Representation of Molecular Configurations by CAST Coding Method

Hiroko Satoh,*[,§,†,‡] Hiroyuki Koshino,*[,§] Kimito Funatsu,[#] and Tadashi Nakata[§]

RIKEN (The Institute of Physical and Chemical Research), 2-1 Hirosawa, Wako, Saitama 351-0198, Japan, PRESTO, Japan Science and Technology Corporation, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan, and Department of Knowledge-Based Information Engineering, Toyohashi University of Technology, 1-1 Tempaku, Toyohashi, Aichi 441-8580, Japan

Received September 27, 2000

A configurational CAST (CAnonical representation of STereochemistry) coding method, which represents relative and absolute configuration, is described. The configurational CAST codes are constructed by canonical rotation of the dihedral angles of the input structure before the CAST codes are assigned. Using the configurational CAST, configurational differences can be distinguished independently of conformational differences. Representation of enantiomers is also achieved by a mirror image conversion method. The CAST representation shows the distinctive characteristics of several diastereomers and conformers that were examined. The method clearly represents the differences in configurations. Applications to organic molecules having complex stereochemistry are also demonstrated.

1. INTRODUCTION

Configuration is an essential attribute of molecular structures. Relative and absolute configurational information is necessary for the identification of diastereomers and enantiomers. The Cahn−Ingold−Prelog system, i.e., ***R*** or ***S*** designation for absolute stereochemistry, is essential for the representation of chiral centers. In general, diastereomeric molecules with the same constitution and connectivity but different configurations, i.e., molecules with the same two-dimensional structure, have different properties, such as biological activities and chemical reactivities. In NMR, diastereomers give different chemical shift values. The representation and recognition of configurations are important for computer programs treating molecular structures such as molecular modeling, synthetic design, reaction prediction, and NMR chemical shift prediction studies.

Several approaches to treat configuration by a computer have been reported.[1−5] Representation of absolute configuration, such as ***R*** or ***S***, on a chiral center is possible in many molecular modeling programs. Some methods to describe the absolute configurations for one-to-one nomenclature have been developed.[1,2] Representations of topological connectivity of atoms in a molecule by some mathematical methods have been studied.[3,4] These methods uniquely describe the absolute configurations. However, they cannot adequately represent stereochemical environments around an arbitrary atom, including an atom of nonchiral center. A specific method for nucleic acids and proteins has been developed, but it cannot easily be applied to organic molecules, generally.[5]

We have developed the CAST (CAnonical representation of STereochemistry) notation as a new canonical coding method to recognize and distinguish molecular three-dimensional structures in a computer system.[6] CAST is suitable for conformational representations. A unique CAST representation is assigned to one conformer or corresponding conformational group. For example, a unique CAST representation is assigned to one conformer of L-tartaric acid, which occurs commonly in nature with *2R, 3R* stereochemistry, and the diastereomeric *meso*-tartaric acid has *2RS, 3SR* stereochemistry. The CAST representation of one of the conformers of L-tartaric acid may not be matched, completely, with CAST representation of any conformer of diastereomeric *meso* form and even with that of enantiomeric D-tartaric acid with *2S, 3S* stereochemistry. In general, a flexible molecule can take too many conformations. Even if we divided 360 degrees into 12 groups for dihedral angle areas around C−C single bond and the symmetric nature of L-tartaric acid are considered, 864 ($12 \times 12 \times 12/2$) conformers and the same number of the CAST representations are obtained, even for this simple organic acid. Using the CAST coding method, the differences in these conformations can be distinguished. However, for the purpose of distinction of the differences in configuration, for example, for distinction of three isomers of L-, D-, and *meso*-tartaric acid molecules, the CAST representation for conformational information is not suitable, because several conformational CAST notations are possible depending on the variety of conformations, even though configurational information is included. Thus, for application of the CAST method to configurational representation independently of the conformational differences, development of new canonical treatment of three-dimensional structure is required.

This paper describes a configurational representation based on the CAST method by the canonical manipulations of the input molecular structure before the CAST coding procedures. A mirror image manipulation for representation of

* Corresponding authors phone: +81-3-4212-2579; e-mail: hsatoh@nii.ac.jp (Satoh) and phone: +81-48-462-1111 (ext. 3112); e-mail: koshino@postman.riken.go.jp (Koshino).

† PRESTO, Japan Science and Technology Corporation.

‡ Present address: National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan.

§ RIKEN (The Institute of Physical and Chemical Research).

Toyohashi University of Technology.

