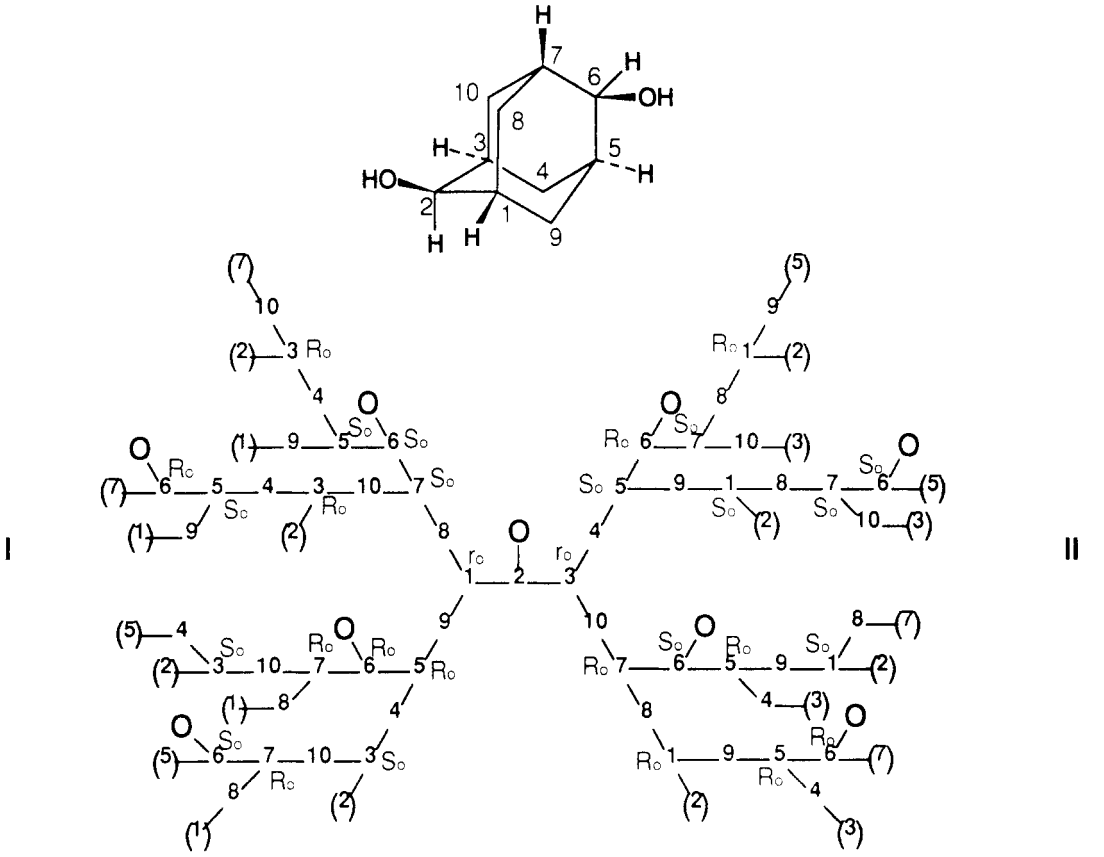


Figure 11. Auxiliary descriptors for the specification of C-2 in adamantane-2*S*,6*S*-diol.

auxiliary descriptors are used and not the definitive ones assigned during the process.

For the analysis of each stereogenic center the molecule is represented by its hierarchical digraph (center-digraph) and stereogenic centers in the ligands are specified comparing their appendages as they appear in the digraph. As cyclic molecules are represented by an acyclic model, this guarantees that all stereogenic centers can be specified.

In the previous description just the cases for which the auxiliary descriptors are equivalent to the definitive ones were considered. In the next paragraphs the implementation of the CIP system for molecules whose definitive descriptors cannot be assigned without the use of auxiliary descriptors will be described.

This process can be very sluggish for most molecules, as the number of auxiliary descriptors to assign can be very high. For example for the molecule in Figure 11, as many as 30 auxiliary descriptors are used for the specification of the center C-2 and the specification of 286 centers is required for the classification and specification of all chiral, prochiral, and pseudoasymmetric centers in the molecule. For the inositol molecule (Figure 10b) the specification of 66 centers is required for a complete analysis.

