131. *Specification of Configuration about Quadricovalent Asymmetric Atoms.*

By R. S. Cahn * and C. K. Ingold.

A convention, termed the sequence rule, is proposed whereby the groups a, b, c, and d attached to a quadricovalent asymmetric atom (Xabcd) can be correlated, each to each, with the OH, CHO, CH_2·OH, and H of glyceraldehyde. According as this correlation leads to dextrorotatory or lævorotatory glyceraldehyde, so the asymmetric atom X is assigned a D- or a L-configuration, and a D or a L symbol is used in the name of the compound. The resulting configurations embrace, as specific cases, previous uses of these prefixes in carbohydrate and amino-acid chemistry; they clarify nomenclature where this has been in dispute, notably when more than one asymmetric atom is present in the molecule, and they provide unambiguous definitions where this was not previously possible. Some representative examples are discussed.

The most extensive single class of optically active molecules consists of those whose dissymmetry can be completely factorised into the asymmetry of quadricovalent atoms. The configuration of molecules of this class may be fully specified by stating the configuration about each asymmetric atom. The purpose of the present paper is to suggest conventions through which this can be done compactly enough for inclusion in the name of a substance.†

In the present state of knowledge, the most satisfactory procedure is, first, to choose, as

---

* This paper presents proposals which are not to be construed as already settled editorial policy.

† The following types of dissymmetry can lead to optical activity, but are not covered in this discussion, namely, (1) that due to asymmetric atoms of covalency other than four, for instance, six, (2) that associated with atomic asymmetry depending on configurational differences in the adjoining groups, as in dissymmetry of the inositol type, of the allene type, and of related semicyclic-unsaturated, and spiran types, and (3) that unconnected with atomic asymmetry, as in diphenyl derivatives and other structures which show restricted rotation about single bonds.

" standard substance," a single optically active compound, and to assume a configuration for it; and then to adopt a " correlation rule," that is, a formula by means of which the groups about any asymmetric quadricovalent atom may be severally correlated with those about the asymmetric atom in the molecule of the standard substance.

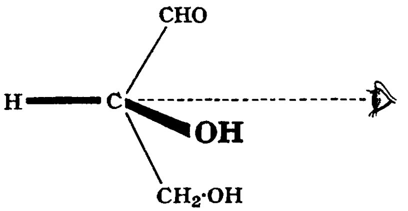
The time will come when absolute configuration can be determined with certainty; and then " standard substance " will be a redundant concept. Anticipating this, we shall express the " correlation rule " in a generalised form, calling it a " sequence rule," so that it can be used without change after the concept of " standard substance " has been discarded. But for the time being we retain this concept.

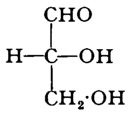
Both in choosing a standard, and in correlating other configurations with it, we are seeking to connect (1) a direction of rotation, (2) a substance, (3) a configurational symbol, (4) a three-dimensional model, and (5) a two-dimensional graphic formula. The last of these is inessential, because the model describes the structure, and the symbol should identify the model; the graphic formula is an additional convenience. The connexion (1)–(2) is established by observation, (2)–(3) by definition, (3)–(4) by an unproven assumption, and (4)–(5) by a convention. None of these determinations can conceivably be taken out of our hands, except the assumption. If, when absolute configurations can be ascertained, that should prove to be wrong, we should still not have to change the configurational symbols : this is what we gain by retaining the concept of " standard substance." But we should have to interchange the models of enantiomers; and we should have either to interchange the two-dimensional formulæ, or to alter the convention by which they are derived from the three-dimensional models.

In order to conform to an already agreed convention, we select glyceraldehyde for the standard substance, but not now as fiducial merely for hydroxy-acids and carbohydrates : we shall seek to describe all configurations that can be expressed entirely in terms of configuration about asymmetric quadricovalent atoms.

In establishing the above connexions for glyceraldehyde itself, $C(H)(OH)(CH_2{\cdot}OH)(CHO)$, we start with the general concept of dextrorotation, observe that a certain isomer of glyceraldehyde has it, define that substance as having the configuration labelled D, assume that its model (represented in Fig. 1) contains the groups OH, CHO, $CH_2{\cdot}OH$ in clockwise order as viewed from the side remote from H (this being one of four similar and equivalent descriptions *), and agree that this model may be represented by the formula (I). We similarly connect lævorotation with the other isomer, the configuratonal designation L, the enantiomeric model, and the formula with the lateral groups interchanged.

FIG. 1.





$$\begin{array}{c} CHO \\ | \\ H-C-OH \\ | \\ CH_2{\cdot}OH \end{array}$$

(I.)

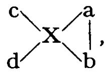
The standard being thus set up, it is necessary next to formulate a rule by which one Group of the four bound to any given asymmetric quadricovalent atom is unambiguously selected for correlation with the Group H of glyceraldehyde, another for correlation with OH, another with $CH_2{\cdot}OH$, and the remaining one with CHO. The final step will be to assign to the given asymmetric atom the configurational symbol, D or L, associated with the particular glyceraldehyde produced when each of the four Groups directly bound to the asymmetric atom is replaced by that Group of glyceraldehyde with which it is correlated.

