RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances

The dynamics of the conformational changes in the hexopyranose ring: a transition path sampling approach

Wojciech Plazinski[1+], Mateusz Drach[2]

[1] *Institute of Catalysis and Surface Chemistry, Polish Academy of Sciences, ul. Niezapominajek 8, 30-239 Cracow, Poland*

[2] *Department of Theoretical Chemistry, Faculty of Chemistry, UMCS, pl. M. Curie-Sklodowskiej 3, 20-031 Lublin, Poland*

[+] Corresponding author. Tel.: +48815375685; fax: +48815375685.
E-mail addresses: wojtek_plazinski@o2.pl; wojtek@vega.umcs.lublin.pl

Abstract

The ring conformation of the hexopyranose-based carbohydrate molecules is one of the central issues in glycobiology. We report the results of the transition path sampling simulations aimed at the detailed description of the dynamic features of these conformational changes. We focused on the α-D- and β-D-glucopyranose molecules (GlcA and GlcB, respectively), treated as model systems. A large number of unbiased dynamic trajectories leading from the $^4C_1$ conformation to the $^1C_4$ one has been collected and subjected to analysis. The results allowed for: (i) identifying the distinct local minima of the free energy corresponding to the states intermediate for the $^4C_1 \rightarrow {}^1C_4$ transitions; (ii) assigning the time-characteristics to these transitions and intermediate states; (iii) searching for the optimal reaction coordinate based on the Peters-Trout approach (likelihood maximization, LM). Additionally, the structures corresponding to the $^4C_1 \rightarrow {}^1C_4$ transition states (TS) have been found; surprisingly, in the case of GlcA, the water dynamics has very little influence on the probability of the TS evolution either to $^4C_1$ or to $^1C_4$. The differing result obtained for GlcB (large influence of water dynamics on the behavior of TS as well as the poor applicability of the LM approach for calculation of the reaction coordinate) speaks for slightly different mechanisms of the $^4C_1 \rightarrow {}^1C_4$ puckering in the molecules of GlcB and GlcA.

Introduction

Even a monosaccharide is a relatively complex molecule which can be associated with the relatively poorly explored conformational degrees of freedom. The three-dimensional structures of oligo- and polysaccharides are defined by the following conformational determinants: (i) glycosidic linkage geometry; (ii) pyranose ring conformation (pucker); (iii) orientation of the exocyclic groups[1]. The dynamics of the exocyclic groups and the glycosidic-linkage-related conformations are better understood, due to their timescale characteristics (tens of pico- and nanoseconds, respectively). On the other hand, the timescale-related issues make conformational rearrangement in the pyranose ring more problematic for both experiments (NMR spectroscopy) and simulations.

The ring conformations of the hexopyranose-based carbohydrates which differ from the standard $^4C_1$ may be significantly populated in a number of systems, including: (i) specific hexopyranoses (the $^1C_4$ conformation is even predominant in the case of e.g. α-D-altrose or α-L-guluronic acid)[2-4]; (ii) sterically crowded hexopyranose derivatives[5-19]; (iii) hexopyranoses in AFM-stretched poly- or oligosaccharides[20-28], cyclodextrins[29-31]; (iv) hexopyranoses in the complex with proteins[32-36]. Moreover, even if the non-$^4C_1$ confomers are non-predominant they still can influence the biological features of carbohydrate-containing systems[37]; furthermore, as it has recently been shown, the hydrodynamic properties of carbohydrate polymers are significantly affected by ring puckering[37].

The main aim of the computational study reported here was to provide the insight into carbohydrate ring puckering in the hexopyranose molecules at the molecular level. Contrary to most of the theoretical studies[38-44], our efforts were focused not on recovering the free energy landscapes associated with the ring conformers but on producing and analyzing the set of transition paths connecting the two chair conformers. The most straightforward approach that can be proposed for the former aim of the study is the application of the plain molecular dynamics (MD) and subsequent analysis of the resulting trajectories. Such method appeared to be applicable in the case of e.g. *N*-acetyl-D-glucosamine (GlcNAc) or iduronic acid and their derivatives[1,45] when the GPU-based hardware was used for MD simulations. However, the free energy barriers reported for most of the non-substituted carbohydrates are significantly higher (e.g. ~ 40 and 52 kJ/mol for α-D- and β-D-glucose, respectively; the data based on the present study) than those corresponding to the $^4C_1 \rightarrow {}^1C_4$ inversion in e.g. the

iduronic acid molecule (25 kJ/mol; data taken from ref. [45]). Moreover, the unbiased MD simulations, depending on the system, required over 3-5 μs [1] or 0.25 μs [45] of the computer time for equilibration. It is hard to expect the same efficiency in the case of e.g. glucose, as well as for numerous other carbohydrates, characterized by much higher free energy barriers separating particular puckers. For instance, we were not able to observe even a single event of the $^4C_1 \rightarrow {}^1C_4$ puckering during over 3 μs unbiased MD simulations of the β-D-glucose molecule. The commonly used, MD-based techniques of enhanced sampling (e.g. metadynamics[46], umbrella sampling-based approaches[47]) have successfully been applied to determine the free energy landscapes of interest[38-43] but the information of the related dynamic features of the puckering is lost. Thus, to collect a sufficiently large number of trajectories leading from the $^4C_1$ conformer to the $^1C_4$ one, the transition path sampling (TPS) technique[48,49] was applied in combination with the standard MD engine.

In theory, the MD trajectories harvested during the TPS procedure should allow for evaluation of the puckering mechanism, associated free energy landscape and the transition state ensemble. Apart from that characteristics, we also made some attempt to calculate the optimal reaction coordinate (RC), expressing the $^4C_1 \rightarrow {}^1C_4$ inversion. A good RC would be able to predict the commitment probability to reach a given configuration by the system (the so-called committor)[50-53]. As the previous cases of applying the procedures aiming at estimating RC were focused on the complex systems, containing large biomolecules (protein-containing systems, mainly), it would be interesting to apply the same methods for the (relatively simpler) carbohydrate system to find similarities and differences.

Our simulations were confined to the molecules of α-D- and β-D-glucose (GlcA and GlcB, respectively), treated here as model systems. Note that the methods applied and described in the present manuscript can be equally well applied for any other hexopyranose-containing system. The manuscript is organized as follows. After the description of computational methodologies, we report the results related to the following aspects of the carbohydrate ring puckering: (i) brief description of the intermediate states (transition paths) of the $^4C_1 \rightarrow {}^1C_4$ rearrangment and their discussion in the context of the canonical conformations of the six-membered rings; (ii) the time characteristics of the $^4C_1 \rightarrow {}^1C_4$ transition; (iii) searching for the optimal reaction coordinate describing the $^4C_1 \rightarrow {}^1C_4$ transition in terms of the Peters-Trout

(Likelihood Maximization, LM) approach[53]; (iv) discussion on the found coordinates, potential transition states and their features. We end with concluding remarks.

Methods

Molecular dynamics. The simulations were carried out using the GROMOS 56a6CARBO force field, dedicated to the hexopyranose-based carbohydrates and designed with special emphasis on reflecting the ring conformational properties[38]. Some hand-written scripts were applied to implement the GROMOS $56a6_{CARBO}$ force field into the GROMACS[54] package (version 4.55), which included the proper calculation of the 'exceptional' Lennard-Jones parameters for selected 1-4 and 1-5 atom pairs. The exemplary $56a6_{CARBO}$ force field files and the GROMACS version of the force field entries for GlcB (with the script for correcting the topology) are given on www.gromacs.org. The simulation systems contained one carbohydrate molecule (GlcA or GlcB) placed in the cubic ($30 \times 30 \times 30$ Å^3) simulation box containing about 900 SPC water[55] molecules. Before the TPS protocol, the standard, unbiased MD simulations were performed for ~ 10 ns to fully equilibrate the system. The details of the TPS simulation protocol are described in the subsequent section. The equations of motion were integrated using the leapfrog scheme[56] with a timestep of 2 fs. During the MD runs the LINCS algorithm[57,58] was applied to constrain all bond lengths. The simulations were carried out under periodic boundary conditions and under the NPT conditions. The temperature was maintained close to its reference value (298 K) by applying the V-rescale thermostat[59], whereas the Parrinello-Rahman barostat[60] was used to control the pressure (1 bar). The centre of mass motion was removed every step. Nonbonded interactions were computed using a twin-range scheme[61], with the short- and long-range cutoff distances of 0.8 and 1.4 nm, respectively, and a frequency of 5 timesteps for the update of the short-range pairlist and intermediate-range interactions. To account for electrostatic interactions beyond the long-range cutoff radius, a reaction field correction was applied using a relative dielectric permittivity of 61[62]. The additional enhanced sampling simulations were performed to provide some insights into the full free energy landscape and to test the correctness of the $56a6_{CARBO}$ force field implementation in GROMACS. To study the free energy landscape as a function of the Cremer-Pople puckering coordinate θ[63] we apply parallel-tempering combined with metadynamics[46]. This combination improves the accuracy of both methods. On one hand, parallel-tempering allows sampling all degrees of freedom (improving metadynamics

accuracy) and on other hand, metadynamics improves exploration of low probability regions. The calculations were performed by GROMACS combined with PLUMED 1.3[64]. Several metadynamics simulations were calculated in parallel at different temperatures (298, 310, 323, 335, 348, 362, 375, 389, 404, 418, 433, 448 K) with configuration exchange according to the replica exchange scheme. The starting deposition rate was set to 0.01 kJ/mol/ps with the Gaussian width equal to 0.05 nm. Molecular dynamics parameters (timestep, cut-offs, etc.) in free energy calculation and in TPS simulations had the same values.

