Supplementary information

Collision-induced dissociation of xylose and its applications in linkage and anomericity identifications

Hock-Seng Nguan[1,2], Shang-Ting Tsai[1,2], Jien-Liang Chen[1], Po-Jen Hsu[1], Jer-Lai Kuo[1] and Chi-Kung Ni*[1,3]

[1] Institute of Atomic and Molecular Sciences, Academia Sinica, P. O. Box 23-166, Taipei 10617, Taiwan

[2] Equal contribution to this work

[3] Department of Chemistry, National Tsing Hua University, Hsinchu 30013, Taiwan.

*Email address: ckni@po.iams.sinica.edu.tw

Supplementary:

1. Puckering index

The puckering form of a molecule with ring structure can be characterized quantitatively by Cremer-Pople (CP) index [1]. For a 6-member ring molecule, such as the xylose in ring form, the CP index is given be the 3-dimensional spherical coordinate of (q, θ, ϕ), where q quantify the magnitude of deviation from a planar structure of the molecule, θ and ϕ indicate the form of the puckering. The q, θ and ϕ are given by the equations below:

$$q = \sum_{i=1}^{6} z_i^2, \qquad (1a)$$

$$q\cos(\theta) = \left(\frac{1}{6}\right)^{\frac{1}{2}} \sum_{i=1}^{6} z_i \cos[2\pi(i-1)/3], \text{and} \qquad (1b)$$

$$q\sin(\phi) = \left(\frac{1}{6}\right)^{\frac{1}{2}} \sum_{i=1}^{6} z_i \sin[2\pi(i-1)/3]. \qquad (1c)$$

where z_i represents the deviation of i-th atom on the ring from its sugar ring mean plane, which is defined according to reference [1]. To construct the initial structures of our pyranose xyloses, we applied the method of Hill and Reilly [2]. The group of puckering forms consists of two chair forms (C), six boat forms (B), six skew forms (S), twelve half boat (H) and twelve envelop forms (E). Each form is denoted by an alphabet and number (numbers) according to the type of the form and atom (or atoms) on the ring that deviate from the ring mean plane, respectively. Total of 38 puckering forms are displayed in the Figure 1S.

