Interdisciplinary Sciences: Computational Life Sciences (2019) 11:485–495
https://doi.org/10.1007/s12539-018-0293-4

ORIGINAL RESEARCH ARTICLE

Quantum Calculations on Plant Cell Wall Component Interactions

Hui Yang[1] · Heath D. Watts[2] · Virgil Gibilterra[3] · T. Blake Weiss[3] · Loukas Petridis[4] · Daniel J. Cosgrove[1] · James D. Kubicki[2]

Received: 11 October 2017 / Revised: 13 March 2018 / Accepted: 21 March 2018 / Published online: 26 March 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract
Density functional theory calculations were performed to assess the relative interaction energies of plant cell wall components: cellulose, xylan, lignin and pectin. Monomeric and tetramer linear molecules were allowed to interact in four different configurations for each pair of compounds. The M05-2X exchange-correlation functional which implicitly accounts for short- and mid-range dispersion was compared against MP2 and RI-MP2 to assess the reliability of the former for modeling van der Waals forces between these PCW components. Solvation effects were examined by modeling the interactions in the gas phase, in explicit H_2O, and in polarized continuum models (PCM) of solvation. PCMs were used to represent water, methanol, and chloroform. The results predict the relative ranges of each type of interaction and when specific configurations will be strongly preferred. Structures and energies are useful as a basis for testing classical force fields and as guidance for coarse-grained models of PCWs.

Keywords Interaction energies · Quantum chemistry · Cellulose · Xylan · Lignin · Pectin

Abbreviations

DFT	Density functional theory
PCW	Plant cell wall
CMF	Cellulose microfibril
C	Cellulose constrained to the crystalline form
G	Cellulose unconstrained
X	Xylan
L	Lignin
P	Pectin

✉ James D. Kubicki
jdkubicki@utep.edu

[1] Department of Biology, The Pennsylvania State University, University Park, PA 16802, USA

[2] Department of Geological Sciences, University of Texas at El Paso, El Paso, TX 79968, USA

[3] Department of Geological Sciences, The Pennsylvania State University, University Park, PA 16802, USA

[4] Center for Molecular Biophysics, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

1 Introduction

Plant cell walls (PCWs) are found in diverse forms and serve diverse biological functions that depend on their physical and mechanical properties [1, 2]. These properties are of practical significance in the food, textile, wood products and biofuels industries and are believed to depend on the types, arrangement and covalent crosslinking of polymers within the cell wall as well as on abundant noncovalent bonding interactions between wall components. For example, primary cell walls typically consist of scaffolds of cellulose microfibrils embedded in a hydrated matrix of pectins (such as homogalacturonan), hemicelluloses (such as xyloglucan) and glycoproteins. Important noncovalent interactions include direct cellulose–cellulose junctions [3], cellulose–xyloglucan junctions [4], pectin–cellulose contacts [5–9] and physical entanglements of pectin with xyloglucan [10]. A different set of polymers (predominantly cellulose, lignin and xylan or mannan) make up the secondary cell walls of woody tissues with different kinds of noncovalent interactions [11, 12]. Water dynamics are also important for polymer interactions in both primary and secondary cell walls [7, 13].

Current understanding of these physical interactions within native cell walls is limited and consequently current PCW models are based predominantly on biochemical

characterization and microscopy, with limited insights drawn from computational approaches. Our long-term goal is to create a predictive model for PCWs that can be revised as new information is gained and modified for various types of PCWs depending on their structure. With details included from molecular to continuum scales, the model could be validated by its ability to predict experimental observations such as spectroscopic and mechanical measurements. Previously, various PCW components such as cellulose [14–17] and lignin [18, 19], as well as component interactions [20–22], have been modeled with a similar approach. This work represents a first step toward systematic modeling of component interactions by simulating pairwise interactions between cellulose, xyloglucan, lignin, and pectin components. The endless variations of possibilities make a comprehensive study impractical; however, modeling a subset of possible interactions can capture many of the basic interactions. The main goal of this study is to explore the range of interaction energies and determine which are likely to be the strongest thermodynamically.

