

Novel Canonical Coding Method for Representation of Three-Dimensional Structures

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,
Tempaku, Toyohashi, Aichi 441-8580, Japan

Received December 11, 1999

A new canonical coding method for representation of three-dimensional structures, CAST (CAnonical representation of STereochemistry), is described. CAST canonically codes stereochemistry around an atom in a molecule. The same CAST notations are given for atoms of molecules in the same conformation. The CAST code is based on the dihedral angles of four atoms that are uniquely defined by a molecular tree structure. CAST has successfully represented similarities and differences between several conformers.

1. INTRODUCTION

Stereochemical information is important in treating molecular structures. Computer-assisted synthetic design, reaction prediction, and NMR chemical shift prediction studies need to discriminate similarities and differences in stereochemistry around specific sites, such as strategic or reactive sites, and nuclei where NMR chemical shifts are to be predicted.

Some program systems for NMR chemical shift prediction are based on spectral databases widely used especially in chemical companies.^{1–4} The spectral databases store chemical shift values with the corresponding structures represented by some coding method such as HOSE (Hierarchically Ordered Spherical Description of Environment)⁵ and CANOST (autoCANOnicalization system for organic STructures),⁶ which represent connectivity relationships between atoms without stereochemistry. The NMR prediction systems often successfully predict chemical shifts but do not distinguish some stereoisomers. Thus, for diastereoisomers, just the same chemical shift values are given by these systems, although in general, NMR data of diastereoisomers have different chemical shift values. The predicted NMR chemical shifts ignoring stereochemistry for the interested molecule are not helpful for distinction of possible diastereoisomers despite its importance. NMR chemical shift predictions for each diastereoisomer need a new coding method that can represent stereochemistry, including conformational and configurational information.

Several computer description methods for molecular structures including stereochemistry have been described. One way is to represent the position of each atom by Cartesian or internal coordinates that are generally used in molecular modeling software. The methods that use coor-

dinates are appropriate for accurately describing molecular structure, but direct use for comparing similarities and differences of stereochemistry in molecules is very difficult because they do not canonically represent structures.

Some mathematical description methods have been developed as one-to-one nomenclature by computer.^{7–9} These methods describe a molecule based on topological relationships between atoms in a molecule using mathematical techniques such as the graph theoretical approach. By some approaches, one-to-one description of absolute configuration such as *R* or *S* on a chiral carbon is available. However, these methods are not sufficient for application to practical prediction of NMR chemical shifts or chemical reactions with accuracy, where the stereochemical environment between arbitrary atoms in a molecule is important.

For proteins and nucleic acids, another method for canonically representing stereochemistry was reported and developed for easier communication of NMR data and structures using dihedral angles.⁹ That method is useful for representing proteins and nucleic acids but hard to apply to arbitrary organic molecules in various structures.

This paper introduces a new coding method CAST (CAnonical representation of STereochemistry) that canonically represents stereochemistry around a specific site in a molecule. Applications of CAST to several conformers are also described.

2. METHOD

2.1. Model Structures. Some conformers of D-glucose are used (Figure 1). ⁴C₁ (**1**) and acyclic (**2**) forms of D-glucose¹⁰ are used for description of the CAST method. For demonstration of CAST applications, ⁴C₁, ¹C₄, and ¹⁴B conformations with acyclic form in some conformations are used.

The numbers on carbon atoms in Figure 1 are the serial numbers that correspond to those usually used for carbon atoms in saccharides. Hereafter, carbon atoms are represented as the atomic symbol with the serial number (e.g., C-1), oxygen and hydrogen atoms that connect to carbon atom are represented as the atomic symbol with the serial number of

* Corresponding authors.

† Telephone: 048-462-1111 (ext. 3449). E-mail: hsatoh@postman.riken.go.jp.

‡ RIKEN.

§ PRESTO.

|| Telephone: 048-462-1111 (ext. 3031). E-mail: koshino@postman.riken.go.jp.

⊥ Toyohashi University of Technology.

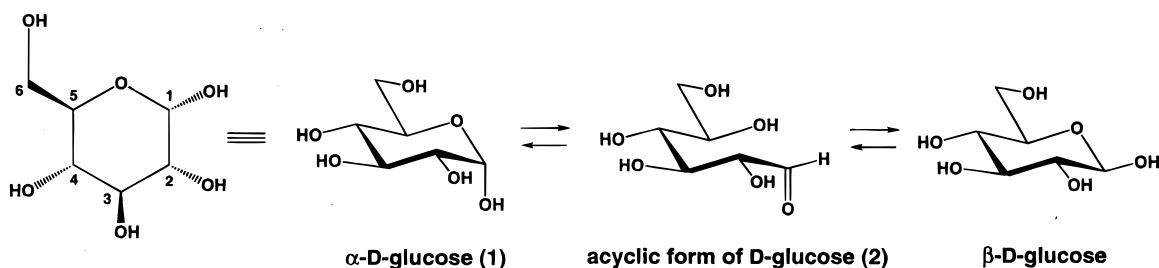


Figure 1. Equilibrium of D-glucose. α -D-Glucose (1) and the acyclic form of D-glucose (2) were used as model structures.

Table 1. CANOST (Modified) Code List

no.	code	substructure	no.	code	substructure	no.	code	substructure
1 ^a	&nc	code for ring closing	12 ^b	Y		22	LF	—F
2	T	—C \equiv	13 ^c	Y1	H—C \equiv	23	LC	—Cl
3	T1	H—C \equiv	14 ^d	YTT		24	LB	—Br
4	DD	=C=				25	LJ	—I
5	DS		15	V	C=O	26	X	other atom
6	D1		16	V1		27 ^e	XR	
7	D2		17	VD	=C=O	28	XW	
8	C		18	QD	=O	29	XD	=X
9	C1		19	Q	—O—	30	XN	=X=
10	C2		20	Q1	—O—H	31	XT	\equiv X
11	C3		21	Q0	—O ⁻	32 ^f	Xnp	XH _{np}
						33	H	—H

^a nc is a serial number of the closing site. ^b Aromatic carbon without hydrogen. ^c Aromatic carbon with hydrogen. ^d —C(OH)—C(O)— in troponoid. ^e Non-carbon atom in aromatic structure. ^f np is the number of connecting hydrogens.

the carbon atom (e.g., O-1 and H-1), and a hydrogen atom of a hydroxyl group is represented as OH with the serial number of carbon atom that connects to the hydroxyl group (e.g., OH-1).