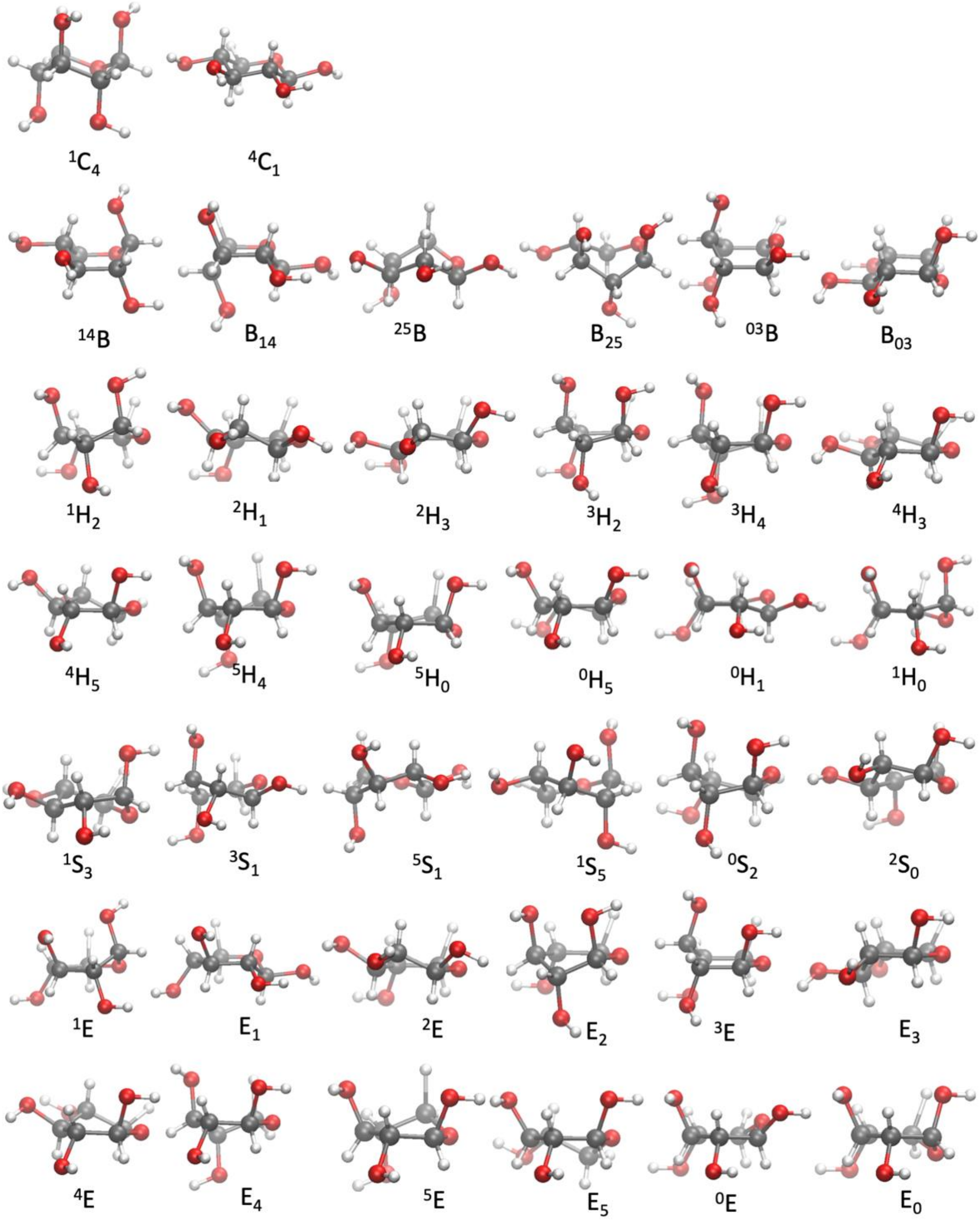

Figure 1S. The schematic drawing of the β-xylose in five types of puckering forms, namely chair forms (C), boat forms (B), skew forms (H), half boat (S) and envelop forms (E).

2. Metadynamics

Metadynamics is an enhanced sampling method introduce by Alessandro Laio and Michele Parrinello [3]. The metadynamics method provides a simple way to accelerate the simulation of the rare events in the molecular system, where those rare events are local minima that separated by rather high energy barrier in its potential energy surface. We used this method to sample the possible conformations of the sodiated linear xylose, each varies by sodium position and the four O-C-C-O dihedral angles (see the structure in Figure 2S). This was achieved by setting relevant collective variables (CV), i.e. the functions of selected atom position. In this case, five CVs are applied in the metadynamics simulation: one CV of coordination number denoted as $C(r_{Na}, r_{O1}, r_{O2}, r_{O3}, r_{O4}, r_{O5})$ and four dihedral angles denoted as $D1(r_{O1}, r_C, r_C, r_{O2})$, $D2(r_{O2}, r_C, r_C, r_{O3})$, $D3(r_{O3}, r_C, r_C, r_{O4})$ and $D4(r_{O4}, r_C, r_C, r_{O5})$. The first CV is defined by the following equation

$$C(r_{Na}, r_{O1}, r_{O2}, r_{O3}, r_{O4}, r_{O5}) = \sum_{i=1}^{5} \frac{1 - \left(\frac{|r_{Oi} - r_{Na}|}{r_0}\right)^8}{1 - \left(\frac{|r_{Oi} - r_{Na}|}{r_0}\right)^{14}} , \qquad (2)$$

which measures the number of oxygen atoms that are close to the sodium ion by some cut off distance r_0. In this work we set the r_0 as 2.4 Angstrom. The results of the sampling using the metadynamics simulation over 43 ps are displayed in Figure 3S. Figure 3S (a) and (b) show the five CVs varies over the simulation time, which shows the coordination number ranging from 0.5 upto 2.5, meanwhile the four dihedral angles fluctuate between -180 to 180 degrees. The position of the sodium ion over the simulation was shown in Figure 3S (c) The preferred configurations are those with 2 and 3 oxygen atoms attached by a sodium ion, the number of oxygen atoms with distance less than 2.5 Angstrom to the sodium ion by is given in Figure 3S (d).

