THE JOURNAL OF Organic Chemistry

VOLUME 51, NUMBER 25

© Copyright 1986
by the American Chemical Society

DECEMBER 12, 1986

On the Distinction of Diastereomers in the Cahn-Ingold-Prelog (*RS*) Notation

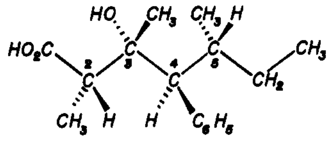
James H. Brewster

Department of Chemistry, West Lafayette, Indiana 47907

Received June 23, 1986

A general and flexible stereochemical notation, balanced in its emphasis on geometry (here, relative configuration) and topography (here, absolute configuration) and consistent with the distinction between stereogenicity and chirotopicity, is developed from the *RS* system by use of the *ul* notation of Prelog and co-workers, the bipartite format of the *RS*(R^*S^*) notation and the concept of external referencing.

There remains the need for a general stereochemical notation[1] that balances and distinguishes geometry[2–4] and topography[2–4] and also gives due recognition to the distinction between stereogenicity[5–9] and chirotopicity.[9] It is proposed here that a blend of the general *l*,*u* notations of Prelog and co-workers[4,10] with the *bipartite* *R*,*S*(R^*S^*) system of Chemical Abstracts,[11,12] all cemented with the concept of an *ideal external reference*,[9] meets that need.



2S,3R,4S,5R

$2S(2R^*,3S^*,4R^*,5S^*)$

u u u

2S(2ref,3u,4l,5u)

2,3-pref,3,4-pref,4,5-pref

S(2l,3u,4l,5u)

I

---

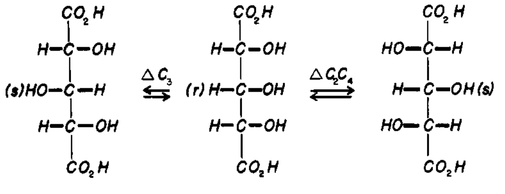
(1) "Parochial notations" (as: D,L—for carbohydrates and amino acids; *cis,trans*—for olefins and monocycles; α,β—for cyclics; *syn,anti*—for acyclics), although not general, continue to be used because they are simple to apply and evoke vivid and immediate images of stereostructures. They are of interest in their own right, and we hope to consider them in a separate paper.

(2) We use these terms in the same way as Prelog and Helmchen,[3,4] following F. Klein (op. cit.). *Geometry* relates to those aspects of stereostructure that are invariant to reflection (as, *relative* configuration); *topography* relates to those that are variant to reflection (as, *absolute* configuration).

(3) Prelog, V.; Helmchen, G. *Helv. Chim. Acta* 1972, *55*, 2581–2598.

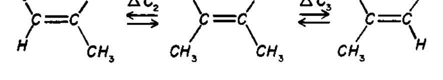
(4) Prelog, V.; Helmchen, G. *Angew. Chem., Int. Ed. Engl.* 1982, *21*, 567–583.

(5) McCasland, G. E. *A General System for the Naming of Stereoisomers*; Chemical Abstracts Service: Columbus, OH, 1953; pp 2,3 defines a stereogenic atom: "a) An atom of such nature and bearing groups of such nature that it can have two different configurations. b) An atom bearing several groups of such nature that an interchange of any two groups will produce an isomer (stereoisomer)." It is time to expand on the concept of stereogenicity. An asymmetric carbon atom with four ligands that differ in composition or constitution is a stereogenic monad; its configuration (*R* or *S*)[6] is independent of that of any other stereogenic atom. A "pseudoasymmetric" atom[3] (as C_3 in the two nonchiral forms of trihydroxyglutaric acid (a)) is not so independent in its stereogenicity; it can go from one configuration to the other ($r \rightarrow s$) by a single ligand exchange at C_3 *or* by simultaneous ligand exchanges at C_2 and C_4. The



a

stereogenic atoms of olefins, such as those in the 2-butenes (b), are not independent at all; C_2 is stereogenic only because C_3 is (and vice versa) and an exchange of ligands at one is equivalent to an exchange of ligands at the other. This stereogenic dyad behaves like a stereogenic entity (or *stereogen*) in having two configurations (*E* and *Z*[7] or *lk* and *ul*)[10] and in that any exchange of ligands that is not trivial (H for H or CH_3 for CH_3) or does not generate a constitutional isomer (2-methylpropene) would produce a stereoisomer. In the same way, the system: C_2–C_4 of 2,3-



b

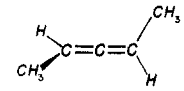
0022-3263/86/1951-4751$01.50/0 © 1986 American Chemical Society