Transition path sampling. The TPS method was applied to sample the 'reactive' trajectories, i.e. those connecting predefined initial and final states (referred further to as the stable states and determined by the free energy minima). The random walk through the trajectory space is performed by generating new trajectories from old ones by a shooting move algorithm combined with the Metropolis rule[48,49]. The randomly chosen timeframe (the shooting point) of the old trajectory is the starting point for creating a new trajectory by integrating the equations of motion forward and/or backward in time by using the conventional MD. Initially, we have used the TPS algorithm as described in ref. [65], i.e. the one-way, flexible path length algorithm, which was also applied in the case of our previous study[66]. Due to very low efficiency of the sampling, another TPS algorithm (biased shooting) has been applied as well, according to its description in ref. [67]. The most important difference in comparison to refs. [65] is that the V-rescale (Bussi) thermostat[59] was used instead of the Andersen one as a generator of stochastic noise. Therefore, the momenta of the shooting point did not have to be changed. The other details of generating and accepting the new trajectories were based on the procedures described in refs. [65,66] and on the MD parameters listed above. The first (input) trajectory connecting the initial ($^4C_1$) and final ($^1C_4$) states was generated by running unbiased MD simulation at 498 K which allowed for observing the $^4C_1 \rightarrow {}^1C_4$ inversion within a relatively short period of time (tens of nanoseconds). In the case of all sampled trajectories, the data were collected every 0.1 ps, as the longer time space does not allow to observe the system evolution in the vicinity of the free energy barriers; furthermore, the application of the larger time space would lead to problems in diverging the MD trajectories in, in turn, to the poor sampling.

Stable states and order parameters. The TPS sampling included the energy barriers between two chair conformers ($^4C_1$ and $^1C_4$), i.e. the 'stable states'. Several different order parameters can be proposed to distinguish between them, e.g. Cremer-Pople (CP)[63], Picket-Strauss (PS)[68], Berces (B)[69], Hill-Reilly (HR)[70] or Zefirov[71] coordinates. For the sake of simplicity, we decided to use the Cremer-Pople θ coordinate, as this choice allows for using

only one parameter to control developing MD trajectories on-the-fly. Based on the standard MD simulations, we defined the stable states as follows: $^4C_1$: $\theta < 20$ deg and $^1C_4$: $\theta > 160$ deg. Such choice determines that the generated TPS trajectories covered only the 'intermediate' region of phase space, lying between two chair conformations and are ceased when $\theta < 20$ deg or $\theta > 160$ deg. The definitions of the dihedral angles associated with the exocyclic groups are as follows (atom numbering is in accordance with Fig. 1(A)): tor1: $H^{(1)}$-$O^{(1)}$-$C^{(1)}$-$O^{(5)}$; tor2: $H^{(2)}$-$O^{(2)}$-$C^{(2)}$-$C^{(1)}$; tor3: $H^{(3)}$-$O^{(3)}$-$C^{(3)}$-$C^{(2)}$; tor4: $H^{(4)}$-$O^{(4)}$-$C^{(4)}$-$C^{(3)}$; tor5: $O^{(6)}$-$C^{(6)}$-$C^{(5)}$-$C^{(4)}$; tor6: $H^{(6)}$-$O^{(6)}$-$C^{(6)}$-$C^{(5)}$.

During the TPS simulations, we monitored several order parameters, intending to use them during the procedure of reaction coordinate analysis. These parameters included: (i) the four types of coordinates used to describe the six-membered ring conformation (i.e. CP, HR, B, PS, see the previous paragraph for definitions and references); (ii) some atom-atom distances, characteristic of the ring conformation of GlcA and GlcB molecules (the $O^{(1)}$-$O^{(4)}$ and $O^{(2)}$-$C^{(6)}$ distances for GlcB and the $O^{(2)}$-$C^{(6)}$ distance for GlcA, see Fig. 1(A) for definition). Note that except for the Cremer-Pople θ, none of the parameter is directly correlated with the progress of the $^4C_1 \rightarrow {}^1C_4$ transition and, thus, can not be treated as a valuable order parameter. We have used the following transformation:

$$v = \sqrt{x_1^2 + x_2^2 + x_3^2} \quad (1)$$

to obtain the order parameters which vary smoothly from 0 to 1 as the ring conformation changes from an ideal $^4C_1$ to an ideal $^1C_4$ chair. v denotes the 'collective' order parameter being a function of the three coordinates (x_i) describing the ring conformation (i.e. HR, B or PS). The value of v can be interpreted as the length of a vector placed in the three-dimensional space, defined by the set of coordinates used in eq. 1. The length of this vector corresponds to the progress of puckering. Note also that the particular assignments of the non-chair conformations are quantitatively different for each of the tested coordinates which was checked after calculating v values for all canonical conformers. Summarizing, the six (GlcB) or five (GlcA) different order parameters were used in the subsequent reaction coordinate analysis: Cremer-Pople θ, the three different v functions (based on the Picket-Strauss, Berces and Hill-Reilly coordinates) and the one or two exocyclic atom-atom distances.

Reaction coordinate. The reaction coordinate was studied on the basis of Likelihood Maximization approach (LM)[53] which requires only the data associated with the trial shots

of the TPS simulation. The LM approach extracts a linear combination of order parameters $v(x)$ that best describes the reaction coordinate (r):

$$r(v(x)) = \sum_{j=1}^{N} a_j v_j(x) + a_0 \ . \tag{2}$$

(The Cremer-Pople θ is treated as another $v(x)$ parameter). The input for this procedure is the ensemble of shooting point configurations belonging to the accepted trajectories ending in the final state $^1C_4$ (i.e. B(x_{sp} → B)) and the rejected shooting points ending in the $^4C_1$ initial state (i.e. A(x_{sp} → A)). By using these configurations, the LM optimizes the likelihood:

$$L = \prod_{x_{sp} \to B} p_B(r(x_{sp})) \prod_{x_{sp} \to A} (1 - p_B(r(x_{sp}))) \ . \tag{3}$$

See details of the procedure in the Supporting Information of ref. [65]. We tested all possible linear combinations of up to three order parameters. The committor p_B (i.e. the commitment probability of a given conformation of the system to $^1C_4$) was expressed by the following function:

$$p_B(r(x)) = \frac{1}{2} + \frac{1}{2}\tanh\left[r(v(x))\right]. \tag{4}$$

According to the Bayesian criterion, adding more v variables to the RC model is significant only if ln L increases by at least 1/2 ln M, where M is the total number of shooting points in the ensemble.

The practical use of the LM procedure can be described as a best-fit procedure during which the values of a_i parameters (eq. 2) are adjusted to maximize the value of L (eq. 3). L is calculated during each iteration based on: (i) the configuration of the shooting points (x, required to calculate v's and $r(v)$ according to eq. (2)); (ii) the accepted functional form of the committor (eq. 4). The highest possible value of L is associated with the 'optimal' values of the a_i coefficients.

To test the obtained reaction coordinates we explicitly computed the committor for selected configurations characterized by different values of p_{Comm}, including those that were identified as putative transition states. The committor for a given configuration is the fraction of unbiased MD trajectories, initialized with random velocities that reach the $^1C_4$ state.

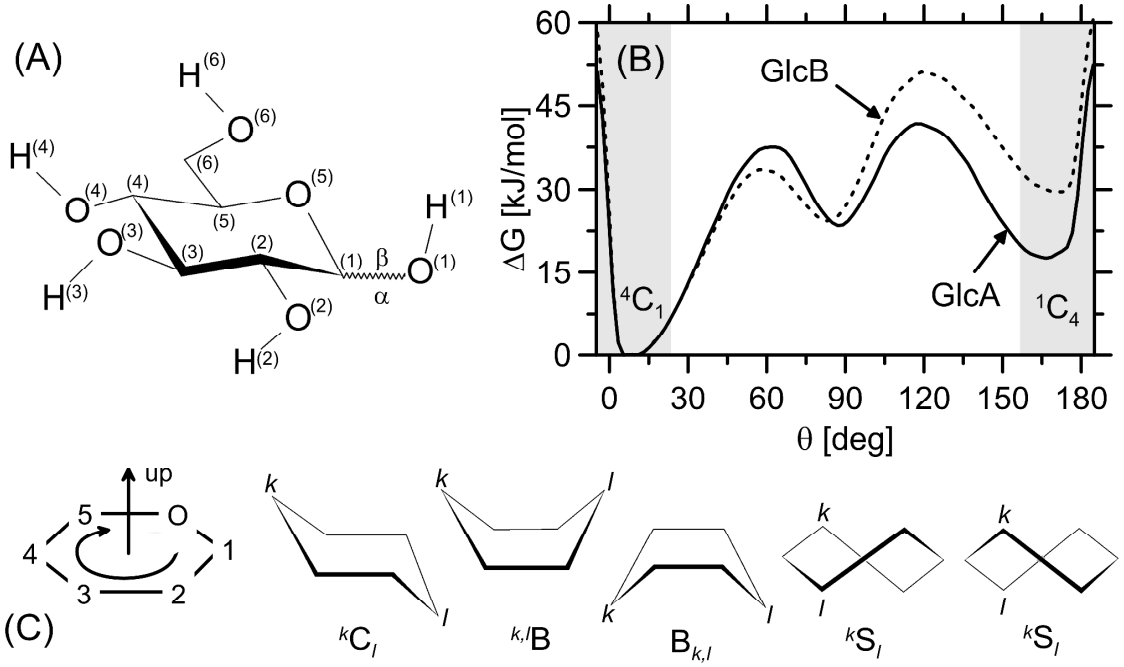


Fig. 1. (A) The chemical structure of the D-glucose molecule. The atom numbering is indicated. (B) The free energy profiles for the molecules of GlcA and GlcB, associated with the θ parameter of the Cremer-Pople coordinates. The grey regions are those which were treated as 'stable states' (i.e. the $^4C_1$ and $^1C_4$ conformers) and not sampled during the TPS simulations. (C) Schematic representation and nomenclature of idealized pyranose ring conformations according to ref. [38]. Conformations are labeled with the type of conformation (chair C, boat B, or skew-boat S) and indices *k* and *l* represent the atom numbers of two atoms pointing upward (superscript) or downward (subscript). The set of all conformers is shown in Fig. 2(A) and 2(B) as points on the plane defined by the Cremer-Pople coordinates.