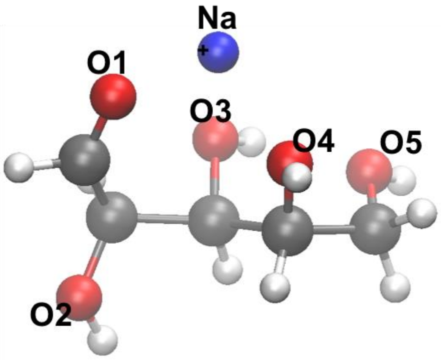

Figure 2S. A schematic drawing of a sodiated linear xylose molecule.

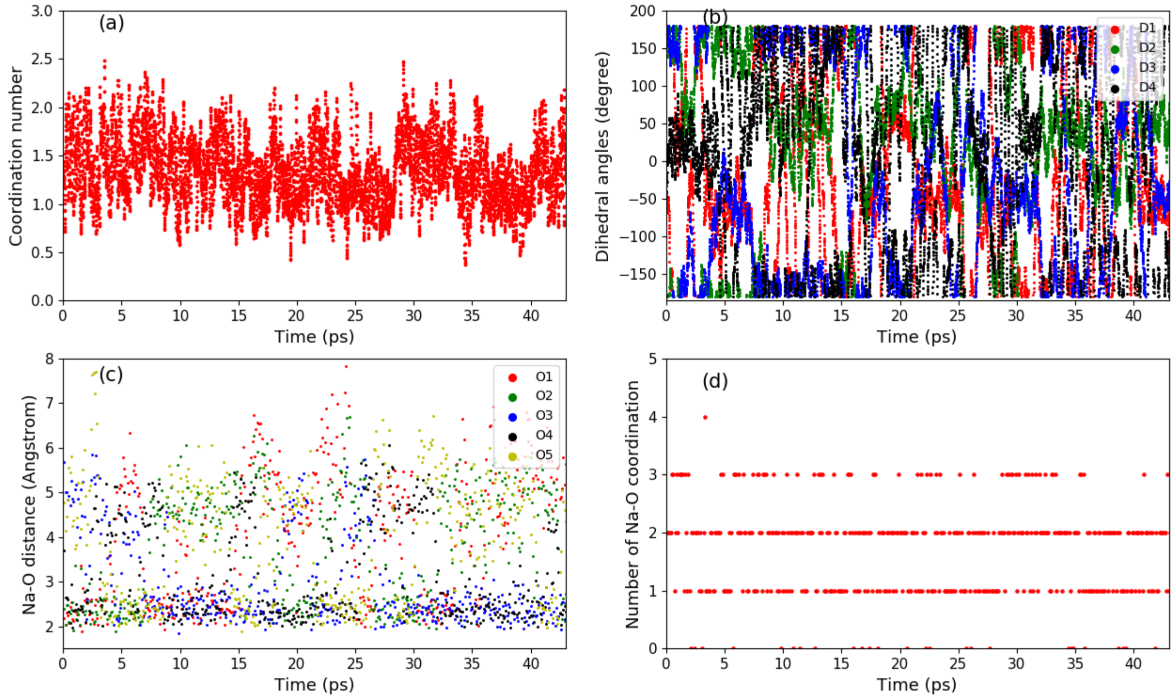

Figure 3S. Results from the metadynamics simulation of sodiated linear xylose molecule, where (a) is the coordination number over the simulation time, (b) is the evolution of the four dihedral angles over time, (c) is the Na^+-O distances over the simulation and (d) is the number of oxygen atoms with distance less than 2.5 Angstrom from the sodium ion.

