This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal’s standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.
Mechanism and Theory of D-Glucopyranose Homogeneous Acid Catalysis in the Aqueous Solution Phase

Manik Kumer Ghosh, a,b Mícheál Séamus Howard, a,b Karla Dussan, c Stephen Dooley a,b

A detailed systematic theoretical study of the mechanism of the homogeneous Brønsted -acid catalysis of D-glucose in aqueous solution phase ("acid hydrolysis") is reported. G4MP2 with the SMD solvation model at B3LYP/6-31G(2df,p) are employed to compute the free energies of the molecular and ionic species pertaining to the isomerization, protonation, hydrogen cation transfer and decomposition processes of D-glucopyranose in aqueous solution phase. This information is used to hypothesise a reaction mechanism that is of improved accuracy and completeness from the existing art. It is found that rotation of the D-glucose alkyl carbon-carbon bond is a facile process and is very important to the subsequent catalytic mechanism. This rotation produces two rotameric isomers which are of notably different thermodynamic stability and reactivity, even with regard to the products of this acid catalysis. As a low energy process \( \Delta G^+ \approx 3.8 \text{ to } 6.7 \text{ kcal/mol} \), the alkyl carbon-carbon bond may rotate toward the hydroxyl group at the adjacent "4" position reducing the energy required to protonate that position by 3.0 – 7.2 kcal/mol (or 15 – 30 %). The combination of two rotomeric isomers with the six structural isomers owing to the oxygen atoms, means that protonated D-glucose cations embark on a complex competition of interconversion and decomposition that is both thermodynamically and kinetically influenced. The calculations support the hypothesis that the acid-catalysed hydrolysis of D-glucose may yield a number of platform chemicals that have not previously been suggested. These include the prospect of three isomers of 5-hydroxymethylfurfural (HMF); 5-(hydroxymethyl)furan-2-carbaldehyde, 5-(hydroxymethyl)furan-3-carbaldehyde and 5-(hydroxymethyl)furan-4-carbaldehyde. Vibrational spectra of these HMF isomers are also computed and compared to experimentally determined infrared spectra of "humins". On this basis, it is cautiously speculated that the alternative HMF isomers, may be monomeric constituent of the polymeric "humins".

1 Introduction

The depletion of oil reserves and global environmental concerns has motivated research and development to the production of fine chemicals and fuels from sustainable resources. The production of valuable platform chemicals from lignocellulosic biomass (cellulose, hemicellulose and lignin) is of key interest. Hemicellulose and cellulose can be readily decomposed to their constituent monomeric carbohydrates through acid and enzymatic hydrolysis. The acid-catalysed conversion of cellulose, which is the most abundant carbohydrate in biomass, to D-glucopyranose ("D-glucose") results in the formation of platform chemicals such as 5-(hydroxymethyl)furan-2-carbaldehyde (HMF) and levulinic acid (LA). The mechanism of how this transformation occurs is of great importance.

Formation of HMF from D-glucopyranose requires three dehydration steps and is experimentally found to have an apparent activation energy of 20 – 38 kcal/mol. In an aqueous phase, D-glucopyranose (<0.5 M) is converted to HMF in low yields (<8 mol%) after short reaction times (<10 min) using \( \text{H}_2\text{SO}_4 \) or \( \text{HCl} \) as acid catalysts at 180 – 200 °C. At longer reaction times levulinic (LA) and formic acid (FA) are produced by rehydration of HMF showing an apparent activation energy of 22 – 26 kcal/mol. The formation of levulinic acid and furfuryl alcohol is generally improved with the use of lower reaction temperatures, reaching yields of over 50 mol% at temperatures lower than 160 °C. However, such processes are currently hampered by low product yields and the formation of unwanted side products, such as soluble and insoluble polymeric molecules, commonly referred to as "humins". After complete conversion of D-glucopyranose, van Zandoort et al. determined that humins accounted for 28–36% of the initial substrate at 180–247 °C (0.05–0.1 M acid). Other products, speculated to be formed through reversion reactions, have also been identified during the acid-catalysed conversion of D-glucopyranose. These include disaccharides (6-O-β- and 6-O-α-D-glucopyranosyl-D-glucose), and anhydrosugars (levoglucosan and 1,6-anhydro-β-D-glucofuranose). These degradation products, however, have been detected only at low D-glucopyranose conversions and short residence times. In their experimental kinetic study, Yang et al. also show that the formation of furfuryl alcohol and HMF are kinetically accessible during the D-glucopyranose dehydration. Reports of
the formation of other by-products from the dehydration of D-glucopyranose include furfural, lactic and acetic acid, and retroaldol products (such as pyruvaldehyde, dihydroxyacetone and glyceraldehyde). All this points to a complicated reaction mechanism.

Select solvents other than water can stabilise HMF in the acid-catalysed conversion of D-glucopyranose, therefore improving the selectivity of the reaction toward levulinic acid and away from humins. Chheda et al.\(^7\) found that the addition of equal parts of dimethyl sulfoxide to water (1:1) under acidic conditions (pH=1.0) can improve the formation of HMF from 2 to 10 mol% at 443 K. Pagán-Torres et al.\(^3\) show that a tandem system of both Lewis (AlCl\(_3\)) and Brønsted (HCl) acids can convert D-glucopyranose to HMF at a yield of 62 mol%. The impressive yield is achieved by isomerizing D-glucopyranose to fructose, followed by dehydration to HMF in a biphasic system of water in an alkyl phenol solvent. Mushrif et al.\(^9\) found that the carbonyl carbon in HMF can strongly coordinate with dimethyl sulfoxide, which in turn does not behave as a hydrogen bond donor. This coordination protects HMF from undergoing dehydronation and thus improves the selectivity of HMF formation from carbohydrates.

From theory, the consensus on the mechanism of the acid-catalysed conversion of D-glucopyranose is that the reaction is initiated by the protonation of one of the hydroxyl (OH) groups adjacent to the pyranose ring, rather than of the methylene hydroxyl group. However, disagreement exists over which OH group is the most preferred protonation position. Yang et al.\(^10\) examined the hydrogen cation-catalysed conversion of α-D-glucopyranose using B3LYP/6-311+G(d,p) including the Polarizable Continuum Model (PCM) to capture the solvent effect. They suggested that the OH group at position “1” (see Figure 1) is the most thermodynamically favoured position for the initial protonation (\(\Delta G^\ddagger_{\text{protonation}}\) = -5.5 kcal/mol relative to reactants). The protonated position is immediately dehydrated resulting in a stable carbocation that forms an oxonium cation by resonance. Protonation of the other OH positions is calculated to produce cations that are thermodynamically less stable and therefore less favoured.

In total contrast, Assary et al.\(^11\) found that the OH group at positions “3” and “5” are the most preferred positions in the protonation of D-glucopyranose in aqueous solution phase. Notably, Assary et al. compute \(\Delta G^\ddagger_{\text{protonation}}\) = +15 kcal/mol, relative to reactants, using G4M2 including a solvation model based on the dielectric current density (SMD). This finding is of very different meaning to the stability of the resulting protonated D-glucopyranose, as infers the cation to be less stable than reactants.

However, Qian et al.\(^12\) corroborate the finding of Assary et al., also reporting that the OH group at position “2” is the most preferred for the initial protonation with a similar computed free energy (\(\Delta G^\ddagger_{\text{protonation}}\) = +15 kcal/mol) calculated by Car-Parrinello ab initio molecular dynamics simulations (CPMD).\(^13\) Following the initial protonation, all of these works state that the protonated D-glucopyranose dehydrates immediately at the site of protonation, so called "decomposition". Protonation at the "2" position and subsequent dehydration (decomposition) is typically considered to be the exact reaction pathway resulting in the formation of HMF from D-glucopyranose\(^5\) -\(^11\), apparently explaining the HMF formation observed in experiment. The decomposition mechanism of β-D-glucose has been studied previously considering initiation of the general reaction by protonation of the various different OH positions of β-D-glucose.\(^10\) -\(^12\) -\(^14\) -\(^16\) From gas and solution phase simulations, these papers can generally summarised to find that: protonation of the OH at position “2” of β-D-glucose leads to the formation of 5-(hydroxymethyl)furancarbaldehyde (HMF) through a 2,5-anhydride intermediate. The protonation of the OH at “3” and “4” positions leads to two different five-membered ring intermediates which may further polymerize to form non-cellulosic materials that are observed in experiment.\(^16\) The β-D-glucose condensation reaction is initiated by the protonation of OH at position “1” through the formation of a carbocation at the carbon atom of position “1”, whereas the protonation of the OH at position “5” does not lead to any degradation products.\(^14\) However, it is also recognised that the protonated D-glucopyranose may undergo several other competing processes; deprotonation, dehydration, and structural rearrangements which stabilise the cation.\(^5\) -\(^11\) -\(^17\) It is of course also possible for the hydrogen cation to transfer between OH sites, isomerizing to other cations before the dehydration and decomposition steps. However, this hydrogen cation transfer mechanism has not been explored.

Obviously, the electronic configuration of D-glucopyranose plays a vital role in the reactivity of the molecule in acid hydrolysis. In attempting to fundamentally explain the reaction product observations noted above, it is crucial to realise that the D-glucopyranose starting material may exist in a number of anomeric forms. In this regard, numerous experiments with nuclear magnetic resonance (NMR)\(^18\) -\(^20\) and various theoretical studies\(^21\) -\(^27\) on the conformational behaviour of D-glucopyranose in both gas and aqueous solution phases have been performed. The conventional numbering of heavy atoms of D-glucopyranose is shown in Figure 1. The hydroxyl group attached to the carbon at the position “1” (C1) can be either axial or equatorial with respect to the ring, yielding α- and β-D-glucopyranose anomers, respectively, in a typical ratio of ~36:64.\(^25\) In a kinetic study of D-glucopyranose mutarotation, Le Barbey et al.\(^29\) experimentally determined the equilibrium constant between α and β-D-glucopyranose to be 1.503 at 30 °C, corresponding to an α- and β-form ratio of 40:60.