The crux of the present proposal is thus the " rule; " and, as stated above, we prefer to formulate it, not simply as a " correlation rule," but in the generalised form of a " sequence

---

* The four descriptions result from viewing the model in four tetrahedrally inclined directions. The description given in the text is the most convenient one from which to generalise to other molecules. But another of the four descriptions, namely, that H, OH, $CH_2{\cdot}OH$ stand clockwise as viewed from the side remote from CHO, is the immediate basis of the projection formula (I). Any of the four descriptions can be reduced to projection formulæ on the usual convention that lines drawn upwards and downwards from the asymmetric atom represent bonds which project behind, and lines drawn to left and right bonds which project in front of, the plane of the diagram. (Naturally, the equivalent representations thus derived are not necessarily interconvertible by rotations in the plane of projection.)

rule," that is, a procedure placing any four different Groups in order. By applying such a rule both to the four Groups bound to the given asymmetric atom, and to the four bound to the asymmetric atom of glyceraldehyde, a one-to-one correlation between the two series of Groups is effected.

The terms Group and Set, written with an initial capital letter, will be used with a definite significance. A Group is an atom or an assembly of atoms bound to the asymmetric atom by a single valency; that is, it is the a, b, c, or d of the Xabcd under consideration. If the asymmetric atom is part of a ring, as in , there are Groups –ab and –ba containing the same atoms but in a different order. The term Group has then not always the significance normally attached to a "group" in organic chemistry. A Set comprises certain atoms, specified by the rule, which are considered in association when applying the rule.

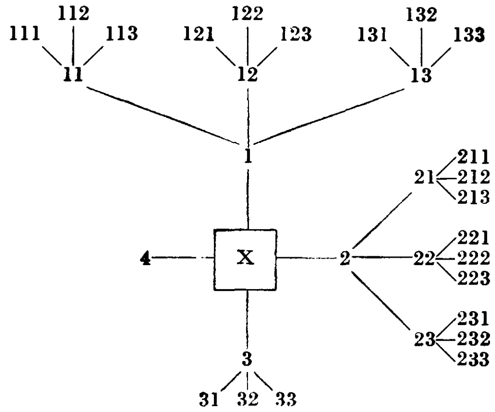
"SEQUENCE RULE," OR METHOD FOR ORDERING GROUPS BOUND TO A QUADRICOVALENT ASYMMETRIC ATOM.

General Procedure.—The Groups are placed in order of decreasing atomic numbers of certain selected Sets of their atoms. These Sets are selected successively by the method, and following as far as is necessary the order of priorities, given hereunder.

Method and Priorities of Selection of Sets of Atoms.—The Sets are selected by reference to a single valency structure of the molecule, that is, one structure containing only localised bonds of integral multiplicity. Methods for the resolution of possible ambiguities in the choice of such a structure are noted later.

(First Set.) To the asymmetric atom let four atoms, 1, 2, 3, and 4 be directly bound. Then the atoms 1, 2, 3, 4 constitute the First Set of atoms.

Fig. 2.



(Second Set.) Let 1 and 2 be any two atoms of identical atomic number in the First Set. To 1 let there be directly bound atoms 11, 12, 13, . . ., here arranged as far as possible in order of decreasing atomic number (see Fig. 2). To 2 let there be directly bound atoms 21, 22, 23, . . ., similarly here arranged. If any of the atoms 11, 12, . . ., 21, 22, . . ., be directly bound to 1 or 2 by a double or triple bond, let that atom be counted twice or thrice, as the case may be, in the series of atoms 11, 12, . . ., or 21, 22, . . . If, after any such augmentation, one of the series of atoms, 11, 12, . . ., and 21, 22, . . ., contains fewer atoms than the other series, let imaginary atoms of atomic number zero be added to the series containing the fewer until each contains the same number of atoms. Then the Second Set contains 11 and 21, unless these have the same atomic number, when it contains instead 12 with 22, unless these also have the same atomic number, when it contains instead 13 and 23, and so on. The complete Second Set of atoms is to be composed by applying the same procedure to every pair of Groups whose atoms of the First Set have identical atomic number.

(Third Set.) Let the atoms 1 and 2 have identical atomic number, and let the atoms 11, 12, . . ., and 21, 22, . . ., be identical in atomic number in pairs, 11 with

21, 12 and 22, and so on throughout. Then the atoms of the Third Set are selected from the atoms 111, 112, . . ., directly bound to atom 11, and the atoms 211, 212, . . ., directly bound to atom 21, as were the atoms of the Second Set from the atoms, 11, 12, . . ., directly bound to atom 1, and the atoms 21, 22, . . ., directly bound to atom 2; unless, however, all the atoms from which the atoms of the Third Set might thus have been selected are identical in atomic number in pairs, each of the series 111, 112, . . ., with one of the series 211, 212, . . ., when the atoms of the Third Set are similarly selected from the atoms 121, 122, . . ., directly bound to atom 12, and the atoms 221, 222, . . ., directly bound to atom 22; and so on. The complete Third Set of atoms is to be composed by generalising the procedure illustrated to every pair of Groups having relationships of identity similar to those assumed for the Groups to which atoms 1 and 2 belong.

(Other Sets.) The method for selection of the $(n + 1)$th Set of atoms is to be developed from that of the nth Set as was that of the Third Set from that of the Second Set.

Supplementary Procedure for taking into Account Dissymmetry dependent on Isotopic Differences.—The General Procedure having failed to yield a complete order of the Groups, it is repeated with respect to those Groups only which are not yet given an unambiguous order, but with the difference that the Groups are now placed in order of decreasing atomic masses of Sets of their atoms selected by the method and according to the priorities specified above.