3. Procedures of generating the guessed transition states

The transition state (TS) searches were performed by selecting the TS candidates in the first step. The TS candidates were selected from the DFT optimized structures which satisfy the following two criteria. First, the structures had to have at least one hydroxyl group that bound to Na^+, where the distance between the oxygen atom of those hydroxyl groups and Na^+ was less than 2.5 Å. These oxygen atoms were denoted O_D, with D stands for donor. This criterion was satisfied by all of our DFT-optimized structures. Second, the structures had to have at least one oxygen atom that was not bound to Na^+, and the distance between this oxygen atom and at least one O_D atom had to be less than 3.0 Å. These oxygen atoms were denoted O_A, where A stands for acceptor. The second criterion ensured that the barrier of H atom transfer from O_D to O_A was not too high. After the TS candidates were selected, the guessed TS structures were generated from the structures of the candidates using the following two steps. In step 1, the H atom of O_D was moved to a new position 1.2 Å away from O_A. In step 2, the C–O_A bond was increased to certain lengths. To increase the chance of obtaining various TSs, three guessed TSs were generated from each selected stable structure for a possible reaction pathway. The three guessed TSs mainly differed through either the H–O_D–C angle or increased C–O_A length. The details of how the three guessed structures were generated are provided in the supplementary information. In cases in which the TS structure optimization failed, another four TS guessed structures with a wide range of C–O_A distances (from 1.65 to 1.9 Å) were generated and optimized again.

From a selected stable structure of xylose+cation complex the guessed transition state (TS) structures are generated through the two steps. First, moves the H atom of O_D to a new position that is 1.2 Angstrom distant from the O_A. Second, elongates the C-O_A bond. For a particular reaction such as dehydration or ring-opening, three guessed TSs are generated from three combinations of these two steps.

For the dehydration, the first combination is the step one followed by the step two. In this combination, the newly formed angle of H-O_A-C bonds is designated to be 109 degrees, a typical value for a hydroxyl group, in step one; meanwhile the new H-O_A-H angle is between 90 to 140 degrees depending on orientation of the hydroxyl group of the O_A atom. In the step two, the O_A is moved outward direction of the C-O_A bond such that C-O_A bond distant becomes 1.8 Angstrom, the same displacement is also applied to the two hydrogen atoms connecting to O_A such that the O_A-H distances and H-O_A-H angle remain the same. As a result, the first combination gives a structure that closes to a dehydration reaction TS. The second and third combinations start with step two where the C-O_A bond length is increased to 1.8 Angstrom and 2.0 Angstrom, respectively, by moving the O_A atom outward of the C-O_A bond, and followed by step one where the H atom is moved along the O_D-O_A direction until its distant from the O_A become 1.2 Angstrom. In some cases where the TS optmized IRC path fail, other TS guessed structures were with wider range of the C-O_A from 1.65 Angstrom to 1.9 Angstrom were generated.

For the ring-opening reaction, first combination involve step one such that newly formed angle of H-O_A-C_1 bonds is about 90 degrees, while in step two the O_A-C_1 bond is elongated to 1.8 Angstrom. The second and third combinations are the same as the dehydration. All the operations that generate the guessed TS structures of both the dehydration and the ring-opening were performed by our in-house program.

For the C-C bonds cleavage of linear xylose cation adduct, the guessed TSs for retro-aldol reactions mechanism [4] are generated manually using Avogadro software molecular modelling package [5]. The generation of TS guesses involves moving the H atom from an O_D atom to a distance that is 1.2 Angstrom distance from O_A atom, meanwhile the C-C bond to be broken is stretched to 1.8 Angstrom.