Similar for fructose, \(^13\)C NMR studies by both Kimura et al.\(^30\) and Maple et al.\(^31\) also show that the α-D-glucofuranose, β-
D-glucufuranose, the open chain form, and the gem-diol species can also exist, however, in low quantities (<2%). Thus, with regard to structural isomers, the pyranose form is known to dominate, and be biased toward the β-form. β-D-glucopyranose therefore presents the fairest proxy for detailed study of D-glucopyranose in aqueous solution phase.

Explaining what elementary reaction events lead to the observed products of HMF, furfuryl alcohol and humins is further complicated when one considers the molecularity of the system to also extend to rotational isomers. Three major rotamers have been reported when the torsional angle (ω in Figure 1) changes from −180° to +180°. The rotamers at ω = −60°, +60°, and 180° are referred to as gauche-gauche (gg), gauche-trans (gt), and trans-gauche (tg), respectively. In a recent quantum mechanical/molecular mechanical molecular dynamics (QM/MM-MD) simulation study, the gauche-gauche and gauche-trans rotamers are suggested to be energetically equivalent, but of ~1.0 – 2.0 kcal/mol lower free energy than the trans-gauche rotamer. So, while all of this isomeric complexity has previously been noted, the influence of the C–C rotation on the hydrogen cation-initiated (acid catalysis) reaction of D-glucopyranose in aqueous solution phase has never been explored.

The major reaction pathways in the Brønsted acid-catalysed D-glucopyranose dehydration reported in the literature is summarised in Scheme 1. The majority of studies suggest that the OH group at position “2” is the most preferred position of protonation and subsequent decomposition. Yang et al. proposed two different reaction pathways through two different intermediates to the formation of furfuryl alcohol and HMF, both reactions initiated from the D-glucopyranose protonated at the “2” position. Their microkinetic study suggests that the reaction pathway to the formation of HMF dominates at low temperatures, whereas at high temperatures the reaction pathway involves the formation of furfuryl alcohol. Assary et al. also proposed a decomposition reaction pathway for the formation of furfuryl alcohol from D-glucopyranose after the protonation of the OH group at position “2”. Weingarten et al. developed an empirical kinetic model for aqueous phase D-glucopyranose dehydration where HMF is produced by removing three molecules of water, in parallel with the formation of humins by decomposition reactions. Once HMF is formed, in the presence of water, a rehydration reaction takes place with two molecules of water to produce levulinic and formic acid. This mechanism is also suggested by ab-initio electronic structure calculations of Assary et al. In a parallel step, HMF also can decompose to form humins. This mechanism is summarised in Scheme 1.

However, all five hydroxyl groups of the D-glucopyranose molecule can be protonated in acidic solutions and the protonated sites may undergo several other competing processes. So, it is obvious that the overall reaction could be affected by protonation of the other hydroxyl groups in the D-glucopyranose molecule. It is therefore also possible that there are other by-products contributing to the formation of humins. But the formation of additional final products and intermediates initiated by the protonation of alternative hydroxyl groups has not been explored in detail.

In this study, computational chemistry is utilised, to further understand the elementary, catalytic reaction mechanism that accounts for the acid hydrolysis of D-glucopyranose in the aqueous solution phase. Free energies in aqueous solution phase are computed for:

(i) α/β-D-glucopyranose.
(ii) rotation of the D-glucopyranose alkyl C-C bond through 360° in 10° increments.
(iii) α/β-D-glucopyranose protonated at the ether and each OH position.
(iv) transition state geometries of the hydrogen cation transfer to the nearest neighbour oxygen atom of each protonated β-D-glucopyranose rotamer.
(v) transition state geometries connecting selected protonated β-D-glucopyranose rotamers through rotation of the alkyl C-C bond.
(vi) the stationary points and transition states of the potential energy surfaces for the decomposition of the most thermodynamically and kinetically accessible protonated β-D-glucopyranose isomers, to an end-point of HMF isomers and/or furfural alcohol platform chemicals.

Thus, this present work sets out to provide a consistent, detailed and more comprehensive informing of the elementary reaction mechanism responsible for the acid catalysis of D-glucopyranose hydrolysis in the aqueous solution phase. Harmonic vibrational frequencies of the hypothesised isolated molecules are also calculated. These are used to evaluate the computational results against experimental infrared spectroscopic characterizations of the complicated humin species observed in experiment.

2 Methodology

2.1 Computational methods

All calculations presented herein are executed with the Gaussian09 program. Equilibrium and transition state structures along the reaction pathways are computed using gradient-corrected density functional theory (DFT) with the Becke three-parameter exchange functional and the Lee-Yang-Parr correlation functional (B3LYP) in combination with
the all electron 6-31G(2df,p) basis set. To date, different computational methodologies (M06-2X, M05-2X, B3LYP, etc.) have been employed for the study of carbohydrate chemistry. Hu et al. employed M06-2X/6-311+G(d,p) method for the pyrolysis of glucose. Csonka et al. performed a benchmark study of fifteen conformers of different carbohydrates including β-D-glucopyranose using a variety of methods and basis sets. They suggested to use M05-2X/6-311+G(d,p)/M05-2X/6-31+G(d) model chemistry for carbohydrate conformational space studies.

However, the standard B3LYP method has been extensively studied for the conformational analysis and decomposition mechanism of glucose. Assary et al. calculated the reaction energies, reaction energies with zero-point energies, reaction enthalpies and reaction free energies for the conversion of glucose to levulinic acid using a variety of methods, i.e. B3LYP/6-31G*, B3LYP/6-31+G*, B3LYP/6-311++G**, B3LYP/6-31G(2df,p), B3LYP/6-311+G(3df,2p)//B3LYP/6-31+G*, B3LYP/6-311+G(3df,2p)//B3LYP/6-31G*, MP2/6-311+G(3df,2p)//B3LYP/6-31G*, G3MP2B3, G4MP2, and G4. Among the DFT methods, B3LYP/6-31+G* and B3LYP/6-31G(2df,p) perform best with regard to accuracy, where a maximum difference of ~2 kcal/mol in dehydration energies is noted compared to the results of the very accurate G4 method. Consequently, B3LYP/6-31G(2df,p) is utilized in this work for geometry optimization and for the determination of potential energy surfaces.

Minimum energy reaction paths are determined by optimizing the geometries of the energy minima and transition states. To follow the minimum energy path, also called intrinsic reaction coordinate (IRC), the Gonzalez-Schlegel second-order method is used. To characterize all stationary points, the Hessian (matrix of energy second derivatives) is calculated and diagonalized at each stationary point, which also yields the zero-point energy (ZPE) correction. The nature of the stationary points is evaluated from the harmonic modes and no imaginary frequencies are found for the optimized structures corresponding to local minima on the potential energy surfaces. The normal modes corresponding to transition-state structures are examined with molecular visualization to verify the nuclear motion that tends to deform the transition-state structure along the pertinent reaction coordinate. In all cases reported, a single imaginary frequency is found, ensuring that each transition state corresponds to different reaction intermediates.

Then, all the structures are calculated with the highly accurate G4MP2 theory, which is of comparable accuracy to G4, offering energetics comparable with experimental results to ~1 kcal/mol accuracy, as well as reaction barriers comparable to CCSD(T) with a quadruple zeta basis set. This method is based on a series of quantum chemistry calculations, combined assuming additivity of the energy terms. The six-step G4MP2 series of calculations starts with a geometry optimization followed by vibrational frequencies calculations with B3LYP/6-31G(2df,p). Further single point correlation energies are calculated with (i) CCSD(T,E4,FrzG4), (ii) MP2(FrzG4)/GTMP2LargeXP, (iii) HF/GFHFB3, and (iv) HF/GFHFB4 levels of theory. The potential energy surfaces of every reaction are first calculated in gas phase which yields enthalpies of reactions and enthalpies of activation.

To account for the effects of an aqueous environment, the universal solvation model (SMD) in a water dielectric medium with B3LYP is utilized, using the same basis set as used for the geometry optimizations. This model applies the integral equation formalism polarizable continuum model (IEFPCM) protocol, solving the non-homogeneous Poisson equation using a set of optimized atomic Coulomb radii. A parameterized function which includes terms for atomic and molecular surface tensions as well as the solvent accessible surface area has been considered for the non-electrostatic contributions. The alternative to this approach is to use DFT methods by including a number of solvent molecules coordinated to the solute explicitly. However, such modelling techniques have limitations, in computational time and tractability, difficulties in locating transition states as a result impossible to describe comprehensive potential energy surfaces. Therefore, in order to address the solvent medium, an implicit solvation model (SMD) has been employed which is adequate to describe the energetics and is computationally less expensive. The implicit solvation models are applied consistently to each geometry of the potential energy surfaces and has been extensively used for the decomposition mechanism of sugar molecules in solution.

The free energy change associated with protonation of each and every possible position of the D-glucopyranose ring, i.e. each hydroxyl group and each ether position, for both α and β anomers and for the two major rotamers of β-D-glucopyranose are first determined. This gives the equilibrium scenario of which position the protonated D-glucopyranose ought to decompose from. The kinetic propensity for each protonated D-glucopyranose to interconvert is also considered by determination of the transition state geometry connecting each cation. Considering both the calculated free energy (stability) of the protonated D-glucopyranose and the magnitude of the calculated free energy changes for hydrogen cation transfer, a set of two protonated D-glucopyranose are selected for decomposition studies. All of the geometries and their vibrational frequencies are tabulated as Supporting Information.