A *general* notation must be in and of the Cahn–Ingold–Prelog (CIP) system,[6] under which descriptors of topography (*R* or *S*) can be asigned to the alternative configurations of stereogens that are reflection variant (chirogens).[5] In an extension of this system, Prelog and his co-workers[4,10] have proposed that the geometric (and conformation invariant) relationship of two such stereogens be described as *l* ("like") if their topographic descriptors are the same (*RR* or *SS*) or *u* ("unlike") if they are different (*RS* or *SR*).[13] Seebach and Prelog[10] make *serial* comparisons so that in 1[14] the pairings would be as follows: 2,3; 3,4; 4,5. Locants are not used, on which basis 1 would be designated the *uuu* diastereomer. This notation was developed to allow a convenient, conformation-independent means of distinguishing the diastereomerism at a particular bond, as in the products of an aldol condensation. Thus, a condensation linking C_2 to C_3 would give the *uuu* and *ulu* isomers if it followed one geometric course and the *luu* and *llu* isomers if it followed the other. Whatever the value of this notation for such purposes, it is flawed as a general notation, not so much by the absence of locants (a serious defect, but one that could be remedied) as by its *serial* nature. Particularly in more complex compounds, it would be awkward to deduce relationships betweeen nonadjacent stereogens. For that, it would be required to determine whether they were separated by an odd or an even number of *u*'s; thus, in 1, C_2 and C_5 are separated by three *u*'s, whence they are *u* to one another. This notation is overloaded with geometry and it is difficult to recover topography from it. It would be utterly useless in the case where the configuration at some central stereogen was still unknown.

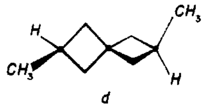
This defect could be avoided by adapting the notation of Prelog and Helmchen[4] (which was devised as an aid to distinguishing ligands that differ in geometry or topography) to the more general purpose considered here. One stereogen—generally the first encountered—is taken as a reference and all others are compared (*l* or *u*) to it. Locants are used for both stereogens; on this basis 1 would be identified as 2,3-*u*,2,4-*l*,2,5-*u*, or, perhaps 2-*ref*,3*u*,4*l*,5*u*. Such *internal referencing* has been used in other notations,

---

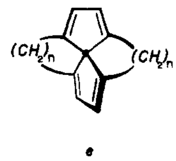
pentadiene (*c*) is a stereogenic dyad (or triad, if double bonds are treated as "two-membered rings"). The analogous dimethylspiroheptane (*d*) is



c

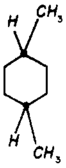


d

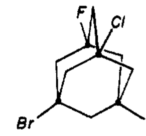


e

indisputably a stereogenic triad. (Note that the ligands of the stereogenic spiro atom are identical, in pairs, via rotation about the C_2 axis; the four ligands at the stereogenic central atom of a vespirene[8] (*e*) are identical through D_2 symmetry. *An sp*^{3} *stereogenic atom is not always asymmetric.*) The atoms to which the interchangeable ligands are attached may be separated by nonstereogenic atoms, as with the dyad in f, the triad in d, and the tetrad in g. The latter case shows that asymmetric

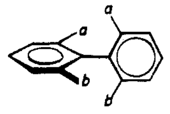


f

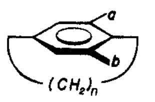


g

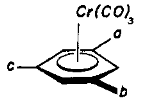
atoms at the bridgeheads of atom-bridged polycycles may be *mechanically* interdependent. They may be described, formally, as monads, but *individual* inversion at such atoms need not be considered seriously. In some cases, mechanical hindrance to rotation about single bonds produces systems that can *also* be described as stereogens (h,i). In the latter case there are only two exchangeable ligands: there are only three in j. The



h

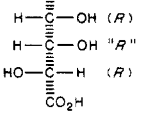


i



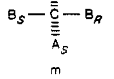
j

term *chirogen* (whence, *chirogenic* and *chirogenicity*), meaning *a stereogen that is variant to reflection*, should prove useful. This definition generally overlaps that of Mezey (Mezey, P. G. *J. Am. Chem. Soc.* **1986**, *108*, 3976–3984): "*a tetrahedral carbon center C* (is) *a chirogenic center if and only if no two substituents of C are mirror images of one another*", but there are important differences. Thus, the central atom of k is a chirogenic center under Mezey's definition and would be labeled *R* since " it is the center of a chiral molecule". It would receive no label under



$$\begin{array}{l} \quad\ \mathrm{CO_2H} \\ \mathrm{H-C-OH}\ \ (R) \\ \mathrm{H-C-OH}\ \ "R" \\ \mathrm{HO-C-H}\ \ \ (R) \\ \quad\ \mathrm{CO_2H} \end{array}$$

