Stereo Numbers, Cosets, and the Configuration Symmetry Group[†]

JOHN W. MCLEOD[1]

Computer Department Building D40A, Department of National Defense, FMO Halifax, Nova Scotia,
Canada B3K 2X0

Received October 15, 1985

The old (1959) stereo number designation of stereoisomers and the more recent configuration symmetry group method for identifying equivalent stereo number codes are described. Application of hexadecimal stereo number codes to families of aldoses, alditols, substituted inositols, and unsubstituted inositols is outlined. A novel matrix definition of configuration called the Configured Component Connectivity Matrix (CCCM), an extension of the Component Connectivity Matrix (CCM), leads to mathematical evidence for the soundness of these procedures.

INTRODUCTION

In 1959, Alfred Feldman introduced the stereo number system to code stereoisomers with binary numbers.[2] He had observed that no more than two symbols were ever needed to describe the configuration of a position such as an asymmetric carbon. By choosing "0" and "1" as the two symbols and then concatenating the codes for all asymmetric centers in a molecule, he achieved a binary number code for molecular configuration. He then went further, converting the binary code into more compact decimal and octal numbers. Figure 1 demonstrates the extraction of the decimal code 12 from D-mannose with a variation on his method.

S. Neelakantan[3] and H. A. Klein[4] have applied the stereo number system to inositol and carbohydrate nomenclature. Neelakantan saw that cis-inositol could be assigned the equivalent decimal stereo numbers 0 and 63 but that a unique code was available through choosing the lowest or highest of these. Figure 2 demonstrates how he arrived at the unique code "63".

In 1978, Nourse introduced a method for determining equivalence of stereo numbers called the configuration symmetry group.[5] This is, in essence, a group of actions on binary stereo number codes that preserve configuration. Each configuration symmetry is a permutation, written in cyclic form, of the digits of the binary stereo number. In addition to permuting the digits, the configuration symmetry may also change the sense of some digits. An inversion of configuration at digit n is then symbolized by replacing n where it appears in the permutation by n'. For example, the possible configuration symmetry $(1\ 2'\ 3\ 4')$ on a four-digit binary stereo number could be expressed by saying: 1 goes to 2 and inverts configuration; 2 goes to 3; 3 goes to 4 and inverts configuration; 4 goes to 1. Examples of this action on binary stereo numbers would include

$$[1100](1\ 2'\ 3\ 4') = [0011]$$

$$[0000](1\ 2'\ 3\ 4') = [0101]$$

$$[0010](1\ 2'\ 3\ 4') = [0100]$$

Composition of two possible configuration symmetries proceeds left to right, with the understanding that two inversions will cancel each other out. Thus, for example, $(1\ 2'\ 3')(4)\cdot(1\ 2'3\ 4') = (1\ 3'\ 2'\ 4)$. Nourse's group[5b] and, more recently, Abe and co-workers[6] have used stereo numbers and configuration symmetries to generate exhaustive lists of stereoisomers.

The group nature of this set of allowable operations on stereo numbers leads to some powerful results. Consider, for example, an arbitrary molecule with n asymmetric centers. The configuration symmetry $\mu = (1')(2')(3')...(n')$, which reverses

the configuration at each asymmetric center, is equivalent to the operation of ones complementing a binary stereo number, where the sense of each binary digit is reversed. If we let $\bar{a}$ be the ones complement of an n-digit binary stereo number a, then by definition $\bar{a} = a\mu$. The operation μ may not be in the configuration symmetry group, but it is easy to see that it commutes with all the possible values for a configuration symmetry group element. By this we mean that $\pi\mu = \mu\pi$ for any configuration symmetry π. Now suppose a and b are equivalent binary stereo numbers of the molecule. Then there must be a configuration symmetry group element π such that $b = a\pi$. But then $\bar{b} = b\mu = a\pi\mu = a\mu\pi = \bar{a}\pi$, and $\bar{b}$ is equivalent to $\bar{a}$. We see from this that a class of equivalent stereo numbers must contain either no pairs of ones complements (an enantiomer) or only pairs of ones complements (a mesoform). In the case of an enantiomer, the ones complements of all the enantiomers contained stereo numbers form another distinct enantiomer, completing the enantiomeric pair.

Suppose we were to choose for our unique stereo number code one of the form a/b, where a is the smallest and b is the largest stereo number for a particular isomer. Then mesoforms would code in the form $a/\bar{a}$, and pairs of enantiomers would code as $(a/b, \bar{b}/\bar{a})$.

Stereo numbers by themselves do not code the absolute configurations of asymmetric centers. The binary designation at each center is always relative to a freely chosen reference, or "zero", configurational isomer. Furthermore, the order of digits in the stereo number will conform to an arbitrary, predetermined ordering of the centers. For example, in assigning the decimal number 12 to D-mannose in Figure 1, I took the zero molecule to be D-allose and chose the numbering of centers proceeding from top to bottom.

In order to achieve truly unique and unambiguous codes for stereoisomers, it is necessary and sufficient to establish rules for choosing the zero molecules and their centers' orderings. This is, properly speaking, a still unresolved problem in coding molecular constitution (see footnote 13 on choosing a "representative CCCM") and is thus beyond the scope of this paper.

APPLICATIONS

If a family of asymmetric carbon configurational isomers has the trivial configuration symmetry group $\{(1)(2)(3)...(n)\}$, then it will have 2^n members. It is natural to identify these isomers with the 2^n binary stereo numbers of n digits and to follow Feldman's example and convert these numbers to a higher base.