Conventions for Selection of a Valency Structure.—The sequence rule is formulated on the assumption that the molecule to which it is applied is being represented by a unique valency structure, that is, one structure having all its bonds localised and of integral multiplicity. Since many molecules cannot thus be represented accurately and for all purposes, it becomes necessary to adopt conventions by which a unique valency structure may be assigned to a molecule for the purpose of applying the rule. The following suggested conventions are intended to provide the more obvious necessities of this kind; they do not exhaust every rarely occurring possibility.

(1) Hyperconjugation is neglected. (For example, the bonds of an alkylacetylene or of a propargyl halide are taken as single or triple when applying the sequence rule.)

(2) Participation of atomic d orbitals in bonding is neglected; that is, octet structures are adopted whenever possible. (For example, all bonds of the sulphur atoms in sulphoxides and sulphones are taken as single for the purposes of the rule.)

(3) To any mesomeric molecule its usual valency structure is assigned, that is, the structure considered to be most closely related to its normal mesomeric state. (For example, the structure C:C·C:C is assigned to a buta-1 : 3-diene, Cl·C:C to a vinyl chloride, C:C·C:O to an acraldehyde or a vinyl ketone, O:C·OR to a carboxylic ester, O:C·NR_2 to a carboxyamide, the non-dipolar two-double-bond structure to a furan or other five-membered aromatic ring compound, and a Kekulé or Kekulé-like structure to any benzene, pyridine, or other six-membered aromatic ring compound.)

(4) When alternative usual valency structures for aromatic molecules exist, one is selected in which the aromatic double bonds are introduced in the way determined by the following procedure. The aromatic ring is considered as composed of chains branching from the point to which is attached the side chain containing the asymmetric atom; these branches are compared according to the priorities of selection set out above, all bonds of the aromatic ring structure being at first counted as single bonds; double bonds of Kekulé type are then inserted into the ring structure starting at, or as close as possible to, the branching point, and proceeding along that branch which according to the priorities of selection would be cited first.

Assignment of Configuration.—The sequence rule places the four Groups bound to the asymmetric atom of glyceraldehyde in the order

OH CHO CH_2·OH H

(for detailed analysis see below). Therefore, if the rule places the four Groups directly bound to a given asymmetric quadricovalent atom in the order

$$\alpha \quad \beta \quad \gamma \quad \delta$$

then α is correlated with the Group OH, β with CHO, γ with CH_2·OH, and δ with H of glyceraldehyde; and the configurational symbol D or L is assigned to the asymmetric atom in the manner already stated.*

---

* The case of asymmetric tercovalent atoms could readily be covered by adding to such atoms an imaginary atom of atomic number zero. This would automatically become correlated with the H of glyceraldehyde.

By the assumption made concerning models, the Groups OH, CHO, $CH_2{\cdot}OH$ stand in clockwise order in the model of D-glyceraldehyde, as viewed from the side remote from the H. Therefore, according as the configurational symbol D or L is assigned to the given asymmetric atom, so in the model of the molecule which contains it, α, β, γ will stand in clockwise or counterclockwise order as viewed from the side remote from δ.

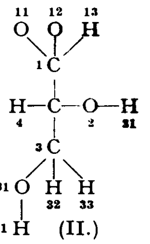
APPLICATION OF THE PRIORITY CONVENTIONS.

In principle the sequence rule prescribes that the atoms attached to the asymmetric atom shall be arranged in order of decreasing atomic number, and that, if two or more of them are the same, one " works outwards " until a decision is possible. This simple principle assumes the somewhat involved form of the rule set out above because of the need to define unambiguously how to " work outwards " as the chains ramify. In this section, the procedure will be applied to a few structures chosen to illustrate priority features which may not be obvious at first sight.

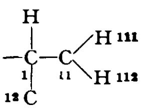
Analysis of the simple substance glyceraldehyde according to the rule is represented in (II). Comparison of the First Set gives the order of atoms, 2, (1, 3), 4. Decision between the CHO and the $CH_2{\cdot}OH$ rests on comparison of the Second Set; atom 11 is identical with atom 31, but atom 12 (O) has a higher atomic number than atom 32 (H). The complete sequence, in order of decreasing atomic number, is therefore 2, 1, 3, 4, that is, OH, CHO, $CH_2{\cdot}OH$, H, as stated above.

This example illustrates the representation of a multiple bond (C═O) by single bonds , but a further point about such representations requires exposition.

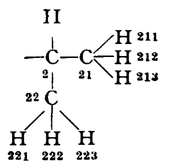
Suppose that a decision is required between Groups $^-CH{:}CH_2$ and $^-CH(CH_3)_2$. Analysis gives the structures (III) and (IV). The rule prescribes that one atom attached by a double bond shall be counted twice, in this case the carbon atom of the CH_2 group which appears in (III) as numbers 11 and 12; but this atom only is counted twice, and not also the atoms (H) attached to it. Then, when comparing the Set 11 and 21, or the Set 12 and 22, the valency difference must, according to the rule, be compensated by imaginary atoms of atomic number zero, denoted by " o " in (III*a*). The order of decreasing atomic numbers is thus settled at the three-figure Sets: (IV) precedes (III). Imaginary atoms of atomic number zero are also used in other cases, *e.g.*, to establish precedence of $^-NO_2$ over ^-NO, of $^-NH_3{}^+$ over $^-NH_2$, of $^-SO_3H$ over SO_2H, etc.; such priorities would not otherwise have been defined.