4. Xylose lithium adduct zero-point corrected energies of TSs and their corresponding reactants

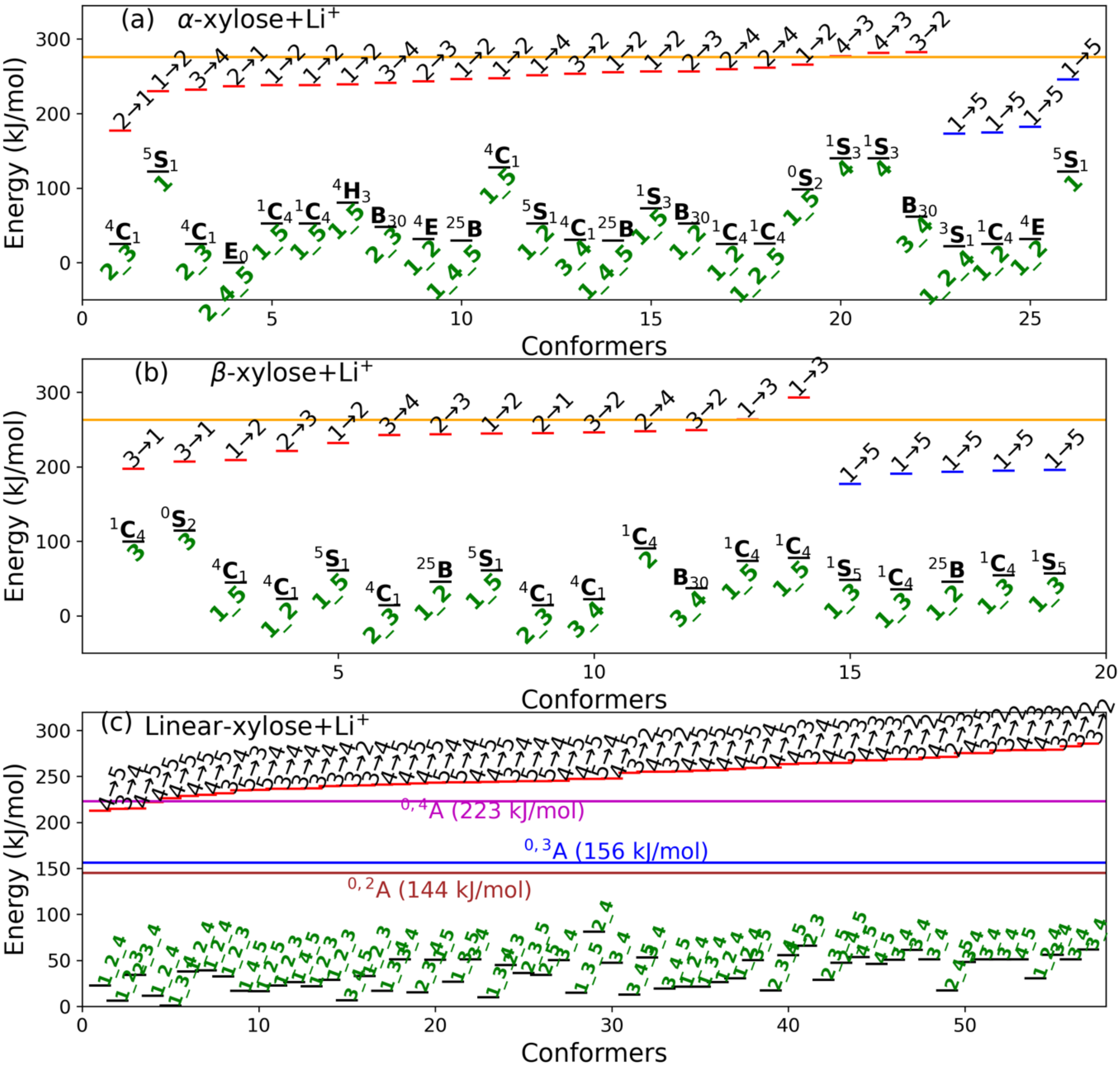

Figure S4. The full results of zero-point corrected energies of xylose lithium adduct TSs and their corresponding reactants.

