H-Bond Isomerization in Crystalline Cellulose III_I: Proton Hopping versus Hydroxyl Flip-Flop

Pan Chen,[*,†,¶] Mateusz Marianski,[*,‡,¶] and Carsten Baldauf[*,‡]

†*Aachener Verfahrenstechnik & AICES Graduate School, RWTH Aachen University, D-52062 Aachen, Germany*

‡*Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-14195 Berlin, Germany*

¶*Both authors contributed equally.*

E-mail: pan.chen@avt.rwth-aachen.de; marianski@fhi-berlin.mpg.de; baldauf@fhi-berlin.mpg.de

Abstract

Based on density-functional theory calculations, we discuss three forms of cellulose III_I that are characterized by different inter-sheet H-bonding patterns. Two alternative mechanisms can facilitate the interconversion between these H-bonding patterns: the rotation of hydroxy groups ("flip-flop") or a concerted proton transfer from one hydroxy group to the other ("proton hopping"). Both mechanisms have energy barriers of very similar height. Electronic structure theory methods allow us to study effects that involve the breaking/forming bonds, like the hopping of protons. In many of the force field formulations, in particular the ones that are typically used to study cellulose, such effects are not considered. However, such insight at the atomistic and electronic scale can be key to find energy-efficient means for cellulose deconstruction.

Cellulose is the richest source of biomass on Earth and abundant enough to eventually replace the limited resources of fossil fuels.[1–3] However the usage of cellulose to generate basic chemicals for chemical industry or energy conversion is hindered by its rather inert nature.[1,4] In order to facilitate energy-efficient cellulose deconstruction, the understanding of the crystalline structure of cellulose is indispensable.

The first structure study on native cellulose can be traced back to 1913,[5] just one year after the discovery of the X-ray diffraction technique by Laue. 100 years of exploration with various experimental techniques slowly unraveled the peculiarities of cellulose structure.[6,7] A hallmark in the field was the discovery of two distinct crystalline forms (α and β) of cellulose I by Atalla and VanderHart[8] using solid state NMR measurements. Especially the combination of synchrotron X-ray and neutron diffraction on highly crystalline samples[9–13] has provided high-resolution structure data, but the reconstruction of the complete hydrogen bonding networks in cellulose crystals is still a challenge that relies on the interpretation of the experimental data by modeling and simulation.[14]

Cellulose is a linear polymer of D-glucose subunits connected by $\beta(1\rightarrow4)$ linkages. Besides the I_α and I_β forms, other crystal allomorphs are known. All these allomorphs differ in relative chain direction, chain stacking, and unit-cell parameters as well as in the orientation of hydroxy- and hydroxymethyl groups and the resulting hydrogen-bonding networks. The hydrogen-bonding networks in cellulose I_α and I_β form two-dimensional sheets, which in turn are held together by van der Waals forces. Alternatively, in cellulose III_I the three-dimensional hydrogen-bonding network spans multiple sheets, effectively rendering them more cooperative than those in two-dimensional Cellulose I_α and I_β (Figure 1).[15]

Recently, Chen and colleagues[16] have proposed an alternative hydrogen bonding pattern for cellulose III_I that differs from the experimentally-derived structure in the orientation of two polar hydrogen atoms that facilitate inter-chain hydrogen bonds.[17] Hints for a greater stability of the alternative hydrogen pattern come from force field molecular dynamics and density-functional theory (DFT) calculations. In order to set the two alternative hydrogen

bonding structures of cellulose III_I into relation to each other and to investigate possible interconversion mechanisms between both, we present here a first-principles study of crystalline cellulose III_I.

The initial structures of cellulose III_I with the two distinct hydrogen-bonding patterns *A* and *B* were taken from a previous study by one of us.[16] The monoclinic unit cell of cellulose III_I contains two β-D-glucose subunits with two hydrogen bonding chains perpendicular to the cellulose chain and oriented antiparallel to each other (Figure 1). In pattern *A* the OH2 hydroxy group is oriented *trans* to the hydrogen at C2 carbon, whereas in pattern *B* the group exists in *cis* orientation. The OH6 group at the methyl carbon is arranged accordingly. Moreover, transition-path calculations identified a stable intermediate structure, in which two H-bonding chains are arranged parallel to each other. This newly identified structure (called pattern *P*) has been included in our analysis.