10.1021/ci000136g CCC: $20.00
Published on Web 08/31/2001

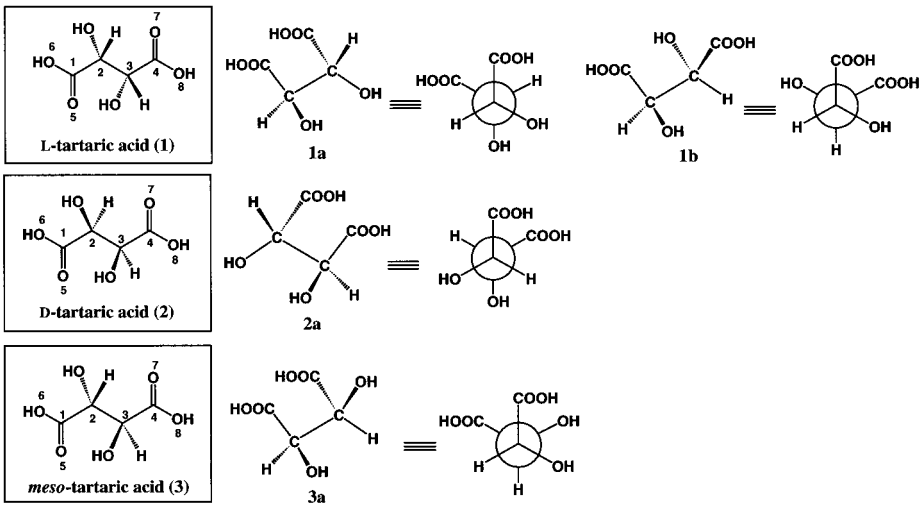

Figure 1. Conformers **1a** and **1b** of L-tartaric acid (**1**), a conformer **2a** of D-tartaric acid (**2**), and a conformer **3a** of *meso*-tartaric acid (**3**) are used as model structures.

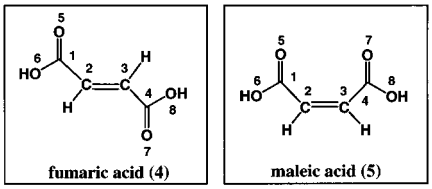

Figure 2. Fumaric acid (**4**) and maleic acid (**5**) are used as model structures for sp^2 systems.

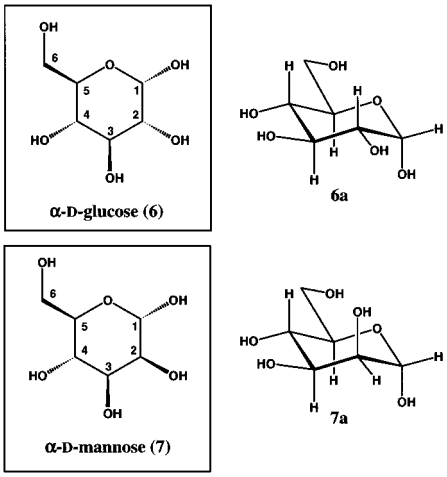

Figure 3. α-D-Glucose (**6**) and α-D-mannose (**7**) with $^4C_1$ conformation **6a** and **7a**, respectively, are used as model structures.

enantiomers is also described. Some applications to representations of configurational information are demonstrated.

2. METHOD

2.1. Model Structures. Model structures used for descriptions of the methods and the executions are shown in Figures 1, 4, and 5. Conformers **1a** and **1b** of L-tartaric acid (**1**), conformer **2a** of D-tartaric acid (**2**), and conformer **3a** of *meso*-tartaric acid (**3**) (Figure 1) are compared as conformers (**1a** and **1b**), diastereomers (**1a** and **3a**), and enantiomers (**1a** and **2a**). Fumaric acid (**4**) and maleic acid (**5**) (Figure 2) are compared as geometrical isomers. The $^4C_1$ conformation[8] **6a** of α-D-glucose (**6**) and **7b** of α-D-mannose (**7**) (Figure 3) are also examined for distinction of diastereomers.

Hereafter, a carbon atom is described with the number (e.g., C1), and oxygen and hydrogen atoms are described with the number of a carbon where the oxygen/hydrogen atom is connected (e.g., O1 and H1). When more than two oxygen/hydrogen atoms are connected to a carbon atom, the other numbers are used (e.g., O6 in L-tartaric acid (**1**)).

2.2. Canonical Rearrangement Rules for Deriving Configurational Information. The input structure is rearranged according to canonical rules. After the manipulations, the CAST codes are assigned to give a configurational CAST notation. Hereafter, the CAST notation for a molecule that is not rearranged is called conformational CAST notation, which includes conformational information of the input structure.

The rearrangements are performed before the CAST codes are assigned. The definition of the CAST codes and the procedure of construction of the conformational CAST notation are referred to in a previous paper.[6] The rearrangement rules for sp^3 and sp^2 systems are described as follows.

For sp^3 Systems. For sp^3 systems, the dihedral angle is rotated, and an atom that has the highest CANOST priority is taken place at the **si** position. Namely, a **si** CAST code is assigned to the most CANOST prior atom.