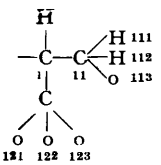
(II.)



(III.)



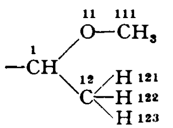
(IV.)

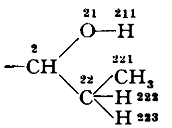


(III*a*.)

In the rule, the atoms 11, 12, 13, . . ., 21, 22, 23, . . ., etc., are arranged in order of decreasing atomic number (second sentence of the paragraph relating to the Second Set), and where necessary the atoms 111, 112, . . ., are compared with the atoms 211, 212, . . . before proceeding to a comparison of 121, 122, . . . with 221, 222, . . . The necessity for this is best explained by an example such as (V) and (VI). No decision is possible at the Second Set (two-figure numbers). In the Third Set, comparison of 111 with 211 gives the order (V) before (VI), and this is the decision of the rule since atoms 11 and 21 (O) have a higher atomic number than atoms 12 and 22 (C). Without such a provision in the rule, comparison of 121 with 221 might have reversed the priority.

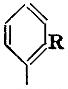
(V.)





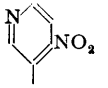
(VI.)

The method for selection of bond structures for aromatic rings has been arbitrarily made the same as that for determining priorities by the general rule, solely in order to have only one method in use in the procedure as a whole. Some structures thus selected are illustrated in the examples (VII)—(XV). The absence of any common chemical significance attaching to these particular structures is not considered a disadvantage. The structures are derived from the order of decreasing atomic number, and thus they as well as the rest of the sequence rule are divorced from such chemical theory as may change with the progress of science; it has been a main object to place the sequence rule on a non-ephemeral physical basis.



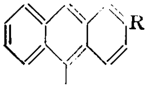
(VII.)

(VIII.)

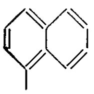


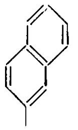
(IX.)

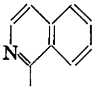
(X.)



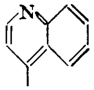
(XI.)


(XII.)


(XIII.)

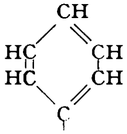


(XIV.)

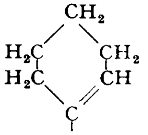


(XV.)

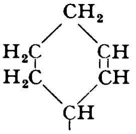
The priorities assigned by the rule to cyclic structures may be illustrated by selected examples. The order (XVI)—(XIX) is clearly as shown. So also is the order (XX)—(XXIII), followed by (XVI), etc., the ring junctions of α-naphthyl (XXI) being represented as in (XXI*a*). *cyclo*Hexyl (XIX) precedes *n*-hexyl (XXVII), the decision being taken already at the first carbon atom; *cyclo*hexyl also precedes 1-*n*-hexyl-*n*-heptyl (XXV) and longer radicals, the decision being taken at the " seventh " carbon atom which in *cyclo*hexyl is the original point of entry into the ring where the " chain " branches; for an " opened ring " such as (XXVI) the decision is reached at an earlier stage; *tert.*-butyl, however, precedes *cyclo*hexyl; the complete priority order of these Groups is as given in the row of formulæ. In all these examples also the absence of a chemical logic is noticeable, and again no apology is made for this.



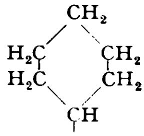
(XVI.)



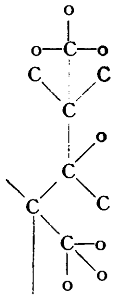
(XVII.)



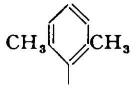
(XVIII.)



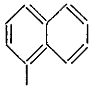
(XIX.)

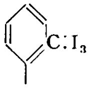


(XXI*a*.)

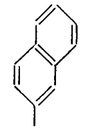


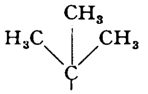
(XX.)


(XXI.)

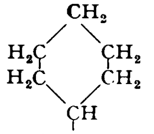


(XXII.)

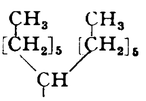

(XXIII.)



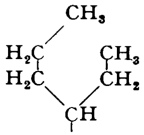
(XXIV.)



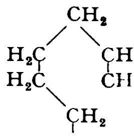
(XIX.)



(XXV.)



(XXVI.)



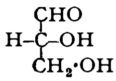
(XXVII.)

Other features of the sequence rule are, it is hoped, self-explanatory.

EXAMPLES OF THE SEQUENCE RULE.

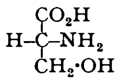
Nomenclature.—The symbols D and L chosen for use with the sequence rule should invariably be preceded by the numeral or letter used to designate the appropriate asymmetric atom; this symbol should be attached to the name of a substituent present on the asymmetric atom or, if no such substituent is named, to that part of the name which embraces the asymmetric atom in question. Examples are αL-alanine, βD-amino-αL-hydroxybutyric acid, 4D-hydroxy-2D-proline, and 9D : 10D-α-decalone. The attached positional numeral or letter supplies differentiation from previous uses of the letters D and L.