5. Geometries of sodiated xylose

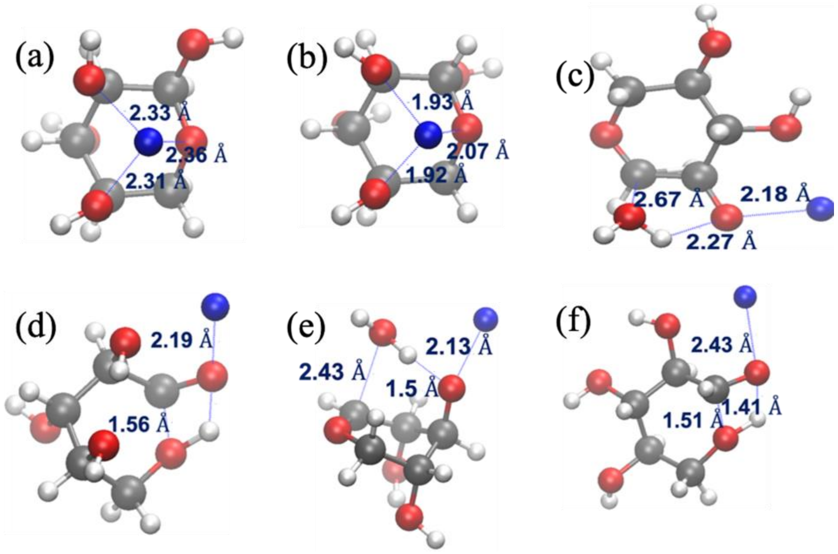

Figure S5. The geometries of most table (a) α-xylose sodium adduct, (b) β-xylose sodium adduct, and the geometries of the transition state with the lowest energies for (c) dehydration for α-xylose sodium adduct, (d) ring-opening for α-xylose sodium adduct, (e) dehydration for β-xylose sodium adduct, and (f) ring-opening for β-xylose sodium adduct.