2.2. Coding. CAST gives a linear notation that represents stereochemical relationships between atoms in a molecule. The CAST notation is used in parallel with a notation of two-dimensional structure represented by connectivity relationships between atoms, as follows.

2.2.1. Two-Dimensional Structural Notations. In the current execution, two-dimensional structures are represented by a substructural coding method, CANOST,⁶ modified to encode every atom in a molecule for CAST notation. The modified CANOST codes are listed in Table 1.

The first step of construction of the CANOST notation is to draw a molecular tree structure from each atom in a molecule. A molecular tree from C-6 of D-glucose in acyclic

form (2) is shown at A-I in Figure 2. In a molecular tree, atoms are arranged into each level that is defined as 1 plus the number of bonds from the starting atom. For example, C-5 in A-I, which has one bond from C-6, and the next C-4 are arranged into the second and the third levels, respectively. Hereafter, the level of the starting atom is defined as the highest level.

For representation of a ring system, an ampersand code "&" with numbering is used at the ring closure point. When the number of atoms at the lowest level for the ring is 2, i.e., in the case of an odd membered ring, a bond between the atoms is cut at first. Then, an & code is added to the next lower level and connected to each of the cut point atoms. When the number of atoms at the lowest level is 1, i.e., in the case of an even membered ring such as α -D-glucose (1), a bond between the lowest level atom and an atom at the higher level is cut. After duplication of the lowest atom, the

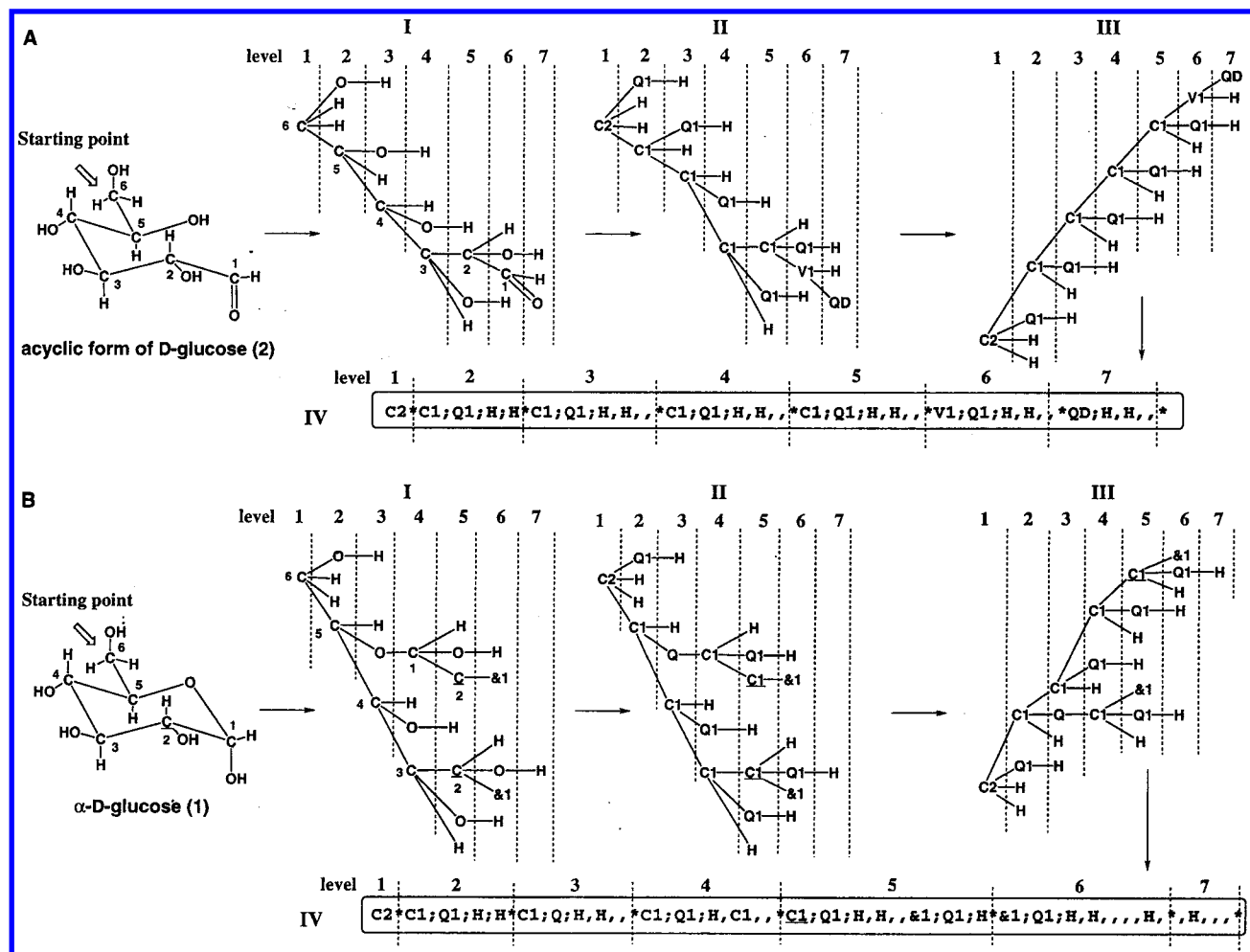


Figure 2. CANOST coding method that has been modified for CAST. A-I and B-I are molecular trees by atoms for **2** and **1**, respectively. A-II and B-II are molecular trees by modified CANOST codes. A-III and B-III are molecular trees ordered according to the CANOST priorities. A-IV and B-IV are linear notations called CANOST notations.

duplicate atom is connected to the higher atom, and then an & code is added to both. Duplicated atoms in α -D-glucose (**1**) are underlined in a molecular tree shown at B-I in Figure 2.

The second step is assignment of a CANOST code for each atom (A-II and B-II in Figure 2). A list of modified CANOST codes is shown in Table 1, which is almost the same as the original ones, but some codes were added, such as the "H" code, which stands for a hydrogen atom. The original CANOST code is assigned to a substructure, but the modified CANOST code is assigned to every atom in the molecule for comparison with the CAST code that is assigned to every atom in the molecule.

The third step is ordering the atoms in each of the levels according to the priority of the CANOST codes (A-III and B-III in Figure 2). The priority is defined as the order of the codes listed in Table 1. The ordering is performed from level 1 to the lowest level. For a molecular tree of a ring system, a duplicated atom at lower priority is deleted after the ordering (B-III in Figure 2).