Figure 3 demonstrates hexadecimal codes for the aldoses up to six carbons. Standard names are given for comparison.[7] I have chosen the order and sense of the digits so that D isomers will have even codes and L isomers will have odd ones. The extension to higher sugars is easy. For example, the hexadecimal code 67 specifies 2,3,4-D-lyxo-5,6,7,8-L-altro-nonose
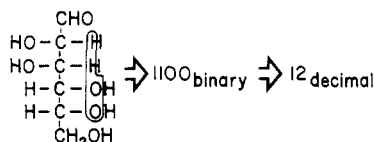
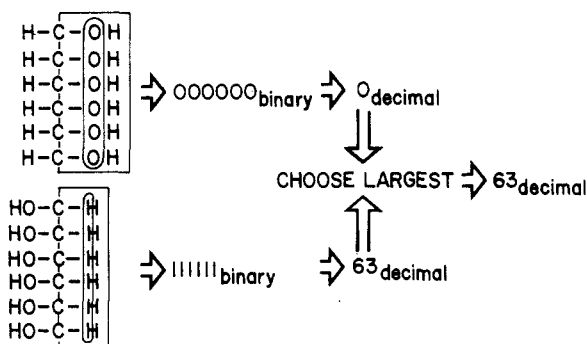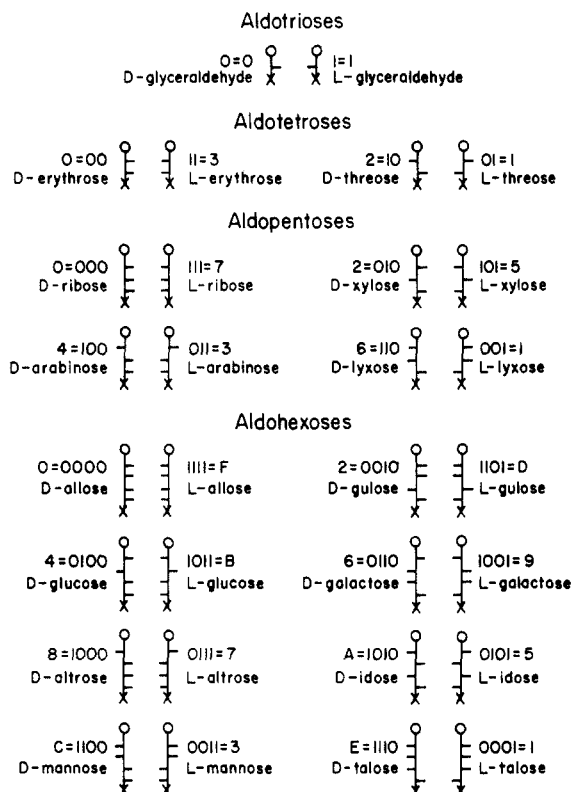**Figure 1.** Decimal stereo number of D-mannose.

Figure 2. Decimal stereo number of *cis*-inositol.

Figure 3. Aldose hexadecimal stereo numbers to six carbons.

with the digit 6 coding D-lyxo and the digit 7 coding L-altro (see Figure 4). Note that the three odd asymmetric carbons correspond to the *most* significant digit in the hexadecimal code: *1100*111 = 67. The standard IUPAC[8] name for the molecule is 6,7,8-L-*ribo*-2,3,4,5-D-*manno*-nonose, which corresponds to the binary numbering systems of Neelakantan[3] and Klein.[4] I have chosen to use a less standard ordering to preserve the correspondence between even/odd numbers and D/L isomers.

The sugar alcohols each have two configuration symmetry group members, the trivial one and one other where the order and sense of all digits are reversed. The groups are {(1), (1')}, {(1)(2), (1' 2')}, {(1)(2)(3), (1' 3')(2')}, {(1)(2)(3)(4), (1' 4')(2' 3')}, etc. These groups generate some pairs of equivalent stereo numbers. With the "*a/b*" or "min/max" code of the previous section, the alditols to six carbons are easily coded. (See Figure 5. Again, standard names are given for comparison.[7])
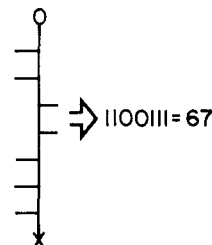

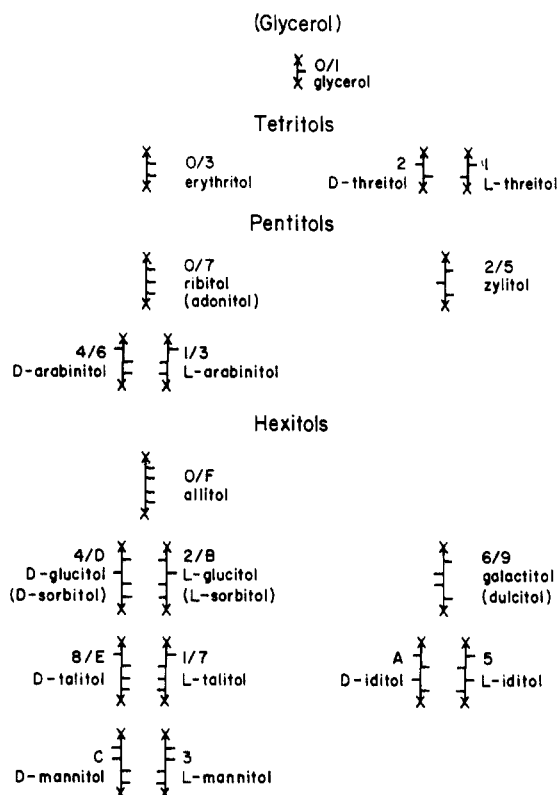**Figure 4.** Hexadecimal stereo number of an aldononose.

Figure 5. Hexadecimal codes of alditols to six carbons.

A more interesting group of configuration symmetries is the one of 12 elements corresponding to inositol:

{(1)(2)(3)(4)(5)(6), (1 2 3 4 5 6), (1 3 5)(2 4 6),
   (1 4)(2 5)(3 6), (1 5 3)(2 6 4), (1 6 5 4 3 2),
   (1')(2' 6')(3' 5')(4'), (1' 2')(3' 6')(4' 5'),
   (1' 3')(2')(4' 6')(5'), (1' 4')(2' 3')(5' 6'),
      (1' 5')(2' 4')(3')(6'), (1' 6')(2' 5')(3' 4')}

Figure 6 demonstrates the calculation of inositol stereoisomers equivalent to hexadecimal code 2C. This calculation reveals B/32 to be the min/max code of this isomer. Figure 7 lists the inositol min/max codes together with the corresponding IUPAC[9] names.