Brewster, Hughes, Ingold, and Rao (*Nature*, 1950, **166**, 178) have rigidly completed the genetic chain between D-glyceraldehyde (I) and D-serine (XXVIII); the sequence rule also



$$\begin{array}{c} \mathrm{CHO} \\ \mathrm{H{-}C{-}OH} \\ \mathrm{CH_2{\cdot}OH} \end{array}$$

(I.)



$$\begin{array}{c} \mathrm{CO_2H} \\ \mathrm{H{-}C{-}NH_2} \\ \mathrm{CH_2{\cdot}OH} \end{array}$$

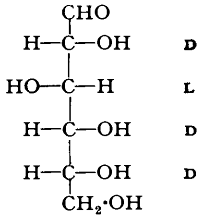
(XXVIII.)

ascribes the D-configuration to (XXVIII), so that the prefix D deduced by the sequence rule and based on D-glyceraldehyde as " standard substance " corresponds also to the convention for serine. It is this correspondence, indeed, the large overall area of agreement between the proposed general notation and accepted sectional notations, that makes it advisable to use D and L to denote configurations deduced by the sequence rule (which was, in fact, framed so that it should yield a maximum of such agreement).

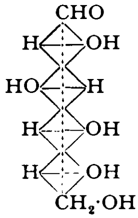
Thus the way is cleared for a rational nomenclature applicable both to carbohydrates and to amino-acids, without serious disturbance of existing conventions, and to a much wider field. The need for such a joint nomenclature follows from the fact that the prefix D, or more specifically D$_g$, for carbohydrates refers to the configuration at the highest-numbered asymmetric carbon atom, whereas the prefix D, or more specifically D$_s$, for amino-acids refers to the configuration at the α-carbon atom (corresponding to $C_{(2)}$ of an aldose sugar). Thus L$_s$-threonine is 2-amino-2 : 4-dideoxy-D$_g$-threonic acid. It will be shown below how such anomalies are overcome by use of the sequence rule.

For hexoses, pentoses, and their derivatives the universally accepted trivial names define the stereochemistry of each asymmetric atom; α-D-glucofuranose, for instance, can be only one substance, and there is no need for further prefixes specifying the D- or L-configuration of carbon atoms 2—5. Similarly the universally accepted names for amino-acids need not be modified, except by the addition of the positional numeral or letter required for use with the sequence rule (as in αL-alanine). But there are unresolved nomenclature problems and some confusion in both these fields, and the sequence rule and nomenclature based on it offer rational solutions.

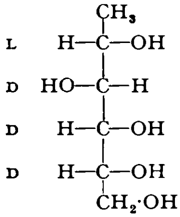
Carbohydrates.—Application of the sequence rule to the *aldehydo*-form of D-glucose (XXIX*a* = *b*) shows that the carbon atoms 2, 3, 4, and 5 are respectively D, L, D, and D, *i.e.*, D-atoms all have OH to the right in the conventional formula (XXIX). Owing to the representation through which double bonds are treated by the sequence rule, these results remain the same in any semiacetal form of D-glucose, as well as in the derived glucosides. The same prefixes for the individual asymmetric atoms apply to the glucose derivatives in which



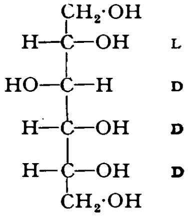
(XXIX*a*.)



(XXIX*b*.)

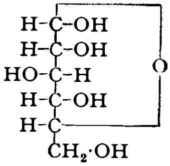


(XXX.)

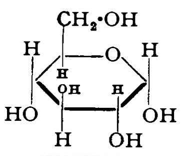


(XXXI.)

CHO is replaced by groups such as CN, CO_2H, $CO{\cdot}NH_2$, which precede $-CH(OH){\cdot}C$ by the sequence rule. If, however, as in (XXX), the CHO group is replaced by CH_3, which follows $-CH(OH){\cdot}C$ by the sequence rule, the configurations at $C_{(2)}$ and $C_{(3)}$ are reversed. It may be considered that the pentol (XXX) can barely be classed as a sugar derivative, but a similar



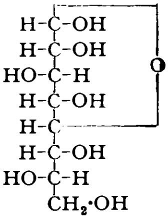
(XXXII*a*.)

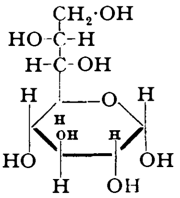


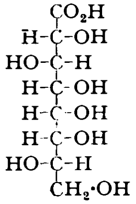
(XXXII*b*.)

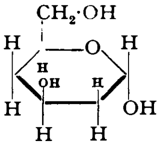
" reversal " is observed with D-sorbitol (XXXI). It is clear, therefore, that the relative positions of the H and OH in the conventional formulæ are no guide to the prefix to be attached to the separate atoms and that each case must be considered individually.

In α-D-glucopyranose (XXXII*a* = *b*) carbon atom number 1 has the L-configuration by the sequence rule; this substance might therefore be unambiguously named as 1L-D$_g$-glucopyranose or as 2L : 3D : 4L : 5D-tetrahydroxy-6D-hydroxymethylpyran, but there seems no need to disturb the generally accepted and simpler α-D-glucopyranose. Similar considerations apply to glycosides. L. F. Fieser (*J. Amer. Chem. Soc.*, 1950, **72**, 623) has suggested that the configuration of individual atoms of pyranose and furanose sugars and their derivatives can be defined as α or β according to whether the hydrogen atom lies below or above the plane of the ring according to the conventional cyclic sugar formula. This convention is however restricted to cyclic carbohydrates and is thus less generally useful than the sequence rule.