Figure 4 shows the rearrangement for the dihedral angle around a $C(sp^3)-C(sp^3)$ bond of a conformation of L-tartaric acid (**1**). In this example, CAST codes are assigned for a carbon atom C4 of a carboxyl group, an oxygen atom O3 in a hydroxyl group, and a hydrogen atom H3 based on the dihedral angle around the C2−C3 bond against to the other carboxyl group's carbon C1, which is shown in gray in Figure 4. CANOST codes for C4, O3, and H3 are **V**, **Q1**, and **H**, respectively, where the highest priorty code is **V**. Thus, the dihedral angle around the C2−C3 bond is rotated and the atom **V** is taken place at the **si** position against the C1 atom (**1a**-i to **1c**-i in Figure 4). Then, the CAST codes **si**, **tw**, and **te** are assigned to the atoms C4, O3, and H3, where the CANOST codes are **V**, **Q1**, and **H**, respectively (**1c**-ii in Figure 4). In general, for sp^3 systems, the dihedral angle around a L_{n-1}-L_{n-2} bond is rotated, and the most CANOST prior L_n atom is taken place at the **si** position against the L_{n-3} atom, where L_n is an atom at the nth level from a starting atom.[6]

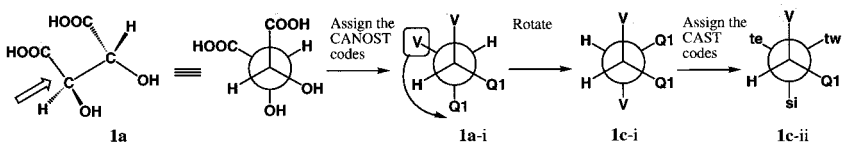

Figure 4. The rearrangement rule for sp^3 system. A conformer **1a** of L-tartaric acid (**1**) is rearranged for the CAST coding from C1. Newman projection of **1a** with CANOST codes is **1a**-i. The dihedral angle around C2−C3 is rotated, and the most prior CANOST code **V** is set at the **si** position to give **1c**-i. Then the CAST code **si**, **tw**, and **te** is assigned to C4, O3, and H3, respectively, to give **1c**-ii.

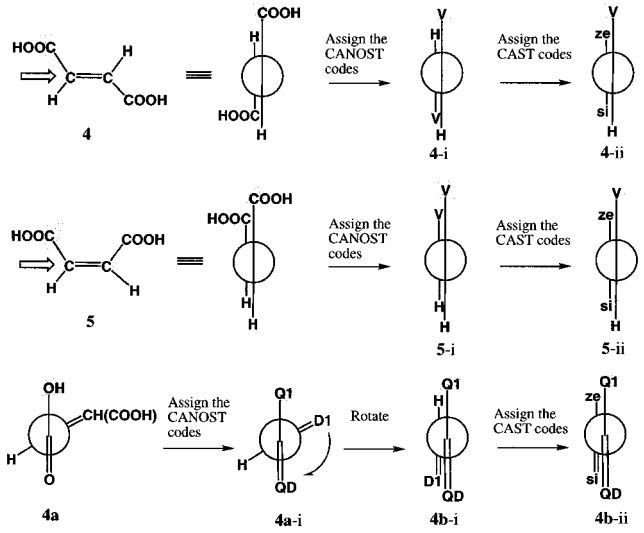

Figure 5. The rearrangement rule for sp^2 system. The planar structure around the unrotatable Csp^2-Csp^2 bond, such as double and aromtic bonds, is retained, and the CAST codes are assigned shown as **4**-ii and **5**-ii. For the rotatable $C(sp^2)-C(sp^2)$ bond, such as C1−C2 bond of **4**, the same rule as that for the sp^3 system is applied to give **4b**-ii.

For sp^2 Systems. The planar structure around double and aromatic bonds is retained, and the dihedral angle around a single bond of $C(sp^2)-C(sp^2)$ is rotated according to the same way as that for sp^3 systems.

Figure 5 shows the CAST code assignment for fumaric acid (**4**) and maleic acid (**5**). When a CAST code is assigned to a carbon atom C4 of a carboxyl group and a hydrogen atom H3 against to the other carboxylic group's carbon C1, the dihedral angle around C═C double bond is not rotated and the planar structure is retained, and the CAST codes **si** and **ze** are assigned to C4 and H3, respectively (**4**-ii in Figure 5). For maleic acid, **ze** and **si** are assigned to the corresponding carbon and hydrogen atoms C4 and H3, respectively (**5**-ii in Figure 5). Thus, the *E/Z* structures are also differently represented in the configurational CAST notations. When a CAST code is assigned to the C3 and H2 atoms against the hydroxylic group's oxygen atom O6 connecting to the C1 atom of **4a**, the dihedral angle around the C1−C2 bond is rotated and rearranged according to the same rule as that for sp^3 systems. Namely, CANOST codes **D1** and **H** are assigned to the C3 and H2, respectively (**4a**-i in Figure 5), the dihedral angle around the C1−C2 bond is rotated, and the C3 atom takes place at the **si** position (**4b**-i in Figure 5), because **D1** is prior to **H** in the CANOST rule. Then, the CAST codes **si** and **ze** are assigned to C3 and H2 atoms, respectively (**4b**-ii in Figure 5).