The gas phase harmonic frequencies of selected isolated product molecules are calculated at B3LYP/6-31G(2df,p) level of theory. As experimental data is not available, all calculated gas phase harmonic frequencies are scaled by a uniform scaling factor of 0.9668 (see section 3.7.1), which was determined by comparison of the calculated frequency to the absorption transition factor of 0.9668 (see section 3.7.1), which was determined by comparison of the calculated frequency to the absorption frequency measured by experiment, for the gas phase IR spectra of furfuryl alcohol. The optimum scaling factor for vibrational frequencies is determined by a least-squares procedure minimizing the residuals as:

\[ s = \sum_{i} \frac{v_{i}^{\text{theor}}}{v_{i}^{\text{exp}}} / \sum_{i} (v_{i}^{\text{theor}})^{-2} \] (1)
where, $v_i^{\text{theo}}$ and $v_i^{\text{exp}}$ are the $i$th theoretical harmonic and $i$th experimental fundamental frequencies (in cm$^{-1}$), respectively. The optimized scaling factor $s$ is used to calculate the minimized residual, $\Delta_{\text{min}}$, and the root-mean-square error, RMS:

$$\Delta_{\text{min}} = (s \cdot v_i^{\text{theo}} - v_i^{\text{exp}})^2$$  \hspace{1cm} \text{(2)}$$

$$\text{RMS} = \left( \sum_i \Delta_{\text{min}} / n \right)^{1/2}$$  \hspace{1cm} \text{(3)}$$

2.2 Experimental Details

To test the calculations, and due to the absence of experimental data, the infrared (IR) spectra of 5-(hydroxymethyl)furfuraldehyde (HMF) was measured experimentally. HMF (CAS Registry Number 67-47-0, 97% purity) is purchased from Carbosynth and used without further purification. Liquid phase infrared (IR) spectra are measured using an Agilent Cary 630 FTIR spectrometer fitted with a ZnSe Attenuated Total Reflection accessory. The resolution of the instrument is set to 8 cm$^{-1}$. The IR spectrum of air is recorded as background. The HMF sample is then placed on a ZnSe crystal in order to determine the IR spectrum. The reported spectra are the average of 32 scans for both the background and the HMF sample performed at 293 K ($\pm$ 5 K).

3 Results and Discussion

3.1 Conformational analysis of $\beta$-D-glucopyranose

Firstly, a detailed rotational analysis of the rotation of the C5–C6 bond, the only free rotar, along the $\omega$ dihedral angle (O6–C5–C6–O5) is carried out. The results are presented in Figure 2. Four minima are found, two of which are located at $\pm 60^\circ$ with slightly different stabilities. The global minimum located at $\omega = -60^\circ$, is more stable than the other local minimum at $\omega = 60^\circ$ by 0.3 kcal/mol. One can expect that the rotamer at $\omega = +60^\circ$ and $\omega = -60^\circ$ would give similar results for protonation and decomposition as the methylene OH makes a hydrogen bond with the cyclic ether oxygen atom in both rotamers and the geometries of the other parts of the rotamers remain unchanged. Thus, only the geometry at $\omega = -60^\circ$, which is the

<table>
<thead>
<tr>
<th>Conformer</th>
<th>Gas Phase</th>
<th>Aqueous Phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$-D-glucopyranose</td>
<td>Rotamer-1</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>Rotamer-2</td>
<td>0.00</td>
</tr>
<tr>
<td>$\beta$-D-glucopyranose</td>
<td>Rotamer-1</td>
<td>1.40</td>
</tr>
<tr>
<td></td>
<td>Rotamer-2</td>
<td>0.94</td>
</tr>
</tbody>
</table>

global minima in the potential energy surface is considered for the rest of the calculations. It is referred to as "Rotamer-1". The other two minima are located at $\pm 170^\circ$ with identical energies. Both these minima are slightly less stable than "Rotamer-1" by 1.5 kcal/mol. The rotamers at $\omega = +170^\circ$ and $\omega = -170^\circ$ can be assumed to show similar behaviour for protonation and decomposition, thus only the minimum at $\omega = -170^\circ$ is considered for detailed analysis. It is referred to as "Rotamer-2". In aqueous solution, the peak energy of the rotational transition state connecting Rotamer-1 and Rotamer-2 is computed to be 3.8 $\pm$ 6.2 kcal/mol. These transition state barriers to rotation along the C5–C6 bond are also calculated by Appel et al.\textsuperscript{21} in the gas phase to be $\approx 3.7 - 5.8$ kcal/mol at B3LYP/6-31+G(d,p), which is in agreement with the results herein. As this free energy change is small, and also lower than the free energy changes of protonation, it is considered that the sugar rapidly interconverts to each rotamer easily, even at room temperature.

The relative stabilities of different rotamers of both $\alpha$- and $\beta$-D-glucopyranose in gas and aqueous phases are presented in Table 1. In the gas phase, Rotamer-2 of $\alpha$-D-glucopyranose is more stable than the other rotamers by 0.61 $\pm$ 1.40 kcal/mol, while in solution, Rotamer-1 of $\beta$-D-glucopyranose is more stable than the other rotamers by 0.12 $\pm$ 1.30 kcal/mol. This correlates with experimental results, by which the $\beta$-D-glucopyranose isomer is more favourable in an aqueous solution.\textsuperscript{28} However, as both the $\alpha$ and $\beta$ anomers show similar stabilities in the range of 1.3 kcal/mol, it is expected that they both exist in comparable populations. This has important implications when protonation of both anomers is discussed in Section 3.2.

3.2 Protonation of D-glucopyranose

Hydrogen cation addition (protonation) is the first elementary step in the reaction mechanism of the acid hydrolysis of sugars. In the literature, this quantity is also referred to as the "hydrogen cation affinity", the "free energy of protonation", "protonation energy" or the "proton affinity" etc. This quantity is formally defined by Equation 4 as:

$$\Delta G^{\text{protonation}} = G(AH^+) - G(A) - G(H^+)$$  \hspace{1cm} \text{(4)}$$

It is equal but opposite to the deprotonation affinity given by Equation 5:

![Figure 2 free energy profile (at 298.15 K) of rotational analysis of $\beta$-D-glucopyranose along the torsional angle $\omega$(O6-C5-C6-O5) computed at B3LYP/6-31G(2df,p) using the SM4 solvation model.](image)

This journal is © The Royal Society of Chemistry 20xx

Phys. Chem. Chem. Phys., 2018, 00, 1-3 | 5
\[ \Delta G_{\text{protonation}}^f = G_f(A) + G_s(H^+) - G_f(AH^+) \]

Where, AH\(^+\) is the protonated form of species A, G\(_s\) is the free energy in the solution phase. For A and AH\(^+\) the values for the free energy in the solution phase are calculated with G4MP2 combined with the SMD solvation model as outlined previously. However, the free energy of the hydrogen cation in solution, G\(_s\) \((H^+)\) is calculated using the following equation:

\[ G_f(H^+) = G_A(H^+) + \Delta G_{\text{solvation}}(H^+) \]

Where G\(_f\) \((H^+)\) is the free energy of the hydrogen cation in the gas phase calculated with G4MP2 and \(\Delta G_{\text{solvation}}(H^+)\) is the free energy change of solvation for a hydrogen cation. As \(\Delta G_{\text{solvation}}(H^+)\) cannot be calculated accurately with the SMD solvation model, this study uses the standard value of \(-262.5\) kcal/mol, as indicated from recent high-level theoretical and experimental studies.\(^{46,47}\) Equation 4 is solved for the addition reaction of a hydrogen cation to each hydroxyl group of both \(\alpha\)- and \(\beta\)-D-glucopyranose. As internal rotation of the C5–C6 bond shows a free energy barrier of \(\sim 3.8 – 6.2\) kcal/mol, this is expected to be a facile process resulting in the interconversion of the most stable rotamers. As the rotational free energy barrier is so low, it is obvious that the C5–C6 bond rotation will also occur for the protonated D-glucopyranose. This allows for the potential conversion of protonated Rotamer-1 of D-glucopyranose to protonated Rotamer-2 of D-glucopyranose. This process must therefore also be considered.

The computed free energies of protonation are presented in Figure 3. The calculations suggest that protonation at any site of the D-glucopyranose molecule is thermodynamically less stable by \(\sim 16.6 – 27.6\) kcal/mol relative to reactants, in an aqueous solution at 298 K. It is important, as it allows the rotational free energy barrier converting Rotamer-1 to Rotamer-2 to be competitive with protonation. Thus, acid-catalysed hydrolysis of D-glucopyranose should be interpreted as occurring from two rotameric reactants. It appears this has not been realised in the literature to date,\(^{3,10,11,21,22}\) and has not been incorporated into the developing micro-kinetic models.\(^5\)

Figure 3 shows that the magnitude of the protonation energy depends but moderately on the particular anomer (\(\alpha\)(red)/\(\beta\)(blue)). On Rotamer-1, the “3” position has the lowest protonation energy in both anomers (\(\alpha\) \(= 18.9\) kcal/mol, \(\beta\) \(= 20.5\) kcal/mol). This may occur because the position of the adjacent carbon atom (C3) is far away from the electronegative oxygen atom in the D-glucopyranose ring which is expected to have a poorer electron-withdrawing effect on the carbon atom at the “3” position. The “2” and “4” positions have slightly higher protonation energy than the “3” position, since their adjacent carbon atoms are at smaller distances from the cyclic ether oxygen atom. On the basis of electronegativity, the highest protonation energy should be at the ethereal oxygen site which is less electronegative than any of the oxygen atoms of the hydroxyl groups. However, when the ethereal oxygen in Rotamer-1 is protonated, the hydrogen cation makes a strong hydrogen bond with the methylene hydroxyl group (1.799 Å) and becomes stabilized (see Figure 4a). Consequently, the “1” position has the highest protonation energy in Rotamer-1 where the adjacent carbon atom “C1” is directly bonded with the electronegative oxygen atom of the D-glucopyranose ring. Importantly, much larger differences in protonation energy are found due to the facile rotation of the (O6–C5–C6–O5). This is due to the orientation of the anomeric hydroxyl group. Unlike Rotamer-1, the lowest protonation energy of the anomers of Rotamer-2 occurs at the “4” position instead of the “3” position, as in Rotamer-1. Due to the rotation, the energy required to protonate at the “4” position is reduced by 3.0/7.2 kcal/mol for the \(\alpha/\beta\) stereoisomers of Rotamer-2 relative to those of Rotamer-1. This means that the minimum protonation energy occurs in Rotamer-2, at the “4” position and not in Rotamer-1, at the “2” position as has been previously thought.\(^{13}\)