The last step is conversion to a linear notation (hereafter, called a CANOST notation) according to almost the same rule as that of the original CANOST coding method (A-IV and B-IV in Figure 2). An asterisk "*", a semicolon ";", and a comma "," are separators in the CANOST notation. The asterisk and the semicolon are code and level separators,

respectively. The comma is a group separator. The group consists of atoms that connect to the same atom at one higher level.

2.2.2. CAST Notations. A CAST code is defined for each atom on the basis of the dihedral angle *uniquely* selected by four atoms along the molecular tree, from the atom to be given a CAST code, going back to the higher level (Figure 3). When a CAST code is assigned to an atom at the n th level, denoted by L_n in Figure 3, the dihedral angle is defined by the atom and the three sequential atoms at the $(n-1)$ th, $(n-2)$ th, and $(n-3)$ th levels along the molecular tree; namely, a dihedral angle of $L_n-L_{n-1}-L_{n-2}-L_{n-3}$ is used for the coding, as shown in Figure 3. For the atoms at the first three levels of every molecular tree, CAST codes of **A**, **B**, and **C** are assigned, respectively. For some structures that contain an undefinable dihedral angle such as alkynes, specific codes are defined. A detailed definition of CAST codes is described below.

Twelve types of CAST code are defined to twelve areas of the dihedral angle, where 360° are equally divided into twelve areas and CAST codes are defined to the areas as the first two letters of numbers *zero* to *eleven* as shown in Figure 4. For example, a CAST code of **tw**, **si**, or **te** is assigned to each atom in the staggered conformation (I in Figure 5), and **ze**, **fo**, or **ei** is assigned to that in the eclipsed conformation (II in Figure 5). These twelve types of CAST

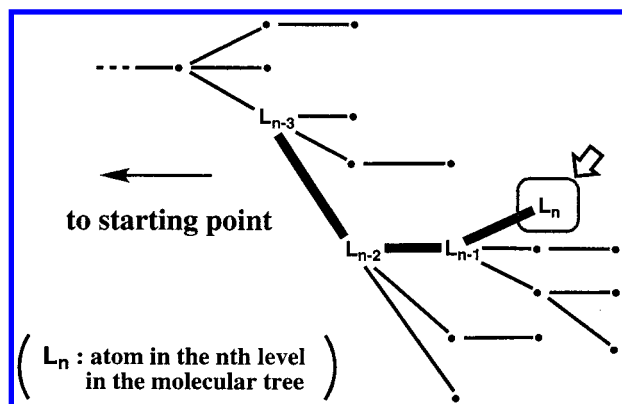


Figure 3. Dihedral angle used for definition of CAST codes. The dihedral angle is uniquely selected by four atoms along the molecular tree, from an atom of L_n going back to the starting point. A CAST code is defined for the L_n atom based on the dihedral angle around a bond between L_{n-1} and L_{n-2} .

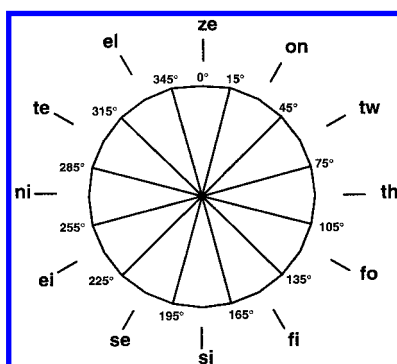


Figure 4. CAST codes defined for dihedral angles. Twelve types of CAST code are defined to twelve areas of the dihedral angle. The CAST codes are defined as the first two letters of clockwise numbering from *zero* to *eleven*.

code for 30° are defined here to distinguish differences and similarities in conformations in the same molecules examined in this paper. If necessary, the number of codes, the range of angles, and weighted divisions using two or more kinds of the range as well as the equal division can be freely defined according to the purpose.¹¹

This *clock-type* CAST codes are also used for double bond systems. For double bond systems, the dihedral angle around the double bond is restricted by its planarity character; thus a CAST code of *ze* is assigned to an allylic atom of L_n at the *cis* position against the L_{n-3} atom and *si* is assigned to one at the *trans* position (Figure 6). For conjugated systems, if sequential bond types are $L_{n-3}=L_{n-2}-L_{n-1}=L_n$, the clock-type CAST codes are assigned to the atom L_n based on the dihedral angle around the $L_{n-2}-L_{n-1}$ single bond, because the single bond in a conjugated system is rotatable. For a $L_{n-3}-L_{n-2}=L_{n-1}-L_n$ sequence in a conjugated system, a CAST code of *ze* or *si* is assigned to the atom L_n , which is the same as that for a simple double bond system.

For triple bond systems, specific codes are defined because a usual dihedral angle around the triple bond cannot be defined by its linearity character. If sequential bond types are $L_{n-3}-L_{n-2}-L_{n-1}\equiv L_n$, a CAST code of *SP* is assigned to an atom of L_n (I in Figure 7), and if sequential bond types are $L_{n-3}-L_{n-2}\equiv L_{n-1}-L_n$, a CAST code of *TT* (Terminal of Triple bond(s)) is assigned to an atom of L_n (II in Figure 7). If sequential bond types are $L_{n-3}\equiv L_{n-2}-L_{n-1}-L_n$, a clock-type CAST code is assigned on the basis of the dihedral angle