In general, for sp^2 systems, when the CAST codes are assigned to L_n and L_n' in a system of L_{n-3}-L_{n-2}═L_{n-1}-L_n(-L_n'), the CAST codes of **si** and **ze** or **si** and **ze** are assigned to L_n and L_n', respectively, according to the input *E/Z* structure. The resulting notations represent differences

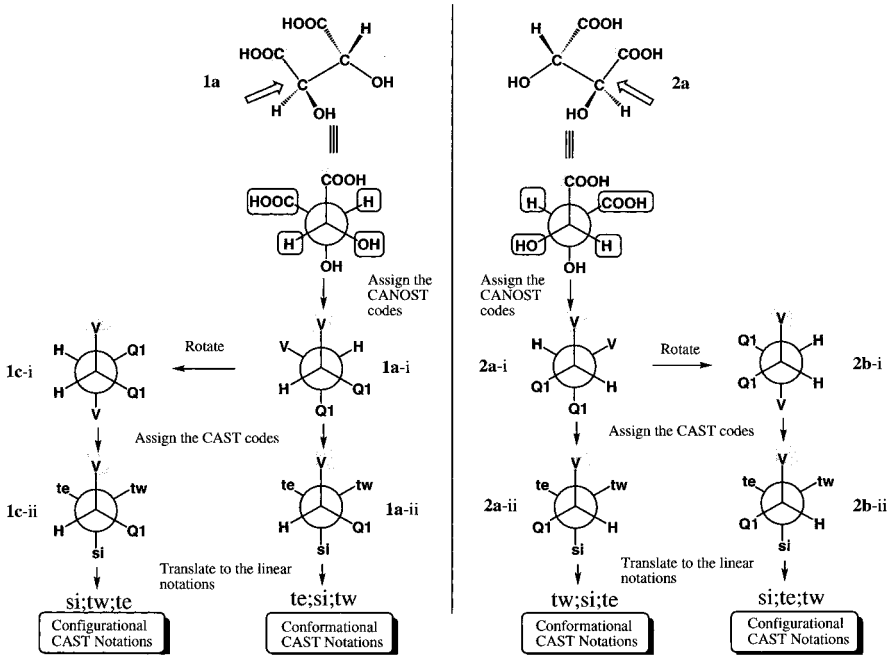

Figure 6. Mirror image conversion of the CAST codes to represent the enantiomer and recognize the same relative configuration based on mirror image structure-inversion of the enantiomer. The CAST codes that are converted against the plane defined by the **ze**-**si** axis and the corresponding bond around which the dihedral angle is defined. A conformer **1a** is recognized as the enantiomer of **2a** by the converted conformational and configurational CAST notations.

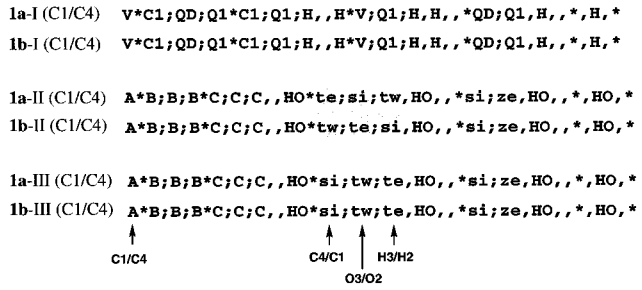

Figure 7. The CANOST (**1a**-I and **1b**-I), conformational CAST (**1a**-II and **1b**-II), and configurational CAST (**1a**-III and **1b**-III) notations from C1 and C4 for **1a** and **1b**, which are conformers of L-tartaric acid (**1**). Linear notations labeled with **1a**-I and **1b**-I are CANOST notations, for **1a** and **1b**, respectively, and the starting atom is described in the parentheses. In this case, the notation from C1 is the same as that from C4, thus one of the notations is described and "C1/C4" is denoted in the parentheses. Hereafter, the notations are labeled by the same way. The CANOST notations given here represent that **1a** and **1b** have the same two-dimensional structure. The conformational CAST notations represent conformational differences between **1a** and **1b**. The configurational CAST notations represent that **1a** and **1b** possess the same absolute configuration. Namely, **1a** and **1b** are recognized as conformers with the same absolute configuration by both the conformational and configurational CAST notations.

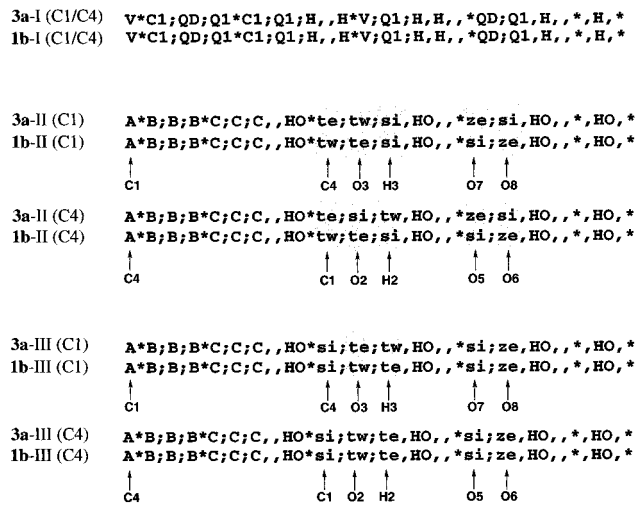

Figure 8. The CANOST (**3a**-I and **1b**-I), conformational CAST (**3a**-II and **1b**-II), and configurational CAST (**3a**-III and **1b**-III) notations from C1 and C4 for **3a** and **1b**, which are diastereomers of tartaric acid. The same CANOST notation represents the same two-dimensional structure of **3a** and **1b**. In the conformational CAST notations, differences of conformations and configurations between **3a** and **1b** are represented, while only the configurational difference on C3 chiral center is represented in the configurational CAST notations.

in geometric isomers. When the CAST codes are assigned to L_n and L_n' in a system of L_{n-3}=L_{n-2}-L_{n-1}=L_n(-L_n'), the same rearrangement rule as that for the sp^3 system is applied to the dihedral angle around the L_{n-2}-L_{n-1} bond.