k



$$\begin{array}{c} A_R \\ B_S-C-B_R \\ A_S \end{array}$$

m

our definition because it is not stereogenic. On the other hand, the central atom of m, which now has *two sets* of mirror image substituents (contrast a, which has only one pair), is stereogenic when A ≠ B and inverts on reflection. It would, therefore, be chirogenic under our definition [*R* as shown, since ($A_R > A_S$), ($B_R > B_S$) regardless of the sequence rule ranking of A and B] but not under Mezey's. Should this term prove useful, the invention of a second term, for *stereogens that do not invert on reflection*, might become necessary. It would seem best to allow Necessity a full term of gestation on this.

(6) Cahn, R. S.; Ingold, C.; Prelog, V. *Angew. Chem., Int. Ed. Engl.* **1966**, *5*, 385–415. The notational system developed in this and earlier papers (op. cit.) is commonly refered to as CIP.

(7) Blackwood, I. E.; Gladys, C. L.; Loening, K. L; Petrarca, A. E.; Rush, I. E. *J. Am. Chem. Soc.* **1968**, *90*, 509–510.

(8) Haas, G.; Prelog, V. *Helv. Chim. Acta* **1969**, *52*, 1202–1218. Haas, G.; Hulbert, P. B.; Klyne, W.; Prelog, V.; Snatzke, G. *Helv. Chim. Acta* **1971**, *54*, 491–509.

(9) Mislow, K.; Siegel, J. *J. Am. Chem. Soc.* **1984**, *106*, 3319–3328. A chirotopic atom is one that lies in a chiral environment. For our purposes the importance of this concept is that it makes it clear that chirality does not reside in stereogens and that it is the chiral sense of the *whole* molecule that matters. This in turn suggest that the holistic chiral sense has the significance of the signature ("right- or left-handed") of a set of Cartesian coordinates. Since a given construction can be described in terms of *either* right- or left-handed coordinates, it follows that "absolute" configuration is purely a matter of convention and is, in fact, relative to some external ideal chiral figure. This, finally, serves to demystify topography, which, it is now seen, is *not* more fundamental than geometry.

(10) Seebach, D.; Prelog, V. *Angew. Chem., Int. Ed. Engl.* **1982**, *21*, 654–660.

(11) *Chemical Abstracts, 9th Collective Index, Index Guide*, **1977**, pp 185I–188I.

(12) For an early use of the bipartite format, see: McCasland, ref 5, p 8 ff.

(13) It is a useful convention[3,6] that descriptors of topography be printed in capitals and descriptors of geometry in lower case. It is unfortunate that the *E,Z* notation,[7] which is a general CIP-based notation for the geometry of olefins, does not adhere to this convention (as, *e,z*).

(14) Maehr (Maehr, H. *J. Chem. Educ.* **1985**, *62*, 114–120) addresses the same problem we do here, as it relates to structural formulas. It is our belief that his convention will prove awkward and confusing and that it might prove simpler to attach labels to numerical signatures. A stereoformula necessarily suggest one enantiomer. The enantiomer shown here as 1, then, could be referred to as 1 or, for emphasis, as *sic*-1, its mirror image would be *ent*-1 and the racemate *rac*-1. If the absolute configuration were unknown (often still the case with X-ray structures for natural products), the drawing could be labeled 1*; this would allow usages such as (±)-1*, (+)-1*, and (−)-1* for racemic and nonracemic samples of substances for which only the geometry was known.

as *cis*, *trans* (*c*,*t*)[15] and α,β,[11] and has many advantages. It is a simple matter to compare centers not explicitly paired; thus C_3 and C_5 are both *u* to C_2 and are, therefore, *l* to one another. Lack of information on one stereogen has no effect on the designations of others. Topography is readily recovered if the absolute configuration of the reference stereogen is shown, as in the *bipartite* formulation: *S* (2-*ref*,3*u*,4*l*,5*u*).

Chemical Abstracts Services has abundantly demonstrated the value of such bipartite notations.[11,12] In their usage, *RS* descriptors are first assigned to each of the chirogens in a specific enantiomer. If the first center is *R* then the descriptors are simply starred to convert them to geometric descriptors; if the first center is *S* (as in 1) they must be inverted before being starred. The topographic descriptor for the reference stereogen is shown, with a locant if necessary: 2*S*-(2*R**,3*S**,4*R**,5*S**). This notation has many strong points. Each stereogen has one descriptor and one locant. No one kind of pairwise relationship is stressed over another; it is immediately evident that both C_3:C_4 and C_2:C_5 are *u*. The recovery of topography may, however, require a curious double negative logic; thus, in *sic*[14]–1: C_5, being *S**, is *S* if C_2 is *R*—but C_2 is *S*, so C_5 is *R*. The principal disadvantages of this notation are conceptual. Topographical symbols, albeit somewhat modified by starring, are used to designate geometry; this tends to confuse the two and to suggest that differences in geometry are in some way less basic than differences in topography, indeed, something of an afterthought.

