Carbohydrate Research 401 (2015) 51–57

Contents lists available at ScienceDirect

Carbohydrate Research

journal homepage: www.elsevier.com/locate/carres

A tree-step computational approach to simplify conformational determination of cellobiose and lactose

Dong Chen *, Zhichao Wei, Yuheng Yao, Bo Liu *

Institute of Photo-Biophysics, Physics and Electronics Department, Henan University, 475004 Kaifeng, China

ARTICLE INFO

Article history:
Received 11 August 2014
Received in revised form 27 October 2014
Accepted 29 October 2014
Available online 4 November 2014

Keywords:
Tree-step computational approach
Conformational determination
Cellobiose
Lactose
Quantum mechanical calculation

ABSTRACT

Great theoretical attentions have been paid on the conformational preference of individual molecular building blocks of carbohydrates because it is helpful for assignments of the experimental signals and explorations of the biological implications. A tree-step approach is applied here to simplify the conformational determination of phenyl β-cellobioside and benzyl β-lactoside, for which 35 and 23 initial structures are built, respectively. After the high-level calculations, low-energy conformers are determined and then compared with previous experimental and theoretical results. The low-energy conformers are reconstructed in our work for both cellobiose and lactose and the results show a quantitative agreement between the experimental signature and the predicted IR vibration assignment. In addition, two low-energy conformers, which are predicted in our work, have not been reported by the previous work using the traditional method. The tree-step computational approach provides an alternative timesaving and accurate method to focus on determining the preferred conformations of disaccharides.

1. Introduction

Molecular conformations and molecular shape of oligosaccharide chains are responsible for their biological activities in glycoproteins or glycolipids.[1,2] Many specific interactions in cell involve such glycoconjugates. Even in the monosaccharide, the specificity involved in molecular recognition is already presented. Therefore, knowledge of the conformational landscapes of oligosaccharides is the first step to understand the origins of their molecular specificities.

In recent years, the conformational research of individual molecular building blocks of carbohydrates in the gas phase has been extensively studied.[3–6] Experimentally, it is of challenge to study the structural motif of a single molecule in the gas phase due to the difficulty in removing the effect of the environment. One method to solve this problem is the spectroscopic measurements involving conformer specificity and mass-selection, which can probe the vibrational signatures of small oligosaccharides initially free of the environment.[7–10] By combining with *ab initio* computational chemistry, these signatures can be assigned to conformational structures. In general, the traditional theoretical study relays on conformational searches to generate the initial structures.[8,9,11] The advantage of conformational search is to generate a large number of reasonable starting structures without assumptions. However, the number of the starting structures is often so numerous that optimization consumes computational costs a lot.[12–14] Alternatively, an application of energy cutoff in the Monte-Carlo Multiple Minimum (MCMM) conformational search can reduce the starting conformers before the optimization process. However, some conformers with high energies could be stable with low energies after structural optimizations and quantum mechanical (QM) calculations. A similar phenomenon also can be seen in our previous work for studying the structures of dimannose. This difficulty becomes obvious especially for the complex and flexible conformations.[11,15]

Fortunately, our previous works inspire a potential approach to overcome this difficulty. As we all know, the structural motif can be attributed to a delicate balance among various interactions through covalent bonds, such as those hydrogen bonds that determine their skeletal structures and the hydrogen bonding network that are formed among multiple hydroxyl groups. A detailed investigation of the linkage constraint and hydrogen-bonded network determining the structural preferences of dimannosides was conducted in our previous work.[16] The results suggest that a tree-step computational approach is feasible to build up reasonable initial structures manually so as to simplify the conformational determination. During the building processes, the main factors contain glycosidic linkage, inter-ring hydrogen bonds, and cooperative intra-ring hydrogen bonds. The number of starting structures generated by this approach is much smaller than that by the traditional

* Corresponding authors. Tel.: +86 371 3881602.
E-mail addresses: dongchen@henu.edu.cn (D. Chen), boliu@henu.edu.cn (B. Liu).

http://dx.doi.org/10.1016/j.carres.2014.10.024
0008-6215/

conformational approach, which can reduce at least 90% of the total calculation time of the traditional strategy and at the same time obtain the accurate lowest-energy conformations.[15]

Cellulose disaccharide and its C-4′ epimer are common in the core oligosaccharide in N-linked glycoproteins and especially critical to the future prospects of the glycoprotein. Cellulose disaccharide joins two glucose units through a β1,4-linkage. Lactose is the C-4′ epimer of cellobiose, where OH4′ is oriented axially rather than equatorially. The changed hydroxyl configuration at C-4′ exhibits a dramatically different hydrogen bonding network and the rigidity of the β1,4-glycosidic linkage.[8] For these reasons, phenyl β-cellobioside and benzyl β-lactoside have been studied both experimentally and theoretically.[8,9] These efforts provide us the reliable data to examine our results. Therefore, cellulose disaccharide and its C-4′ epimer are treated as model molecules in our calculations and revealed the process of determining the preferred conformations using the tree-step computational approach. In experiments, the addition of the phenyl and benzyl groups is used as photo-tag. In order to compare with the experimental and theoretical results, the two disaccharides are attached by a phenyl or benzyl group in our calculations. Moreover, their presence does not disturb the conformational preferences of the original disaccharides.