(XXXIII*a*.)


(XXXIII*b*.)


(XXXIV.)


(XXXV.)

In the above and similar cases the present carbohydrate nomenclature is simpler than that which follows from the sequence rule, and need not be changed; but we shall next mention briefly three classes of compounds where the sequence rule can be usefully employed. There is not yet agreement on trivial names for sugars containing seven or more carbon atoms. A nomenclature based on the sequence rule is sufficiently exemplified by the names α-D-gluco-6D : 7L-octapyranose for (XXXIII*a* = *b*), and D-gluco-6D : 7L-octonic acid for (XXXIV); the α-D-gluco defines the configurations at positions 1—5, and those of the remaining asymmetric atoms are specifically added thereafter. In these names, it is at once evident that the 6D and 7L are derived by the sequence rule, since here the D and L are associated with positional numerals; and it should be clear that the D-gluco derives from the carbohydrate convention, because the D is attached to gluco and not associated with a position symbol: such names, involving mixed usage, are really abbreviations of, *e.g.*, α-D$_g$-gluco-6D : 7L-octapyranose, and the fuller names should be employed if the distinction in usage is not considered to be obvious otherwise.

Conversion of a CH(OH) group into CH_2 leaves open a choice of sugar names from which to derive that of the deoxy-sugar. According to the sequence rule, (XXXV) may be unambiguously termed 2L : 4L-dihydroxy-6L-hydroxymethyltetrahydropyran or, if it is desired to retain a sugar terminology, 2 : 4-dideoxy-1L : 3L : 5L-hexapyranose or perhaps 2 : 4-dideoxy-α-3L : 5L-hexapyranose.

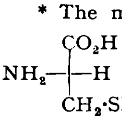
Current research on amino-sugars and amino-deoxy-sugars is leading to substances which can be considered as lying in the field of amino-acid chemistry, so that interpenetration of the D$_g$ and D$_s$ systems may become serious. The discussion of amino-acids which follows will show how the sequence rule can here too be used with advantage.

Amino-acids.—It has already been pointed out that the sequence rule and the standard convention accord the same configuration to serine; and the same is true for almost all the optically active amino-acids with only one (α-)asymmetric atom, for the order of priority will be NH_2, CO_2H, R, H for these substances.* It remains only to show how the sequence rule can be of benefit when more than one asymmetric atom is present, the previous conventions having led to difficulties in such cases. We shall use δ-hydroxylysine, β-hydroxyaspartic acid, 4-hydroxyproline, and threonine as examples.

The δ-hydroxylysine (XXXVI) has the L$_s$: D$_g$-configuration; we are concerned also with its diastereoisomeride (XXXVII) and the enantiomorphs of these two. According to the sequence rule, (XXXVI) has the L-configuration at the α- and at the δ-position and can

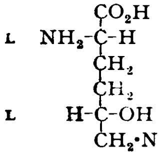
---

* The main exception known to us is cysteine (and, of course, cystine). The natural form, L$_s$-cysteine (inset), becomes αD-cysteine when named in accordance with the sequence rule. Such cases appear to be rare. In the literature of specialist fields, subscripts are often elided from the prefixes D$_g$, L$_g$, and D$_s$, L$_s$; but for cysteine, and in any other such examples, should they arise, it will be unwise either to drop the subscript when naming by a specialist convention, or, of course, to omit the position symbol when using the more general system here suggested.



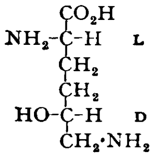
$$\begin{array}{c} CO_2H \\ NH_2-\!\!\!-|-\!\!\!-H \\ CH_2{\cdot}SH \end{array}$$

accordingly be termed δL-hydroxy-αL-lysine. Its diastereoisomeride (XXXVII) can similarly be named δD-hydroxy-αL-lysine. The enantiomorphs become δD- and δL-hydroxy-αD-lysine.



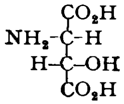
$$\begin{array}{c} \mathrm{CO_2H} \\ \mathrm{NH_2\text{-}C\text{-}H} \\ \mathrm{CH_2} \\ \mathrm{CH_2} \\ \mathrm{H\text{-}C\text{-}OH} \\ \mathrm{CH_2{\cdot}NH_2} \end{array}$$

(XXXVI.)



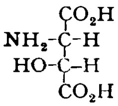
$$\begin{array}{c} \mathrm{CO_2H} \\ \mathrm{NH_2\text{-}C\text{-}H} \\ \mathrm{CH_2} \\ \mathrm{CH_2} \\ \mathrm{HO\text{-}C\text{-}H} \\ \mathrm{CH_2{\cdot}NH_2} \end{array}$$

(XXXVII.)



$$\begin{array}{c} \mathrm{CO_2H} \\ \mathrm{NH_2\text{-}C\text{-}H} \\ \mathrm{H\text{-}C\text{-}OH} \\ \mathrm{CO_2H} \end{array}$$

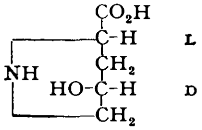
(XXXVIII.)