Substituted inositol (one substituent on carbon 1) is easily dealt with. The configuration symmetry group becomes {(1)(2)(3)(4)(5)(6), (1')(2' 6')(3' 5')(4')} and the number of min/max codes jumps to 32 from 9. These are displayed in Figure 8.

The remainder of this paper presents evidence for the soundness of the stereo number and configuration symmetry group systems. The treatment is mathematical and depends on some elementary definitions and results in matrix theory and group theory. These are summarized in the Appendix.

A MATRIX DEFINITION OF CONFIGURATION

Kudo and co-workers have represented molecular constitution using the Component Connectivity Matrix (CCM).[10] This square matrix lists the labels of molecular components
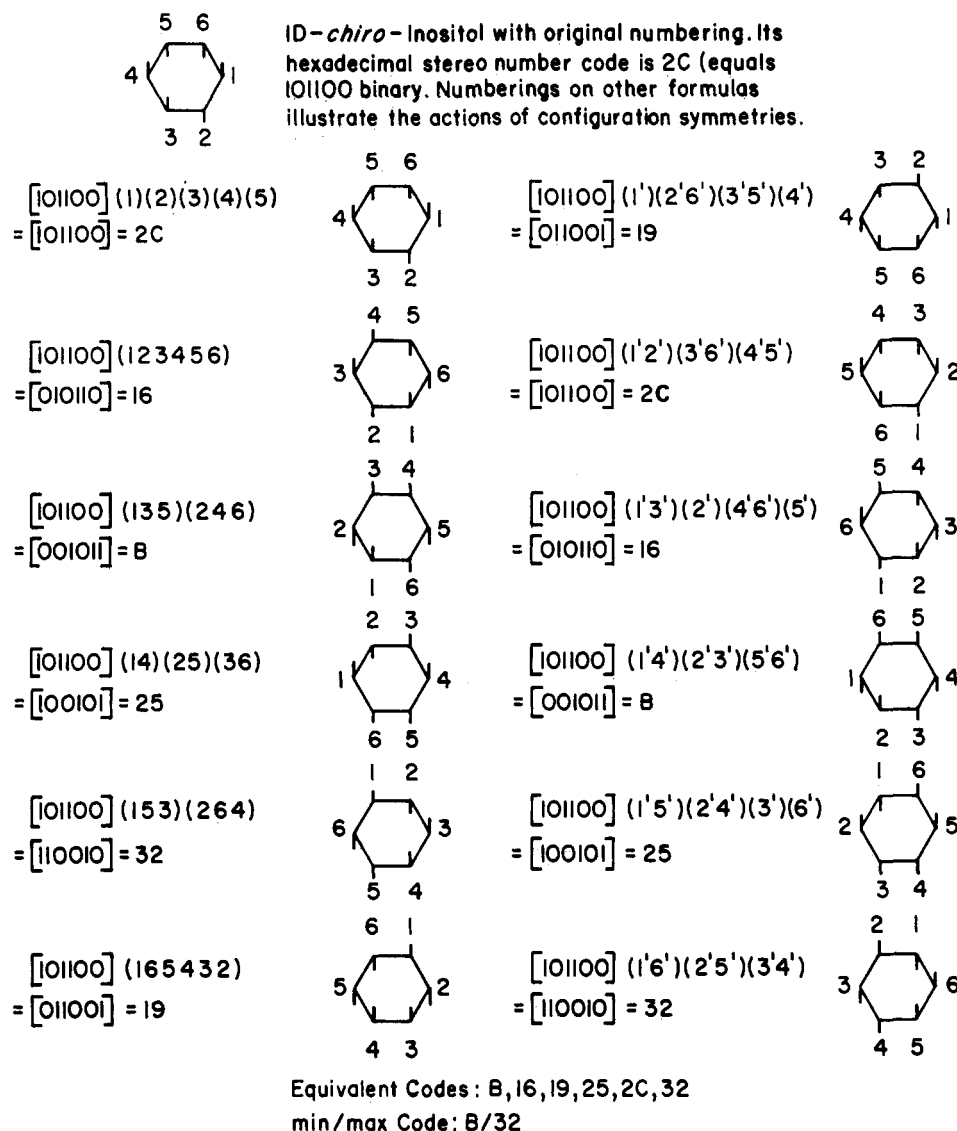
ID-*chiro*-Inositol with original numbering. Its hexadecimal stereo number code is 2C (equals IOIIOO binary. Numberings on other formulas illustrate the actions of configuration symmetries.

$[\overline{I}OIIOO]$ (1)(2)(3)(4)(5)
= $[\overline{I}OIIOO]$ = 2C

$[\overline{I}OIIOO]$ (1')(2'6')(3'5')(4')
= $[OIIOO\overline{I}]$ = 19

$[\overline{I}OIIOO]$ (123456)
= $[OIOIIO]$ = 16

$[\overline{I}OIIOO]$ (1'2')(3'6')(4'5')
= $[\overline{I}OIIOO]$ = 2C

$[\overline{I}OIIOO]$ (135)(246)
= $[OOIOI\overline{I}]$ = 8

$[\overline{I}OIIOO]$ (1'3')(2')(4'6')(5')
= $[OIOIIO]$ = 16

$[\overline{I}OIIOO]$ (14)(25)(36)
= $[\overline{I}OOIOI]$ = 25

$[\overline{I}OIIOO]$ (1'4')(2'3')(5'6')
= $[OOIOI\overline{I}]$ = 8

$[\overline{I}OIIOO]$ (153)(264)
= $[\overline{I}IOOIO]$ = 32

$[\overline{I}OIIOO]$ (1'5')(2'4')(3')(6')
= $[\overline{I}OOIOI]$ = 25

$[\overline{I}OIIOO]$ (165432)
= $[OIIOO\overline{I}]$ = 19

$[\overline{I}OIIOO]$ (1'6')(2'5')(3'4')
= $[\overline{I}IOOIO]$ = 32

Equivalent Codes : 8,16,19,25,2C,32
min/max Code: 8/32

Figure 6. Calculation of inositol hexadecimal stereo numbers equivalent to 2C.