Results and Discussion

1. Remarks on the sampling results

The number of harvested trajectories was equal to 2032 (GlcA), 842 (GlcB, unbiased shooting algorithm) and 2409 (GlcB, biased shooting algorithm). The acceptance ratios were equal to 9.7%, 1% and 5.9 %, respectively. The unbiased shooting algorithm appeared to be ineffective in the case of GlcB due to very low probability of randomly choosing such MD frame that would lead to producing the whole reactive MD trajectory. The reason for that is the low fraction of such 'effective' frames in the 'old' reactive path.

The issue of the correct TPS sampling may be problematic in the case of complex molecules for which distinct reactive paths lie on the poorly known free energy landscape. Then, if energetic barriers separating them are very high, some essential paths may be missed during the TPS-based harvesting of the MD trajectories. The scenario is much simpler in the case of carbohydrate ring puckering as the related free energy profiles can be calculated separately by using an independent method (e.g. metadynamics[40] or umbrella sampling[38]) and compared with the TPS results. On the basis of: (i) inspection into the TPS path density plots (Figs. 2(A) and 2(B) show that there are no unexplored regions of the phase space defined by the Cremer-Pople coordinates); (ii) comparison of the results with the relative populations of the canonical conformers reported in ref. [38] and estimated using the same force field, we concluded that the sampling was efficient enough to reveal all possible reactive paths maintaining their relative weights.

2. Free energy landscapes

The free energy landscapes associated with the ring puckering in the molecules of GlcA and GlcB have been studied quite extensively by using different computational methods. Nevertheless, a brief characteristic of the issue may be useful from the point of view of the interpretation of the TPS results. Furthermore, the comparison of the results obtained by applying the relatively new force field GROMOS $56a6_{CARBO}$ with those obtained previously is worth some remarks.

First of all, note that we did not sample the whole conformational space during TPS simulation but only these parts of it which correspond to the non-chair conformation. Thus,

one can not expect to directly compare the TPS results with the full free energy landscapes obtained by using different computational techniques. However, the TPS-related trajectory density map (as shown in Figs. 2(A) and (B)) can be qualitatively viewed as a free energy landscape representation for the 'intermediate', non-chair conformers. The results can be summarized as follows: for both GlcA and GlcB only two local minima of the free energy have been found, none of them corresponding to a single specific 'canonical' conformation of the six-membered rings [63]. Furthermore, these minima always correspond to the boat and skew-boat conformers, lying along the $\theta = 90$ deg line.

In the case of GlcB, the two observed minima (of different sizes and depths) are separated by the two free energy barriers: one of them centered around $\phi \approx 260$ deg (lower barrier) whereas the second one (higher) roughly covering the region 80 deg $< \phi <$ 140 deg. The relative heights of these barriers are the reason why the overwhelming majority (>99 %) of the events of GlcB system switching from one local minimum of free energy to another involves crossing the lower barrier; this corresponds to temporarily adopting the conformations close to $^{1}S_{5}$ and $^{1,4}B$. The ranges of these two local minima cover roughly the following canonical conformations: (shallower minimum) $B_{3,O}$ and $^{1}S_{3}$; (deeper minimum, preferable intermediate state): $B_{2,5}$, $^{O}S_{2}$, $^{3,O}B$, $^{3}S_{1}$ and $B_{1,4}$. Very similar conformational preferences (two local minima assigned to $^{1}S_{3}$ and $^{O}S_{2}$/$^{3,O}B$ conformers) have been reported by Spiwok et al.[40], using the previous version of carbohydrate-dedicated GROMOS force field[72].

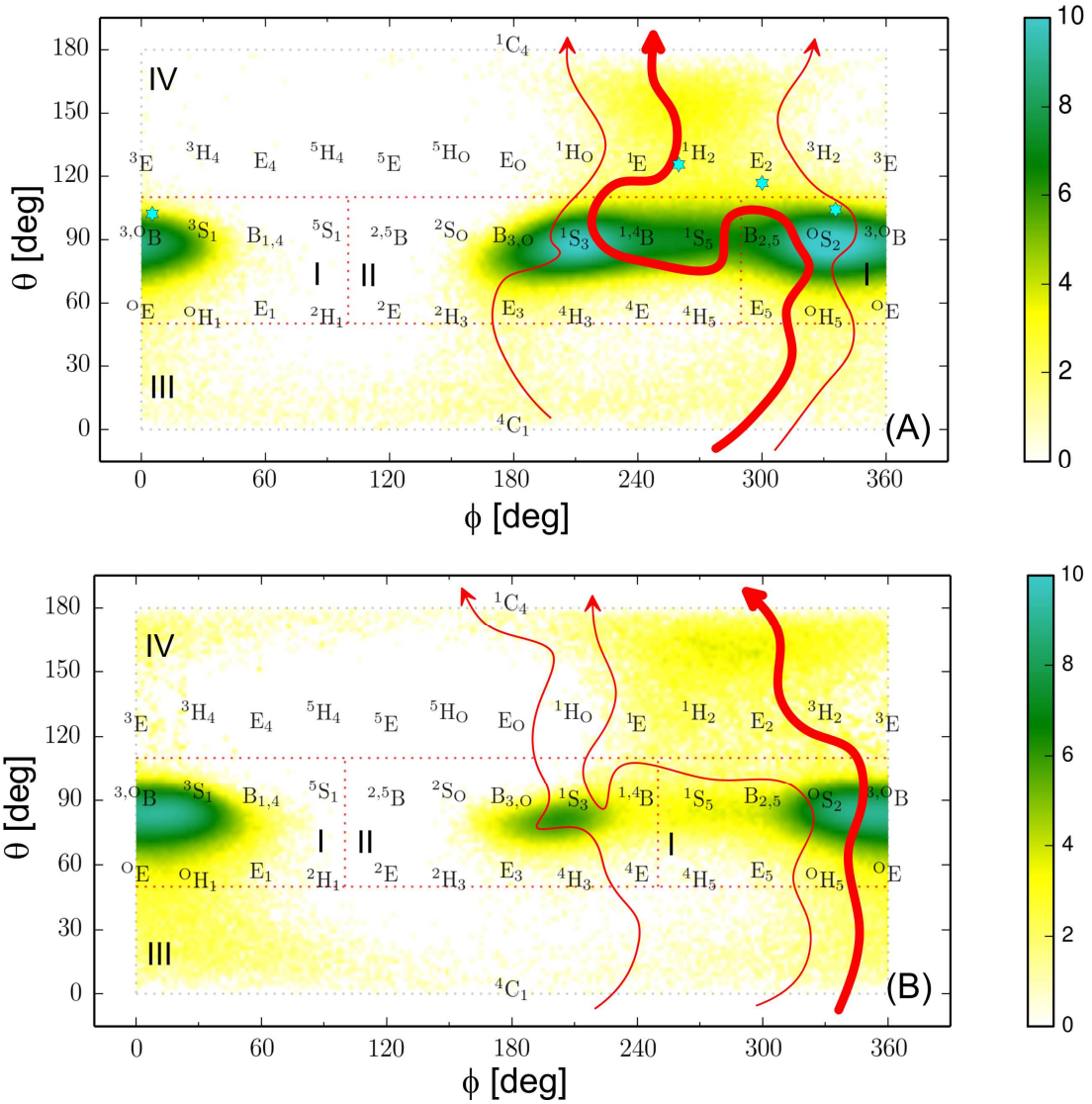


Fig. 2. Path density plots obtained from the reactive trajectories corresponding to the process of the $^4C_1 \rightarrow {}^1C_4$ transition in the molecules of GlcA (A) and GlcB (B) represented in the Cremer-Pople (ϕ,θ) set of coordinates. The scale refers to the natural logarithm of the number of trajectories per one bin (1 deg^2). The borders defining the phase space regions (I, II, III and IV) used in subsequent analyses are given as red dotted lines. The blue stars (panel (A)) denote the four exemplary transition state structures identified for GlcA. The red arrows represent the different possible paths of the conformational transition; the thickness of each line indicates the fraction of MD trajectories which can be ascribed to the given path type: > 70% (thick lines), < 15% (thin lines). The three path types were distinguished: (i) paths visiting both regions I and II; (ii) paths visiting only region I; (iii) paths visiting only region II. Further details are given in the text.