One would not expect to be able to model a complete PCW in atomic detail, so coarse-grained and continuum models will be necessary. Molecular-level interaction energies are the basis for the mesoscale behavior; however, estimates of these parameters could be fed into larger scale models and used as a benchmark for developing complete PCW models. In the interim, classical atomistic models can extend the spatial and compositional range of PCW simulations, so we test our density functional theory (DFT) [23, 24] results against CHARMM [25] calculations on the same set of tetramers. These comparisons help assess the accuracy of the classical force field and provide targets for further refinement of force field parameters. Components of larger scale simulations can also be extracted and quantum calculations conducted to evaluate the how well the short-range energies are estimated in the large-scale models [22].

Although we do not assume that the PCW is an equilibrium assemblage of the various components contained therein, when specific interactions occur, such as xyloglucan–cellulose, the bonding energetics control the strength of these local associations. Additionally, one would not expect plants to work directly against thermodynamic favorability and that nature would look to find favorable interactions to create a strong and stable PCW.

2 Methods

This project was intended to evaluate the interactions of different plant cell wall components to help elucidate the molecular structure of the plant cell wall. The four components that were evaluated in their monomeric and tetramer forms are cellulose (C, constrained to the crystalline form found in microfibrils, or G, unconstrained), xylan (X), lignin (L), and pectin (P). Monomer and tetramer units of each of the former components were built in Material Studio 8 [26] (Fig. 1). Several types of calculations were conducted to elucidate favorable/unfavorable interactions. Tetramers of 1–4 β-linked glucose (i.e., cellotetraose) and xylose were built to represent cellulose and xylan, respectively. For lignin, coniferyl alcohol was used to represent the monomeric subunit linked with the β-*O*-4 ether bonds to form tetramers.

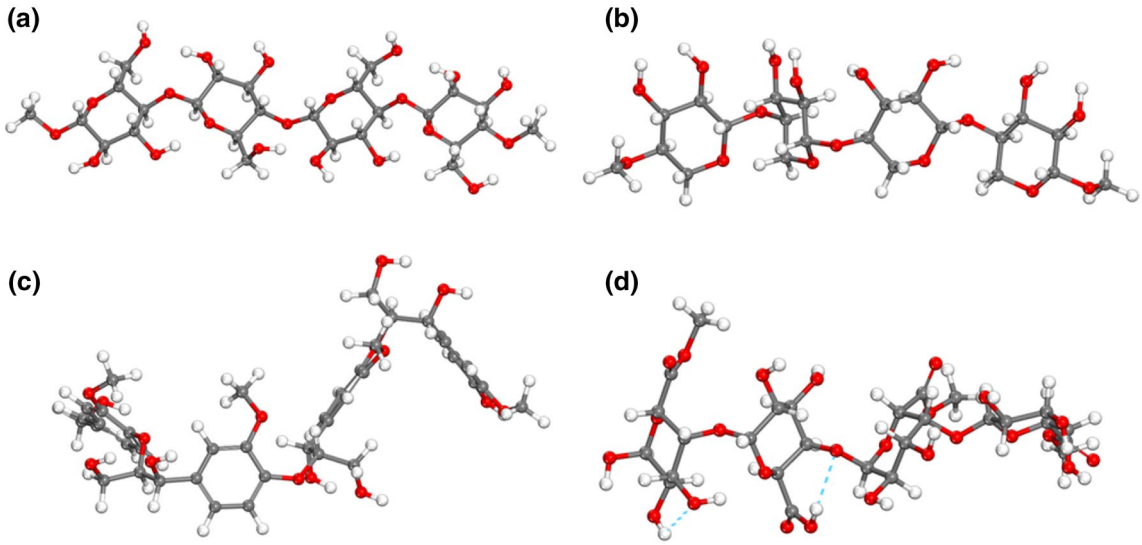


Fig. 1 Structures of the tetramers included in this study: **a** glucose tetramer–amorphous cellulose, **b** xylose tetramer–hemicellulose, **c** coniferyl alcohol tetramer–lignin, and **d** galacturonic acid tetramer–pectin. C atoms are shown in gray, O in red and H in white. Blue dashed lines represent H bonds. Methods for generating, optimizing and depicting each model are discussed in the text