mesoforms

0/3F
cis-Inositol

1/3E
epi-Inositol

3/3C
allo-Inositol

5/3A
myo-Inositol

7/38
neo-Inositol

9/36
muco-Inositol

15/2A
scyllo-Inositol

enantiomeric pair

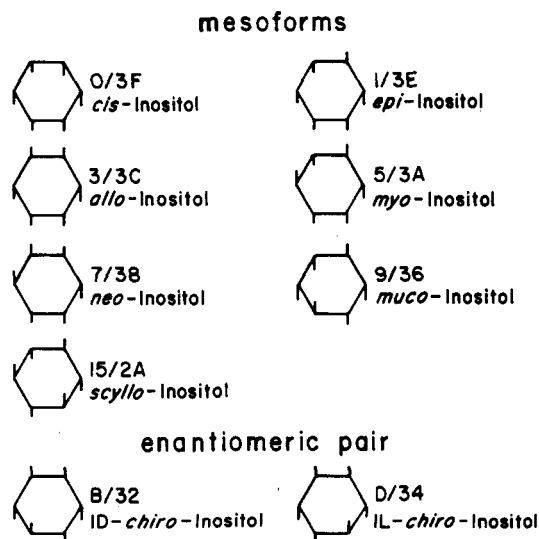B/32
ID-*chiro*-Inositol

D/34
IL-*chiro*-Inositol

Figure 7. Hexadecimal codes of unsubstituted inositols.

down its main diagonal and bond orders in its off-diagonal elements. If off-diagonal element (i,j) is zero, it means the ith and jth components are not bonded together; however, if (i,j) is nonzero, it means there is a bond of order (i,j) connecting these components. Figure 9 shows one way to represent tartaric acid by this method. For two such CCMs X

and Y, we say that X is *constitutionally isomorphic* to Y when there exists a permutation matrix U such that

$$X = U^{-1}YU \qquad (1)$$

The matrix U is then called a *constitutional isomorphism* from X to Y. Figure 10 demonstrates such a relationship involving tartaric acid. This relation is just the matrix version of graph isomorphism[11] as applied to chemical molecules.

Constitutional isomorphism, like graph isomorphism, is an equivalence relation. To see why this is the case, suppose X, Y, and Z are CCMs. Then $X = I^{-1}XI$, where I is the identity matrix, so X is constitutionally isomorphic to itself, and we have the reflexive property. If X is constitutionally isomorphic to Y, then there exists a constitutional isomorphism U satisfying eq 1, so $Y = (U^{-1})^{-1}U^{-1}YU(U^{-1}) = (U^{-1})^{-1}X(U^{-1})$ and Y is constitutionally isomorphic to X with constitutional isomorphism U^{-1}, and the symmetric property follows. Finally, suppose X is constitutionally isomorphic to Y with constitutional isomorphism U, and Y is constitutionally isomorphic to Z with constitutional isomorphism V. Then $X = U^{-1}YU = U^{-1}(V^{-1}ZV)U = (VU)^{-1}Z(VU)$, so X is constitutionally isomorphic to Z with constitutional isomorphism VU, and the transitive property also holds. The equivalence classes of CCMs with respect to constitutional isomorphism are termed *constitutional isomers*.

To proceed from constitutional isomerism to the modeling of configuration, we must first deal with the configuration
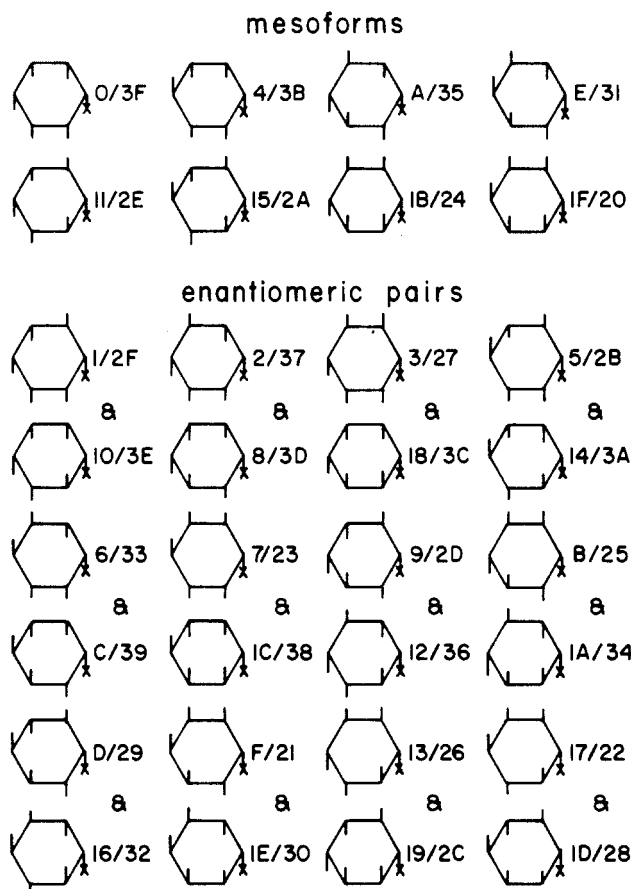
mesoforms

O/3F   4/3B   A/35   E/31

II/2E   I5/2A   IB/24   IF/20

enantiomeric pairs

I/2F   2/37   3/27   5/2B
&   &   &   &
IO/3E   8/3D   I8/3C   I4/3A

6/33   7/23   9/2D   B/25
&   &   &   &
C/39   IC/38   I2/36   IA/34

D/29   F/2I   I3/26   I7/22
&   &   &   &
I6/32   IE/30   I9/2C   ID/28

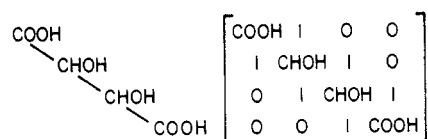**Figure 8.** Hexadecimal codes of 1-substituted inositols.

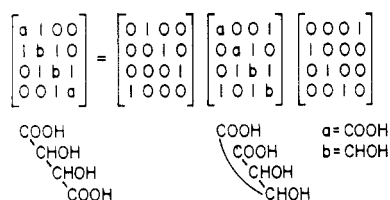**Figure 9.** Component connectivity matrix (CCM) for tartaric acid.