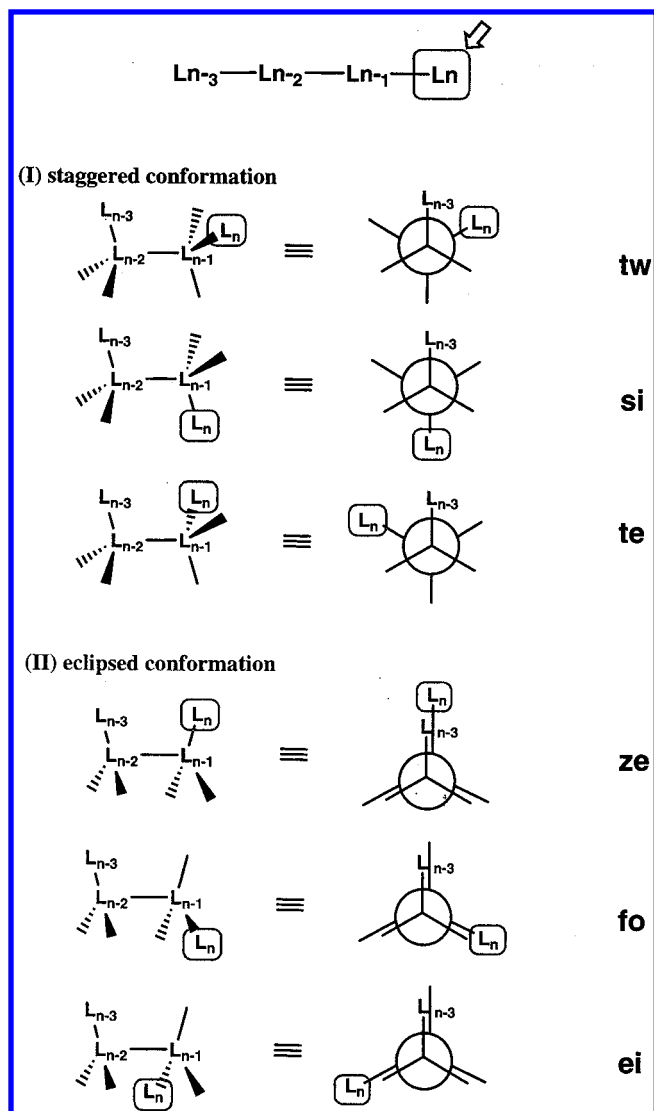


Figure 5. Assignment of CAST codes for alkanes. The CAST code of *tw*, *si*, or *te* is assigned to the staggered conformation (I), and *ze*, *fo*, or *ei* is assigned to the eclipsed conformation (II).

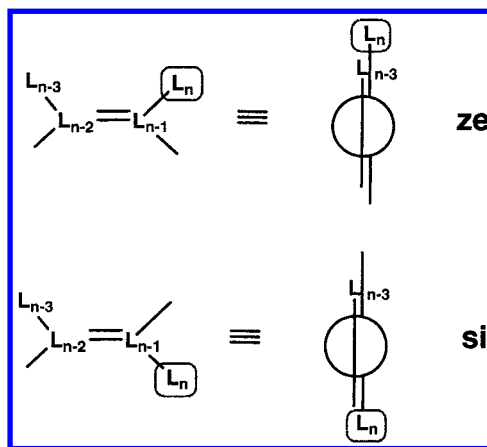


Figure 6. Assignment of CAST codes for alkenes. For double bond systems, a dihedral angle around the double bond is restricted by its planarity character. The CAST code of *ze* is assigned to an allylic atom of L_n at the *cis* position against the L_{n-3} atom, and *si* is assigned to one at the *trans* position.

around the line-bond structure of $L_{n-4}-L_{n-3}\equiv L_{n-2}-L_{n-1}$, namely, on the basis of the dihedral angle defined by L_{n-5} , L_{n-4} , L_{n-1} , and L_n , and the clock-type CAST code is set with

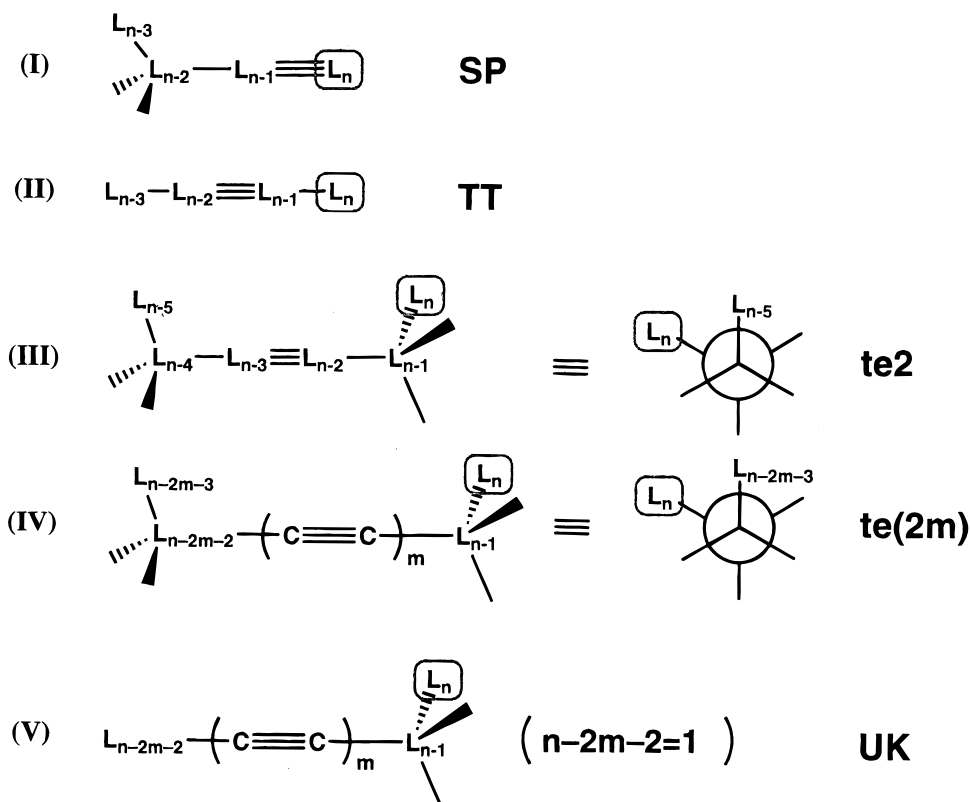


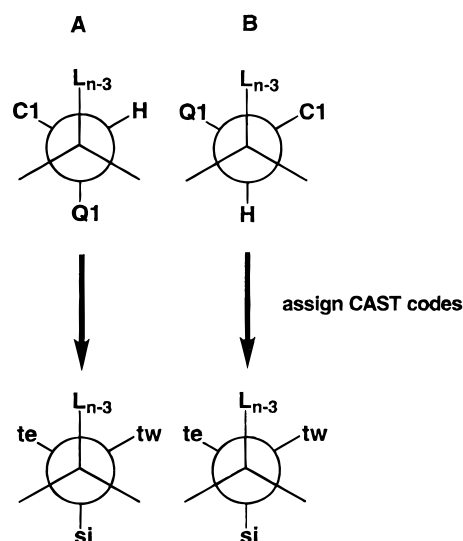
Figure 7. Assignment of CAST codes for alkynes. For triple bond systems, specific codes are defined because the usual dihedral angle around the triple bond cannot be defined by its linearity character (I, II, and V). The *clock-type* CAST codes are assigned on the basis of the dihedral angle around the line-bond structure of sequential *sp* atoms with the number of *sp* atoms (III and IV).

the number of the skipped *sp* atoms, such as **te2** (III in Figure 7). In general, the clock-type CAST code based on the dihedral angle defined by L_{n-2m-3} , L_{n-2m-2} , L_{n-1} , and L_n with the number of skipped *sp* atoms is assigned to L_n in $L_{n-2m-3}-L_{n-2m-2}-(C\equiv C)_m-L_{n-1}-L_n$ structure, where m is the number of triple bonds (IV in Figure 7). In the structure of $L_{n-2m-2}-(C\equiv C)_m-L_{n-1}-L_n$, if L_{n-2m-2} is the starting atom (namely, $n-2m-2=1$), a CAST code of **UK** (*UnKnown*) is assigned to L_n (V in Figure 7).