The implementation of this process is made in the subprogram INRINGS. This module assigns, on the basis of the center-digraph, the auxiliary descriptors and compares ligands according to rules 4 and 5.

The first stage of this process consists of assigning the auxiliary descriptors and storing them. All the center-digraph nodes are successively analyzed by increasing order of priority to guarantee that there is enough information for the comparison of their ligands. If any of the nodes do not have all information required for an exhaustive analysis, the process is stopped, as specification of the centers in the molecule is impossible. Otherwise the node is examined, in a way similar to the one described above, to check if it is a stereogenic center and, if so, to assign an auxiliary descriptor to it.

It should be noted that different descriptors can be assigned to nodes corresponding to the same atom, but appearing in different positions in the center-digraph (e.g. node 6 in Figure 11), and thus each node must be individually analyzed by considering its characteristics in the center-digraph.

The information about the definitive descriptors assigned to each center is stored in sets (STATMR and STATMS (for *R* and *S* centers, respectively) and STATPR and STATPS (for *r* and *s* centers, respectively)). However, the information about auxiliary descriptors cannot be stored in this format, as different descriptors can be assigned to nodes corresponding to the same atom. It is necessary to store this information in an array in which each element corresponds to one node in the center-digraph. Thus auxiliary descriptors are stored in the array CLA, in which each element corresponds to an element of WAY. One of the following values is stored in each element, according to the characteristics of the corresponding node: 1, *R* chiral; 2, *S* chiral; 3, *r* pseudoasymmetric; 4, *s* pseudoasymmetric; 5, nonstereogenic.

However, there are certain limitations in the digraph generated for the analysis of the different nodes (node-digraphs), particularly in the identification of duplicated nodes. In fact, in the node-digraph there are nodes (Figure 12) corresponding to the same atoms that can appear twice and should not be considered duplicated and nodes that should be considered duplicated (as they are so in the center-digraph) but that correspond to atoms that had not yet appeared in previous levels of the node-digraph. For node-ligands starting in levels of the center-digraph external to the one containing the node being analyzed, the duplicated atoms are corectly identified. However, for those node ligands that start in