2.3. Conversion Rules for Identification of Enantiomers. Enantiomers are represented by different CAST notations. To identify enantiomers, namely, to recognize two structures with the same relative configuration but different absolute configurations, a mirror image inversion is applied to one of the structures, and the converted CAST notations are compared with the other's CAST notations. If the CAST notations are the same, these two structures are identified as the enantiomers of each other.

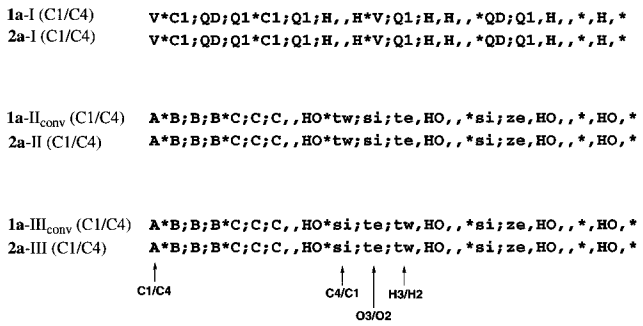

Figure 9. The CANOST (**1a**-I), enantiomerically converted conformational CAST (**1a**-II$_{conv}$), and converted configurational CAST (**1a**-III$_{conv}$) from C1 and C4 for **1a** and the CANOST (**2a**-I), conformational CAST (**2a**-II), and configurational CAST (**2a**-III) notations from C1 and C4 for **2a**, which is the enantiomer of **1a**. The same two-dimensional structure of **1a** and **2a** is represented by the same CANOST notations. The converted conformational and configurational CAST notations for **1a** are the same as those for **2a**, and it is represented that **1a** is the enantiomer of **2a** having the same relative configuration.

The enantiomer's CAST notation is obtained by assignment of the CAST codes after inversion of all atoms in a molecule against the plane including both of the **ze**-**si** axis and the rectangular axis around which the dihedral angle is defined. For example, a conformer **1a** of L-tartaric acid (**1**) is inverted to its enantiomeric conformer **2a** of D-tartaric acid (**2**) by inversion of all atoms of **1a** against the plane including both of the C2−C3 bond and its rectangular **ze**-**si** axis (Figure 6).

For the conformer **1a**, the conformational CAST codes for a carbon atom C4 of a carboxylic group, an oxygen atom O3 of a hydroxyl group, and a hydrogen atom H3 against the other carboxylic group's carbon atom C1 are **te**, **si**, and **tw**, respectively (**1a**-ii in Figure 6), while **tw**, **si**, and **te** are assigned to the corresponding atoms for **2a**, respectively (**2a**-ii in Figure 6). In the conformational CAST codes, when **te** and **tw** for **1a** are converted to **tw** and **te**, respectively, the conformational CAST codes for **2a** are obtained. The configurational CAST codes for **2a** is obtained from the configurational CAST codes for **1a** by the same conversion procedures (**1c**-i and **1c**-ii for **1a**; **2b**-i and **2b**-ii for **2a** in Figure 6). In general, mirror image conversion of all of the conformational/configurational CAST codes against the **ze**-**si** axis gives its enantiomer's conformational/configurational CAST codes. The enantiomer is generally represented by converted CAST notations from two atoms that are apart from each other through more than three bonds.

2.4. Execution in Computer. All procedures of the conformational/configurational CAST coding are executable on a computer, automatically. The program is running on irix OS of SGI workstation.

3. RESULTS AND DISCUSSIONS

3.1. Distinction of Conformers and Identification of Absolute Configuration for L-Tartaric Acid. The conformers **1a** and **1b** of L-tartaric acid (**1**) have different dihedral angles around the C2−C3 bond.

The CANOST, conformational CAST, and configurational CAST notations from C1 and C4 for the conformers are shown in Figure 7. The notations started from C1 are the same as those from C4.

```
6a-I (C6)   C2*C1;Q1;H;H*C1;Q;H,H,,*C1;Q1;H,C1,,*C1;Q1;H,H,,&1;Q1;H*&1;Q1;H,H,,,,H,*,H,,,*
7a-I (C6)   C2*C1;Q1;H;H*C1;Q;H,H,,*C1;Q1;H,C1,,*C1;Q1;H,H,,&1;Q1;H*&1;Q1;H,H,,,,H,*,H,,,*

6a-II (C6)  A*B;B;H;H*C;C;C,HO,,*si;tw;te,si,,*si;te;tw,HO,,si;te;tw*si;tw;te,HO,,,,HO,*,HO,,,*
7a-II (C6)  A*B;B;H;H*C;C;C,HO,,*si;tw;te,si,,*si;te;tw,HO,,si;te;tw*si;te;tw,HO,,,,HO,*,HO,,,*
                ↑                                                          ↑  ↑
                C6                                                        O2 H2
```

Figure 10. The CANOST (**6a**-I and **7a**-I) and configurational CAST (**6a**-II and **7a**-II) from C6 for **6a** and **7a**, which is the epimer of **6** at the C2 chiral center. The same two-dimensional structure is represented by the same CANOST notations. The configurational differences in the C2 chiral center are represented by the configurational CAST notations.