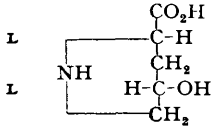
$$\begin{array}{c} \mathrm{CO_2H} \\ \mathrm{NH_2\text{-}C\text{-}H} \\ \mathrm{HO\text{-}C\text{-}H} \\ \mathrm{CO_2H} \end{array}$$

(XXXIX.)

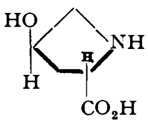
The β-hydroxyaspartic acid (XXXVIII) can be semisystematically named α′L-hydroxy-αL-aspartic acid (or, quite systematically, αL-amino-α′L-hydroxysuccinic acid), and its diastereoisomeride (XXXIX) is then α′D-hydroxy-αL-aspartic acid. The necessity for special "trivial" prefixes, such as Dakin's para and anti (*J. Biol. Chem.*, 1921, **48**, 273), is avoided by use of the general rule.



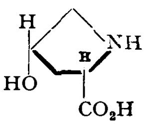
(XL*a*.)



(XLI*a*.)



(XL*b*.)



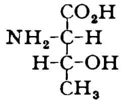
(XLI*b*.)

Difficulty has attended attempts to reach internationally an agreed nomenclature for the diastereoisomerides (XL*a* = *b*) and (XLI*a* = *b*), which are often termed 4-hydroxy-L-proline and *allo*-4-hydroxy-L-proline, respectively. According to the sequence rule, (XL) has the L-configuration at position 2 and the D-configuration at position 4, and is accordingly named 4D-hydroxy-2L-proline. Similarly (XLI) is 4L-hydroxy-2L-proline. The enantiomorphs are 4L- and 4D-hydroxy-2D-proline respectively. The *allo*-prefix may thus be abandoned (cf. below); the new names are precise and short in these and all similar cases, and conform to a general rule.

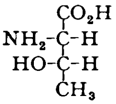
It may be noted in further illustration of a previous remark, that in the conventional representation (XL*a*) the 4-hydroxyl group is on the left although the D-configuration attaches to this position. This contrasts with a novel suggestion (offered to the Subcommittee on Biochemical Nomenclature of the National Research Council, U.S.A.) which would name both the NH and the OH of (XL*a*) zephos (Z; west), and the OH in (XLI*a*) eos (E; east); such a convention would have quite limited applicability.

L$_s$-Threonine (XLII) and *allo*-L$_s$-threonine (XLIII) conform to the same pattern so far as the sequence rule is concerned. The former can be systematically named αL-amino-βD-hydroxybutyric acid or αL : βD-threonine, and the latter αL-amino-βL-hydroxybutyric acid or αL : βL-threonine. The names for the enantiomorphs are obvious. The names L-threonine and L-*allo*threonine, in which *allo* denotes the reversed configuration at the β-position, may not be willingly abandoned as they are rather generally used. Nevertheless, whereas the partial names gluco, ido, xylo, etc., used for carbohydrates, are quite satisfactory as they indicate precisely a definite steric relation of all the asymmetric atoms in each case, the prefix *allo* (meaning "the other") refers back to the configuration of the "normal" substance; * this might have been satisfactory if *allo* were used only for one pair of substances, or if the asymmetric atoms of all the "normal" substances bore the same relation to one another; as it is, the prefix *allo* has no systematic significance, and the international difficulties mentioned above have centred mainly around constructing a generalisation as to which of a pair of diastereoisomerides is the normal substance. It seems more satisfactory to abandon the prefix *allo* as a means of stereochemical differentiation—if not for threonine, then certainly for other cases and for the more remote threonine derivatives.

(XLII.)



$$\begin{array}{c} \mathrm{CO_2H} \\ \mathrm{NH_2\text{-}C\text{-}H} \\ \mathrm{H\text{-}C\text{-}OH} \\ \mathrm{CH_3} \end{array}$$



$$\begin{array}{c} \mathrm{CO_2H} \\ \mathrm{NH_2\text{-}C\text{-}H} \\ \mathrm{HO\text{-}C\text{-}H} \\ \mathrm{CH_3} \end{array}$$

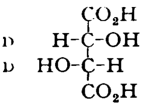
(XLIII.)

Some Special Cases.—Tartaric acid, obviously on account of its historical importance, was

---

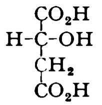
* If (XLII) is to be termed L$_s$-threonine, (XLIII) would have been more aptly named L$_s$-erythronine; but it seems to us preferable to place such names on a purely systematic basis.

chosen as a configurational standard by Fischer (*Ber.*, 1891, **24**, 2683). Freudenberg adopted the symbol D for the dextrorotatory acid (XLIV), and assigned as D-compounds the malic acid (XLV), the *iso*serine (XLVI), and the lactic acid (XLVII), which are genetically related to



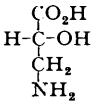
$$\begin{array}{rc} & CO_2H \\ D & H\text{-}C\text{-}OH \\ D & HO\text{-}C\text{-}H \\ & CO_2H \end{array}$$

(XLIV.)



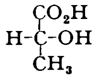
$$\begin{array}{c} CO_2H \\ H\text{-}C\text{-}OH \\ CH_2 \\ CO_2H \end{array}$$

(XLV.)



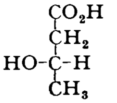
$$\begin{array}{c} CO_2H \\ H\text{-}C\text{-}OH \\ CH_2 \\ NH_2 \end{array}$$

(XLVI.)