CAST masks hydrogen atoms in methylene and methyl groups where a specific CAST code of **H** is assigned, because assignment of different CAST codes to the atoms would be caused by initial ordering atoms in the input file (MDL-molfile format). The masking function can be used to ignore insignificant stereochemical information. For example, a specific CAST code of **OH** can be assigned to hydrogen atoms in hydroxyl groups where the orientation of the hydrogen atoms would not be discussed in some cases.

After recognition of CAST codes for every atom in a molecular tree, a linear notation is constructed by the same way as that for CANOST notations. For parallel comparison between CAST and CANOST notations, the molecular tree is ordered according to the priority of the CANOST codes. Figure 8 shows parts of CANOST and CAST notations, where differences in stereochemistry between the structures, which have the same CANOST notations, are represented by CAST notations.

2.3. Representation of Conformers. Differences and similarities in some equilibrium conformers of α -D-glucose (1) and acyclic form of D-glucose (2) were represented by CAST.



CANOST notation for A, B	C1; Q1; H,
CAST notation for A	te; si; tw
CAST notation for B	tw; te; si

Figure 8. Parallel comparison of parts of CANOST and CAST notations. Differences in stereochemistry are represented by CAST notations for structures that have the same CANOST notations.

For an acyclic form of D-glucose, a chair-form-type conformation (**2a** in Figure 9) and a zigzag conformation (**2b** in Figure 9) are examined. Figure 9 shows the Newman projection of **2a** from C-1 (**2a-1**) and from C-6 (**2a-2**), and that of **2b** from C-1 (**2b-1**) and from C-6 (**2b-2**), respec-

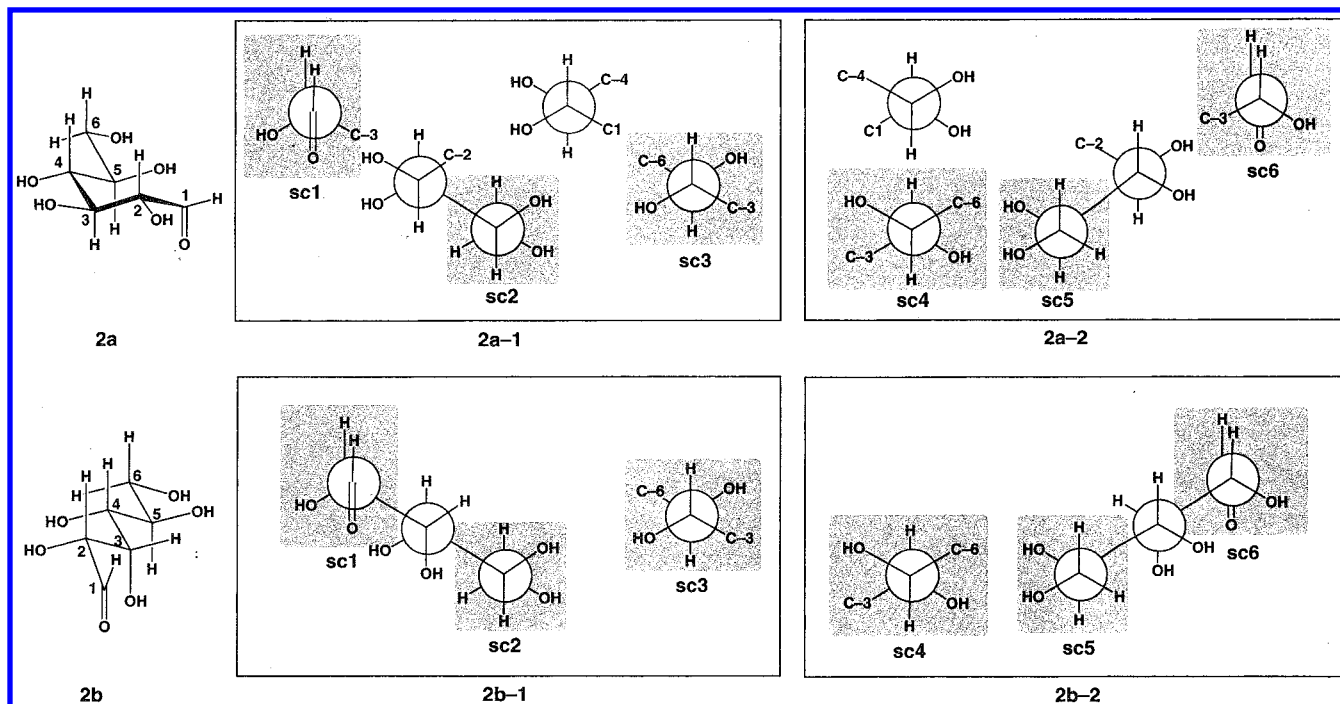


Figure 9. Two kinds of conformations of acyclic form of D-glucose (2) with Newman projections. One is a chair-form-type conformation (2a) and the other is a zigzag conformation (2b). The same conformations between 2a and 2b are denoted by the same labels of sc (the same conformation), 1, 2, 3, 4, 5, and 6.

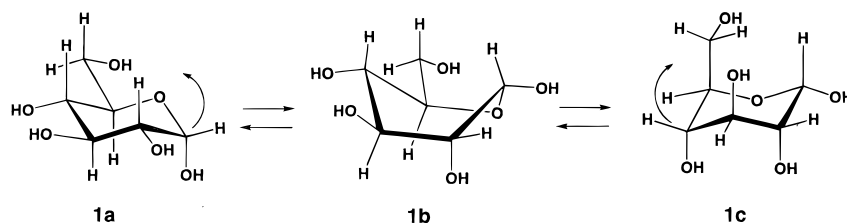


Figure 10. Three kinds of conformations of α -D-glucose (1). Two chair conformations (1a and 1c) and a boat conformation (1b). Flipping of C-1 of 1a up to the β -side achieved conversion from 1a to 1b, and flipping of C-4 of 1c up to the β -side achieved conversion from 1c to 1b.

tively. In the Newman projections, substructures with the same conformations are denoted by the same labels as sc (the same conformations) 1, 2, 3, 4, 5, and 6.