In the following, the tree-step computational approach[15,16] is employed to determine the preferred structures of cellobiose and lactose, especially to illustrate the details of the building processes. This work will help the readers understand the tree-step computational approach and apply it to another biochemical system. The computational details are presented in Section 2. The H-bond networks of each structure and the building methods of the initial structures are analyzed in Sections 3.1 and 3.2. A comparison of vibrational properties between our work and others is discussed in Section 3.3.

2. Computational methods

A tree-step computational method can provide us an effective building procedure to simplify the conformational determination through the generation of few initial structures. The strategy generated new starting structures according to the glycosidic linkage, the categories of inter-ring hydrogen bonds (H-bonds) and cooperative H-bonds. The details are as follows. (i) for the glycosidic linkage, we only consider two cases: *cis*[17,18] conformation with *anti*/*syn* (φH, ψH) ≈ 180°, 0° and *trans*[19,20] conformation with *syn*/*syn* (φH, ψH) ≈ 0°, 0°. For example, a *cis* conformation of cellulose disaccharide is shown in Figure 1. (ii) All reasonable inter-ring hydrogen bonds are considered as keys contributing to the complex inter-ring hydrogen bonding network. Actually, the orientations of the hydroxymethyl group of ring M′ and M are rotated to form inter-ring H-bonds in the building process. Intermolecular H-bonds are favorable to the stabilization of a molecule. (iii) The cooperative intra-ring H-bonds are included in the building process. The three hydroxyl groups can interact together in a clockwise chain of intramolecular H-bond of OH2′→OH3′→OH4′, which is labeled 'c', or in a counterclockwise chain of intramolecular H-bond of OH4′→OH3′→OH2′, which is labeled 'cc'. In this paper, an arrow points from a proton donor to its acceptor. For example, the H-bond of A→B indicates A is the proton donor and B is its acceptor. A H-bond of A↔B means two directions.

Both A→B and B→A can exist. For the *cis* structure of cellulose disaccharide, the orientations of the hydroxyl groups OH2′ and OH3 cooperatively form OH3→OH2′→OH3′→OH4′. Moreover, two different orientations of the cooperative hydrogen bonds can be built as the solid arrows labeled in Figure 1. According to the important role of H-bonds, the basic rule of building process is to form as many as H-bonds in the stable structures and the orientations of phenyl and benzyl are not considered in the building process.

Considering all the adjustable structural parameters, a lot of conformers can be built. Actually, some of them are ruled out based on chemical intuition and the reason will be illustrated in the results and discussion section. Then, all the initial structures were submitted for density functional theory (DFT) geometry optimization using the Gaussian 03 program package (Revision A.1)[21] with B3LYP/6-311+G* basis sets. Then the conformations were submitted for further high-level calculations of MP2/6-311++G** to determine their relative energies, which were corrected by zero point energy using the frequency calculations at the B3LYP level with 6-311+G* basis sets. The frequencies of the OH stretch modes were corrected for anharmonicity using the multiplier of 0.9734.

3. Results and discussion

3.1. Building trees of cellobiose and lactose

In Figure 2, we present the building tree of cellobiose. In the region of SI, the glycosidic linkage is considered firstly because the glycosidic linkage can restrain the type of the inter-ring H-bonds to a great extent.[16] Therefore, conformers are classified into the *cis* and *trans* configurations. For the *cis* and *trans* structure, all categories of the inter-ring H-bonds network are considered in SII and they are labeled as A, B, C et al. In the range of SIII, the type of the cooperative intra-ring H-bonds is added into the building process. Similarly, the building tree of lactose is established and provided in the Supporting information.

In help for reading the categories of inter-ring and cooperative H-bonds, the conformational notation system is demonstrated here. For example, conformer A1 in Figure 2 is labeled as ccG−g+G−g+. The first two letters of 'cc' or one letter of 'c' indicate the configuration of the cooperative H-bonds in a counterclockwise or clockwise orientation, correspondingly. The following two pairs of letters (the capital letter and the lower-case one) are used to

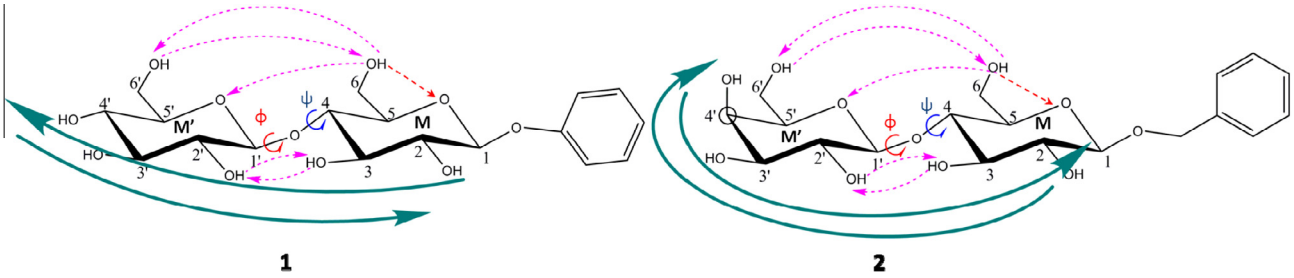

Figure 1. Structural representations of cellobiose and lactose, the cellulose disaccharide (**1**) and its C-4′ epimer (**2**). The dashed arrows indicate the potential H-bonds. The solid arrows refer to the possible orientations of the cooperative hydrogen bonding networks. The dihedral angles are clearly labeled. Φ = H1′–C1′–O–C4, Ψ = C1′–O–C4–H4.