Finally, tetramers of galacturonic acid, where three of the four COOH groups were methyl esterified, were used to represent pectin (Fig. 1).

Interactions between monomers and tetramers were arranged into four different categories (stacked, stacked180, side-by-side, and side-by-side180—Fig. 2). To arrange the interactions appropriately, one molecule in the interaction was the reference point, while the other molecule was positioned to create the desired starting configuration. Figure 2 shows the G–G interaction. The red/white tetramer was rotated around the yellow-highlighted tetramer, which was held in a constant orientation. For the "C" interactions, the tetramer was extracted from the Iβ cellulose structure [15, 27], and the five glycosidic O atoms were frozen during energy minimizations.

To obtain initial starting configurations, energy minimizations and molecular dynamics simulations were run in Materials Studio 7 utilizing the "Fine" quality convergence parameters (Energy = 4.184×10^{-4} kJ/mol; Force = 2.092×10^{-2} kJ/mol/Å; Displacement = 5×10^{-5} Å) and the COMPASS II force field [28]. MD simulations were run on the gas-phase dimers for 100,000 steps (time step = 1 fs) at 298 K using the NVT ensemble for each type of cell wall component interaction (except for the "C" models that were extracted from the cellulose crystal structure model). After MD simulations, the dimers were subjected to energy minimization using the COMPASSII force field. The resulting structures were used as input for quantum mechanical geometry optimizations.

The cellulose microfibril (CMF) is a key building block of the plant cell wall. Due to the central role of the CMF, we modeled the cellotetraose molecule in both its fully relaxed state (the "G" tetramer) and as constrained to approximate four glucan monomers in a cellulose fiber (the "C" tetramer). These models were evaluated for their interactions with all the other PCW components we considered, including the "G" and "C" tetramers. These G–G and C–C interactions represent the energies holding cellulose polymers together within the CMF and, therefore, serve as a basis of comparison for the other dimeric interactions.

Quantum mechanical energy minimizations using DFT methods [M05-2X/6-31G(d,p)] [29, 30] in Gaussian 09 [31] were conducted starting from the geometry optimized interaction confirmations, to investigate other possible favorable interactions as well as compare the accuracy of the COMPASSII force field against a more accurate method. Although other methods may be available, this exchange-correlation functional with dispersion corrections performs reasonably well for carbohydrate and carbohydrate–aromatic (e.g., lignin) interactions [32]. Interaction energies were calculated as in Eq. (1) and compared to those from COMPASSII. The energy difference between the dimer and the two individual monomers was calculated as the energy of the interaction via:

$$\Delta E_{\text{interaction}} = E_{\text{dimer}} - E_{\text{monomer1}} - E_{\text{monomer2}}. \quad (1)$$

The reported ΔE values are divided by the number of monomeric units present to normalize the monomer versus