The bond dissociation energies of intra molecular hydrogen bonds of protonated D-glucopyranose molecules cannot be accurately calculated. It is well established that, bond dissociation energy is proportional to bond length, and so bond distance is proposed as a proxy for the undeterminable dissociation energy. To explain the origin of the stability of the D-glucopyranose anomer protonated at the “4” position of Rotamer-2, an analysis of the computed bond lengths of each protonated D-glucopyranose anomer is presented in Table S1. When the OH group at the “4” position is protonated in both \(\alpha\) and \(\beta\) anomers, Rotamer-2 becomes more stable than the Rotamer-1 by 3.0 and 7.2 kcal/mol, respectively. When the OH at the “4” position is protonated, the C4–O4 bond of Rotamer-1 and Rotamer-2 is elongated by \(-0.12\) and 0.08 Å, respectively in both \(\alpha\)- and \(\beta\)-D-glucopyranose anomers. Other bonds are essentially similar, as in the other anomers. For Rotamer-2, however, a new six-membered ring is formed with the existing D-glucopyranose ring by sharing of the hydrogen cation between the OH groups at the “4” and “5” positions (See Figure 4b). This strong hydrogen bonding plays the key role of stabilizing the hydrogen cation at the “4” position of Rotamer-2. In the case of Rotamer-1, this kind of hydrogen bonding is absent as the OH group at the “5” position is far away from the OH group at the “4” position. Similarly, when the ethereal oxygen (O6) is protonated, Rotamer-1 is more stable than Rotamer-2 in both the \(\alpha\) and \(\beta\) anomers by 2.3 and 5.5 kcal/mol, respectively. This may be due to the presence of a hydrogen...
bond between the hydrogen cation and the methylene hydroxyl group in Rotamer-1 (See Figure 4a).

It is not possible to calculate the protonation energy of the "S" position of Rotamer-2. During the geometry optimization, the hydrogen cation is transferred from the "S" position to the comparatively much more stable "4" position (see Figure 4b). The minimum energy geometry corresponding to the protonation of the "S" position of Rotamer-2 is unable to be found. This kind of behaviour is also noted by Feng et al.48 in their gas phase hydrogen cation affinity calculations of cyclic D-glucopyranose, and cyclic and acyclic fructopyranose. Yang et al.10 also reported this issue when investigating the protonation of the "5" position for Rotamer-2 of α-D-glucopyranose.

Protonation energies of D-glucopyranose, fructopyranose and fructofuranose have also been studied extensively by Assary et al.11 with G4MP2 including an SMD model calculated at B3LYP/6-31G(2df,p) level of theory, using a water dielectric medium. The calculations of Assary et al. are compared to the present findings for four different anomers (α/β and Rotamer-1/Rotamer-2) of D-glucopyranose in Figure 3. It is important to highlight that the anomeric and rotameric form of the actual D-glucopyranose studied is not specified by Assary et al.11. They found that the free energy of protonation of any oxygen site is in the range of 15.0 – 17.1 kcal/mol, the lowest energy is ascribed to the "3" and "5" positions (15.0 kcal/mol) at 298 K. This disagrees with the present study; where the findings are summarised in Figure 3. The free energy of protonation of any oxygen site of Rotamer-1 is in the range of 18.9 – 24.7 kcal/mol and Rotamer-2 is in the range of 20.5 – 25.2 kcal/mol for α- and β-D-glucopyranose, respectively. Another study by Yang et al.10 found that the most favourable protonation site is at the "1" position of the α-D-glucopyranose molecule. The computed ΔG = −5.5 kcal/mol, is more thermodynamically favoured than that of the other protonation sites computed (6.7 – 9.6 kcal/mol). This is in total contradiction to the findings herein, and those of Assary et al., which suggest a much higher magnitude of protonation energy. Yang et al.10 modelled the Brønsted-acid as the Zundel complex (H3O+2) to approximate the hydrogen cation in water. Another study by Qian et al.12 found that the free energy barrier for the protonation of OH at position "2" is ~15 kcal/mol, this is 6.0 kcal/mol lower than our calculated free energy of protonation. This discrepancy may be due to the use of different instances of hydrogen cation solvation energy; other than those derived from experiment, as is standard in the field.

Following an extensive literature review, all previous studies have only considered the protonation of Rotamer-1 of D-glucopyranose and the rotational isomerization has not been considered previously. In the present study, protonation at the "4" position of Rotamer-2 is shown to yield the lowest free energy change among all protonated D-glucopyranose anomers. Therefore, these findings may have ramifications for the comprehension of this important reaction mechanism.

3.3 Hydrogen cation transfer reactions (isomerization) of protonated β-D-glucopyranose

Once protonated, the protonated D-glucopyranose can (1) decompose by covalent bond cleavage, (2) deprotonate or (3) isomerize to different protonated D-glucopyranose. Decomposition pathways yielding a series of water molecules and different oxygenated hydrocarbons have already been studied from different protonation sites.10 From that work, it appears that these reactions are consistent with experiment in that they occur through large activation energy barriers, in the range of ~25–35 kcal/mol. As the dehydration processes require large amounts of energy, the alternative possibilities must also be considered as competitive pathways. The protonation energies of the different rotamers of α- and β-D-glucopyranose have been calculated as discussed in the previous section. From Equation 4 and Equation 5 it can be seen that, at equilibrium conditions, the free energy change for deprotonation reactions for any molecule are equal in magnitude to the free energy change of protonation.

For each protonated D-glucopyranose isomer, there are five alternative positions for protonation. The energetics of the possible hydrogen cation transfers have not been previously determined. On the assumption that some of these transformations will be of at least comparable kinetic accessibility (similar free energy barrier) to the decomposition pathways, the free energy of the transition states that connect all isomers are also calculated. Only the hydrogen cation transfer to neighbouring OH groups has been considered, as this is a 5-centred process. Hydrogen cation transfers other than to the nearest neighbour oxygen atom involves more complex transition states. It is therefore reasonable to neglect these processes, as they are likely to be unfavoured.

Notation:

The general nomenclature of the protonated isomer is represented as "RxGy+" where "Rx" corresponds to Rotamer-1 (R1) or Rotamer-2 (R2), and y designates the position of the protonated oxygen as previously described. For example, R1G4+ denotes Rotamer-1 of β-D-glucopyranose protonated at the "4" position. Similarly, the transition states are represented as TSNy,uvm,n, where u/v corresponds to R1 (Rotamer-1) or R2 (Rotamer-2), and m/n is the protonated position (Gy+) of the associated rotamers (u/v), respectively. For example, TSN5,1–G4+ represents the transition state between the R1G5+ and the R2G4+ of D-glucopyranose molecules. The calculated free energy
profile of the hydrogen cation transfer reactions is presented in Figure 5. All the free energies reported in Figure 5 are relative to the sum of the free energy of the most stable solvated anomer (Rotamer-1 of β-D-glucopyranose, 430,757.38 kcal/mol) and that of the solvated hydrogen cation (\(G_0(H^+)\), – 6.28 kcal/mol + \(\Delta G_{\text{solvation}}(H^+)\), –262.5 kcal/mol).

### 3.3.1 Intramolecular hydrogen cation transfer in Rotamer-1

For brevity, the results are discussed as a hypothetical process of sequential hydrogen cation transfers, starting from the most stable anomer. Obviously, each reaction is reversible, and the quoted transition state reaction energy is presented on the directional nature of the discussion. The most stable protonated sites of Rotamer-1 are at the "3" and "6" positions, labelled as R1G**+ (\(\Delta G_1^{\text{protonation}}\) = 20.5 kcal/mol) and R1G** (\(\Delta G_1^{\text{protonation}}\) = 20.6 kcal/mol), respectively. There is only a 0.1 kcal/mol difference in free energy between the two species. The hydrogen cation can be transferred from R1G** to the neighbouring OH sites at both the "2" or "4" positions through the respective transition states TS12G** or TS14G** with free energy barriers of 8.9 and 6.6 kcal/mol to respectively form R1G**2 (\(\Delta G_2^{\text{protonation}}\) = 21.6 kcal/mol) and R1G**4 (\(\Delta G_4^{\text{protonation}}\) = 23.8 kcal/mol). Both reactions are endothermic by 1.1 and 3.3 kcal/mol respectively. The formation of R1G**4 is kinetically more favourable than the formation of R1G**2.

From R1G**+ the hydrogen cation transfer to R1G**+ (\(\Delta G_1^{\text{protonation}}\) = 25.2 kcal/mol), through TS12G** or TS14G**, has a free energy barrier of 9.1 kcal/mol. From here, the hydrogen cation transfer to R1G**+ proceeds through TS12G** or TS14G** and has a much higher barrier of 23.2 kcal/mol. This process (R1G**+ -> R1G**+) is also exothermic by 4.6 kcal/mol but may be regarded as uncompetitive on kinetic grounds. The hydrogen cation can be transferred from R1G**+ to the neighbouring OH sites, either "1" as already discussed, or to the "5" position through the transition state TS15G**+ with a free energy barrier of 3.2 kcal/mol. The process is kinetically accessible but the resultant product, R1G**+ (\(\Delta G_5^{\text{protonation}}\) = 21.6 kcal/mol) is slightly less stable than R1G** by 1.0 kcal/mol. Although R1G**+ and R1G** are the most stable protonated isomers of Rotamer-1, their interconversion is unlikely, due to the high activation free energy barrier (27.8 kcal/mol) required (R1G**+ -> R1G**). The hydrogen cation transfer between R1G**+ and R1G** is not possible as the OH at the "4" and "5" positions are far away from each other.