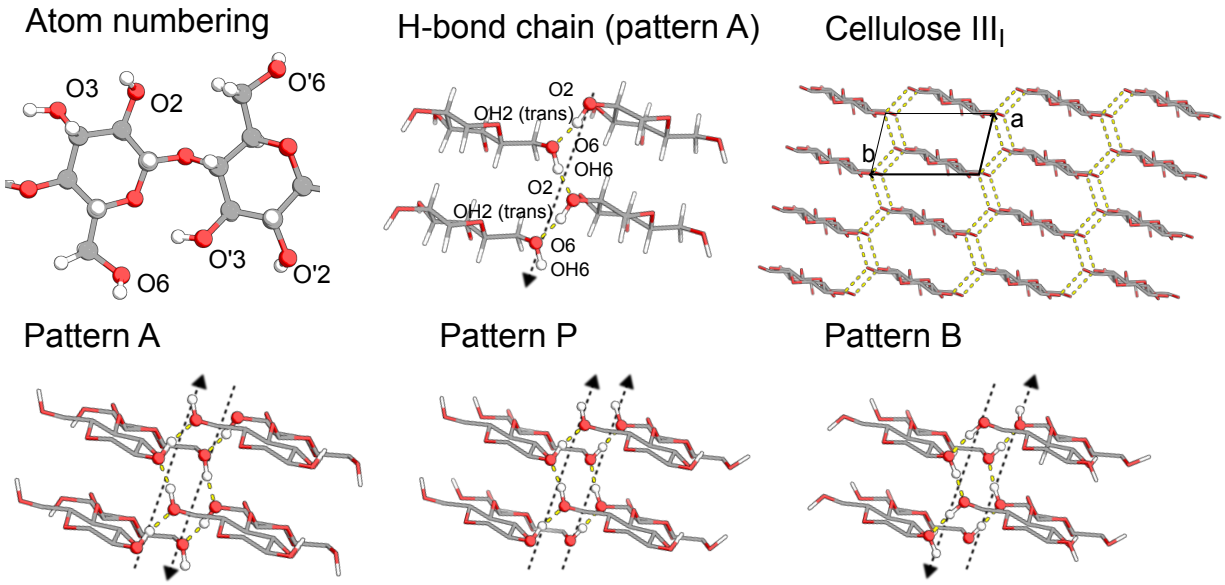

Figure 1: Top: the atom numbering, example of the H-bonding chain and the crystal packing of the cellulose III_I. Bottom: three distinct hydrogen boding patterns studied in this work

We rely on the first principles of DFT to study the energetics of and the transitions between the different H-bonding patterns of cellulose III_I. All geometry optimizations were carried out using the generalized gradient approximation (GGA) functional PBE[18] aug-

mented with the pairwise Tkatchenko-Scheffler[19] (vdW) correction for long-range van der Waals dispersion. Geometry and unit-cell optimizations were performed with `k_grid` set to $6 \times 6 \times 6$, `light` convergence settings, and a `tier-1` basis as implemented in the all-electron numeric atom-centered orbitals code FHI-aims.[20] The energetics of the optimized structures were further refined with the hybrid functional PBE0[21] augmented by a many-body dispersion scheme (MBD)[22] and with `tight` convergence criteria and a `tier-2` basis. In addition to PBE and PBE0, we compare the relative energies of local minima for different density-functional approximations, the GGA functionals BLYP[23,24] and revPBE[25] as well as for the hybrids HSE06[26] and B3LYP.[27] Finally, we investigate the impact of phonon vibrations. Calculations in the harmonic approximation were performed with $2 \times 2 \times 2$ super cells with the codes `phonopy`[28] and FHI-aims[20] (PBE+vdW with `tier-1` basis).

The relative energies of different isoforms calculated at different levels of theory are presented in Table 1. All tested functionals consistently predict pattern *B* to be energetically more stable than the experimentally determined pattern *A*.[16] The GGA functionals render pattern *B* from 0.9 (revPBE and BLYP) to 1.9 kcal/mol (PBE) per cellobiose subunit more stable than pattern *A*. The contribution from exact exchange in hybrid functionals increases the difference to 2.3 (PBE0 and HSE06) and 2.7 kcal/mol (B3LYP). Neither the refinement with a denser integration grid and tighter convergence criteria ("tight settings") nor geometry and unit-cell relaxations affect the relative energy of both structures. Furthermore, the many-body dispersion treatment does not change the relative energetics of two structures compared to the pairwise treatment of long-range van der Waals interactions. However, considering harmonic vibrations influences the relative stability of the allomorphs noticeably by decreasing the gap between patterns *A* and *B* by 0.4 kcal/mol (see Table 1).