In spite of the similar picture emerging in the case of GlcA, there exists a number of significant differences, mentioned below. The free energy barrier separating two minima is significantly lower than that corresponding to GlcB. Thus, the system can interchange the non-chair conformations assigned to each particular minimum more freely. This, combined with the lowered free energy barriers separating the chair and inverted chair conformers from the non-chair ones, may be one of the reasons for different behavior exhibited by GlcA in comparison to GlcB (as explained in the further parts of the paper). The two minima have a similar location to those calculated for GlcB, with the notable differences of: (i) shift one of the minima towards $B^{2,5}$ conformation, leaving $^{3}S_{1}/B_{1,4}$ region less intensively explored; (ii) more similar depth and sizes of the two energy wells. As a result, the lower of the two barriers located around θ = 90 deg, is placed in the vicinity of the $^{1}S_{5}$ canonical conformation (i.e. ϕ = 270 deg).

These results remain in agreement with those obtained during the enhanced sampling studies, reported in ref. [40]. In spite of minor discrepancies in the number and location of the local minima (which can be ascribed to differences in the force field parameters) the main conclusion remains the same, confirming that the non-chair canonical conformers do not necessarily reflect the actual conformational preferences of the glucose molecule. As the authors[40] used different force-fields (e.g. GROMOS45a4[72], GLYCAM[73], OPLS[74]) and computational methodologies (e.g. implicit or explicit solvation) we can safely assume that such scenario is force field-independent. Furthermore, the metadynamics simulations based on the *ab initio* potentials and the Car-Parrinello MD formalism gave the related results[43], confirming the canonical conformations do not match the local minima of free energy in the GlcB molecule. The canonical conformations have always been treated as natural reference points when calculating the conformational properties of six-membered rings. They are also useful when quantitatively comparing the conformational features of different compounds or designing the carbohydrate-dedicated force fields[38]. The fact of existing the highly-populated conformers not matching the canonical minima leads also to proposing several assignment schemes aiming at associating an actual ring conformation to a canonical conformation[75].

The observations discussed above are very significant for the interpretation of the TPS data. Note that attempts of applying some of the existing assignment schemes to describe the dynamic characteristics of the studied conformational rearrangements would give, as an

output, the data of doubtful physical sense. In other words: as the system can cross only one significantly large free energy barrier during exploration of the non-chair region of the phase space, there is no need to create additional distinctions (e.g. equivalent to assignments of the system to canonical ring conformations). Thus, we decided to introduce a very simple (but somewhat arbitrary) division of the phase space, based on the θ and ϕ Cremer-Pople coordinates. The related regions are shown in Fig. 2 and are defined by the following (limiting) values of ϕ coordinate in the non-chair (i.e. those for which the 20 deg < θ < 160 deg relation is fulfilled) area of phase space: (for GlcA) region I: 50 deg < θ < 110 deg, ϕ < 100 deg and ϕ > 290 deg; region II: 50 deg < θ < 110 deg and 100 deg < ϕ < 290 deg; region III: θ < 50 deg; region IV: θ > 110 deg; (for GlcB) region I: 50 deg < θ < 110 deg, ϕ < 100 deg and ϕ > 250 deg; region II: 50 deg < θ < 110 deg and 100 deg < ϕ < 250 deg; region III: θ < 50 deg; region IV: θ > 110 deg. Such type of assignment will be used in the subsequent time characteristic of the conformational rearrangements. Note that for both GlcA and GlcB regions II and III can be associated with the previously described local minima of free energy lying in the region θ ~ 90 deg.

3. Time characteristics and transition paths

The $^4C_1 \rightarrow {}^1C_4$ rearrangement can be classified as a typical 'rare event', i.e. the process which occurs fast but one may wait for a very long time to observe even single conformational transition. Figs. 3(A) and (B) show the histogrammed time lengths of the reactive trajectories harvested for GlcA and GlcB. The overwhelming majority of each trajectory corresponds to the non-chair conformations located around θ = 90 deg which is rather obvious when keeping in mind the free energy landscape presented in Fig. 1(B). The rest of the conformations (i.e. those located in regions III and IV) roughly represent the barriers of free energy separating the wells located around θ = 0, θ = 90 and θ = 180 deg. The crossing of these barriers is a very rapid process, lasting, on average, less than 1 ps (this actually was the main reason for saving coordinates every 0.1 ps which is rather unusual in the standard MD simulations). A more detailed analysis is focused on particular regions of phase space, listed in the previous section and roughly representing the free energy wells of the non-chair conformations and the barriers separating them from the chair conformers.

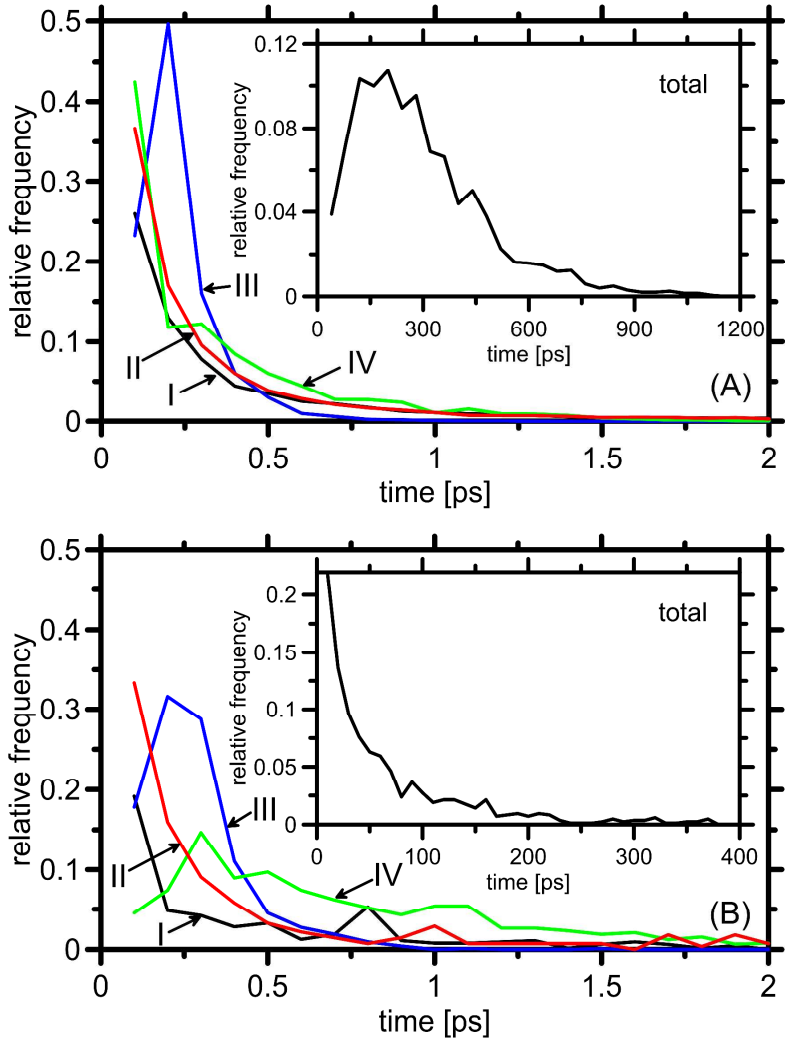


Fig. 3. The histogrammed lengths of the reactive MD trajectories including the contribution of the particular regions of the phase space (defined in the text and presented in Fig. 2) for GlcA (A) and GlcB (B). The histograms corresponding to the particular regions are colored in the following manner: black (region I); red (region II); blue (region III); green (region IV). The inset panels correspond to the lengths of the complete reactive trajectories, leading from $^{4}C_{1}$ to $^{1}C_{4}$. Further details are given in the text.

The average time which is spent by the system in each region varies from 0.2 to 20.5 ps (for GlcA) and 0.3 to 52 ps (for GlcB). The average length of the total trajectory leading from $^4C_1$ to $^1C_4$ is much longer (277 and 65 ps for GlcA and GlcB, respectively) which makes a significant difference, especially for GlcA. The main reason for that is the characteristic shape of the free energy landscape (Fig. 1(B)). The large free energy barriers separating the chair conformation from the non-chair ones make the system 'trapped' in the local minima around $\theta = 90$ deg. Each structure exhibiting significant deviation from the basin of attraction assigned to the $\theta \approx 90$ deg value (i.e. crossing the free energy barrier) is likely to end either as $^4C_1$ or $^1C_4$ conformer. This explains short time spent by the systems in regions III and IV. Similar values of the average trajectory length observed for regions I and II correspond to frequent 'oscillations' of the conformation between regions I and II (being the results of the relatively low free energy barrier separating these regions). When calculating the trajectory lengths corresponding to the combined regions I and II one would obtain nearly the same length distributions as those plotted for the total trajectories (i.e. Fig. 3, panels 'total'). Furthermore, there exists a small number of trajectories not leaving regions I or II for much longer time period than the average value (this is depicted by the standard deviation of the trajectory lengths: 40.7 and 36.7 for GlcA, for regions I and II, respectively and: 75.2 and 11.9 ps for GlcB, regions I and II, respectively). Such diversity was not observed for regions III and IV (standard deviations of time length characteristic of regions III and IV are about half of their mean values for both GlcA and GlcB) which confirms again the minor contribution of these regions to the overall trajectory length.