### 3.3.2 Intramolecular hydrogen cation transfer in Rotamer-2

The most stable protonated site of (all anomers and of) Rotamer-2 is at the "4" position, labelled as R2G** (\(\Delta G_4^{\text{protonation}}\) = 17.9 kcal/mol). This hydrogen cation can be transferred to the neighbouring OH sites, at either the "3" or "5" position. The free energy barrier of the hydrogen cation transfer between R2
G\text{2}^+ \text{and } R2G\text{3}^+ \text{ is 11.1 kcal/mol. } R2G\text{2}^+ (\Delta G^{\text{protonation}} = 20.4 \text{ kcal/mol}) \text{ is less stable than } R2G\text{3}^+ \text{ by 2.5 kcal/mol and thus the process is marginally unfavoured, but kinetically accessible.}

From the "3" position, the hydrogen cation can further transfer to the "2" position, to form R2G\text{2}^+ (\Delta G^{\text{protonation}} = 22.3 \text{ kcal/mol}), through the transition state TS\text{G1}^+\rightarrow\text{G2}^+. \text{ With a free energy barrier of 7.6 kcal/mol, the hydrogen cation transfer to the "1" position to form } R2G\text{1}^+ (\Delta G^{\text{protonation}} = 27.0 \text{ kcal/mol}), \text{ has a net free energy barrier of 9.7 kcal/mol and proceeds through the transition state, TS\text{G2}^+\rightarrow\text{G1}^+. \text{ From here (R2G\text{1}^+), the hydrogen cation transfer to the "6" position to form } R2G\text{6}^+ (\Delta G^{\text{protonation}} = 24.2 \text{ kcal/mol}) \text{ proceeds through TS\text{G5}^+\rightarrow\text{G6}^+ \text{ and has a net free energy barrier of 22.5 kcal/mol.}

The hydrogen cation may also transfer from the "4" to the "5" position of Rotamer-2. Though this pathway has been considered very carefully, no stable structure of β-D-glucopyranose protonated at the "5" position of Rotamer-2 is found. There is no transition state between these two protonated species. Once the protonated species is formed at the "5" position, the hydrogen cation is quickly transferred back to the "4" position and stabilized by strong hydrogen bonding as discussed earlier. Each of these protonated species are thermodynamically less stable than R2G\text{4}^+, \text{ making these channels unfavourable in an absolute sense. However, as the range of protonation energies of each species is 17.9 − 27.0 kcal/mol, one should reasonably expect a distribution of each isomer according to their kinetic accessibilities.

### 3.3.3 Conversion between protonated Rotamer-1 and Rotamer-2

\text{R1G}^+ (\Delta G^{\text{protonation}} = 20.5 \text{ kcal/mol}), \text{R1G}^+ (\Delta G^{\text{protonation}} = 20.6 \text{ kcal/mol}) \text{ and } \text{R2G}^+ (\Delta G^{\text{protonation}} = 17.9 \text{ kcal/mol}) \text{ are the most stable anomers among all the protonated species of Rotamer-1 and Rotamer-2 (see Figure S5). } \text{R1G}^+ \text{ and } \text{R2G}^+ \text{ can be interconverted by C5−C6 bond rotation and hydrogen cation transfer, i.e. } \text{R1G}^+ \rightarrow \text{R1G}^+ \rightarrow \text{R2G}^+ \text{ and } \text{R1G}^+ \rightarrow \text{R2G}^+ \rightarrow \text{R2G}^+. \text{ The former process involves hydrogen cation transfer between } \text{R1G}^+ \text{ and } \text{R1G}^+ \text{ through the transition state } \text{TSF1}^+\rightarrow\text{F2}^+. \text{ With a free energy barrier of 6.6 kcal/mol, as discussed in the previous section. Following this, the C5−C6 bond may rotate converting } \text{R1G}^+ \text{ to } \text{R2G}^+ \text{ through the transition state } \text{TSF2}^+\rightarrow\text{F1}^+. \text{ With a free energy barrier of just 3.3 kcal/mol. Similarly, the } \text{R1G}^+ \rightarrow \text{R2G}^+ \rightarrow \text{R2G}^+ \text{ process requires the C5−C6 bond rotation converting } \text{R1G}^+ \text{ to } \text{R2G}^+ \text{ through the transition state } \text{TSF3}^+\rightarrow\text{F2}^+ \text{, and has a free energy barrier of just 2.8 kcal/mol.}

Once R2G\text{4}^+ \text{ is formed, the hydrogen cation can be transferred to the neighbouring hydroxyl groups (at the "2" or the "4" position), yielding R2G\text{2}^+ \text{ or } R2G\text{4}^+ \text{ (the most stable protonated anomer) through the transition states } \text{TSF2}^+\rightarrow\text{F1}^+ \text{ or } \text{TSF4}^+\rightarrow\text{F2}^+. \text{ The free energy barriers for } \text{TSF2}^+\rightarrow\text{F1}^+ \text{ and } \text{TSF4}^+\rightarrow\text{F2}^+ \text{ are 7.6 and 8.6 kcal/mol, respectively. The formation of R2G\text{4}^+ \text{ from R2G\text{3}^+ \text{ is thermodynamically more favourable than the formation of R2G\text{2}^+ \text{ from R2G\text{3}^+ \text{ as the free energy is lower by 2.5 kcal/mol. However, the formation of R2G\text{3}^+ \text{ from R2G\text{4}^+ \text{ is kinetically less favourable than the formation of R2G\text{2}^+ \text{ from R2G\text{3}^+ \text{ as the free energy transition state for the formation of R2G\text{4}^+ \text{ is higher by 1.0 kcal/mol. Therefore, both channels are plausibly competitive.}
R1G^\text{6} and R2G^\text{4} can also be interconverted through the pathway R1G^\text{6} \rightarrow R1G^\text{5} \rightarrow R2G^\text{4}. This firstly involves a hydrogen cation transfer from the "6" position to the "5" position, which has a free energy barrier of only 3.2 kcal/mol. The C5–C6 bond is rotated and the concurrent hydrogen cation transfer from R1G^\text{6} \rightarrow R2G^\text{4} proceeds through a transition state (TS\text{16,15}) of 2.9 kcal/mol free energy barrier. This reaction is also exothermic by -3.7 kcal/mol, and thus kinetically and thermodynamically favoured. Another possible pathway from R1G^\text{6} to R2G^\text{4} is through the initial hydrogen cation transfer of R1G^\text{6} \rightarrow R1G^\text{5} and then subsequent hydrogen cation transfers and C5–C6 bond rotation. However due to the large initial free energy barrier (27.8 kcal/mol) for the initial hydrogen cation transfer, this pathway may be neglected. It is important to note that the rotational barrier is comparable to even the lowest hydrogen cation isomerization barrier, thus these channels are significant and warrant consideration in the developing microkinetic models of the acid hydrolysis process.

To summarise, it is found that the "4" position of Rotamer-2 is the most stable protonated isomer of all D-glucopyranose conformers. When other sites are protonated, hydrogen cation transfer to the "4" position of Rotamer-2 requires very low free energy barriers (2.9 – 8.6 kcal/mol) for both Rotamer-1 and Rotamer-2. Thus, R2G^\text{4} is expected to be found in comparatively higher populations than other protonated species in aqueous media and thus available to decompose to valuable platform chemicals. Note, as there are eleven different protonated anomers, this system is very complicated. Therefore, it is not feasible to include all of these decomposition processes in one manuscript. Instead, the decomposition of the most thermodynamically stable isomer R2G^\text{4} is studied. Currently, in the existing literature the decomposition of R1G^\text{5}_\text{ex} is the only proposed reaction pathway to form the experimentally observed HMF. As, the interconversion between Rotamer-1 and Rotamer-2 is a facile process, the decomposition of R2G^\text{4} is also examined in this work as a potential additional pathway to form HMF.

3.4 Decomposition of Rotamer-2 of β-D-glucopyranose at the "4" protonation site

As discussed in Section 3.2, the OH at the "4" position of Rotamer-2 is the most thermodynamically stable and kinetically accessible protonation site of β-D-glucopyranose. The complete reaction mechanism for the decomposition pathways from this site have never been previously explored. This computed potential energy surface is presented in Figure 6. Furfuryl alcohol is proposed as the thermodynamically favourable decomposition product, following protonation of the "4" position of Rotamer-2 of β-D-glucopyranose. The protonated C4–OH group is encountered by C2 crossing a free energy barrier of 31.6 kcal/mol. This yields a five-membered ring intermediate (II) with free energy of +14.6 kcal/mol. Qian et al. \textsuperscript{16} studied the decomposition reaction pathway of the β-D-glucopyranose by Car-Parrinello molecular dynamic simulations in the gas phase at 500K. They observed that the reaction initiated from the protonation of OH at "4" position leads to a five-member ring intermediate on the time scale of <1 ps, which may further polymerize to form the non-cellulosic materials.
However, they did not observe the formation of five-member ring intermediate in solution phase.\textsuperscript{15}  Although it can undergo either (i) a hydrogen cation shift to the neighboring C2–OH position or (ii) dehydration; (i) The first pathway (I) has a free energy barrier of 24.9 kcal/mol where the hydrogen cation is shifted from the C3–OH to the C2–OH group. The C2–OH group may protonate, forming the intermediate I2, which is more stable than I1 by 4.6 kcal/mol. Subsequent elimination of water and formic acid, and deprotonation (I2 → I3 → I5) yields a very stable intermediate, 2,5-dihydro-2-hydroxy-5-(hydroxymethyl)furan (I5) with a free energy of \(-18.6\) kcal/mol. I5 could also be formed following dehydration and deprotonation of I2 and subsequent elimination of formic acid, though the free energy barrier is large at \(-32.2\) kcal/mol and can probably be neglected. Once, the stable I5 is formed, the intermediate can be protonated at the C1–OH group to form I6, with an increase in free energy of 11.3 kcal/mol. Subsequent dehydration and deprotonation of I6 results in the formation of furfuryl alcohol with a free energy change of \(-33.0\) kcal/mol, significantly more stable than the parent \(\beta\)-D-glucopyranose molecule. Assary et al.,\textsuperscript{15} also calculated the free energy of the formation of furfuryl alcohol from D-glucopyranose to have a value of \(-33.0\) kcal/mol, which is in excellent agreement with the value calculated in this work. This is the first instance a competitive pathway to furfuryl alcohol has been identified from D-glucopyranose protonated at the “4” position. Previously, the only pathway considered for furfuryl alcohol formation from D-glucopyranose is through protonation at the “2” position and subsequent decomposition from the same site.\textsuperscript{5, 11} According to the DFT calculations of Yang et al.,\textsuperscript{9} the largest free energy barrier for this pathway is 31.1 kcal/mol, for the first dehydrogenation from R1G\textsuperscript{2}. This is essentially equivalent to the largest free energy barrier for the furfuryl alcohol formation pathway proposed in this work (31.6 kcal/mol). Thus, the new pathway established here in ought to be competitive to the previously known pathway, but the starting protonated species is suggested to be more stable.