For α -D-glucose (1), two kinds of chair conformations (1a and 1c in Figure 10) and a boat conformation (1b in Figure 10) were examined.

3. RESULTS AND DISCUSSIONS

3.1. CAST Notations for Conformers of the Acyclic Form of D-Glucose. Figure 11 shows the CANOST and CAST notations for chair-form-type (2a) and zigzag conformations (2b), where the same sequences of codes in the notations for 2a as those for 2b are denoted by the same labels as sn (the same notations) 1, 2, 3, 4, 5, and 6.

Starting from C-1, the same CANOST notations were given for 2a and 2b as shown by sn1 in Figure 11, while two subsequences (sn2 and sn3) were the same between the CAST notations for 2a and those for 2b (II in Figure 11). The sn3 corresponds to the sc3 conformation (Figure 9), which is determined by the dihedral angle around a bond between C-4 and C-5, and the sn2 corresponds to the sc2 in Figure 9, which is the conformation determined by the dihedral angle around a bond between C-5 and C-6. The same conformation denoted by sc1 conformation (Figure 9),

which is determined by the dihedral angle around a bond between C-1 and C-2, is hidden under CAST codes of A, B, and C.

Starting from C-6, the same CANOST notations were also given for 2a and 2b as shown by sn5 in Figure 11, while two subsequences (sn4 and sn6) were the same between the CAST notations for 2a and those for 2b (IV in Figure 11). The sn4 corresponds to sc4 in Figure 9, which is the conformation determined by the dihedral angle around a bond between C-5 and C-4, and the sn6 corresponds to sc6 in Figure 9, which is the conformation determined by the dihedral angle around a bond between C-2 and C-1, which was hidden in the CAST notation started from C-1. The same conformation denoted by sc5 in Figure 9, which is the conformation determined by the dihedral angle around a bond between C-6 and C-5, is hidden under CAST codes of A, B, and C.

3.2. CAST Notations for Conformers of the Cyclic Form of α -D-Glucose. Figure 12 shows CANOST and CAST notations for two chair conformations (4C_1 (1a) and 1C_4 (1c) in Figure 10) and for a boat conformation (^{14}B (1b) in Figure 10) of α -D-glucose (1). In Figure 12, CANOST notations started from C-1 and C-6 are shown in I and III, respectively, and CAST notations started from C-1 and C-6

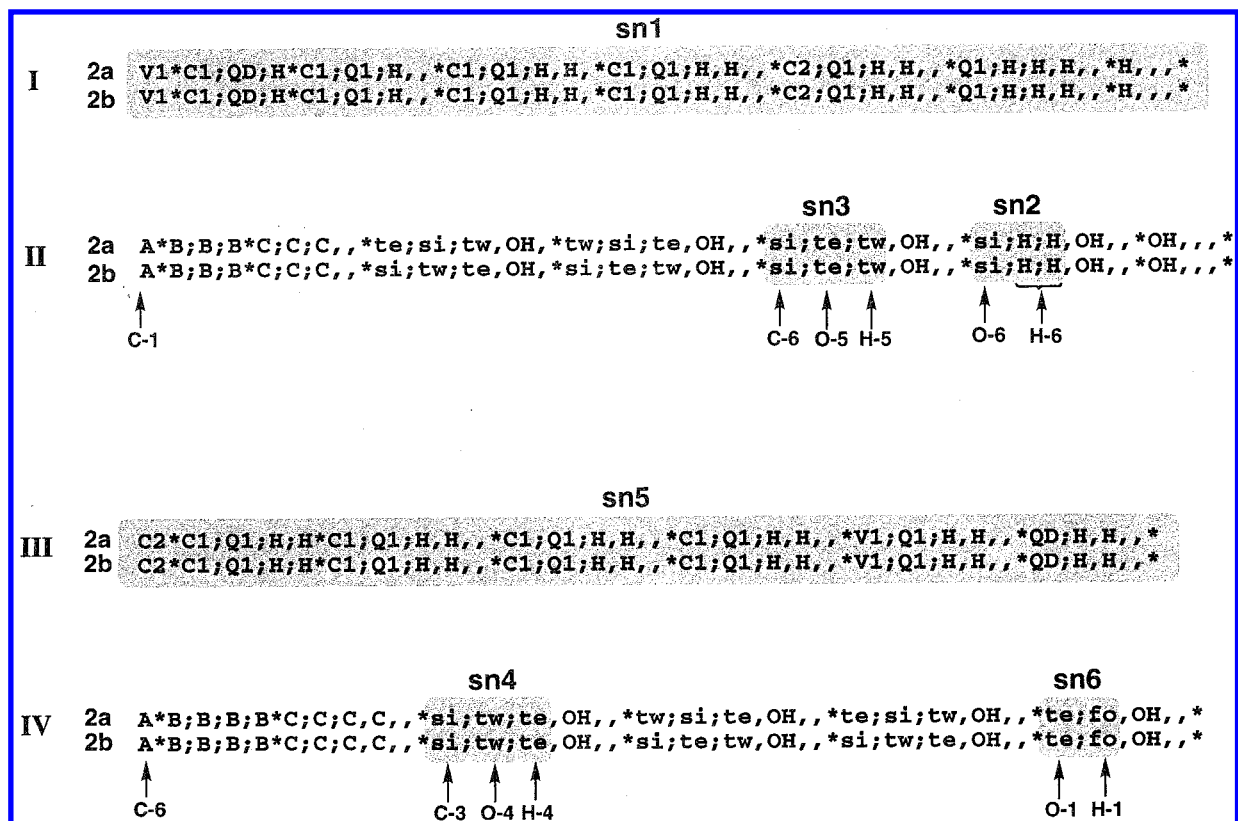


Figure 11. CANOST and CAST notations for a chair-form-type conformation (2a) and a zigzag conformation (2b): (I and II) CANOST and CAST notations from C-1, respectively; (III and IV) CANOST and CAST notations from C-6, respectively. The same sequences of codes in the notations for 2a as those for 2b are denoted by the same labels as sn (the same notations), 1, 2, 3, 4, 5, and 6.