Furthermore, the probability of visiting both or only one of the two local minima (regions I and II) by the system was examined (visiting regions III and IV is obligatory for all collected MD trajectories). The free energy barrier separating the two main local minima in the region of non-chair conformers ($\theta \approx 90$ deg) is lower for GlcA than that characteristic of GlcB. This is reflected by the ratio of trajectories which visit both regions I and II: it is equal to 90% for GlcA and only 14% for GlcB. The higher free energy barrier located approximately at the border between regions I and II of GlcB-related free energy landscape makes the conformational rearrangement limiting to the non-chair regions less favourable. Thus, in 86% of the cases, one can distinguish two distinct pathways for GlcB: considering the non-chair conformers, 15% of the trajectories visit only region I while 71% visit only region II.

The reactive pathways for GlcA are more uniform: only 4.5% visit region I not entering into region II while the fraction of trajectories leading only through region II is equal to 5.5%. According to such characteristics, the mechanism of ring puckering associated with the diffusive barrier crossing dynamics can be ascribed rather to GlcA than to GlcB which may be meaningful in the context of further findings (see sections 4. and 5.).

Summarizing, the transition paths characteristic of the $^{4}C_{1} \rightarrow {}^{1}C_{4}$ transition involve the following, consecutive stages:

1. For GlcA:
   i. The system leaves the $^{4}C_{1}$-related basin of attraction through two possible paths. The more probable one involves adopting the E_5, $^{O}H_5$ and ^{O}E conformations while the second one is associated with the E_3 conformer mainly. Transitions through other regions of phase space (e.g. those corresponding to the $^{2}H_3$, $^{4}H_3$, ^{4}E, $^{4}H_5$, $^{O}H_1$) is also possible but even less likely.
   ii. The system enters the basin of attraction located around $\theta = 90$ deg. The most likely path visits two main local minima of free energy located there (i.e. those centered at $^{1}S_3$ and $^{O}S_2$). and covering a large number of boat and skew-boat conformers ($B_{3,O}$, $^{1}S_3$, $^{1,4}B$, $^{1}S_5$, $B_{2,5}$, $^{O}S_2$, $^{3,O}B$ and $^{3}S_1$). The less probable scenario involves visiting only one local minimum.
   iii. The system leaves the intermediate basin of attraction and crosses the large energetic barriers separating it from the $^{1}C_4$ conformation. There exist only one broad path, covering the region of the ^{1}E, $^{1}H_2$, E_2 conformers.
2. For GlcB:
   i. In the majority of the cases the system leaves the $^{4}C_{1}$-related basin of attraction through the path covering the $^{O}H_5$, ^{O}E and $^{O}H_1$ conformations and goes to the larger of the two minima of the free energy located around $\theta = 90$ deg. The less probable path involves adopting the $^{4}H_3$, ^{4}E conformations when the system goes to the smaller of the two minima.
   ii. The system explores the basin(s) of attraction located around $\theta = 90$ deg (i.e. $B_{3,O}$, $^{1}S_3$, $^{1,4}B$, $^{1}S_5$, $B_{2,5}$, $^{O}S_2$, $^{3,O}B$ and $^{3}S_1$). Contrary to GlcA, the most likely path visits only one local minimum of the free energy, i.e. that centered around $^{3,O}B$). Small fraction of the trajectories visits two local minima (i.e.

both $^{3,O}B$ and $^{1}S_3$). Even less probable scenario involves visiting only one, smaller local minimum, centered around $^{1}S_3$.

iii. The system leaves the intermediate basin(s) of attraction and crosses the large energetic barriers separating it from the $^{1}C_4$ conformation. As in the case of GlcA, there exist only one broad path, covering the region of the ^{1}E, $^{1}H_2$, E_2, $^{3}H_2$ and ^{3}E conformers.

It should be stressed that the time values depicted in Fig. 3 can not be directly interpreted in terms of rate constants, as during TPS simulation we did not measure the conformational 'flux' of the system from the $^{4}C_1$ conformation to those lying above θ = 20 deg. Such analysis is postponed to our future work.

Finally, we focus on the behavior of the exocyclic groups during the course of the ring puckering. Surprisingly, the probability distributions of all dihedral angles associated with the orientation of the exocyclic moieties are affected by the progress of the ring deformation (expressed as the θ value) to a very minor extent. This is depicted (Figs. S1 and S2, Supporting Information) by the density plots extracted from the TPS trajectories. The most notable (but still relatively minor) differences are limited to: (i) tor1 in GlcA (which corresponds to the orientation of the $C^{(1)}$-$O^{(1)}$-$H^{(1)}$ hydroxyl group); additional peak appears around tor1 = 90 deg, absent in the conformers for which θ < 90 deg; (ii) tor2 in GlcA (defining the orientation of the $C^{(2)}$-$O^{(2)}$-$H^{(2)}$ hydroxyl group); the peak around tor2 = 60 deg is reduced compared to the conformers for which θ < 90 deg; (iii) tor5 in GlcB (orientation of the $C^{(5)}$-$C^{(6)}$-$O^{(6)}$-$H^{(6)}$ hydroxymethyl group); the dimension of the peak around tor5 = 60 deg is increased, comparing the conformers for which θ < 90 deg. Furthermore, the pattern of these angle distributions is conserved above θ = 90 deg, i.e. for all the phase space sampled (20 deg < θ < 160 deg) the distribution of any dihedral angle associated with the orientation exocyclic group remains nearly unchanged.

The lack of correlation between progress of the $^{4}C_1 \rightarrow {}^{1}C_4$ reaction and the probability distribution of the tor1-tor6 dihedral angles speaks for the minor contribution of the exocyclic groups to the process of the ring puckering. This is also discussed in the next section.

4. Reaction coordinates

The mathematical forms of the calculated 'optimal' reaction coordinates are given in Tab. 1, with the related ln *L* values. Below, we briefly summarize the output of the LM procedure.

1. Order parameters which were found to be the most essential components of RC are defined on the basis of only ring coordinates; no coordinates involving the orientation of exocyclic groups (e.g. the distance between the $O^{(1)}$ and $O^{(4)}$ atoms) contribute to the calculated RC.
2. The most 'effective' appeared to be the collective coordinates (v), defined by eq. (1) and their combinations with the Cremer-Pople θ. Single coordinates (x in eq. (1)), being components of vs, were rejected.
3. In both cases the final RCs consist of linear combination of three (i.e. the maximum allowed number) order parameters. The LM criterion suggesting to reduce the number of order parameters composing RC if ln *L* increases by at least 1/2 ln *M*, (*M* is the total number of shooting points) was not fulfilled in any case.
4. As the consequence of the accepted functional form of RC (eq. (4)), its values can change from -∞ to +∞, and the equal probability (p_B) of reaching either $^1C_4$ or $^4C_1$ (transition state) corresponds to RC = 0. In practice, only the range -2.5 < RC < 2.5 is considered in which the p_B value varies from 2% to 98%.

The lack of exocyclically-related coordinates in the obtained RCs can be considered as surprising, due to undoubtedly significant role of the exocyclic group in the conformational equilibrium in the carbohydrate systems (e.g. the inversion of stereoconfiguration at any of the carbon ring atoms results in a change in a dynamic equilibrium between particular conformers). However, this remains in agreement with the results reported in the previous section, according to which the orientation of the exocyclic groups is not correlated with the progress of the considered ring conformational change. The structural role of the exocyclic substituents is included implicitly, as the set of the three selected order parameters and their relative weights (i.e. the *a* coefficients in eq. (2)) differ for GlcA and GlcB. The most probable reason for selecting the collective coordinates (v) instead of their components (x) is that the single v parameter (as expressed by eq. (1)) is able to differentiate between all possible canonical conformers; therefore it is more sensitive for the gradually changing ring conformation. None of the order parameters defining v exhibits such a feature.

Tab. 1. LM analysis of the TPS ensembles sampled in the two TPS simulations (GlcA and GlcB). All values of order parameters are given in degrees. The lower indices denote the type of coordinates defining given *v* (Hill-Reilly or Berces).

GlcA		
No. of order parameters	ln *L* (eq. (3))	RC
1	-3439	$r^{(1)} = -1.34 + 5.66 \cdot 10^{-3} v_{\mathrm{B}}$
2	-3412	$r^{(2)} = -2.62 + 5.23 \cdot 10^{-3} v_{\mathrm{B}} + 1.50 \cdot 10^{-2} \theta$
3	-3400	$r^{(3)} = -1.84 + 4.16 \cdot 10^{-3} v_{\mathrm{B}} + 2.27 \cdot 10^{-2} \theta - 1.39 \cdot 10^{-2} v_{\mathrm{HR}}$
GlcB		
1	-1063	$r^{(1)} = -11.2 + 0.103 \theta$
2	-1027	$r^{(2)} = -10.4 + 0.104 \theta - 7.03 \cdot 10^{-3} v_{\mathrm{B}}$
3	-1066	$r^{(3)} = -6.39 + 0.115 \theta - 7.30 \cdot 10^{-3} v_{\mathrm{B}} - 4.77 \cdot 10^{-2} v_{\mathrm{HR}}$

The main aim for the calculated RC is to predict the probability of evolution of the system to a given state (i.e. to the $^1C_4$ conformation in the considered case) on the basis of those parameters which appear in the mathematical form of RC. As for both GlcA and GlcB only the ring coordinates appear in their respective RCs, thus, in theory, one can predict the probability of the conformational transition knowing only the specific shape of the carbohydrate ring. To test the quality of the obtained RC the committor (i.e. the p_B values in the function of RC) has been calculated explicitly for the numerous configurations characterized by different values of RC. The results (shown in Fig. 4) have been compared with RC based on the LM procedure, proving that serious inaccuracies exist, especially in the case of GlcB.