Figure 10. Demonstration of constitutional isomorphism between two CCM representations of tartaric acid.

a = COOH   b = CHOH

Figure 11. A configured component connectivity matrix (CCCM) representation of L-tartaric acid.

around each component in the CCM. We will consider two groups associated with the ith component of the CCM: its *rearrangement group* G_i and its *skeletal symmetry group* P_i. Both of these groups are groups of permutations on the set of attachment sites connecting the ith component to other components or rather on the numbering of these sites given a fixed initial numbering for that type of component. The rearrangement group specifies the possible rearrangements of neighboring components to different attachment sites and will usually be one of the symmetric groups of all permutations on n numbers or a direct product of symmetric groups. The latter case will arise when the intercomponent bonds are of more than one order. The skeletal symmetry group specifies numberings that are configurationally indistinguishable from the initial numbering of the attachment sites. These definitions are precisely the same as those proposed in a paper by Nourse,[12] but their application here is a bit different. Basically, we consider each component of the CCM to be a molecular
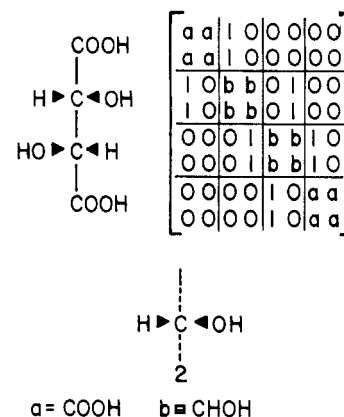
skeleton in its own right. We will consider that a component's label always fixes the component's initial numbering, rearrangement group, and skeletal symmetry group.

We can now model a molecule up to configuration with a *Configured Component Connectivity Matrix* (CCCM). The CCCM is built up from a CCM by the following procedure:

(1) Find n, the maximum number of attachment sites among the components of the CCM.

(2) Expand each diagonal element of the CCM into an $n \times n$ submatrix containing n^2 copies of that element.

(3) Expand each zero element of the CCM into an $n \times n$ submatrix containing n^2 zeros.

(4) Expand each non-zero off-diagonal element (i,j) into an $n \times n$ submatrix consisting of n column vectors. Each of these vectors is a zero vector except for one, the kth, which contains n copies of the value of (i,j). The index k is chosen to indicate that the ith component is connected to the jth component at the jth component's kth attachment site. An example of a CCCM representation of L-tartaric acid is given in Figure 11.

Two CCCMs, $\mathbf{X}$ and $\mathbf{Y}$, that derive from constitutionally isomorphic CCMs can still be represented by a sort of constitutional isomorphism defined by

$$\mathbf{XA} = \mathbf{U}^{-1}\mathbf{YU} \qquad (2)$$

Here, $\mathbf{U}$ is the old constitutional isomorphism of eq 1 expanded in a similar manner to the CCCMs. Each 0 in $\mathbf{U}$ is expanded to an $n \times n$ zero submatrix, and each 1 in $\mathbf{U}$ is expanded to an $n \times n$ identity matrix. The additional matrix $\mathbf{A}$, called a *rearrangement* of $\mathbf{X}$, is added to preserve equality in eq 2. This matrix is composed of zeros except for a sequence of $n \times n$ permutation submatrices located along the main diagonal. The jth permutation matrix along $\mathbf{A}$'s diagonal is the matrix representation of a member of the rearrangement group for the jth component of $\mathbf{X}$. It rearranges the nonzero off-diagonal column vectors in $\mathbf{X}$ that symbolize the pattern of attachments of the jth component to its neighbors. This arrangement is done so that those column vectors will line up with the corresponding column vectors in the $\mathbf{U}^{-1}\mathbf{YU}$ term of eq 2. If it happens that the jth permutation matrix along $\mathbf{A}$'s diagonal is a member of the skeletal symmetry group of the jth component of $\mathbf{X}$, then we can say that the constitutional isomorphism preserves configuration at that component. The expansion of a constitutional isomorphism between *meso*-tartaric acid and L-tartaric acid is given in Figure 12.

The rearrangement of $\mathbf{X}$ is considered to be an essential part of the constitutional isomorphism relationship for CCCMs, so we will speak of the constitutional isomorphism $(\mathbf{A}, \mathbf{U})$ between CCCMs $\mathbf{X}$ and $\mathbf{Y}$. The set of all possible values for the rearrangement of a CCCM $\mathbf{X}$ is then a matrix representation of the Cartesian product $G_1 G_2...$ of the rearrangement

STEREO NUMBERS

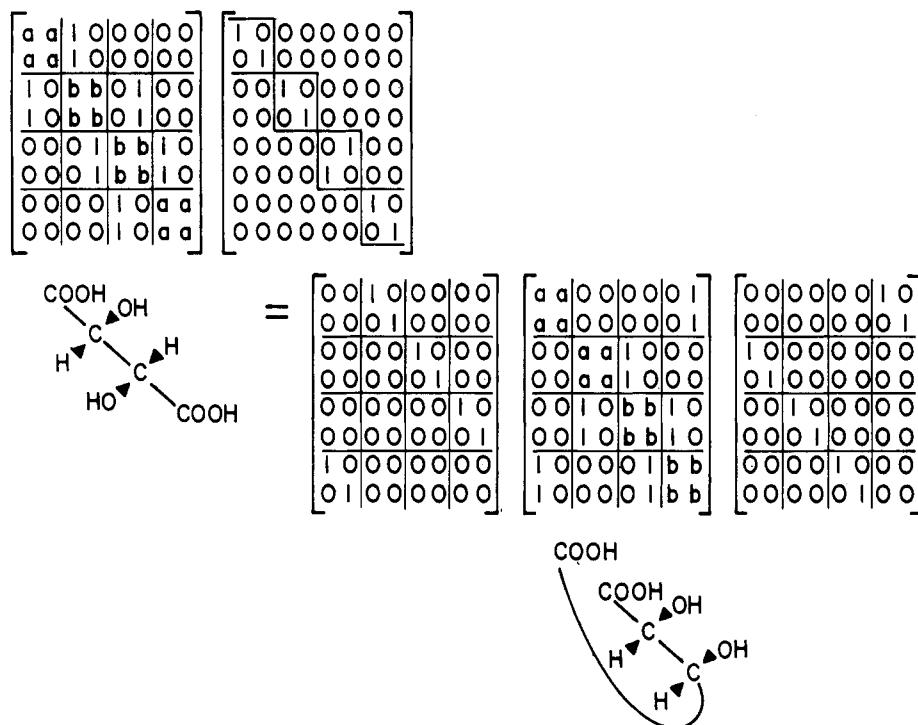*J. Chem. Inf. Comput. Sci., Vol. 26, No. 2, 1986* **81**