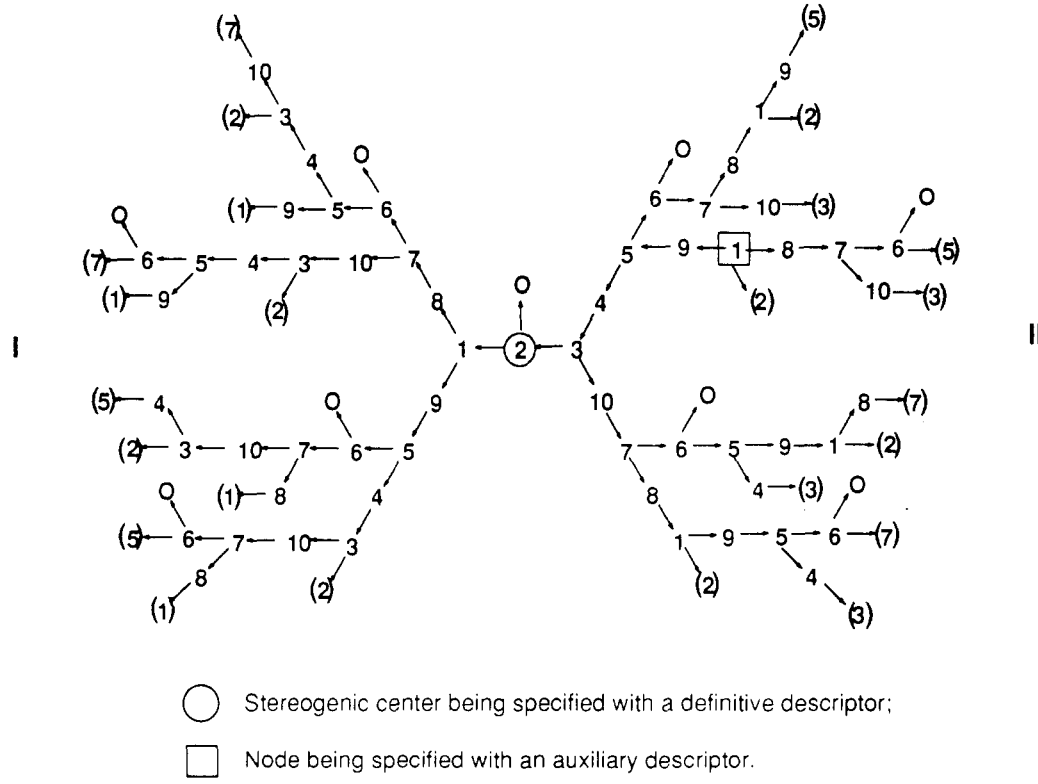


Figure 12. The hierarchical digraph of node 1 in the digraph of the adamantane-2*S*,6*S*-diol C-2 center for the determination of its auxiliary descriptor. (Arrows indicate the direction of the edges in the node digraph.)

internal levels of the center-digraph the cases described above are not identified correctly. It was decided to implement this simplified method for the generation of the node-digraphs since a complete one would be too time consuming and it was impossible to conceive situations for which these limitations could give rise to wrong descriptors. Usually distinction of the ligands is made before arriving at such a situation.

A ligand can also consist just of a duplicated node. In this case the first characteristic to be compared is the atomic number of the corresponding atoms and then, if no difference is encountered, the fact that one is duplicated and the other is not is enough to rank them. This comparison is done by INRINGS before starting the comparison of the ligands as previously described.

As the information about the auxiliary descriptors is stored in CLA referring to the position of the nodes in the center-digraph, it is necessary to reorganize this information for node-ligand comparison purposes. There is a subprogram (FILL_NCLA) that makes the correspondence between the nodes in the center-digraph and the nodes in the node-digraph and fills an array (NCLA) with the reorganized relevant information for node-ligand comparison. Node-ligands are then compared, according to rules 4 and 5, and specified according to the result of the comparison with an auxiliary descriptor which is stored in CLA. This process is made as many times as required until all the potentially stereogenic nodes are analyzed and all the information about them is stored in CLA.

At last the ligands of the potential stereogenic center are compared according to rules 4 and 5, considering the information about the auxiliary descriptors, in a way similar to that described above.

4.3. Double Bonds. In the stereochemical perception module the analysis of double bonds is made by the subprogram PRSTCT. This module was also reformulated. The main changes are related to improvements in the method for ligand comparison (the same as described for stereogenic centers) and in the range of double bonds considered, which was extended from only olefins to include also oximes, oxime ethers, and haloimines.

The process starts with the identification of the double bonds which must be selected for specification. As in the previous implementation, double-bond specification requires just one explicit ligand in each of the atoms involved in the bond.

The atoms comprising the double bond are analyzed, a linear representation is attributed to each one, and their ligands are compared according to the CIP rules. The results of the ranking of the ligands and the linear representation allow specification of the double bond.[16] However, for double bonds in rings with less than eight atoms, the two ring ligands are always considered to be *cis* to each other and this does not depend on the way the user draws them, since *trans* double bonds in these rings are considered to be too instable.

Information about the descriptors attributed to the double bonds is stored in two sets, STRBDE and STRBDZ, respectively, for *E* and *Z* double bonds, and the information about the linear representations is stored in a list (CISLST) to be used by other modules of the program. Results of the specification are presented to the user. If it is impossible to specify a double bond, because there is no information available about stereogenic centers and the comparison of the ligands depends on them, the user is also informed.

5. CONCLUSION AND FUTURE WORK

The implementation of the CIP system described above is quite complete and fulfills all LHASA's stereochemical requirements. The module can identify and specify the prochiral, pseudoasymmetric, and chiral centers having four different ligands and also noncumulated double bonds. It can be easily extended to deal with other stereogenic units, particularly stereogenic axes. The implementation of a module

for the analysis of ambiguous cyclic systems (cf. 4.1.2) would also be an important future improvement.