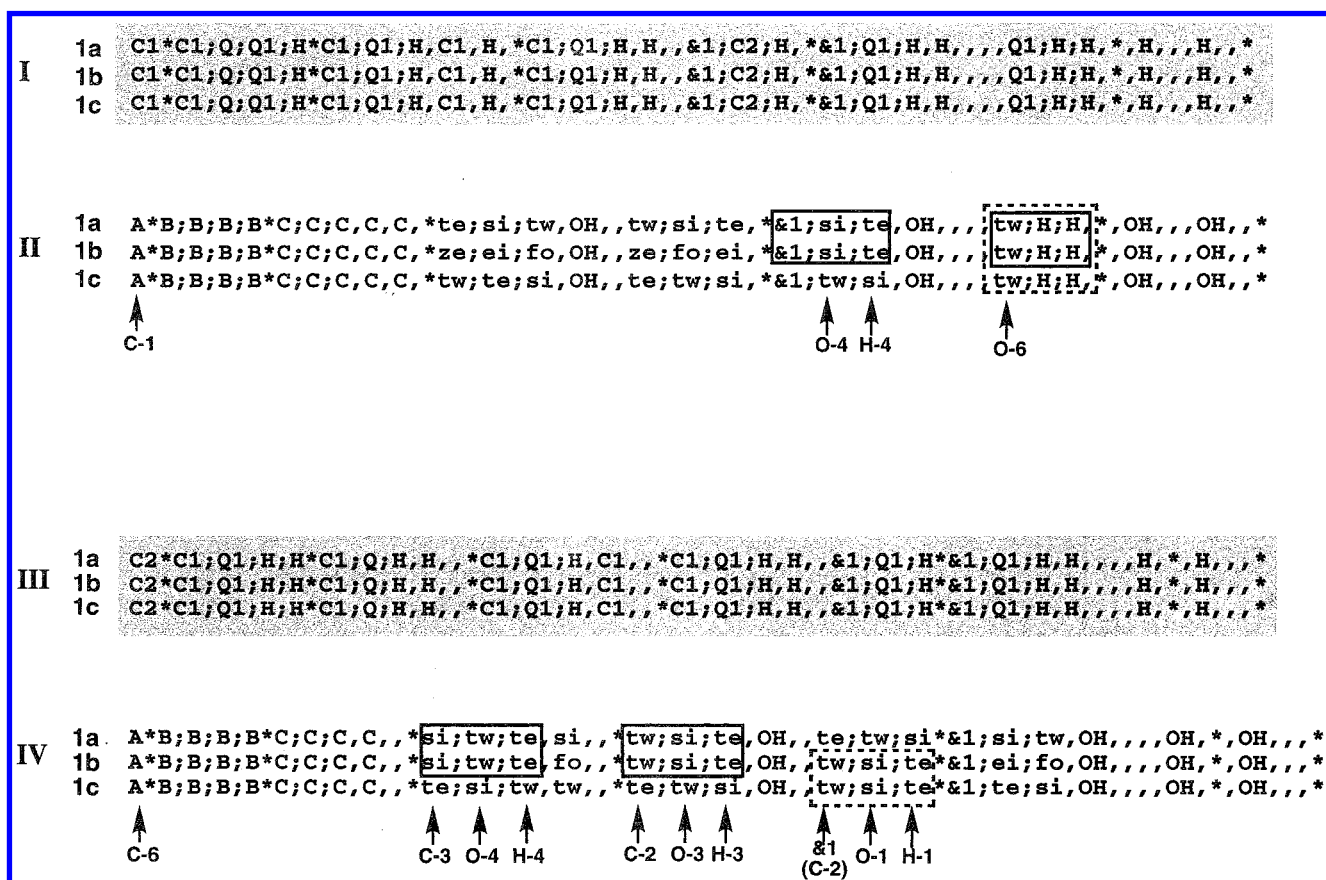


Figure 12. CANOST and CAST notations for two chair conformations of 1a and 1c, and for a boat conformation of 1b of α -D-glucose (I): (I and II) CANOST and CAST notations from C-1, respectively; (III and IV) CANOST and CAST notations from C-6, respectively. The same sequences of codes in notations for 1a as those for 1b are enclosed by solid-line frames, and the same sequences of codes in notations for 1b as those for 1c are enclosed by dotted-line frames.