The same CANOST notations from C1 and C4 were obtained for **1a** and **1b** as shown in **1a**-I and **1b**-I in Figure 7, respectively. This shows that the conformers have the same two-dimensional structure. In Figure 7, **1a**-II and **1b**-II are the conformational CAST notations for **1a** and **1b**, respectively, from C1 and C4, where different CAST codes between **1a**-II and **1b**-II are marked by grayed squares. The conformational differences around the C2−C3 bond between **1a** and **1b** are represented by the CAST codes for C4, O3, and H3 in the notation started from C1 and by the CAST codes for C1, O2, and H2 in the notation started from C4. This shows that the conformational CAST notations distinguish the conformational differences between **1a** and **1b**. In Figure 7, **1a**-III and **1b**-III are the configurational CAST notations for **1a** and **1b** from C1 and C4. The conformational differences are ignored in the configurational CAST notations, and the same notations were given to the conformers. The configurational CAST notations represent that the two conformers **1a** and **1b** possess the same absolute configuration.

3.2. Distinction of Diastereomers for Tartaric Acid. Figure 8 shows the CANOST, conformational CAST, and configurational CAST notations from C1 and C4 for a conformer **1b** of L-tartaric acid (**1**) compared with those for a conformer **3a** of *meso*-tartaric acid (**3**).

In Figure 8, **3a**-I and **1b**-I are the CANOST notations started from C1 and C4 for **3a** and **1b**, respectively. The CANOST notation started from C1 is the same as that from C4, so one of them is shown. The diastereomeric differences between **1b** and **3a** can be distinguished by using both notations from C1 and C4, because, in general, whole molecular stereochemistry is represented by CAST notations started from two atoms separated each other by three or more bonds.

In the conformational CAST notations from C1 and C4, conformational differences around C1−C2, C2−C3, and C3−C4 bonds with the configurational differences for the C3 chiral center are represented. The differences are shown in the set of the CAST codes for (C4, O3, H3) and (O7, O8) in the notation from C1 and in the set of the CAST codes for (C1, O2, H2) and (O5, O6) in the notation from C4 (**3a**-II for **3a** and **1b**-II for **1b** in Figure 8).

The conformational differences are ignored, and only configurational differences are clear out in the configurational CAST notations shown in **3a**-III for **3a** and in **1b**-III for **1b** in Figure 8. In the notations, the configurational differences for the C3 chiral center by the CAST codes for O3 and H3 from C1. The results show the configurational differences were clearly distinguished by the configurational CAST notations.

3.3. Representation of Enantiomers and Identification of Relative Configuration for L- and D-Tartaric Acids.

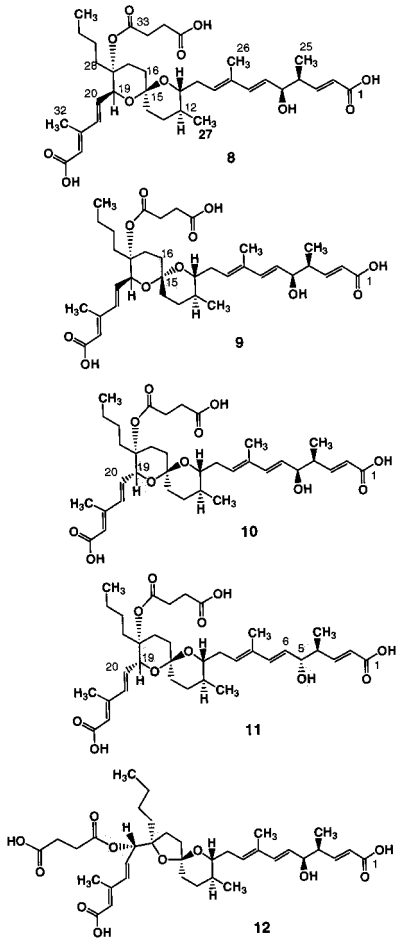

Figure 11. Structures of reveromycin A, the diastereomers, and reveromycin B. Reveromycin A (**8**), 15-epi-reveromycin A (**9**), 19-epi-reveromycin A (**10**), 5,19-epi-reveromycin A (**11**), and reveromycin B (**12**). The different partial structures compared with **8** are marked by gray circles.

Figure 9 shows the CANOST notations from C1 and C4 for **1a** (**1a**-I) and **2a** (**2a**-I), the mirror image converted conformational (**1a**-II_{conv}) and configurational (**1a**-III_{conv}) CAST notations for **1a**, and conformational (**2a**-II) and configurational (**2a**-III) CAST notations for **2a**. The notations from C1 are the same as those from C4.