(ii) Figure 7 shows the pathways following the dehydrogenation of I1 (pathway ii). After dehydrogenation and deprotonation of I1, the stable intermediate tetrahydro-2,3-dihydroxy-5-(hydroxymethyl)furan-4-carbaldehyde (I9) is formed, which is more stable than \(\beta\)-D-glucopyranose by 8.6 kcal/mol. Protonation at the C2–OH group of I9, dehydrogenation and deprotonation results in another stable intermediate, 2,5-dihydro-2-hydroxy-5-(hydroxymethyl)furan-4-carbaldehyde (I12), having a free energy of \(-18.8\) kcal/mol. The two intermediates, I9 and I12, have never been reported either experimentally or theoretically. Again, the protonation of the C1–OH group of I12 and subsequent dehydrogenation and deprotonation yields 5-(hydroxymethyl)furan-4-carbaldehyde (SHMF4) as a final product, with a free energy of \(-33.7\) kcal/mol. The intermediate I1 is a bifurcation point at which D-glucopyranose can follow either the pathway to the formation of furfuryl alcohol or the pathway to the formation of SHMF4. A comparison between these two channels, indicates that both pathways are plausible and the latter process (ii, Figure 7) is kinetically more accessible than the hydrogen cation shift (i,
Figure 6), as it is a barrierless pathway. The possible formation of 5HMF4 is quite significant, as it is the most kinetically favourable pathway from 2R2G2, which is the most favourable site for protonation of D-glucopyranose. The formation of 5HMF4 has not previously been proposed. It is an isomer of 5-(hydroxymethyl)furan-2-carbaldehyde (HMF), which is a highly sought after platform chemical. The implications of this pathway will be discussed at a later stage in this work.

3.5 Decomposition of Rotamer-2 of β-D-glucopyranose at the "2" protonation site

Most literature studies consider only Rotamer-1 as initial reagent. Decomposition from Rotamer-2 protonated at the "2" position has never been examined. The free energy of Rotamer-1 and Rotamer-2 of β-D-glucopyranose protonated at the "2" position are almost equivalent, 21.6 and 21.0 kcal/mol, respectively. Thus, given the low rotational barrier, protonated species of both rotamers are equally likely to exist and are thus important to consider. The decomposition pathways initiated by protonation at the "2" position of Rotamer-2 via cyclic species are presented in Figures 8 and 9. The reaction pathways result in the formation of 5-(hydroxymethyl)furan-2-carbaldehyde (HMF) or 5-(hydroxymethyl)furan-3-carbaldehyde (5HMF3), respectively.

Following protonation of the "2" position, an isomerization process may occur, yielding a five-membered ring intermediate, 115, through a free energy barrier of 20.2 kcal/mol where a new bond (C2–O6) forms by breaking the C1–O6 bond. This reaction is slightly exothermic (0.5 kcal/mol), eliminating one molecule of water and also deprotonating, yielding the intermediate species, tetrahydro-3,4-dihydroxy-5-(hydroxymethyl)furan-2-carbaldehyde (116) (see Figure 8). Qian et al.26 studied the decomposition reaction pathway of β-D-glucose by Car-Parrinello molecular dynamic simulations in the gas phase at 500K. Through in the gas phase, they found that that the reaction initiated from the protonation of OH at "2" position leads to a 2,5-anhydride intermediate which shown to be a precursor for 5-(hydroxymethyl)furan-2-carbaldehyde (HMF). They also observed similar reaction mechanism in solution phase where the carbon cation arrangement occurs rapidly after the protonated –OH group was separated from the carbon ring.35 Subsequent protonation of 116, elimination of one water molecule and deprotonation from 118 results in a molecule intermediate, 4,5-dihydro-5-(hydroxymethyl)furan-2-carbaldehyde (119), which is 15.5 kcal/mol more stable than 116. Importantly, these two important intermediates are also hypothesised without evidence by Robyt et al.49 in the conversion of D-glucopyranose to HMF via the protonation of the ring oxygen atom through the acyclic pathway. The final product HMF is formed from (119) after protonation, elimination of one water molecule and deprotonation. The overall process is highly exothermic and kinetically accessible as HMF is more stable than the reactants, β-D-glucopyranose and a hydrogen cation by −33.7 kcal/mol. From a kinetic point of view, if the protonated β-D-glucopyranose overcomes the initial free energy barrier (41.2 kcal/mol), the formation of HMF will readily occur. This is an important additional mechanistic explanation for HMF production from D-glucopyranose to the
Scheme 2. The approximate model for the reaction mechanism of β-D-glucopyranose acid hydrolysis in water at 298.15 K as recommended by the detailed analysis of; isomerization, protonation, hydrogen cation transfer and decomposition as detailed in this work. Free energies are calculated with G4MP2. The water dielectric medium is calculated at B3LYP/6-31G(2df,p) using the SMD solvation model. The reaction pathways outlined by green and black arrows (highlighted in blue fill) are from this study. Green and blue arrows designate pathways reported in the literature. Green highlight designates pathways apparent from experimental studies in the literature. Free energies are in kcal/mol. ΔG¢ denotes the free energy barrier. *From reference 11 and †from reference 5.

Previously suggested pathways from both computational and experimental empirical kinetic studies. 5, 11, 12, 15

Another decomposition pathway for D-glucopyranose protonated at the "2" position is presented in Figure 9, leading to the formation of 5-(hydroxymethyl)furan-3-carbaldehyde (5HMF3) through an initial free energy barrier of 28.6 kcal/mol, only 8.4 kcal/mol higher than the initial free energy barrier of the pathway to the expected 5-(hydroxymethyl)furan-2-carbaldehyde (HMF) isomer. The reaction proceeds through the transition state T56, where the C3–C4 bond cleaves and the C2–C4 bond forms to yield an intermediate 122. After elimination of one molecule of water and deprotonation, 122 forms a stable intermediate tetrahydro-2,4-dihydroxy-5-(hydroxymethyl)furan-3-carbaldehyde (I23), which is thermodynamically favourable by 16.2 kcal/mol. Subsequent protonation, elimination of water and deprotonation of I22 yields another stable intermediate 4,5-dihydro-4-hydroxy-5-(hydroxymethyl)furan-3-carbaldehyde (I26) which is 14.4 kcal/mol more stable than the initial reactants. Finally, following protonation, elimination of water and deprotonation of I26, 5HMF3, another isomer of HMF, is formed with a free energy of ~35.1 kcal/mol. Interestingly, this product is 1.4 and 2.1 kcal/mol more stable than HMF and furfuryl alcohol (FAL), respectively, but has not been observed experimentally. Assary et al.11 also proposed the formation of 5HMF3 as a decomposition product from D-glucopyranose initiated from protonation at the "2" site. Yang et al.5 have also identified the species I22 as an intermediate in the formation of furfuryl alcohol from R1G2+. Therefore, the intermediate I22 is another bifurcation point at which D-glucopyranose dehydration can follow either the pathway to formation of 5HMF3, as discussed above, or the pathway to formation of furfuryl alcohol as proposed by Yang et al5, and corroborated here.

The free energy barriers for the formation of furfuryl alcohol, 5-(hydroxymethyl)furan-4-carbaldehyde (5HMF4) and 5-(hydroxymethyl)furan-3-carbaldehyde (5HMF3) are 48.2, 48.2 and 49.6 kcal/mol, respectively which are ~7.0 – 8.4 kcal/mol higher than the free energy barrier for the formation of 5-(hydroxymethyl)furan-2-carbaldehyde (HMF) may indicate an unfavourable process at mild conditions. However, the propensity for these reaction pathways to contribute to the overall reaction flux can only be interpreted through the construction of a detailed reaction kinetic model that considers these fundamental thermodynamic quantities.