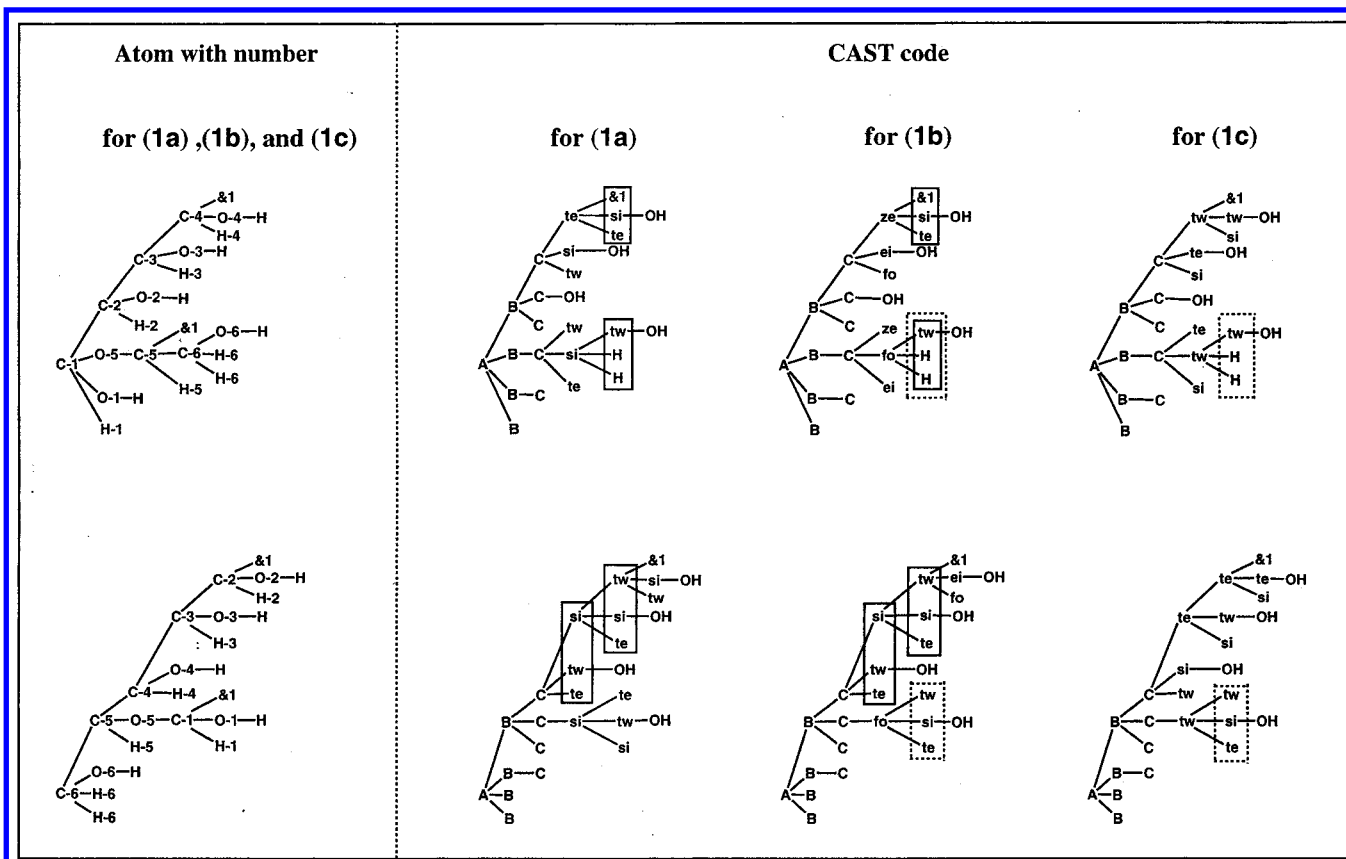


Figure 13. Molecular trees for the conformations of **1a–c** by atoms with number and by CAST codes. Those in the first and second rows are started from C-1 and from C-6, respectively. The same sequences of codes for **1a** as those for **1b** are enclosed by solid-line frames, and the same sequences of codes for **1b** as those for **1c** are enclosed by dotted-line frames.

are shown in II and IV, respectively. Figure 13 shows molecular trees for the conformations **1a–c**. In Figure 13, molecular trees in the first and second rows are started from C-1 and from C-6, respectively. Molecular trees in the first column are denoted by the atoms with the number for **1a–c**. Molecular trees in the second, third, and fourth are denoted by CAST codes for **1a–c**, respectively.

Starting from C-1, the same CANOST notations were given for three of the conformations (I in Figure 12), and starting from C-6, the same CANOST notations were also given for them (III in Figure 12).

To change conformations from **1a** to **1b**, the dihedral angles around bonds between C-3 and C-4, between C-4 and C-5, and between C-5 and C-6 are retained. The retentions are represented in the molecular trees and CAST notations, where the same CAST codes that represent the retention are enclosed by solid-line frames in Figure 12 and Figure 13. The same CAST codes are for nodes of C-2 C-3, O-3, O-4, O-6, H-3, H-4, and H-6. To change conformations from **1b** to **1c**, dihedral angles around bonds between C-1 and C-2, between C-1 and O-5, and between C-5 and C-6 are retained. The same CAST codes that represent the retention are enclosed by dotted-line frames in molecular trees of Figure 12 and Figure 13. The same CAST codes are for nodes of O-1, O-6, H-1, and H-6.

Flipping of C-1 of **1a** up to the β -side achieved conversion from **1a** to **1b**, and flipping of C-4 of **1c** up to the β -side achieved conversion from **1c** to **1b**, as shown in Figure 10. The results demonstrated the similarities and differences caused by the flipping.

As shown in these results, the CAST successfully represented the similarities and differences between different conformations of the same molecules for both acyclic and cyclic molecules.

In general, the whole of a molecule can be generally represented by two CAST notations with CANOST started from two atoms that are apart from each other through more than three bonds.

Furthermore, representation of the configuration is possible by simple modification of the CAST coding method, about which will be reported elsewhere.

4. CONCLUSION

CAST, a novel method for canonical representation of three-dimensional structures has been described. Application of CAST to some conformers of α -D-glucose and acyclic D-glucose demonstrated that CAST with CANOST notations can canonically represent similarities and differences between several conformations in the same structures.

CAST will be applied to the broad field of computer-assisted chemistry, such as in reaction prediction, synthetic design, and NMR chemical shift prediction studies that are now in progress.

ACKNOWLEDGMENT

We thank Dr. Jun Uzawa, the head of the Molecular Characterization Division in RIKEN (The Institute of Physical and Chemical Research), for his helpful advice and encouragement. H.K. gratefully acknowledges financial support by the President's Special Research Grant from RIKEN.

REFERENCES AND NOTES

- (1) Funatsu, K.; Sasaki, S. Recent Advances in the Automated Structure Elucidation System, CHEMICS. Utilization of Two-Dimensional NMR Spectral Information and Development of Peripheral Functions for Examination of Candidates. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 190–204.
- (2) Bremser, W. Structure Elucidation and Artificial Intelligence. *Angew. Chem., Int. Ed. Engl.* **1988**, *27*, 247–260.
- (3) Will, M.; Fachinger, W.; Richert, J. R. Fully Automated Structure Elucidation – A Spectroscopist's Dream Comes True. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 221–227.
- (4) (a) Van Bramer, S. ACD/CNMR and ACD/HNMR: A Software Review. *Concepts Magn. Reson.* **1997**, *9*, 271. (b) ACD/CNMR and ACD/HNMR; Advance Chemistry Development, Inc.: <http://www.acdlabs.com>.
- (5) Bremser, W. HOSE – A Novel Substructure Code. *Anal. Chim. Acta* **1978**, *103*, 355–365.
- (6) 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.
- (7) (a) Wipke, W. T.; Dyott, T. M. Stereochemically Unique Naming Algorithm. *J. Am. Chem. Soc.* **1974**, *96*, 4834–4842. (b) Choplin, F.; Dorschner, R.; Kaufmann, G.; Wipke, W. T. Computer Graphics Determination and Display of Stereoisomers in Coordination Compounds. *J. Organomet. Chem.* **1978**, *152*, 101–109.
- (8) (a) Herndon, W. C.; Leonard, J. E. Canonical Numbering, Stereochemical Descriptors, and Unique Linear Notations for Polyhedral Clusters. *Inorg. Chem.* **1983**, *22*, 554–557. (b) 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.
- (9) 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.
- (10) 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.
- (11) Modified 12 CAST codes are used for our NMR chemical shift prediction system that will be reported elsewhere. The wide range of 40° is given to each of the codes of **tw**, **si**, **te** for a staggered conformation and to each of those of **ze**, **fo**, **ei** for an eclipsed conformation to weight two of the major and important conformations. To the rest of the codes, i.e., **on**, **th**, **fi**, **se**, **ni**, and **el**, the narrow range of 20° is given.

CI990147D