Patterns *A* and *B* share the common feature of an energetically favorable anti-parallel arrangement of dipole moments associated with H-bonded hydroxy groups. In principle, the parallel arrangement of H-bonded chains in pattern *P* should lead to a destabilization of the crystal. However, the relative energy of the pattern *P* is predicted to lie exactly between

Table 1: Relative energies and harmonic free energies [in kcal/mol] of different HB patterns in cellulose III_I crystals. The numbers refer to pattern B lattice parameters and PBE+vdW optimized geometries while the numbers in parentheses refer to the geometries relaxed at the PBE+vdW unit cell parameters for respective hydrogen bonds arrangements. For calculating the free energies PBE+vdW optimized minimum structures were used.

Functional	Dispersion	light settings			tight settings		
	correction	A	P	B	A	P	B
PBE	vdW	1.9 (1.8)	1.0 (0.9)	0.0 (0.0)	1.8	0.9	0.0
BLYP	vdW	0.9 (0.8)	0.5 (0.5)	0.0 (0.0)	0.8	0.4	0.0
PBE0	vdW	2.3 (1.9)	1.2 (1.0)	0.0 (0.0)			
B3LYP	vdW	2.7 (2.6)	1.4 (1.4)	0.0 (0.0)			
revPBE	vdW	1.0 (0.9)	0.5 (0.6)	0.0 (0.0)			
HSE06	vdW	2.3 (1.9)	1.2 (1.0)	0.0 (0.0)			
PBE	MBD	1.7	0.9	0.0			
PBE0	MBD	2.2	1.1	0.0			
F_{harm}(300K), $2 \times 2 \times 2$ unit cell							
PBE	vdW	1.5	0.9	0.00			

pattern A and B (see table 1). This energetic midpoint suggests that the interaction between dipoles is insignificant and that the energetics of the crystal are mainly governed by the local arrangement of the hydroxy groups. To confirm that intrachain interactions are responsible for the stabilization of pattern B, we calculated single-point energies (PBE+vdW, tier-1) of an isolated cellulose chain fixed in either pattern A or B geometries. The predicted energy difference of 2.1 kcal/mol is close to the 1.9 kcal/mol that are observed in the crystal and agrees well with the proposed intrachain stabilization of pattern B.

Formally, we see two mechanisms for a concerted rearrangement of hydrogen bonds from pattern A to pattern B and *vice versa*: proton hopping and H-bond flip-flop.[29,30] The former mechanism occurs by shifting protons from the O6 and O2 to respectively O2 and O6 in the adjacent unit cell (see Figure 2). The latter mechanism features the rotation around the C2-O2 and C6-O6 bonds (see Figure 3). Moreover, the single-bond rotation can occur clockwise or anti-clockwise rendering 4 unique transitions in total. Our results showed that the rearrangement of two H bonding chains occurs sequentially through the P intermediate. Therefore, we split the transition path into two sections: the transition from A to P and the

transition from P to B. The postulated transition mechanisms were investigated with the string method[31] as it is implemented in the `aimsChain` tool.[32] First, a set of structures along a path between initial and final structure connected with spring-like constraints is generated. The path is then locally optimized to yield a minimum-energy path connecting the terminal structures. We have used the lattice parameters of hydrogen bonding pattern B for the transition path calculations. This approximation is valid due to the negligible difference between the relaxed unit cells of pattern A and B (see tables 6 and 8 in the Supporting Information). The transition state search has been carried out at PBE+vdW level of theory and resulting paths were recalculated with the PBE0+MBD method.

Figures 2 and 3 present energy profiles of two possible conversion mechanisms between the patterns A and B connected by the intermediate pattern P. In the proton-hopping mechanism the two structures are separated by a single transition state in which the proton is shared between two adjacent O2 and O6 oxygens (inset "TS" in Figure 2). PBE+vdW predicts the $E^{TS}_{A\rightarrow P}$ and $E^{TS}_{P\rightarrow B}$ transition state energy to be 3.6 and 3.5 kcal/mol per hydrogen bond respectively. The path refinement at the PBE+MBD level of theory somewhat decreases the energy of both transition states by 0.1 kcal/mol. However, it is well known that the GGA functionals suffer from overdelocalization of the electron density due to a self-interaction error.[33] The error tends to strongly affect energetics of transition states that involve elongated bonds.[34,35] In essence, the GGA functionals predict too low barriers of proton-transfer reactions. The incorporation of the Hartree-Fock exchange in the PBE0 functional partly corrects for the self-interaction error and improves the energetics of the transition states.[36] Moreover, augmenting PBE0 with a long-range dispersion correction further amends the resulting barrier heights.[36] Consequently, we refined the proton-hopping mechanism transition path at the PBE0+MBD level of theory, which significantly increased the energy of the transition states to 5.5 and 5.4 kcal/mol per H bond.