Figure 12. Demonstration of constitutional isomorphism between CCCM representations of L-tartaric acid and *meso*-tartaric acid.

groups for the components of X. We call this group the rearrangement group of the CCCM X, and will denote it by G. Similarly, the set of rearrangements of X corresponding to the Cartesian product $P_1P_2...$ of X's components' skeletal symmetry groups is called its skeletal symmetry group P. Since each P_i in X is a subgroup of the corresponding G_i, we have immediately that P is a subgroup of G.

Now let X and Y be CCCMs and let (A,U) be a constitutional isomorphism from X to Y. If A belongs to the skeletal symmetry group P of X, then we say that (A,U) is a *configurational isomorphism* from X to Y.

Configurational isomorphism on CCCMs is an equivalence relation, and we will term the resulting equivalence classes *configurational isomers*. Since the skeletal symmetry group P of a CCCM is always a subgroup of the corresponding rearrangement group G, the configurational isomorphism relation is a refinement of constitutional isomorphism, and each constitutional isomer is partitioned into distinct configurational isomers. We can call this partition of a single constitutional isomer into configurational isomers a *family of configurational isomers*.

MATHEMATICAL BASIS OF STEREO NUMBERS

Suppose we have a CCCM X. Then any CCCM Y that is constitutionally isomorphic to X satisfies eq 2 for some constitutional isomorphism (A,U). We can rewrite eq 2 as (XA)I = U^{-1}YU, where I is the identity matrix. Clearly, (I,U) is a configurational isomorphism from XA to Y. This means that every configurational isomer belonging to X's family of configurational isomers can be represented in the form XA for some rearrangement of X. Given a single fixed representative CCCM for this family,[13] we can therefore characterize all its configurational isomers using only the rearrangements of that CCCM.

Now let X be a CCCM with rearrangement group G and skeletal symmetry group P. We want to consider the *left cosets* of P in G. Such a coset AP for a rearrangement A of X is the set of all rearrangements of the form A**P** with A held fixed and **P** varying through the skeletal symmetry group P. The left cosets of P partition G into equivalence classes. An

Table I. Multiplication Table of Rearrangement Parities

	even	odd
even	even	odd
odd	odd	even

alternative way of saying that two rearrangements A and B of X belong in the same left coset of P is to say that A is *congruent to* B *modulo P*, written A ≡ B (mod P). By definition, this means that $A^{-1}B$ is in P. These definitions follow the treatment of Larsen's book on elementary algebra.[14] A proof that congruence modulo P implies membership in the same left coset and vice versa can be found there. Having built this much machinery, we can now prove our first major result.

Theorem I. Suppose X is a CCCM with rearrangements A and B and skeletal symmetry group P. Further, suppose that A is congruent to B modulo P. Then, XA is configurationally isomorphic to XB.

Proof. Clearly, XA(A^{-1}B) = XB = I^{-1}(XB)I, where I is the identity matrix, but A^{-1}B is in P by hypothesis. Therefore, (A^{-1}B,I) is the required configurational isomorphism from XA to XB.

All the rearrangements of X that lie in a particular left coset of P therefore represent the same configurational isomer. This means that we can code configuration within X's family of configurational isomers by using left cosets instead of rearrangements. The question remains, though, how we will characterize these cosets.

Consider the special case where the only potentially asymmetric components in a CCCM X are asymmetric carbon atoms. If the ith component of X is such an atom, its rearrangement group G_i is, by definition, the symmetric group of all renumberings of that component's attachment sites, and its skeletal symmetry group P_i is the set of all even permutations on the sites. There are exactly two left cosets of P_i in G_i, the set of even permutations and the set of odd permutations. We say a rearrangement has *even parity* if it is in the coset of even permutations and has *odd* parity otherwise. Since P_i is a normal subgroup of G_i, we have that, for any a,b in G_i, $(aP_i)(bP_i) = (ab)P_i$, and this defines a table for combining the parities of rearrangements of X's ith component (see Table

I). An important feature of this table is that for two rearrangements in G_i their product will be even if and only if the two rearrangements have identical parity. Furthermore, a rearrangement in G_i always has the same parity as its inverse.

Now, consider a CCCM **X** with rearrangement group G and skeletal symmetry group P. Recall that for **A** and **B** in G, the left cosets **A**P and **B**P are identical if and only if **B**$^{-1}$**A** is in P. Rearrangements multiply with each other component by component, so a moment's reflection will reveal that this can happen only when the parity of each component in **A** matches that of its corresponding component in **B**.

Thanks to the foregoing argument, we can now characterize the left cosets of the skeletal symmetry group of a CCCM having only asymmetric carbons as potentially asymmetric centers. Quite simply, membership in a particular left coset is defined by the parity of the asymmetric centers. We can assign even centers the value "0" and odd centers the value "1" and concatenate all the parity values into a binary number with each digit standing for the value of a particular center. Each binary number on n digits, where n is the number of asymmetric centers, then codes for one and only one left coset of P. This procedure is just Feldman's stereo number system. What we have shown is that a stereo number codes for all the CCCMs of the form **XA**, where **X** is held fixed and **A** varies through a single left coset of **X**'s skeletal symmetry group. Theorem I assures us that all these CCCMs are configurationally isomorphic, so a stereo number always refers unambiguously to a single configurational isomer. Notice that this is true regardless of the presence of branching or rings in the CCCM.