In the case of GlcA the level of agreement is satisfactory in the region corresponding to RC < 0 but decreases significantly for RC ≥ 1. On the other hand, RC calculated for GlcB has virtually no prediction ability due to extremely large discrepancies observed not only in the region RC > 0 but for nearly all range of RC except of its extremely small and large values. Appreciating the unsatisfactory results obtained for GlcB, one can still note that the GlcA-related RC is able to predict the probability of the system evolving to the $^1C_4$ state fairly well for RC < 0. One can experience the impression that such RC gives the correct results in a half of the cases. Actually, the condition RC < 0 is fulfilled for the majority of the phase space expressed in the Cremer-Pople θ coordinate. The rough estimation shows that the transition state (TS) structures lies around θ = 125-130 deg, thus, the calculated coordinate should work quite well when the GlcA conformation corresponds to $\theta < 125$ deg. When considering the cutoffs of the developed trajectories (20 deg $< \theta <$ 160 deg), this represents over 70-75 % of the possible values of θ. Note also that providing the exact value of θ above which RC fails to calculate p_B correctly is impossible, as RC is also the function of other essential order parameters.

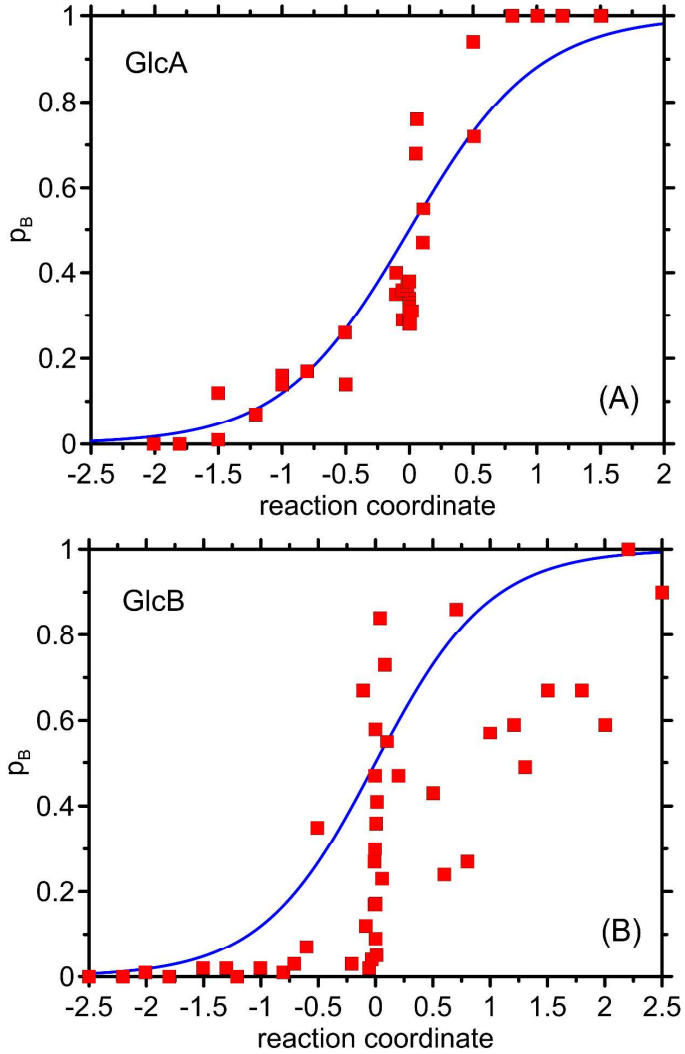


Fig 4. The test of the calculated reaction coordinate for GlcA (A) and GlcB (B). Solid lines represent the theoretically calculated RC composed of three order parameter ($r^{(3)}$ in Tab. 1), whereas each square denotes the p_B value obtained from the explicitly calculated committor. Further details are given in the text.

The reasons for discrepancies between the predictions of RCs and the explicitly calculated commitors may be of various nature. Below the most probable hypotheses are discussed:

i. The procedure of calculating RC based on the LM approach is inapplicable for the GlcB system (but at least partially applicable for GlcA). The procedure used here has been derived under the assumption that the so-called 'diffusive' barrier crossing controls the dynamic behavior of the system[53]. There is no test allowing for unambiguous statment how close is the behavior of the considered system to the ideal 'diffusive' dynamics. Obviously, the free energy landscapes of GlcA and GlcB are less complicated than those of larger biomolecules for which the LM approach has been applied so far (e.g. refs. [65,76]). Judging only on the basis of the height of the free energy barriers one can suppose that GlcB will be further from the ideal 'diffusive' behavior (compared with GlcA) as the very high (~50 kJ/mol) free energy barrier separating the non-chair conformer from the $^{1}C_{4}$ pucker seems to dominate the whole dynamics of the system. Such hypothesis is also supported by the shorter average length of the reactive trajectories obtained for GlcB. The longer trajectories collected in the case of GlcA correspond to more diffusive barrier crossing which, on average, takes more time.

ii. The lack of some essential order parameter(s) which results in the poor (GlcB) or insufficiently good (GlcA) RC. Contrary to conformational transitions in more complex biomolecules (e.g. proteins) the number of order parameters is rather limited. In spite of that we can not exclude the situation in which some crucial order parameter(s) can be missing (e.g. those related to the arrangement and dynamics of water molecules) or some details of the process can not be described by simple parameters (based on the ring puckering-related coordinates and some atom-atom distances) applied in this study. This issue is discussed in more detail in the subsequent part of the paper.

iii. Transition states are located extremely close to one of the basins of attraction ($^{1}C_{4}$). Unlike the issues mentioned in points 1. and 2. which are of rather general nature and the respective questions can be raised in the context of any system subjected to the analogical LM procedure, the present hypothesis seems to be characteristic of only the carbohydrate ring conformational

changes. The reason for that is the combined influence of: (i) the asymmetry in the depths of the two energy wells corresponding to the two stable states; (ii) the asymmetry in the heights of the two largest free energy barriers separating these states (and located around θ = 60 and 120 deg, see Fig. 1(B)). This results in the fact that, independently of the order parameter used, the transition state is located in the direct vicinity of the (final) $^1C_4$ conformer. In practice, conformations corresponding to TS and $^1C_4$ can be separated by only one frame of MD trajectory (saved every 0.1 ps). Thus, nearly all the successful shooting points correspond to RC < 0. Although this does not influence the sampling process, such asymmetric distribution may be a potential reason for artifacts during adjusting RC; this is dictated by the fact that shooting points located around TS have the largest weight in the LM procedure (according to eq. (15) in ref. [53]). Expanding a small piece of the phase space (130 deg < θ < 160 deg) into a large region of RC > 0 may result in inaccuracies located in the latter range.

5. Transition states and water dynamics

The ideal RC would allow for predicting the probability of evolution of the system either to the $^1C_4$ or $^4C_1$ conformation. As our success was rather limited and restricted to the case of GlcA, using any of the equations from Tab. 1 for determining the transition state structures would lead to serious inaccuracies (note that this concerns only the region of reaction progress for which RC > 0; we can still fairly well predict the fate of given conformational state for RC < 0, basing on the algebraically expressed RC). However, thanks to the explicitly calculated committors for both GlcA and GlcB, we can use their values for selecting the potential transition states and for testing the influence of water dynamics on their time evolution.