In the already described proton-hopping mechanism, the protons shift along the H-bonding chain to the adjacent unit cells. Contrary, in the H-bond flip-flop mechanism the

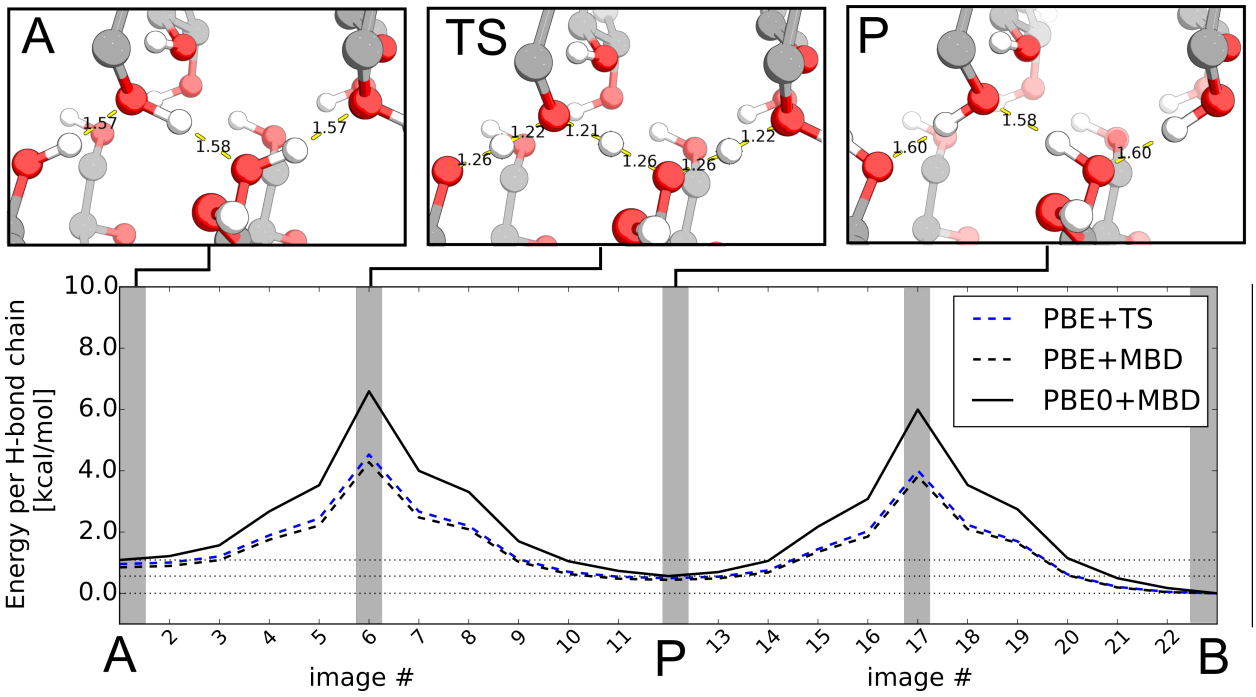

Figure 2: The proton hopping transition path between patterns A and B.

hydrogen rotates around the oxygen atom and remains in the initial unit cell. The H-bond rotation in the zigzag H-bonding chain can occur via smaller (140°) or larger (220°) dihedral angle at two different carbon atoms, resulting in total in four unique paths. Here we report only the rotation with the lowest energy path in which both rotations occur via the smaller angle. The energy profiles of the remaining three possibilities can be found in tables 3 to 5 of the Supporting Information.