3.6 Reaction Mechanism Overview
Scheme 2 summarises the proposed reaction mechanism derived from the computational findings above, also including the established reaction pathways of Brønsted acid-catalysed D-glucopyranose dehydration. In solution, both Rotamer-1 and Rotamer-2 may exist in similar fractions and they may readily interconvert through the C5–C6 bond rotation, which has a low free energy barrier of 3.7 kcal/mol, much lower than the free energy barriers for the hydrogen cation transfer reactions. R1G\textsuperscript{2+} and R2G\textsuperscript{2+} are the most stable protonated species of all rotamers. So, high populations of both species could be expected to form before decomposition. But, the initial decomposition from these two protonated D-glucopyranoses has not been examined before. Only decomposition from R1G\textsuperscript{2+} has been considered in previous studies. The decomposition products, furfuryl alcohol or 5-(hydroxymethyl)furan-2-carbaldehyde (HMF) are formed from R1G\textsuperscript{2+}, as suggested by Assary et al.\textsuperscript{11} and Yang et al.\textsuperscript{5} Aside from these channels, 5-(hydroxymethyl)furan-3-carbaldehyde (SHMF3) or furfuryl alcohol may also be formed from R2G\textsuperscript{2+}, as is outlined by this present study which is in analogous agreement with both Assary et al. and Yang et al. who studied the decomposition from R1 G\textsuperscript{2+}.

Alternatively, R1G\textsuperscript{4+} may convert to R2G\textsuperscript{4+} via the formation of R1G\textsuperscript{4+} by a hydrogen cation shift to the neighbouring hydroxyl group followed by the C–C rotation or via the formation of R2G\textsuperscript{4+} by the C–C rotation followed by a hydrogen cation shift to the neighbouring hydroxyl group. Finally, decomposition products of furfuryl alcohol and/or 5-(hydroxymethyl)furan-4-carbaldehyde (SHMF4) may be formed from R2G\textsuperscript{4+}. As R2G\textsuperscript{4+} is the most stable among all the protonated species, these processes are very significant, but have not apparently been previously considered in mechanistic studies of the Brønsted acid-catalysed D-glucopyranose dehydration.

At the time of writing, all literature studies appear to consider only Rotamer-1, and even then, significant discourse exists over which is the most favourable OH site for protonation. However, there is consensus that protonation and decomposition of β-D-glucopyranose at the “2” position is considered the most likely pathway to form the precursor of HMF. According to the calculations of this work, where the alternative rotamers of D-glucopyranose are considered, the most probable protonation site of β-D-glucopyranose is at the “4” position (R2G\textsuperscript{4+}), but of Rotamer-2. Decomposition of this species does not yield the experimentally observed HMF isomer. Yang et al.\textsuperscript{10} explored reaction pathways through protonation of all hydroxyl groups in D-glucopyranose identifying that protonation and decomposition at the “4” position had a free energy (ΔG\textsuperscript{protonation} = +28 kcal/mol) lower than that of the “2” position (ΔG\textsuperscript{protonation} = +35 kcal/mol). They also stated that the degradation of D-glucopyranose protonated at the “4” position could not lead to the formation of HMF. Instead, they suggest that intermediates, such as I1 and I2 (in the present study), could readily undergo reversion reactions contributing to the formation of humins, thus potentially explaining why the decomposition products, SHMF4 and furfuryl alcohol, have not been observed in experiment. The hydrogen cation transfer pathways explored in this work explain how HMF may be formed from R2G\textsuperscript{2+} despite it not being the most favourable protonation position. As the hydrogen cation transfer processes are more facile than decomposition processes, R2G\textsuperscript{4+} may be converted to R2G\textsuperscript{2+} through the hydrogen cation transfer mechanism (Figure 5) leading to the formation of HMF and/or SHMF3 by decomposition. Further computational studies of the reaction kinetics of the formation of furfuryl alcohol, HMF, SHMF3 and SHMF4 and their reactivity are warranted to further determine the role of these new product candidates in the overall acid-catalysed conversion of D-glucopyranose.

3.7 Rationalisation of computational propositions to experimental evidence

The major products of the conversion of D-glucopyranose in aqueous homogenous acid hydrolysis that are experimentally reported are; 5-(hydroxymethyl)furan-2-carbaldehyde (HMF), levulinic acid, and formic acid.\textsuperscript{1-3} Other reaction products are poorly characterized. However, the presence of other soluble and insoluble products is commonly recognised and attributed

![Infrared spectra of; (a) furfuryl alcohol isolated in an argon matrix at 14 K. (b) furfuryl alcohol computed with B3LYP/6-31G(d,p) at 298.15 K, and c) humins prepared from D-glucopyranose by van Zandvoort et al. ([0.01 M H2SO4 at 180 °C for 6 h). In the calculated spectra, theoretical wavenumbers are scaled by 0.9668 (see text).](image-url)
to a generalised condensation reaction forming a polymeric species known as "humins". Up to 35 mass % of D-glucopyranose can present as humin residues after complete conversion at 180–247 °C in acidic aqueous phase.

Sumerskii et al.\textsuperscript{10} suggested that the humins consist of a network of 60% furan rings linked with 20% aliphatic fragments formed via polycondensation. They also reported that protonating HMF at the carbonyl group could result in dehydration followed by further HMF addition through the formation of acetal and hemiacetal bonds. Patil et al.\textsuperscript{51, 52} suggested that humin formation is a sequential process where D-glucopyranose is firstly converted to HMF and consequently to 2,5-dioxo-6-hydroxyhexanal by a rehydration process. Afterwards, 2,5-dioxo-6-hydroxyhexanal may undergo aldol addition/condensation polymerization with HMF resulting in the formation of humins.

It is thus quite clear that during the acid-catalysed conversion of D-glucopyranose, HMF is the key species in the formation of humins. However, the mechanism of humin formation and growth is less understood, although several plausible mechanisms have been suggested from experimental results.\textsuperscript{4, 50-55} However, theoretical groundings of such mechanistic propositions are very limited,\textsuperscript{10} considering that the "4" position is found to be the most favourable position for initial protonation of D-glucopyranose and this is one of two rotameric forms. The derived protonated D-glucopyranose can be fairly expected to exist in appreciable population in an aqueous acidic medium. It's formation would hypothetically lead to the formation of 5-(hydroxymethyl)furan-4-carbaldehyde (5HMF4) and it may contribute humins formation. The availability of 5HMF4 would produce a more varied arrangement of furanic functionalities in humins that are in addition to the features offered by HMF. However, 5-(hydroxymethyl)furan-3-carbaldehyde (5HMF3) and 5-(hydroxymethyl)furan-4-carbaldehyde (5HMF4), as apparent from the theoretical calculations herein, have not been observed in experimental studies in the literature. The prospect of these propositions from theory as being constituents of humins is now discussed.

Very little information is available regarding the detailed chemical composition of humins because of its intrinsic complexity and recalcitrance to analytical techniques. However infrared spectroscopy is commonly applied. To leverage this characterization of humins, the vibrational frequencies of each of the proposed decomposition products are also calculated. Harmonic vibrational frequencies are determined by the analytical evaluation of second derivatives of the energy with respect to nuclear displacement using B3LYP/6-31G(2df,p). It is known that the so computed quantum mechanical harmonic vibrational frequencies are typically slightly larger than what is observed by experiment.\textsuperscript{56} However, the overestimation is relatively uniform and a good agreement to experimental can be made by applying a uniform scaling factor.\textsuperscript{57, 58} Merrick et al.\textsuperscript{59} determined that the vibrational frequencies scaling factor by comparing theoretically computed harmonic frequencies to the corresponding experimental frequencies for a set of 1066 individual vibrations using more than 100 theoretical methods.

Fig. 11 Experimental IR spectra of (a) HMF, calculated IR spectra of (b) HMF, (c) 5HMF3, and (d) 5HMF4, experimental IR spectra of (e) humins prepared from HMF (0.1 M HCl at 50 – 140 °C) by Tsilomelekis et al.\textsuperscript{53}, humins prepared from D-glucopyranose by Tsilomelekis et al.\textsuperscript{53} and van Zandvoort et al.\textsuperscript{52} (0.01 M H2SO4 at 180 °C for 6 h). In the calculated spectra, theoretical wavenumbers are scaled by 0.9668 (see text).
Their study reported that hybrid density functionals can perform well in determining the fundamental frequencies using a scaling factor of 0.9652 from B3LYP/6-31G(2df,p) computations. In the present work, the calculated harmonic vibration frequencies from B3LYP/6-31G(2df,p) are scaled to the experimental spectra of furfuryl alcohol\(^\text{45}\) isolated in an argon matrix at 14 K, as shown in Figure 10. The uniform scaling factor is determined as 0.9668 with a root-mean-square deviation of 28 cm\(^{-1}\), closely consistent to the scaling factor recommended by Merrick et al. As shown in Figure 10(a) and 10(b), the calculated spectra of furfuryl alcohol agrees quite well with the experimentally observed spectra and many peaks can be assigned to vibrations respectively associated with the furan ring and hydroxymethyl group.\(^\text{45, 60}\) In general, the calculated frequencies are found to be within ±48 cm\(^{-1}\) of experiment.

### 3.7.1 Analysis of infrared spectra of furfural alcohol and D-glucopyranose-derived humins

The furan C=\(C\) stretching vibration peak appears at 1508 and 1487 cm\(^{-1}\) in the experimental and calculated furfuryl alcohol spectrum, respectively. An experimental IR spectrum of D-glucopyranose-derived humins is also presented in Figure 10(c) and also shows a strong absorbance at ~1507 cm\(^{-1}\). The wagging peak of the furfuryl alcohol CH\(_2\) group is observed at 1377 and 1371 cm\(^{-1}\) in the experimental and calculated spectra, respectively, while this peak appears in the humins spectra at 1359 cm\(^{-1}\). The strong peak around 1156 cm\(^{-1}\) observed from humins is commonly attributed to the C–O stretching of the furan ring, this is also observed in furfuryl alcohol at 1168 and 1155 cm\(^{-1}\) in the experimental and calculated spectra, respectively. Another strong peak appears at 1019 cm\(^{-1}\) for the D-glucopyranose-derived humins which can be assigned to the C–O stretching of the hydroxymethyl group, consistent with the experimental and computational signals of furfuryl alcohol at 1018 and 1006 cm\(^{-1}\), respectively. From this analysis it may be concluded that the chemical structure of humins is consistent with that of furfuryl alcohol derived furanic functionalities. This conclusion is in concurrence with Pin et al.\(^\text{61}\) Moreover, it is demonstrated that the appropriately scaled calculated harmonic vibration frequencies from B3LYP/6-31G(2df,p) replicate the experimental measurements quite well and can allow for the analysis of humins vis-a-vis, their hypothetical chemical structure.