Having the structures identified as transition states (10 different structures for both GlcA and GlcB for which $p_B = 0.5 \pm 0.1$, as confirmed by explicit committor calculations) we examined the influence of the instantaneous water structure on the committor to estimate the contribution of the water dynamics to the reaction coordinate. We froze the carbohydrate coordinates and simultaneously disrupted the water structure/dynamics by running short (~ 200 ps) MD simulations at 298 K. Then we checked if the committor for these 'randomized' structures deviates systematically from its initial value, i.e. 0.5; this was done by analogy to the explicit commitor calculations described in the Methods section. In the case of GlcA the

recalculated committors were equal to 0.5 ± 0.1. Therefore we concluded that this committor is independent of the dynamics of the water. This conclusion seems to be surprising in the view of the reports describing the significant role of water in the carbohydrate conformations. There is no contradiction, however, as the water configuration can still play a structural role, influencing the carbohydrate configurations in the free energy wells. On the other hand, the analogous calculations performed for GlcB gave completely different results; it appeared that the recalculated committors deviated significantly from their initial values (varying from 0.08 to 0.4), evidencing that water dynamics is a crucial component of reaction coordinate. This may explain the poor results of searching for an optimal reaction coordinate in the case of GlcB as no water-related order parameters were explicitly included in the LM procedure. There also exist the previously reported cases of water dynamics being irrelevant for reaction coordinates[76] when concerning the systems in which the considered process was driven by the 'diffusive' energy barriers crossing. This observation, combined with the results obtained for GlcA (indicating the larger contribution of 'diffusive' dynamics) allows for speculation if the relevance of water dynamics is correlated with the degree of 'diffusivity' of the system over the free energy landscape accompanying the process of interest. Independently of the specific reasons, one has to appreciate that the dynamics of conformational rearrangement in the molecules of GlcA and GlcB is driven by (at least) slightly different mechanisms. This remains in agreement with the fact that in spite of using exactly the same set of order parameters the resulting RCs differ significantly in their ‘prediction power’; this also implies that not only water structural arrangement but also its dynamics can be an important component of RC. Such scenario may explain why the committors recalculated for GlcB deviated much from their initial values (any order parameter directly related to the dynamics of water molecules has not been tested during the LM procedure). The ultimate explanation of this issue (as well as proper choice from the alternative mentioned in the previous part) would require designing new, water-related, order parameter(s) and testing it (them) in the new LM procedure. However, we have not been able to find any proper order parameters, based on the structural features of water molecules surrounding the GlcB molecule and correlated with the progress of the $^4C_1 \rightarrow {}^1C_4$ transition. Note also that the LM procedure approves only the conclusion that the water dynamics has a little (GlcA) or significant (GlcB) effect on the conformational rearrangements for the system being in the vicinity of the transition state; the reason for that is the TS configurations have the largest weight in the LM procedure (according to eq. 3).

The calculated transition states corresponding to GlcA exhibit relatively similar structures, in spite of the fact that they do not correspond to any single canonical conformation. The ring conformations of the TS structures cover a relatively large area of the Cremer-Pople (ϕ,θ) representation, centered around ($\phi = 300$ deg, $\theta = 105$ deg) which corresponds to an intermediate between the E_2 and $B_{2,5}$ conformers and spreading up to the ^1E and ^3E conformers. Marking several TS on the Cremer-Pople puckering free energy landscape (Fig. 2(A)) clearly shows that such set of coordinates is insufficient to predict the TS structures; TS-related points are rather scattered and do not indicate the existence of some specific 'bottleneck' controlling the flow of the system into the $^1C_4$-related basin of attraction. Furthermore, the lack of the appropriate parameters associated with the free energy landscape is the cause why some of the TS structures are closer to the local minima of the free energy than to the free energy barriers. The orientation of the exocyclic groups varies from one TS structure to another, remaining in agreement with the results described in the previous section. The four exemplary TS structures (those depicted in Fig. 2(A)) are deposited in Supporting Information (PDB format). As we found that water environment and its dynamics are essential for determining if the given structure of GlcB is a transition state, the related analysis based solely on the GlcB structure (being a potential TS) seems to be pointless.

Conclusions

The transition path sampling study described in this paper concerned the conformational rearrangements occurring in the ring of hexapyranose (α-D- (GlcA) and β-D-glucose (GlcB)) molecules. The equilibrium set of unbiased molecular dynamics trajectories reflecting the chair-to-inverted chair transition ($^4C_1 \rightarrow {}^1C_4$) has been collected by applying the transition path sampling method and subjected to further analysis. Based on the data we found that the transition paths, after crossing the free energy barrier separating the $^4C_1$ conformer from the non-chair conformers, lead to the two local minima corresponding to the non-chair (boat and skew-boat) conformers. The next step is the escape of the system from this basin of attraction by crossing another (larger) free energy barrier separating the system from the $^1C_4$ conformation. The time spent by the system in the intermediate, non-chair region of the phase space comprises for most of the time of the $^4C_1 \rightarrow {}^1C_4$ transition (which takes, on average, 277 and 65 ps for GlcA and GlcB, respectively). Crossing of the two largest free energy barriers is a very rapid process, lasting on average less than 1 ps. The inspection into

the ring puckering-associated free energy profiles reveals that the local minima of the free energy do not match particular canonical conformers, which remains in agreement with the results reported previously by other authors. We have found only two distinct minima of the free energy, corresponding to the non-chair conformations when using both the Cremer-Pople coordinates and other order parameters, used in the subsequent reaction coordinate analysis. Furthermore, the orientation of the exocyclic groups (of both the GlcA and GlcB molecules) is nearly independent of the progress of the $^4C_1 \rightarrow {}^1C_4$ transition. In spite of failing to develop the reaction coordinate (RC) describing the conformational transitions in the GlcB ring, we were able to find an analogous RC for GlcA. The GlcA-related coordinate works fairly well, providing the GlcA molecule (undergoing the conformational rearrangement from $^4C_1$ to $^1C_4$) has not reached the transition state yet. In practice, this corresponds to the majority of the phase space expressed by the Cremer-Pople θ coordinate, i.e. $\theta < 130$ deg. For larger values of θ, when the ring conformation becomes close to the final $^1C_4$ state, the inaccuracies inherent in the calculated RC increase dramatically. The calculated RC has the form of simple linear combination of parameters based solely on the ring conformation; no parameters involving the orientation of the exocyclic functional groups or the water environment are included. The explicitly calculated transition states (TS) of GlcA and GlcB molecules (exhibiting equal probability of evolution either to the $^4C_1$ to $^1C_4$ states) were subjected to the analysis, revealing that the water dynamics is significant for the $^4C_1 \rightarrow {}^1C_4$ transition only in the case of GlcB, contrary to GlcA. This may be an explanation for the inapplicability of the RC adjustment procedure (as we were not able to propose any water-related coordinate correlated with the progress of the ring puckering) and is an evidence for different mechanisms of ring puckering in the molecules of GlcA and GlcB.

Acknowledgments

The authors acknowledge the financial support of the Polish Ministry of Science and Higher Education (contract financed in 2012–2015 under Project No. 2011/03/D/ST4/01230 SONATA).

References

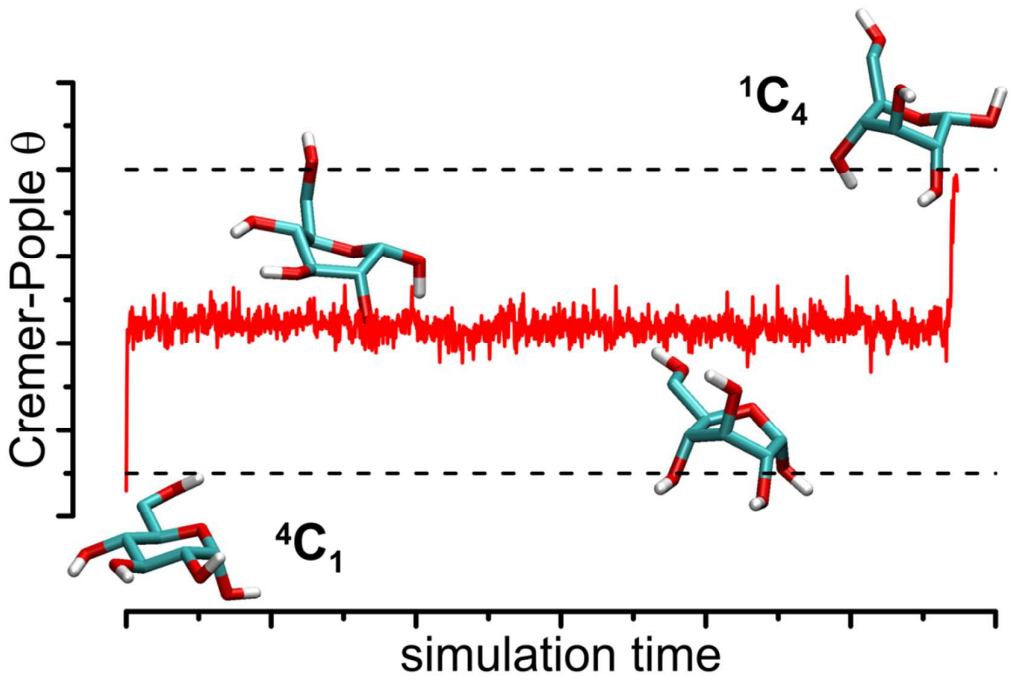
1. B. M. Sattelle and A. Almond, *Glycobiology*. 2011, 21, 1651.
2. E. Autieri, M. Sega, F. Pederiva and G. Guella, *J. Chem. Phys.* 2010, 133, 095104.
3. J.R. Snyder and A.S. Serianni, *J. Org. Chem*. 1986, 51, 2694.
4. O. Smidsrod, R.M. Glover and S. Whittington, *Carbohydr. Res.* 1973, 27, 107.
5. D.A. Horita, P.J. Hadjuk and L.E. Lerner, *Glycoconjugate J*. 1997, 14, 691.
6. V. Spiwok and I. Tvaroska, *J. Phys. Chem. B* 2009, 113, 9589.
7. P.N. Sanderson, T.N. Huckerby and I.A. Nieduszynski, *Glycoconjugate J.* 1985, 2, 109.
8. F.H. Cano, C. Foces-Foces, J. Jimenez-Barbero, M. Bernabe and M. Martin-Lomas, *Carbohydr. Res.* 1986, 145, 319.
9. M. Ragazzi, D.R. Ferro, A. Provasoli and *J. Comput. Chem.* 1986, 7, 105.
10. D.R. Ferro, A. Provasoli, M. Ragazzi, G. Torri, B. Casu, G. Gatti, J.-C. Jacquinet, P. Sinay, M. Petitou and J. Choay, *J. Am. Chem. Soc*. 1986, 108, 6773.
11. P.N. Sanderson, T.N. Huckerby and I.A. Nieduszynski, *Biochem. J*. 1987, 243, 175.
12. R.M. Giuliano, R.F. Bryan, P. Hartley, S. Peckler and M.K. Woode, *Carbohydr. Res.* 1989, 191, 1.
13. D.R. Ferro, A. Provasoli, M. Ragazzi, B. Casu, G. Torri, V. Bossenec, B. Perly, P. Sinay, M. Petitou and J. Choay, *Carbohydr. Res.* 1990, 195, 157.
14. D. Lamba, A.L. Segre, M. Ragazzi, D.R. Ferro and R. Toffanin, *Carbohydr. Res. C* 1991, 209, 13.
15. M.J. Forster and B. Mulloy, *Biopolymers* 1993, 33, 575.
16. B. Coxon and R.C. Reynolds, *Carbohydr. Res.* 2001, 331, 461.
17. M.U. Roslund, K.D. Klika, R.L. Lehtilä, P. Tähtinen, R. Silanpää and R. Leino, *J. Org. Chem.* 2004, 69, 18.
18. D. Navarro and C.A. Stortz, *Carbohydr. Res.* 2005, 340, 2030.
19. S. Karamat and W.M.F. Fabian, *J. Phys. Chem. A* 2006, 110, 7477.
20. M. Bergenstråhle, E. Thormann, N. Nordgren and L.A. Berglund, *Langmuir* 2009, 25, 4635.
21. P.E. Marszalek, A.F. Oberhauser, Y.-P. Pang and J.M. Fernandez, *Nature* 1998, 396, 661.
22. P.E. Marszalek, Y.-P. Pang, H. Li, J.E. Yazal, A.F. Oberhauser and J.M. Fernandez, *Proc. Natl. Acad. Sci. USA* 1999, 96, 7894.
23. H. Li, M. Rief, F. Oesterhelt, H.E. Gaub, X. Zhang and J.C. Shen, *J. Chem. Phys. Lett.* 1999, 305, 197.