The flip-flop transition between patterns happens in a stepwise mechanism with a higher-energy minimum along the path (see Figure 3). First, we observe OH6 rotation toward the proximate O3 by which a weak H bond (1.98 Å) is formed. This new H bond compensates for the broken OH2 $\cdots$ O6 H-bond during the following 140° rotation of the OH2 group. In effect, the OH2 reconstitutes the H-bond with the opposite O6 forming an intermediate structure (I) with two H-bonds, OH6 $\cdots$ O3 and OH2 $\cdots$ O6. The intermediate is a stable minimum along the path. In the final step, the OH6 continues the rotation to form the H-bond with

the opposite O2. In the course of the rearrangement, also the OC-CO6 torsion undergoes a minor rearrangement. The observed minimum decreases the barrier when compared to other paths that lack the intermediate structures (5.1 *vs.* 7.2, 7.4 and 9.0 kcal/mol at PBE0+MBD level of theory).

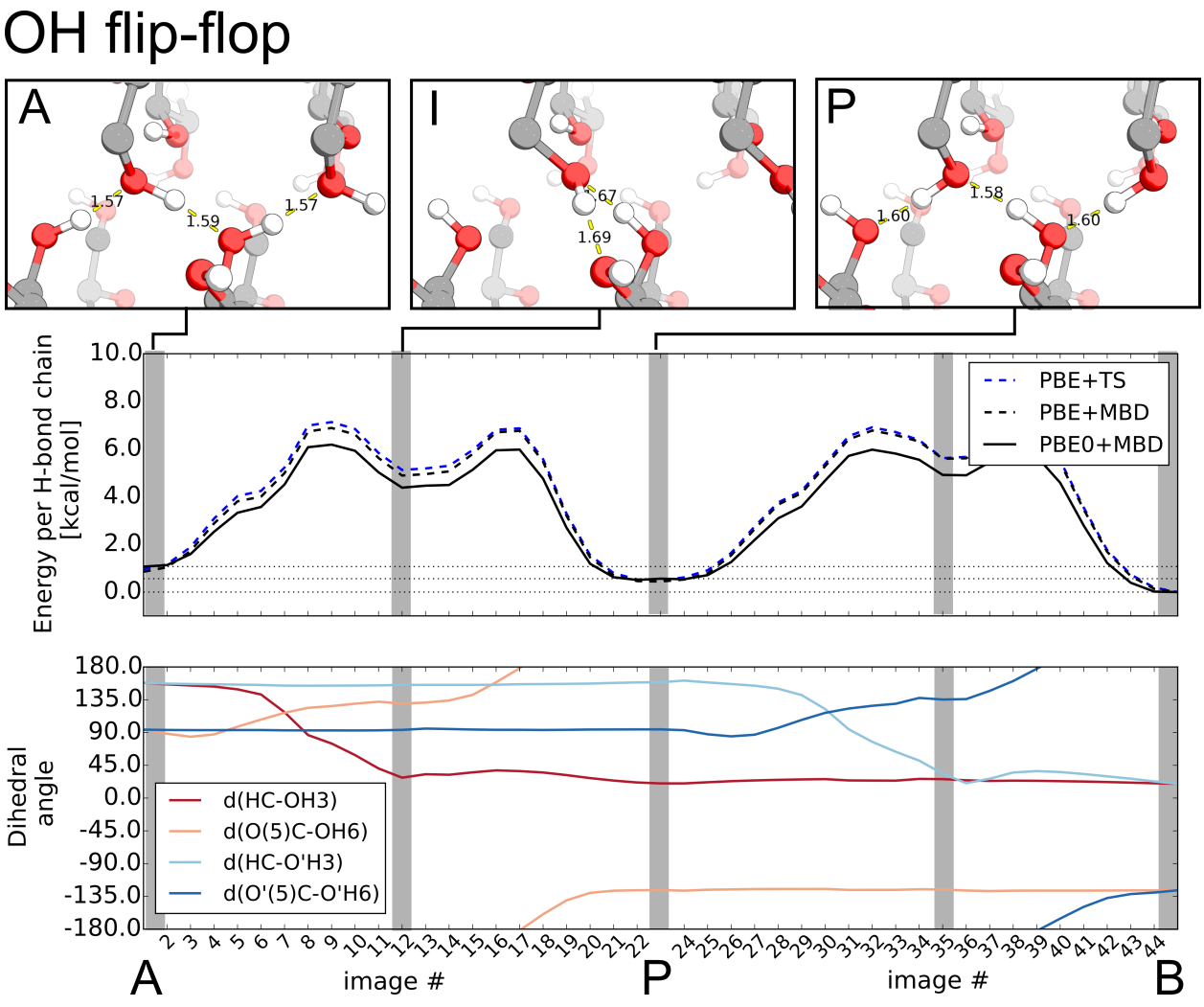

Figure 3: The H-bond flip-flop transition path between patterns A and B.