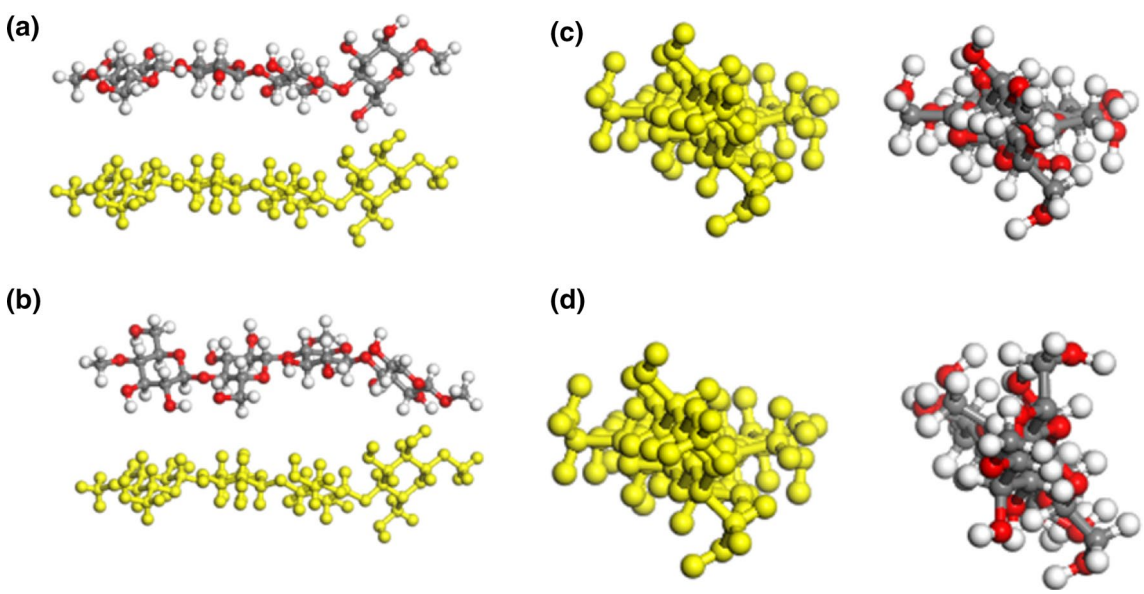


Fig. 2 Glucose–glucose tetramer (cellotetraose) interactions are shown in four orientations (**a** "Stacked", **b** "Stacked180", **c** "Side", **d** "Side180") to explore the range of potential interactions. Stacked configurations are more dominated by van der Waals interaction whereas as side-by-side interaction generally has more H-bonding. The tetramer in red/gray/white has different orientation relative to the tetramer in yellow, rotated by 180° around the *x*-axis

tetramer interaction energies and compare the results with previously reported values [20]. These ΔE values were compared to MP2/6-31G(d,p)//M05-2X/6-311++G(d,p) single-point energies for selected monomers to test the accuracy of the M05-2X method. A test was also run comparing the energies minimized with M05-2X/6-31G(d,p) and 6-311++G(d,p). For the C–C interactions, the 6-31G(d,p) energy minimizations resulted in $\Delta E_{\text{interaction}} = -38$, -38, -39, and -41 kJ/mol and the 6-311++G(d,p) energy minimizations resulted in -30, -27, -26, and -28 kJ/mol. The difference between the more computationally efficient smaller basis sets is within expected uncertainties for this work, so the smaller basis set was used throughout.

MP2 calculations on dimers of tetramers were beyond computational resource limits available, so RI-MP2 [33, 34] calculations were run selected models to compare RI-MP2/def2-TZVP [35] results against M05-2X/6-31G(d,p), M05-2X/6-311+G(d,p), and B3LYP/6-311+G(d,p) [36–43]. The selection of methods tests for the estimate of van der Waals forces between RI-MP2 and M05-2X as well as the effect of basis set and neglecting van der Waals forces using B3LYP. The C–C linear tetramer molecular systems (C–C Side, C–C Side180, C–C Stacked, C–C Stacked180), each having a stoichiometry of $C_{52}H_{92}O_{42}$, have been used as model systems to estimate the accuracy and efficiency of different methods. The RI-MP2 method implemented in ORCA 4.0 [44] was used to calculate intermolecular interaction energies of the four C–C systems, using the def2-TZVP basis set [45], the def2-TZVP/C auxiliary basis, and the RIJK approximation for the HF step [46]. SMD solvation module [47], implemented in ORCA4.0, was also used to estimate interaction energies in water.