24. P. O'Donoghue and Z.A. Luthey-Shulten, *J. Phys. Chem.* B 2000, 104, 10398.
25. G. Lee, W. Nowak, J. Jaroniec, Q. Zhang and P.E. Marszalek, *Biophys. J.* 2004, 87, 1456.
26. Q. Zhang, G. Lee and P.E. Marszalek, *J. Phys. Conden. Mater.* 2005, 17, S1427.
27. Q. Zhang and P.E. Marszalek, *J. Am. Chem. Soc.* 2006, 128, 5596.
28. M.A.K. Williams, A. Marshall, R.G. Haverkamp and K.I. Draget, *Food Hydrocolloids* 2008, 22, 18.
29. B. Hakkarainen, K. Fujita, S. Immel, L. Kenne and C. Sandström, *Carbohydr. Res.* 2005, 340, 1539.
30. M. Anibarro, K. Gessler, I. Uson, G.M. Sheldrick, K. Harata, K. Uekama, F. Hirayama, Y. Abe and W. Saenger, *J. Am. Chem. Soc.* 2001, 123, 11854.
31. G. Uccello-Barretta, G. Sicoli, F. Balzano and P. Salvadori, *Carbohydr. Res.* 2003, 338, 1103.
32. G.J. Davies, V.M.-A. Ducros, A. Varrot and D.L. Zechel, *Biochem. Soc. Trans.* 2003, 31, 523.
33. D.L.Vocadlo and G.J. Davies, *Curr. Opin. Chem. Biol.* 2008, 12, 539.
34. L. Petersen, A. Ardèvol, C. Rovira and P.J. Reilly, *J. Phys. Chem. B* 2009, 113, 7331.
35. M.E.S. Soliman, G.D. Ruggiero, J.J.R. Pernia, I.R. Greig and I.H. Williams, *Org. Biomol. Chem.* 2009, 7, 460.
36. M.E.S. Soliman, J.J.R. Pernia, I.R. Greig and I.H. Williams, *Org. Biomol. Chem.* 2009, 7, 5236.
37. B.M. Sattelle, J. Shakeri and A. Almond, *Biomacromolecules* 2013, 14, 1149.
38. H.S. Hansen and P.H. Hünenberger, *J. Comp. Chem.* 2011, 32, 998.
39. H.S. Hansen and P.H. Hünenberger, *J. Chem. Theor. Comput.* 2010, 6, 2622.
40. V. Spiwok, B. Králová and I. Tvaroška, *Carbohydr. Res.* 2010, 345, 530.
41. M. Sega, E. Autieri and F. Pederiva, *Mol. Phys.* 2011, 109, 141.
42. M. Sega, E. Autieri and F. Pederiva. *J. Chem. Phys.* 2009, 130, 225102.
43. X. Biarnes, A. Ardevol, A. Planas, C. Rovira, A. Laio and M. Parrinello, *J. Am. Chem. Soc.* 2007, 129, 10686.
44. C.B. Barnett and K.J. Naidoo, *J. Phys. Chem. B* 2010, 114, 17142.
45. B.M. Sattelle, S.U. Hansen, J. Gardiner and A. Almond, *J. Am. Chem. Soc.* 2010, 132, 13132.
46. A. Laio and M. Parrinello, *Proc. Nat. Acad. Sci. U.S.A.* 2002, 99, 12562.
47. C.B. Barnett and K.J. Naidoo, *Mol. Phys.* 2009, 107, 1243.

48. C. Dellago, P.G. Bolhuis and P.L. Geissler, *Adv. Chem. Phys.* 2002, 123, 1.
49. P.G. Bolhuis, D. Chandler, C. Dellago and P.L. Geissler, *Ann. Rev. Phys. Chem.* 2002, 53, 291.
50. A. Ma and A.R. Dinner, *J. Phys. Chem. B* 2005, 109, 6769.
51. R.B. Best and G. Hummer, *Proc. Natl. Acad. Sci. U.S.A.* 2005, 102, 6732.
52. W. E, W. Ren and E. Vanden-Eijnden, *Chem. Phys. Lett.* 2005, 413, 242.
53. B. Peters and B.L. Trout, *J. Chem. Phys.* 2006, 125, 054108.
54. S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. Van Der Spoel, B. Hess and E. Lindahl, *Bioinformatics* 2013, 29, 845.
55. H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren and J. Hermans, [in:] Handbook of Intermolecular Forces; B. Pullman, Ed.; Springer-Science+Business Media, B.V.: Israel, 1981.
56. R.W. Hockney, *Methods Comput. Phys.* 1970, 9, 136.
57. B. Hess, *J. Chem. Theory Comput*. 2008, 4, 116.
58. B. Hess, H. Bekker, H.J.C. Berendsen and J.G.E.M. Fraaije, *J. Comput. Chem.* 1997, 18, 1463.
59. G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys*. 2007, 126, 014101.
60. M. Parinello and A. Rahman, *J. Appl. Phys*. 1981, 52, 7182.
61. W.F. van Gunsteren and H.J.C. Berendsen, *Angew. Chem. Int. Ed. Eng*. 1990, 29, 992.
62. T.N. Heinz, W.F. van Gunsteren and P.H. Hűnenberger, *J. Chem. Phys*. 2001, 115, 1125.
63. D. Cremer and J.A. Pople, *J. Am. Chem. Soc*. 1975, 97, 1354.
64. M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. Broglia and M. Parrinello, *Comp. Phys. Comm*. 2009, 180, 1961.
65. J. Vreede, J. Juraszek and P.G. Bolhuis, *Proc. Natl. Acad. Sci. U.S.A.* 2010, 107, 2397.
66. W. Plazinski and M. Drach, *J. Comp. Chem*. 2012, 33, 1709.
67. J. Juraszek and P.G. Bolhuis, *Biophys. J.* 2008, 95, 4246.
68. H.M. Pickett and H.L. Strauss, *J. Am. Chem. Soc*. 1970, 92, 7281.
69. A. Bérces, D.M. Whitfield and T. Nukada, *Tetrahedron* 2001, 57, 477.
70. A.D. Hill and P.J. Reilly, *J. Chem. Inf. Model.* 2007, 47*,* 1031.
71. A.Y. Zotov, V.A. Palyulin and N.S. Zefirov, *J. Chem. Inf. Comput. Sci.* 1997, 37*,* 766.

72. R.D. Lins and P.H. Hünenberger, *J. Comp. Chem*. 2005, 26, 1400.
73. K.N. Kirschner, A.B. Yongye, S.M. Tschampel, J. Gonzalez-Outeirino, C.R. Daniels, B.L. Foley and R.J. Woods, *J. Comput. Chem.* 2008, 29, 622.
74. W. Damm, A. Frontera, J. Tirado-Rives and W.L. Jorgensen, *J. Comput. Chem.* 1998, 18, 1955.
75. H.S. Hansen and P.H. Hünenberger, *J. Comput. Chem.* 2010, 31, 1.
76. J. Juraszek and P.G. Bolhuis, *Proc. Natl. Acad. Sci. U.S.A.* 2006, 103, 15859.

Table of Contents



The transition paths corresponding to the conformational rearrangements in the ring of hexapyranose (α-D- and β-D-glucose) molecules were described by applying the transition path sampling method.