The $E^{TS}_{A\to I}$ and $E^{TS}_{I\to P}$ transitions are equal to 6.2 and 1.7 kcal/mol per H-bond, respectively, at the PBE+vdW level of theory. The consideration of many-body dispersion slightly stabilizes the intermediate state. Finally, the exact exchange in PBE0+MBD decreases the observed transition barriers to 5.1 and 1.6 kcal/mol per H-bond. The $P\to B$ part of the path follows the same mechanism and observed transition states equal 5.4 and 1.0 kcal/mol per

H bond at PBE0+MBD level of theory.

In the present work we show, by means of electronic structure theory methods, that in cellulose III_I: (1) the H-bonding pattern B is energetically more favorable than H-bonding pattern A, which is contradicting the current interpretation of the experimentally-derived diffraction pattern, and (2) the conversion between pattern A and B can occur via two competing mechanisms: proton hopping and H-bond flip-flop. The involved barriers differ only by a few tenths of a kcal/mol per cellobiose subunit. In case of such small differences, other effects like, for example, anharmonic conformational contributions to entropy[37] or quantum effects of the nuclei[38] can become decisive. However, we do not think that both mechanisms, proton hopping and H-bond flip-flop, are mutually exclusive. Instead, we speculate about a possible cooperative mechanism of alternating proton hops and H-bond flip-flops that allows for a movement of protons along the H-bond chain in crystalline cellulose III_I. Such mechanism would explain the possibility to generate deuterated cellulose III_I for neutron diffraction experiments.[12,39] There an exchange of protons must occur between the hydroxy groups of cellulose III_I and deuterated ammonia. However, whether or not proton transfer would also occur inside the inner crystal of cellulose III_I is unknown so far. Distinct H-bond patterns occur also in other forms of cellulose; neutron-diffraction experiments on cellulose I_β at ultra-low temperature indicate different H-bonding patterns in different regions of the fibril.[14] In detail, the one that is observed statistically more often is assigned to the crystalline core, while the other originates from the surface or crystal defects. Such hydrogen-bond disorders might occur in other forms of cellulose as well, often being hinted on by a discrepancy between theoretical and experimental results. We intend to investigate further possible conversion mechanisms in different cellulose allomorphs in the future.

Finally, we note that due to the intrinsic inability of standard force fields to describe bond formation and breaking, the proton-transfer mechanism lies beyond their capabilities. On the contrary, *ab initio* methods do not suffer from this restriction. Furthermore, the accuracy level of electronic structure theory can be adjusted in a systematic way in order to

match requirements of a particular system as it was demonstrated for the relative energetics of allomorphs of organic crystals.[37,40]

Acknowledgement

This work was performed as part of the Cluster of Excellence "Tailor-Made Fuels from Biomass", which is funded by the Excellence Initiative by the German federal and state governments to promote science and research at German universities.

Supporting Information Available

The Supporting Information contains methodological details, tables with energies computed at different levels of theory, and Cartesian coordinates and crystal parameters of cellulose III_I allomorphs *A*, *P*, and *B*. The output of the transition path calculations has been deposited in the NOMAD repository (`http://nomad-repository.eu`, DOI: 10.17172/NOMAD/2015.11.19-1).

This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Himmel, M. E.; Ding, S.-Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W.; Foust, T. D. *Science* **2007**, *315*, 804–807.

(2) Mascal, M.; Nikitin, E. B. *Angew. Chem. Int. Ed.* **2008**, *47*, 7924–7926.

(3) Bornscheuer, U.; Buchholz, K.; Seibel, J. *Angew. Chem. Int. Ed.* **2014**, *53*, 10876–10893.

(4) Chundawat, S. P. S.; Bellesia, G.; Uppugundla, N.; da Costa Sousa, L.; Gao, D.;

Cheh, A. M.; Agarwal, U. P.; Bianchetti, C. M.; Phillips, G. N.; Langan, P.; Balan, V.; Gnanakaran, S.; Dale, B. E. *J. Am. Chem. Soc.* **2011**, *133*, 11163–11174.

(5) Nishikawa, S.; Ono, S. *Proc. Math. Phys. Soc.Tokyo* **1913**, *7*, 131.