CONFIGURATIONAL ISOMERS WITH MULTIPLE STEREO NUMBERS

We can now deal with the general problem of precisely when two rearrangements of a CCCM yield configurationally isomorphic results. To do this we need one more definition. A *constitutional automorphism* (S,U) on a CCCM **X** is a constitutional isomorphism from **X** to itself; that is, it satisfies the condition **XS** = **U**$^{-1}$**XU**. In the following argument, we will use the fact that this equation can be rewritten as **XSU**$^{-1}$ = **U**$^{-1}$**X**. It is now possible to prove the following result.

Theorem II. Suppose **X** is a CCCM with rearrangements **A** and **B** and skeletal symmetry group P. Then, **XA** is configurationally isomorphic to **XB** if and only if there exists a constitutional automorphism (S,U) on **X** such that

$$\mathbf{A} \equiv \mathbf{SU}^{-1}\mathbf{BU} \pmod{P} \tag{3}$$

Proof. Suppose **XA** and **XB** are configurationally isomorphic with configurational isomorphism (T,U) from **XA** to **XB**. Then, **XAT** = **U**$^{-1}$**XBU** = **U**$^{-1}$**XUU**$^{-1}$**BU**, and it follows that **X**(**ATU**$^{-1}$**B**$^{-1}$**U**) = **U**$^{-1}$**XU**. Thus, (**ATU**$^{-1}$**B**$^{-1}$**U**,**U**) is a constitutional automorphism of **X**, and **A**$^{-1}$[(**ATU**$^{-1}$**B**$^{-1}$**U**)**U**$^{-1}$**BU**] = **T** $\in P$, so congruence 3 is satisfied. On the other hand, suppose (S,U) is a constitutional automorphism satisfying congruence 3. Then, **A**$^{-1}$**SU**$^{-1}$**BU** $\in P$, but (**XA**)(**A**$^{-1}$**SU**$^{-1}$**BU**) = (**XSU**$^{-1}$)**BU** = **U**$^{-1}$(**XB**)**U**, and we have that (**A**$^{-1}$**SU**$^{-1}$**BU**,**U**) is a configurational isomorphism from **XA** to **XB**, as required.

Now let us again consider the case of a CCCM **X** with a skeletal symmetry group P all of whose potentially asymmetric centers are asymmetric carbon atoms. Recall that in this case we can identify left cosets of P with unique binary stereo numbers. A typical representative of a left coset has component rearrangements of even parity whenever there is a "0" in the corresponding digit of the stereo number and odd parity whenever there is a "1". Let us consider a particular binary stereo number b of **X**. We construct a typical rearrangement **B** lying in the left coset coded for by b. Now, how does **B** differ

from the term **SU**$^{-1}$**BU** in the right side of congruence 3 for a typical constitutional automorphism (S,U)? First of all, the **U**$^{-1}$**BU** part simply permutes the order of the component rearrangements of **B**. The stereo number for (**U**$^{-1}$**BU**)P will then be the same as b, but with the digits permuted. Now, **S** acts on the **U**$^{-1}$**BU** part through component-by-component multiplication of rearrangements. In particular, wherever **S** has a component rearrangement of odd parity, the parity of the corresponding component rearrangement in the product **SU**$^{-1}$**BU** will be the reverse of that in **U**$^{-1}$**BU**. This means that once we have the permuted stereo number corresponding to (**U**$^{-1}$**BU**)P, we can find the stereo number for (**SU**$^{-1}$**BU**)P by reversing the sense of some of its binary digits. Of course, this is just a configuration symmetry as defined by Nourse. The configuration symmetry group of any CCCM **X** with asymmetric carbons can thus be derived directly from the group of constitutional automorphisms of **X**. Theorem II assures us that in this case the configuration symmetry group will act on any stereo number to generate the list of all its equivalent stereo numbers.

ACKNOWLEDGMENT

APPENDIX: NOTES ON MATRICES AND GROUPS

Permutation Matrices. A permutation on $\{1,2,3,...,n\}$ is a one to one mapping of this set onto itself. A permutation of the type (124)(3) would be read: 1 goes to 2, 2 goes to 4, 3 stays fixed, and 4 goes to 1. Multiplication, or composition, of permutations proceeds left to right: (124)(3)·(1234) = (1342). We are most interested in permutations that act (from the right) on a vector by permuting its elements: [abcd](124)(3) = [dacb]. The matrices that have this effect (through matrix multiplication) on row vectors are called permutation matrices. The matrix version of the action described above of (124)(3) is

$$[abcd]\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} = [dacb]$$

The set of all $n \times n$ permutation matrices forms a group under matrix multiplication (see the note on groups below). The reader can find out more about vectors and matrices by reading from Moore's textbook on the subject.[15]

Equivalence Relations. Let S be a set. If A is in the set, we say A is an element of the set, denoted $A \in S$. A relation $\sim$ on S is a situation where we can say either that A is related to B by $\sim$ or that it is not (written $A \sim B$ or $A \nsim B$) for any A and B in the set. We allow that perhaps A and B could be identical. $\sim$ is called an equivalence relation if for every three A, B, and C in the set we have the following three properties:

(1) $A \sim A$ (reflexive).
(2) if $A \sim B$, then $B \sim A$ (symmetric).
(3) if $A \sim B$ and $B \sim C$, then $A \sim C$ (transitive).

The set of all objects related to A by $\sim$ is called the equivalence class relative to $\sim$ with representative A. It is well-known that sets are broken up, or partitioned, into equivalence classes by equivalence relations. Refer to pages 16–18 in Larsen's book[14] for more information on this. Several equivalence relations have appeared in this paper. The three properties for an equivalence relation were demonstrated for constitutional isomorphism on CCMs but were not demonstrated for other relations such as configurational isomorphism on CCCMs. The reader might find it helpful to try to establish

the three properties in some of the other cases.

Notice that we have avoided constructing a "configurational isomerism" relation in this paper. This strategy avoids confusion that can arise from application of the reflexive property. It is not clear whether it makes sense to say that a molecule is "configurationally isomeric" to itself.

Groups and Subgroups. A group is a set with binary operation (here symbolized by juxtaposition) and identity element I that has the following four properties:

(1) AB is in the set for any A and B in the set (closure).

(2) $(AB)C = A(BC)$ for any A, B, and C in the set (associative law).

(3) $AI = IA = A$ for any A in the set (identity rule).