The same CANOST notation for **1a** and **2a** represents that they possess the same two-dimensional structure. The converted conformational and configurational CAST notations for **1a** are the same as those for **2a**. The results show that **2a** having enantiomeric conformation and configuration for **1a** was represented by the converted conformational and configurational CAST notations. Namely, the same relative configuration between **1a** and **2a** was represented by the converted configurational CAST notations.

8 (C1) A*B;B;B*C;C,,HO*si;ze,,*si;te;tw,*si;tw;te,H;H;H,*si;ze,HO,,,,*si;ze,,*si;ze,*si;ze,H;H;H
*si;H;H,,,,*si;tw;te,,*si;tw;te,si,*si;H;H,H;H;H,,si;tw;te*&2;H;H,,,,,,,si;H;H,si
,,,si;H;H,,,si;tw;te&1;te;tw,,,,si;ze,*,si;H;H,si,si;ze,*si;H;H,,,si;ze,si;ze,
*si;H;H,,,si;H;H,,si;ze,H;H;H*H;H;H,,,si;H;H,,,si;ze,,,,*,,,si;ze,,,,HO*,HO,*

9 (C1) A*B;B;B*C;C,,HO*si;ze,,*si;te;tw,*si;tw;te,H;H;H,*si;ze,HO,,,,*si;ze,,*si;ze,*si;ze,H;H;H
*si;H;H,,,,*si;tw;te,,*si;tw;te,si,*si;H;H,H;H;H,,si;te;tw*&2;H;H,,,,,,,si;H;H,si
,,,si;H;H,,,si;tw;te&1;te;tw,,,,si;ze,*,si;H;H,si,si;ze,*si;H;H,,,si;ze,si;ze,
*si;H;H,,,si;H;H,,si;ze,H;H;H*H;H;H,,,si;H;H,,,si;ze,,,,*,,,si;ze,,,,HO*,HO,*

10 (C1) A*B;B;B*C;C,,HO*si;ze,,*si;te;tw,*si;tw;te,H;H;H,*si;ze,HO,,,,*si;ze,,*si;ze,*si;ze,H;H;H
*si;H;H,,,,*si;tw;te,,*si;tw;te,si,*si;H;H,H;H;H,,si;tw;te*&2;H;H,,,,,,,si;H;H,si
,,,si;H;H,,,si;te;tw&1;te;tw,,,,si;ze,*,si;H;H,si,si;ze,*si;H;H,,,si;ze,si;ze,
*si;H;H,,,si;H;H,,si;ze,H;H;H*H;H;H,,,si;H;H,,,si;ze,,,,*,,,si;ze,,,,HO*,HO,*

11 (C1) A*B;B;B*C;C,,HO*si;ze,,*si;te;tw,*si;te;tw,H;H;H,*si;ze,HO,,,,*si;ze,,*si;ze,*si;ze,H;H;H
*si;H;H,,,,*si;tw;te,,*si;tw;te,si,*si;H;H,H;H;H,,si;tw;te*&2;H;H,,,,,,,si;H;H,si
,,,si;H;H,,,si;te;tw&1;te;tw,,,,si;ze,*,si;H;H,si,si;ze,*si;H;H,,,si;ze,si;ze,
*si;H;H,,,si;H;H,,si;ze,H;H;H*H;H;H,,,si;H;H,,,si;ze,,,,*,,,si;ze,,,,HO*,HO,*

12 (C1) A*B;B;B*C;C,,HO*si;ze,,*si;te;tw,*si;tw;te,H;H;H,*si;ze,HO,,,,*si;ze,,*si;ze,*si;ze,H;H;H
*si;H;H,,,,*si;tw;te,,*si;tw;te,si,*si;H;H,H;H;H,,si;tw;te*&2;H;H,,,,,,,si;H;H,si
,,,si;H;H,,,si;tw;te,,,,si;tw;te,si;H;H*si;ze,si,,si;H;H,,*si;ze,,si;ze,si;H;H,,
*si;ze,,si;H;H,,H;H;H,,*si;ze,H;H;H,si;H;H,,,,,*si;ze,,,,,si;ze,,*,HO,,HO*

Figure 12. The configurational CAST notations for **8**−**12**. The different CAST notations compared with those of reveromycin A (**8**) are marked by gray squares. The stereochemical differences are represented by the configurational CAST.

3.4. Distinction for α-D-Glucose and α-D-Mannose. The CANOST and configurational CAST notations from C6 for a conformer **6a** of α-D-glucose (**6**) are compared with those for a conformer **7a** of α-D-mannose (**7**), which is an epimer of **6** at C2. In Figure 10, the CANOST notations for **6a** and **7a** are shown in **6a**-I and **7a**-I, respectively, and the configurational CAST notations for **6a** and **7a** are shown in **6a**-II and **7a**-II, respectively. The same CANOST notation given for **6a** and **7a** represents that **6a** and **7a** possess the same two-dimensional structure. In the configurational CAST notations, the configurational differences for the C2 chiral center were represented by the different CAST codes for O2 and H2. In other words, diastereomeric differences in α-D-glucose and α-D-mannose were distinguished by the configurational CAST notations.