### 3.7.2 Analysis of infrared spectra of 5-hydroxymethylfurfural isomers at 1800–550 cm\(^{-1}\)

Figure 11 presents the calculated and experimental infrared spectra of 5-(hydroxymethyl)furan-2-carbaldehyde (HMF) in the 1800–550 cm\(^{-1}\) domain. The calculated IR spectrum of HMF has absorbance bands between 1550 cm\(^{-1}\) and 720 cm\(^{-1}\) associated with the furan ring and the hydroxymethyl group, which agrees with the experimentally observed spectrum of HMF. The only difference between the experimental and calculated spectra is observed for stretching vibrations of the C=C and C=O groups. In the experimental spectrum, the C=C
stretching vibrations (~1647 cm\(^{-1}\)) are merged with the C=O stretching vibrations (~1662 cm\(^{-1}\)). Whereas, the corresponding stretching vibrations of the spectrum derived from the DFT calculation (Figure 11b) are found at 1570 cm\(^{-1}\) and 1723 cm\(^{-1}\). It is important to highlight that the experimental spectrum is attained in the solid phase whereas the calculations are of an isolated molecule (corresponding to a low-pressure gas phase) at 298.15 K. This may contribute to the deviation of the calculated stretching vibration bands of C=C and C=O to those observed experimentally.

The computed IR spectra of 5-(hydroxymethyl)furan-3-carbaldehyde (5HMF3) and 5-(hydroxymethyl)furan-4-carbaldehyde (5HMF4) in the 1800–550 cm\(^{-1}\) region are also presented, Figure 11(c) and 11(d). All vibrational frequency bands of 5HMF3 are similar to that calculated for the “regular” HMF (5HMF2) spectra with a maximum deviation of ±25 cm\(^{-1}\). The exception to this rule is the C–H twisting and bending vibrations of the furan ring. These bands are observed in the HMF (5HMF2) spectrum at 861 and 1008 cm\(^{-1}\), respectively. The corresponding vibrational bands for 5HMF3 occur at 784 and 1110 cm\(^{-1}\), which is 77 cm\(^{-1}\) lower, and 102 cm\(^{-1}\) higher than apparent in the computed 5HMF2 spectrum.

Whereas, the vibrational frequencies of 5HMF2 and 5HMF3 are quite similar, those of 5HMF4 do present differences. These include the bands corresponding to (i) the wagging and twisting mode of the furan ring C–H, (ii) the bending vibration of CH\(_2\), and (iii) the stretching vibrations of C–O and C=O. The wagging and twisting modes of the furan ring C–H (i) are at 801 and 861 cm\(^{-1}\) in 5HMF2, whereas they occur at 714 and 812 cm\(^{-1}\) in 5HMF4. The bending vibration of CH\(_2\) (ii) is calculated at 1416 and 1483 cm\(^{-1}\) in 5HMF2 and 5HMF4, respectively. The vibrations corresponding to C–O and C=O (iii) appear at wavenumbers 44 cm\(^{-1}\) higher and 33 cm\(^{-1}\) lower in the 5HMF4 calculation than in the 5HMF2 calculation.

Overall, the frequencies and intensities of the harmonic vibrations of the HMF isomers are generally similar. It could therefore be cautiously argued, that any one of the isomers is an equally plausible constituent of humins, even without considering the reaction kinetics that result in the preferential formation of one HMF over another.

### 3.7.3 Analysis of infrared spectra of 5-hydroxymethylfurfural isomers, 5-hydroxymethylfurfural-derived humins and D-glucopyranose-derived humins at 1800–550 cm\(^{-1}\)

Experimental IR spectra of both HMF- and D-glucopyranose-derived humins\(^{55, 62}\) are also presented in Figure 11(e) and clearly show major similarities.\(^{55, 55, 63}\) The main difference among them is due to the ~1670 cm\(^{-1}\) band assigned to the C=O stretching of the aldehyde group. This band is only observed in HMF-derived humins. The absorbance at ~1668 cm\(^{-1}\) corresponds to an aldehyde group similar to that in the HMF isomers. The C=O stretching bands of 5HMF2, 5HMF3 and 5HMF4 are found in a short frequency range of; 1724, 1733, and 1725 cm\(^{-1}\), respectively, all closely consistent with the major absorbance in humins at ~1720 cm\(^{-1}\).

In the scenario where 5HMF3 and/or 5HMF4 are formed from D-glucopyranose, it would be fair to assume that humins may be derived from the aldol condensation between intermediate compounds such as 2,5-dioxo-6-hydroxyhexanal and 5HMF3/5HMF4, in addition to their reaction with 5HMF2, as commonly proposed. The presence of carbonyl groups in different positions of the furan ring in 5HMF3 and 5HMF4 could also result in the broad spectral envelop observed in the 1665–1760 cm\(^{-1}\) region of the humins spectra derived from D-glucopyranose. However, it is difficult to be conclusive.

### 3.7.4 Analysis of infrared spectra of 5-hydroxymethylfurfural isomers, 5-hydroxymethylfurfural-derived humins and D-glucopyranose-derived humins at 4000–2000 cm\(^{-1}\)

Figure 12 performs the same comparison across the IR spectra range of 4000–2000 cm\(^{-1}\).\(^{52, 62}\) The experimental spectra are aqueous phase, while the computed spectra correspond to an isolated molecule. The experimental O–H stretching band of 5HMF2 is observed between 3400–3300 cm\(^{-1}\), while the corresponding calculations show vibration at 3683 cm\(^{-1}\) for 5HMF2, an offset of ~300 cm\(^{-1}\). 5HMF3 also shows this same vibration at 3683 cm\(^{-1}\) while 5HMF4 shows vibration at 3495 cm\(^{-1}\).

The same O–H stretching bands of the D-glucopyranose and HMF-derived humins occur at 3600–3200 and 3600–3000 cm\(^{-1}\), respectively. This signal is shifted to a lower wavenumber in the HMF-derived humins than that in the D-glucopyranose-derived humins. Notably, the O–H stretching vibration of 5HMF4 is computed at a wavenumber closely consistent to the major IR absorbance observed in both HMF- and D-glucopyranose-derived humins, where no vibration is computed for either 5HMF2 or 5HMF3. From the calculations of the vibrational modes of each of the HMF isomers suggested in this study, there is no straight forward explanation for the broadened absorbance band (3600–3000 cm\(^{-1}\)) of D-glucopyranose-derived humins that is entirely consistent with the O–H stretching of HMF like functionalities. However, of each of the 5HMF2, 5HMF3 and 5HMF4 isomers, the band is most consistent with the presence of the 5HMF4 isomer.

### 4 Conclusions

A detailed systematic theoretical study of the mechanism of acid hydrolysis of β-D-glucopyranose in the aqueous solution phase is performed. Free energy minima of β-D-glucopyranose and of the reaction of β-D-glucopyranose with a hydrogen cation are computed with G4MP2 with the water solution phase incorporated by the SMD solvation model computed with B3LYP/6-31G(2df,p). Specifically, free energies in the aqueous solution phase are computed for;

(i) α/β-D-glucopyranose, (ii) rotation of the β-D-glucopyranose alkyl carbon-carbon bond, (iii) α/β-D-glucopyranose protonated at the ether and each OH position, (iv) transition state geometries for the interconversion of each protonated β-D-glucopyranose rotamer, and (v) the potential energy surface

This journal is © The Royal Society of Chemistry 20xx

Phys. Chem. Chem. Phys., 2018, 00, 1-3 | 17
for the decomposition of the most thermodynamically and kinetically accessible protonated \(\beta\)-D-glucopyranosyl isomers.

Notably, it is found that rotation of the \(\beta\)-D-glucopyranosyl alkyl carbon-carbon bond is a facile process that is dictating to the resulting reaction mechanism. This rotation serves as a gateway to two rotameric \(\alpha/\beta\)-D-glucopyranosyl isomers, which are shown to produce a range of cations that are of notably different thermodynamic stability when protonated. In particular, as a low energy process \((\Delta G^* \approx 3.8 - 6.7 \text{ kcal/mol})\), the alkyl carbon-carbon bond may rotate toward the hydroxyl group at the adjacent "4" position. The free energy required to protonate this position is consequently reduced by 3.0 or 7.2 kcal/mol for the \(\alpha\)- or \(\beta\)-D-glucopyranosyl anomers respectively. These finding are in contrast to previous literature which suggests that the favoured protonation position is at the "1" or "2" positions.

It is further shown that this simple rotation can consequently lead to complicated hydrogen cation transfer and decomposition reaction pathways which may potentially yield a number of platform chemicals that have not previously been suggested. In particular, the computed potential energy surfaces suggest the prospect of three isomers of 5-hydroxymethylfurfural (HMF). That is, in addition to the expected HMF isomer, 5-(hydroxymethyl)furane-2-carbaldehyde, the calculations show that 5-(hydroxymethyl)furane-3-carbaldehyde and 5-(hydroxymethyl)furane-4-carbaldehyde are also possible products from the acid hydrolysis of \(\beta\)-D-glucopyranose in the aqueous solution phase. The computed vibrational frequencies of these HMF isomers are compared to experimentally determined infrared spectra of "humins". On this basis, it is cautiously speculated that the alternative HMF isomers, may be monomeric constituent of the polymeric "humins".

Acknowledgements
This work is supported by Science Foundation of Ireland under 13/SIRG/2185 and 16/ERCD/3685. Computation resources are provided by the Irish Centre for High-End Computing, ICHEC. Both funding sources are gratefully acknowledged.

Notes and references