(4) There is, for any A in the set, an element A^{-1} in the such that $A^{-1}A = AA^{-1} = I$ (inverse rule).

A subgroup of a group is a subset of the group which is still a group under the bigger group's binary operation.

Most of the computational trickery in this paper involves operations with inverses and the identity. The rearrangement and skeletal symmetry groups for a CCCM are subgroups of the group of all permutation matrices (of the size of the CCCM), and they therefore contain the identity matrix $\mathbf{I}$. This matrix, with ones down the main diagonal and zeros elsewhere, shows up in various places. The matrix $\mathbf{I}$ can be inserted freely into a congruence or equation. Notice that by definition $\mathbf{I} = \mathbf{I}^{-1}$. A proof often depends on the judicious insertion of the identity matrix $\mathbf{I}$ in the form $(\mathbf{A}^{-1}\mathbf{A})$ or $(\mathbf{U}^{-1}\mathbf{U})$, for instance.

REFERENCES AND NOTES

(1) Mailing address: 3668 Leaman Street, Halifax, Nova Scotia, Canada B3K 4A1.

(2) Feldman, A. "Stereo Numbers: A Short Designation for Stereoisomers". *J. Org. Chem.* **1959**, *24*, 1556–60.

(3) (a) Neelakantan, S. "Application of Stereo Numbers in Sugar Chemistry". *Curr. Sci.* **1969**, *38*, 353–5. (b) Neelakantan, S. "Stereo Numbers in Sugar Chemistry. Part II: Aldohexopyranoses". *Curr. Sci.* **1970**, *39*, 85–6. (c) Neelakantan, S. "Stereo Numbers in Sugar Chemistry. Part III: Inositols". *Indian J. Chem. Educ.* **1971**, *2*, 15–16.

(4) Klein, H. A. "A Simplified Carbohydrate Nomenclature". *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 15–18.

(5) (a) Nourse, J. G. "The Configuration Symmetry Group and Its Application to Stereoisomer Generation, Specification, and Enumeration". *J. Am. Chem. Soc.* **1979**, *101*, 1210–16. (b) Nourse, J. G.; Carhart, R. E.; Smith, D. H.; Djerassi, C. "Exhaustive Generation of Stereoisomers for Structure Elucidation". *J. Am. Chem. Soc.* **1979**, *101*, 1216–23.

(6) Abe, H.; Hayasaka, H.; Miyashita, Y.; Sasaki, S.-I. "Generation of Stereoisomeric Structures Using Topological Information Alone". *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 216–19.

(7) American Chemical Society "Rules of Carbohydrate Nomenclature". *J. Org. Chem.* **1963**, *28*, 281–91.

(8) IUPAC-IUB "Tentative Rules for Carbohydrate Nomenclature. Part 1, 1969". *Biochemistry* **1971**, *10*, 3983–4004.

(9) IUPAC-IUB "Nomenclature of Cyclitols (1973 Recommendations)". *Pure Appl. Chem.* **1974**, *37*, 285–97.

(10) Kudo, Y.; Yamasaki, T.; Sasaki, S.-I. "The Characteristic Polynomial Uniquely Represents the Topology of a Molecule". *J. Chem. Doc.* **1973**, *13*, 225–27.

(11) Harary, F. *Graph Theory*; Addison-Wesley: Reading, MA, 1969; p 151.

(12) Nourse, J. G. "An Algebraic Description of Stereochemical Correspondence". *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 2385–88.

(13) Finding this representative is equivalent to the notoriously difficult "coding problem". See, e.g., Read, R. C.; Corneil, D. G. "The Graph Isomorphism Disease". *J. Graph Theory* **1977**, *1*, 339–63.

(14) Larsen, M. D. *Introduction to Modern Algebraic Concepts*; Addison-Wesley: Reading, MA, 1969; pp 43–45.

(15) Moore, J. T. *Elements of Linear Algebra and Matrix Theory*; McGraw-Hill: New York, 1968.

ACS Committee on Nomenclature: Annual Report for 1985

KURT L. LOENING

Chemical Abstracts Service, Columbus, Ohio 43210

Nomenclature committees, both national and international, were very active in 1985, resulting in substantial progress in many different fields. A summary of the more important meetings and accomplishments follows.

The *ACS Committee on Nomenclature* held its annual meeting at CAS in November.† Progress of the work of the divisional committees and international commissions was reviewed. The format for the periodic table recommended by the Committee in 1983 and proposed by the IUPAC Commission on Nomenclature of Inorganic Chemistry with very minor modifications is finding acceptance throughout the world, although some objections are still being voiced. Improved communication with high school chemistry teachers is being developed with A. Saturnelli of the Bureau of Science Education of the New York State Education Department. Editors of ACS journals continue as ex officio members of the Committee. To establish closer contact with interested individual ACS members, the Committee held two successful open meetings in Miami Beach and Chicago and plans to continue to hold open meetings regularly on Mondays from

2 to 4 p.m. at the national meetings. Closer liaison with other ACS bodies such as the Committees on Education and Science as well as various Divisions continues to be of concern. Cooperation with nomenclature groups in disciplines related to chemistry continues to be pursued. Contact has been established with the ASTM Committee on Medical Terminology. Efforts to contact appropriate groups in physics and pharmacology are in progress. The promotion of and input into International Union of Pure and Applied Chemistry (IUPAC) recommendations is, as always, a primary objective of the Committee. In the view of the Committee, the nomenclature of biotechnology requires urgent attention. Since this is an interdisciplinary subject, a subcommittee was formed to study the topic and to explore the best way to proceed. The Subcommittee on Chemical Pronunciation continues to be active.

The *IUPAC Interdivisional Committee on Nomenclature and Symbols* (IDCNS) continued to function effectively this year. It held its annual meeting in Lyon, France, in August. In addition to the IUPAC publications listed in the Appendix, specific documents in process and thus not yet recorded in this

† Abbreviations used, not identified in the text, are ACS, American Chemical Society; CAS, Chemical Abstracts Service; JCBN, Joint Commission on Biochemical Nomenclature; and NC-IUB, Nomenclature Committee of International Union of Biochemistry.