3.5. Applications to Reveromycins A and B. As applications to more complicated stereochemical structures, three configurational CAST notations of reveromycins A and B[9−14] are demonstrated. Reveromycin A has seven stereocenters, thus, 2^7 (=128) stereoisomers (including enantiomers) and $2^7/2$ (=64) diastereomers are possible. In the demonstration, reveromycin A (**8**), 15-epi-reveromycin A (**9**), 19-epi-reveromycin A (**10**), 5,19-epi-reveromycin A (**11**), and reveromycin B (**12**) are examined. The structures of **8**−**12** are shown in Figure 11, where the different partial structures compared with **8** are marked by gray circles.

Figure 12 shows the configurational CAST notations started from C-1 for these structures. The different notations compared with those of reveromycin A(**8**) are shown as gray squares. The results show that the configurational CAST explicitly represented the differences in stereochemistry between **8**, **9**, **10**, and **11**. The differences in structures of reveromycin B (**12**) were also distinguished.

Furthermore, some conformers of reveromycin A were examined, and they were identified as conformers each other by comparing the configurational CAST notations. Enantiomers were also successfully identified by the converted notations for enantiomers.

4. CONCLUSION

The canonical rearrangements of CAST notations for characterization of the configurational information have been described. Several applications demonstrated that the configurational CAST notations successfully represent the similarities and differences in absolute and relative configurations.

The configurational CAST notations with the conformational CAST notations are useful to describe three-dimensional structures in a database. They will be applied to broad fields of computer-assisted chemistry such as reaction prediction, synthetic design, and NMR chemical shift prediction studies that are now in progress and will be reported elsewhere.

ACKNOWLEDGMENT

We thank Dr. Jun Uzawa in RIKEN (The Institute of Physical and Chemical Research) for his helpful advice and encouragement. We thank Prof. Nigel Collier of National Institute of Informatics for reading of the manuscript. This research was supported in part by President's Special Research Grant from RIKEN, Shiseido Fund for Science and Technology, and a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

REFERENCES AND NOTES

(1) Wipke, W. T.; Dyott, T. M. Stereochemically Unique Naming Algorithm. *J. Am. Chem. Soc.* **1974**, *96*, 4834−4842.
(2) Choplin, F.; Dorschner, R.; Kaufmann, G.; Wipke, W. T. Computer Graphics Determination and Display of Stereoisomers in Coordination Compounds. *J. Organomatallic Chem.* **1978**, *152*, 101−109.
(3) Herndon, W. C.; Leonard, J. E. Canonical Numbering, Stereochemical Descriptors, and Unique Linear Notations for Polyhedral Clusters. *Inorg. Chem.* **1983**, *22*, 554−557.
(4) Akutsu, T. A New Method of Computer Representation of Stereochemistry. Transforming a Stereochemical Structure into a Graph. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 414−417.
(5) Markley, J. L.; Bax, A.; Arata, Y.; Hilbers, C. W.; Kaptein, R.; Sykes, B. D.; Wright, P. E.; Wüthrich, K. Recommendations for the

Presentation of NMR Structures of Proteins and Nucleic Acids. *J. Biomol. NMR* **1998**, *12*, 1−23.
(6) Satoh, H.; Koshino, H.; Funatsu, K.; Nakata, T. A Novel Canonical Coding Method for Representation of Three-dimensional Structures. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 622−630.
(7) Abe, H.; Kudo, Y.; Yamasaki, T.; Tanaka, K.; Sasaki, M.; Sasaki, S. A Convenient Notation System for Organic Structure on the Basis of Connectivity Stack. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 212−216.
(8) IUPAC−IUB Joint Commission on Biochemical Nomenclature, Conformational Nomenclature for Five- and Six-Membered Ring Forms of Monosaccharides and Their Derivatives. *Pure Appl. Chem.* **1981**, *53*, 1901−1905.
(9) Koshino, H.; Takahashi, H.; Osada, H.; Isono, K. Reveromycins, New Inhibitors of Eucaryotic Cell Growth III. Structures of Reveromycins A, B, C and D. *J. Antibiot.* **1992**, *45*, 1420−1427.
(10) Ubukata, M.; Koshino, H.; Osada, H.; Isono, K. Absolute Configuration of Reveromycin A, an Inhibitor of the Signal Transduction of Epidermal Growth Factor. *J. Chem. Soc., Chem. Commun.* **1994**, 1877−1878.
(11) Shimizu, T.; Masuda, T.; Hiramoto, K.; Nakata, T. Total Synthesis of Reveromycin A. *Org. Lett.* **2000**, *2*, 2153-2156.
(12) Drouet, K. E.; Theodorakis, E. A. Enantioselective Total Synthesis of Reveromycin B. *J. Am. Chem. Soc.* **1999**, *121*, 456−457.
(13) Masuda, T.; Osako, K.; Shimizu, T.; Nakata, T. Total Synthesis of Reveromycin B. *Org. Lett.* **1999**, *1*, 941−944.
(14) Cuzuppe, A. N.; Hutton, C. A.; Lilly, M. J.; Mann, R. K.; Rizzacasa, M. A.; Zammit, S. C. Total Synthesis of (−)-Reveromycin B. *Org. Lett.* **2000**, *2*, 191−194.

CI000136G