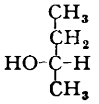
$$\begin{array}{c} CO_2H \\ H\text{-}C\text{-}OH \\ CH_3 \end{array}$$

(XLVII.)



$$\begin{array}{c} CO_2H \\ CH_2 \\ HO\text{-}C\text{-}H \\ CH_3 \end{array}$$

(XLVIII.)



$$\begin{array}{c} CH_3 \\ CH_2 \\ HO\text{-}C\text{-}H \\ CH_3 \end{array}$$

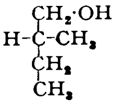
(XLIX.)

D-tartaric acid (" Stereochemie," F. Deuticke, Leipzig, 1932, p. 575). However, dextrorotatory tartaric acid is L_g according to the rule that configurations under the carbohydrate system refer to the highest-numbered asymmetric atom (cf. Wolfrom in Gilman's " Organic Chemistry," Wiley, New York, 1943, p. 1545), although it is D_s under the rule that configurations by the amino-acid system refer to the α-carbon atom. The assignments accorded by the sequence rule are shown in (XLIV), and the acid is best named αD : α′D-tartaric acid. In this way Fischer's and Freudenberg's conventions are preserved. *meso*Tartaric acid might analogously be called αD : α′L-tartaric acid, though the *meso*-prefix does not need alteration. The configurational symbols assigned by the sequence rule to malic acid, *iso*serine, and lactic acid similarly agree with Freudenberg's symbols.

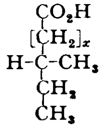
The prefix D or D_g customarily assigned to the lactic acid (XLVII) involves the special convention that it is formed from D-glyceraldehyde by the changes $CHO \longrightarrow CO_2H$ and $CH_2{\cdot}OH \longrightarrow CH_3$, rather than by $CHO \longrightarrow CH_3$ and $CH_2{\cdot}OH \longrightarrow CO_2H$. Since the sequence rule also accords the D-configuration to (XLVII), the special conventions can be allowed to lapse, this being merely one example of the general rule.

The dextrorotatory β-hydroxybutyric acid (XLVIII) and butan-2-ol (XLIX) have on occasions been assigned the L-configuration, although as Freudenberg points out (*op. cit.*, p. 679), there has been no justification for this. The sequence rule, however, accords the L-configuration to (XLVIII) and (XLIX), so that this assignment is now regularised by the general rule. Similar premature assignments can be similarly tested against the sequence rule as occasion arises.

(L.)



$$\begin{array}{c} CH_2{\cdot}OH \\ H\text{-}C\text{-}CH_3 \\ CH_2 \\ CH_3 \end{array}$$

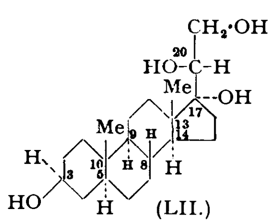


$$\begin{array}{c} CO_2H \\ [CH_2]_x \\ H\text{-}C\text{-}CH_3 \\ CH_2 \\ CH_3 \end{array}$$

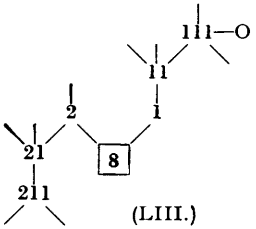
(LI.)

Crombie and Harper (*J.*, 1950, 2685) have recently suggested that the dextrorotatory 2-methylbutanol (L) shall be considered as the parent of a series to be denoted by the prefix D_a, and has proposed rules by which derivatives, notably acids of the type (LI), may be correlated with it and included in the D_a-series. The sequence rule ascribes the D-configuration to (L), so that adoption of the rule would render the subscript " a " superfluous; acids (LI) also have the D-configuration by the sequence rule, as will most derivatives of (L), but in deciding this and similar cases it is essential that the sequence rule and not the rules of Crombie and Harper shall be applied.

Lastly, as a complex example, let us consider Reichstein's substance K on the (unproven) assumption that the orientation (LII) is correct in all its details. Determination of the priority of the Groups is simple except for position 8; there the analysis proceeds as in (LIII), where



(LII.)



(LIII.)

bonds to carbon are indicated but bonds to hydrogen are omitted; the oxygen atom attached to atom 111 (position 17) determines the priority in the sense that the group staring with

$C_{(14)}$ precedes that starting with $C_{(9)}$. It transpires that, on the above assumption, the L-configuration is present at positions 3, 5, 9, 10, 13, and 14, and the D-configuration at positions 8, 17, and 20, and that accordingly Reichstein's substance K may be completely defined as 17-(1D : 2-dihydroxyethyl)-5L : 8D : 9L : 14L-perhydro-10L : 13L-dimethyl-*cyclo*pentanophenanthrene-3L : 17D-diol, or, more briefly, as *allo*pregnane-3L : 17D : 20D : 21-tetrol.

We are indebted to Drs. J. R. Johnson, W. T. Miller, and H. B. Vickery for helpful criticisms of a draft of this paper, and to Dr. L. C. Cross for many valuable discussions.

THE CHEMICAL SOCIETY, BURLINGTON HOUSE,
PICCADILLY, LONDON, W.1.
WILLIAM RAMSAY AND RALPH FORSTER LABORATORIES,
UNIVERSITY COLLEGE, LONDON, W.C.1

[*Received, November* 30*th*, 1950.]