6. Xylose sodium adduct zero-point corrected energies of TSs and their corresponding reactants

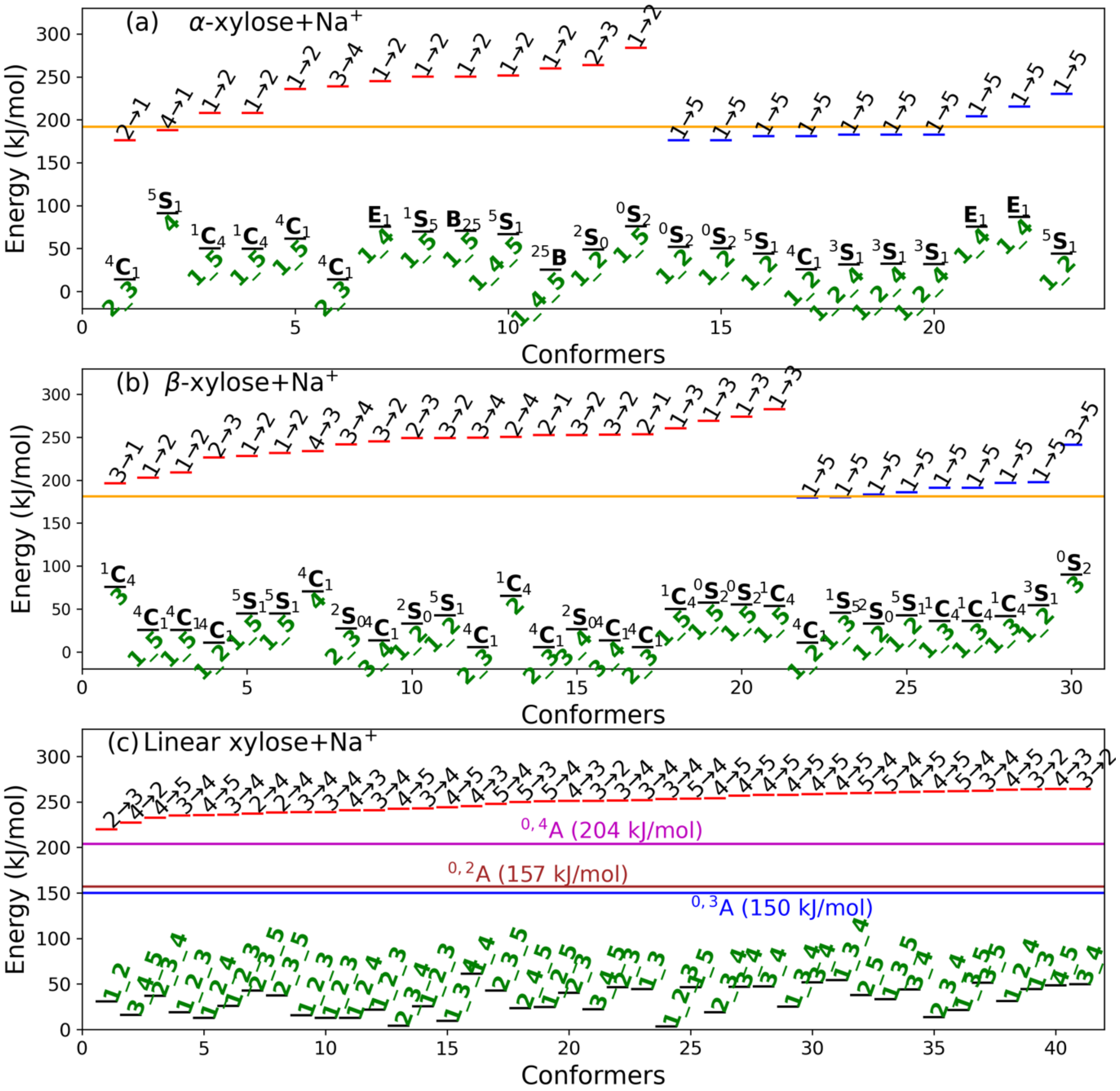

Figure S6. The full results of zero-point corrected energies of xylose sodium adduct TSs and their corresponding reactants.

7. Reaction rate calculations using transition state theory

The dehydration reactions and the ring-opening reaction of xylose cation adduct could happen through different structures of reactants. Considering all these different possible reactants, the reaction rate for a particular reaction can be written as

$$-\frac{d[A]}{dt} = \sum_{i=1}^{M} k_i(T)[A_i]$$

$$= [A]\sum_{i=1}^{M} k_i(T)P_i \qquad (3)$$

where $[A]$ represents the concentration of a xylose attached with a cation, $[A_i]$ represents the concentration of reactant (xylose cation adduct) with structure i, M is the total number of different reactants, P_i is the population probability of reactant i, and $k_i(T)$ is the temperature dependent reaction rate constant of reactant i. At thermal equilibrium, the probability P_i can be calculated from the Boltzmann factor $e^{-(E_i-E_{min})/k_BT}$, such that E_i and E_{min} are the energy of the reactant i and the most stable xylose cation adduct, respectively, and k_B is the Boltzmann constant. Rate constants were calculated using standard transition state theory. Meanwhile, in the calculations of cation ion elimination rate constant, the variational transition state theory (VTST) [6] was used.

References

1. Cremer, D.t. and J. Pople, *General definition of ring puckering coordinates.* Journal of the American Chemical Society, 1975. **97**(6): p. 1354-1358.
2. Hill, A.D. and P.J. Reilly, *Puckering Coordinates of Monocyclic Rings by Triangular Decomposition.* Journal of Chemical Information and Modeling, 2007. **47**(3): p. 1031-1035.
3. Laio, A. and M. Parrinello, *Escaping free-energy minima.* Proceedings of the National Academy of Sciences, 2002. **99**(20): p. 12562.
4. Tsai, S.-T., et al., *Automatic Full Glycan Structural Determination through Logically Derived Sequence Tandem Mass Spectrometry.* ChemBioChem, 2019. **20**(18): p. 2351-2359.
5. https://avogadro.cc/.
6. Klippenstein, S.J., *Variational optimizations in the Rice–Ramsperger–Kassel–Marcus theory calculations for unimolecular dissociations with no reverse barrier.* The Journal of chemical physics, 1992. **96**(1): p. 367-371.