(6) French, A. D. In *Advances in Carbohydrate Chemistry and Biochemistry*; Horton, D., Ed.; Academic Press, 2012; Vol. 67; pp 19–93.

(7) Chanzy, H. *Cellulose* **2011**, *18*, 853–856.

(8) Atalla, R. H.; Vanderhart, D. L. *Science* **1984**, *223*, 283–285.

(9) Langan, P.; Nishiyama, Y.; Chanzy, H. *J. Am. Chem. Soc.* **1999**, *121*, 9940–9946.

(10) Nishiyama, Y.; Langan, P.; Chanzy, H. *J. Am. Chem. Soc.* **2002**, *124*, 9074–9082.

(11) Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P. *J. Am. Chem. Soc.* **2003**, *125*, 14300–14306.

(12) Wada, M.; Chanzy, H.; Nishiyama, Y.; Langan, P. *Macromolecules* **2004**, *37*, 8548–8555.

(13) Langan, P.; Nishiyama, Y.; Chanzy, H. *Biomacromolecules* **2001**, *2*, 410–416.

(14) Nishiyama, Y.; Johnson, G. P.; French, A. D.; Forsyth, V. T.; Langan, P. *Biomacromolecules* **2008**, *9*, 3133–3140.

(15) Parthasarathi, R.; Bellesia, G.; Chundawat, S. P. S.; Dale, B. E.; Langan, P.; Gnanakaran, S. *J. Phys. Chem. A* **2011**, *115*, 14191–14202.

(16) Chen, P.; Ogawa, Y.; Nishiyama, Y.; Bergenstråhle-Wohlert, M.; Mazeau, K. *Cellulose* **2015**, *22*, 1485–1493.

(17) Wada, M.; Chanzy, H.; Nishiyama, Y.; Langan, P. *Macromolecules* **2004**, *37*, 8548–8555.

(18) Perdew, J.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(19) Tkatchenko, A.; Scheffler, M. *Phys. Rev. Lett.* **2009**, *102*, 73005.

(20) Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Ren, X.; Reuter, K.; Scheffler, M. *Comp. Phys. Comm.* **2009**, *180*, 2175–2196.

(21) Perdew, J. P.; Ernzerhof, M.; Burke, K. *J. Chem. Phys.* **1996**, *105*, 9982–9985.

(22) Ambrosetti, A.; Reilly, A. M.; DiStasio, R. A.; Tkatchenko, A. *J. Chem. Phys.* **2014**, *140*, 18A508.

(23) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(24) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

(25) Zhang, Y.; Yang, W. *Phys. Rev. Lett.* **1998**, *80*, 890–890.

(26) Heyd, J.; Scuseria, G. E.; Ernzerhof, M. *J. Chem. Phys.* **2003**, *118*, 8207–8215.

(27) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

(28) Togo, A.; Tanaka, I. *Scr. Mater.* **2015**, *108*, 1–5.

(29) Koehler, J. E. H.; Saenger, W.; Van Gunsteren, W. F. *Eur. Biophys. J.* **1988**, *16*, 153–168.

(30) Saenger, W.; Betzel, C.; Hingerty, B.; Brown, G. M. *Nature* **1982**, *296*, 581–583.

(31) E, W.; Ren, W.; Vanden-Eijnden, E. *J. Chem. Phys.* **2007**, *126*, 164103.

(32) Ghiringhelli, L. personal communication, manuscript in preparation.

(33) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. *Science* **2008**, *321*, 792–4.

(34) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. *Chem. Rev.* **2012**, *112*, 289–320.

(35) Peverati, R.; Truhlar, D. G. *Philos. Trans. A. Math. Phys. Eng. Sci.* **2014**, *372*, 20120476.

(36) Mangiatordi, G. F.; Bre, E.; Adamo, C. *J. Chem. Theory Comput.* **2012**, *8*, 3082–3088.

(37) Reilly, A. M.; Tkatchenko, A. *Phys. Rev. Lett.* **2014**, *113*, 1–5.

(38) Rossi, M.; Fang, W.; Michaelides, A. *J. Phys. Chem. Lett.* **2015**, 4233–4238.

(39) Wada, M.; Nishiyama, Y.; Langan, P. *Macromolecules* **2006**, *39*, 2947–2952.

(40) Reilly, A. M.; Tkatchenko, A. *J. Phys. Chem. Lett.* **2013**, *4*, 1028–1033.

Graphical TOC Entry