This defect vanishes if we admit that the reference atom is *like* itself, or 2*l* in the notation following Prelog and Helmchen.[16] Now a clear geometric descriptor can be applied to each center: *S*-(2*l*,3*u*,4*l*,5*u*). *With no particular center as reference, the topographic descriptor S can now be considered to apply to an external ideal reference*, so that the individual centers are *l* or *u* to it. The same global geometric descriptor (2*l*,3*u*,4*l*,5*u*) can be derived directly from either enantiomer without reversal and topography is readily recovered for each. The use of a single descriptor of topography emphasizes the point that there are only two enantiomers corresponding to this geometry and that their chirality does not derive from localized "chiral centers" but is a property of the whole molecule; all of the atoms of 1 are chirotopic.[9]

The complete separation of topography and geometry that results from the use of a bipartite format *and* external referencing makes this notation concise and flexible in its application to mixtures of stereoisomers and to substances not yet fully characterized as to configuration, that is, to cases often encountered in practice. Thus, the racemate of 1 would be indicated by use of the topographical pair *RS*; the individual enantiomers can correctly and holistically be identified as the "*R* isomer" or the "*S* isomer". A sample with 90% ee of *R* isomer could be fully described as [95*R*:5*S*](2*l*,3*u*,4*l*,5*u*).[17] If the absolute configuration of a substance is not known then that could be indicated by using *X* as the topographic descriptor; in most cases, sign of rotation would be more useful. Thus, for a partially resolved sample of 1, the following obtains: $[75+; 25-]_D$ (2*l*,3*u*,4*l*,5*u*). Configuration at one center might be unknown; that could be indicated by *x*: *R*(2*l*,3*x*,4*l*,5*u*). One center might be partially epimerized: *S*-(2-[85*l*:15*u*],3*l*,4*u*,5*l*). The configurational relationship of two segments of the molecule might not yet be known: *X*-(2*l*,3*u*),*R*-(4*l*,5*u*).

A comment about the *pref–parf* notation of Carey and Kuehne[18] and the corresponding *erythro–threo* notation of Noyori[19] is required. They provide a *direct* geometric description of the relationship between two chirogens and, in that sense, are more fundamental than are the *l*,*u* descriptors, which are derived by comparisons of topographic descriptors assigned individually to the stereogens. Both notations are general, being based on CIP, and they are not conformation-bound as most other notations are. They can be recommended for the common and important case where one particular pairwise relationship is to be specified. For our purposes, however, which include the specification of relationships among more than two stereogens, they are quite unsuitable. At each stereogen only one of the ligands deriving from the main chain (depending on which side the companion stereogen lies) is taken into account. Thus, making the serial comparison that is specified for systems with multiple centers,[18] 1 would be identified as the 2,3-*pref*;3,4-*pref*;4,5-*pref* diastereomer. Stereogens which are *pref* to the same center may be *pref* to one another as C_2 and C_4 are or *parf* as C_3 and C_5 are. This difficulty is not relieved by using C_2 as an internal reference, when 1 becomes 2,3-*pref*;2,4-*pref*;2,5-*parf*. Now stereogens having opposite relationships to the reference center are sometimes *parf* to one another (C_3 and C_5) and sometimes *pref* (C_4 and C_5). External referencing is impossible. It is not possible to obtain the CIP configuration of a companion center *directly*; thus, a stereogen that is *pref* to an *S* center may be either *R* or *S*, whence each CIP assignment must be made de novo.

---

(15) IUPAC Commission on Nomenclature of Organic Chemistry. *Pure Appl. Chem.* **1976**, *45*, 13–27.

(16) This is no more absurd than the point that the identity operation must be included in a symmetry table or that "nothing" merits a numeral (zero).

(17) The normalized enantiomer ratio is more directly related to the observations made in determining enantiomeric composition by NMR or chromatography than is enantiomer excess, which harks back to optical rotation methods for assigning "optical purity".

(18) Carey, F. A.; Kuehne, M. E. *J. Org. Chem.* **1982**, *47*, 3811–3815.

(19) Noyori, R.; Nishida, I.; Sakata, J. *J. Am. Chem. Soc.* **1981**, *103*, 2106–2108. The terms *erythro* and *threo* have become too corrupted to be retained.