The implementation of the capacity to allow the user to define the stereochemistry of the molecule by assigning descriptors to the centers is now possible. The program will then check the validity of the descriptor and attribute stereocharacter to one of the bonds.

ACKNOWLEDGMENT

We thank Junta Nacional de Investigação Científica e Tecnológica (Lisbon), Instituto Nacional de Investigação Científica (Lisbon), and Fundação Calouste Gulbenkian (Lisbon) for partial financial support.

REFERENCES AND NOTES

(1) (a) Wipke, W. T.; Dyott, T. M. Simulation and Evaluation of Chemical Synthesis. Computer Representation and Manipulation of Stereochemistry. *J. Am. Chem. Soc.* **1974**, *96*, 4825–4834. (b) Wipke, W. T.; Dyott, T. M. Stereochemically Unique Naming Algorithm. *J. Am. Chem. Soc.* **1974**, *96*, 4834–4842.

(2) Blair, J.; Gasteiger, J.; Gillespie, P.; Gillespie, D.; Ugi, I. Representation of the Constitutional and Stereochemical Features of Chemical Systems in the Computer Assisted Design of Syntheses. *Tetrahedron* **1974**, *30*, 1845–1859.

(3) Esack, A.; Bersohn, M. Computer Manipulation of Central Chirality. *J. Chem. Soc., Perkin Trans. 1* **1975**, 1124–1129.

(4) Cahn, R. S.; Ingold, C.; Prelog, V. Specification of Molecular Chirality. *Angew. Chem., Int. Ed. Engl.* **1966**, *5*, 385–415.

(5) Prelog, V.; Helmchen, G. Basic Principles of the CIP-System and Proposals for a Revision. *Angew. Chem., Int. Ed. Engl.* **1982**, *21*, 567–583.

(6) Hanessian, S., *et al. CHIRON Program*; Université de Montréal: Montréal, Canada, 1987.

(7) *QUANTA*; Molecular Simulations, Inc.: Waltham, MA, 02154.

(8) (a) Corey, E. J.; Long, A. K.; Rubenstein, S. D. Computer-Analysis in Organic Synthesis. *Science* **1985**, 228, 408–418. (b) *LHASA*, Harvard Chemistry Department, 12 Oxford Road, Cambridge, MA.

(9) Cahn, R. S.; Ingold, C. K. Specification of Configuration about Quadricovalent Asymmetric Atoms. *J. Chem. Soc.* **1951**, 612–622.

(10) Cahn, R. S.; Ingold, C. K.; Prelog, V. The Specification of Asymmetric Configuration in Organic Chemistry. *Experimentia* **1956**, *12*, 81–124.

(11) Cahn, R. S. An Introduction to the Sequence Rule–A System for the Specification of Absolute Configuration. *J. Chem. Ed.* **1964**, *41*, 116–125.

(12) Dodziuk, H.; Mirowicz, M. A Proposal for a Modification of the Cahn, Ingold and Prelog Classification of Chirality. *Tetrahedron Asymmetry* **1990**, *1*, 171–186.

(13) Custer, R. H. Mathematical Statements about the Revised CIP-System. *Match* **1986**, *21*, 3–31.

(14) Mata, P. *Química Orgânica Computacional–Identificação e Especificação de Unidades Estereogénicas*; Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa: Lisboa, Portugal, 1989.

(15) Mata, P.; Lobo, A. M.; Marshall, C.; Johnson, A. P. The CIP Sequence Rules: Analysis and Proposal for a Revision. *Tetrahedron Asymmetry* **1993**, *4*, 657–668.

(16) Corey, E. J.; Howe, W. J.; Pensak, D. A. Computer-Assisted Synthetic Analysis. Methods for Machine Generation of Synthetic Intermediates Involving Multistep Look-Ahead. *J. Am. Chem. Soc.* **1974**, *96*, 7724–7737.