The intermolecular interaction energies were also calculated using M05-2X/6-311+G(d,p), M05-2X/6-31G(d,p), B3LYP/6-311++G(d,p) using Gaussian 09 program. The integral equation formalism for the polarizable continuum model (IEFPCM) solvation model [48] was used to estimate interaction energies in water. The results in Table 1 demonstrate that neglect of van der Waals forces in the stacked configuration (B3LYP/6-311+G(d,p)) leads to discrepancies of approximately 20 kJ/mol and that the small basis set (6-31G(d,p)) causes similar differences with the RI-MP2/def2-TZVP results. The differences between the RI-MP2/def2-TZVP and M05-2X/6-311+G(d,p) energies are significantly less. Consequently, results reported in this study will be based on M05-2X/6-311+G(d,p) calculations, and we caution the reader that errors of at least ± 10 kJ/mol are to be expected.

Another test was run to evaluate the effect of the model polymer length on the calculated interaction energies. Decamers of the C model chain were simulated as an isolated chain with a monolayer of H_2O and as a dimer combining two decamers. The interaction energy calculated was -57 kJ/mol/glucan compared with the -37 kJ/mol/glucan using the decamer. The reason why the interaction energy per glucan should become stronger with length is not clear but likely is a result of the greater flexibility in the decamer compared to the tetramer model. The reader is cautioned that the size effects of these small (compared to actual PCW biopolymers) models on the energetics of interaction could be significant and a source of uncertainty. Future work examining interaction energies as a function of size is desirable.

Table 1 Comparisons of intermolecular interaction energies (kJ/mol/monomer) of C–C and G–L systems in IEFPCM models of water using various quantum methods

Methods	Side180	Side	Stacked180	Stacked
C–C–T				
RI-MP2/def2-TZVP	− 20.1	− 20.1	− 22.7	− 22.7
M05-2X/6-311+G(d,p)	− 30.0	− 29.9	− 29.2	− 28.4
M05-2X/6-31G(d,p)	− 40.4	− 40.4	− 39.6	− 39.6
B3LYP/6-311++G(d,p)	− 21.3	− 21.3	− 1.7	− 1.7
G–L–T				
RI-MP2/def2-TZVP	− 20.2	− 20.2	− 26.8	− 39.5
M05-2X/6-311+G(d,p)	− 16.0	− 16.7	− 23.4	− 35.3
M05-2X/6-31G(d,p)	− 21.6	− 22.0	− 30.5	− 42.6
B3LYP/6-311++G(d,p)	− 6.3	− 8.9	0.6	− 0.9

The gas-phase monomers and dimers were explicitly solvated using Schrödinger Maestro [49, 50] to create a first solvation sphere. In addition to the explicit solvation via H_2O molecules, the monomers and dimers were energy minimized in an implicit solvent using IEFPCM. The default solvent properties of "Water" as defined in Gaussian 09 were used to mimic solvent effects in all cases. In some cases, additional calculations using the "Methanol" and "Chloroform" solvents as defined in Gaussian 09 was used as well to assess the effect of dielectric constant on the model results.

Molecular mechanics calculations were performed with the NAMD software [51] employing the CHARMM carbohydrate [52, 53] and lignin [54] force fields. The coordinates were obtained from the DFT calculations.

3 Results

Van der Waals or dispersion forces are significant components of the interaction energies we are modeling. DFT calculations have long been known to be inadequate for modeling intermolecular interactions, which has led to the developments for including dispersion terms to the DFT energy. One such method is the M05-2X density functional [29]. An alternative method, Møller–Plessett second-order perturbation theory (MP2) [55], is considered reliable, but

MP2 is computationally expensive as it scales as N^5 where N is the number of basis functions. Consequently, MP2 is impractical for models larger than the monomer pairs considered here. Hence, we selected monomer pairs to use as tests of the accuracy of the M05-2X method compared to MP2. Figure 3 compares the interaction energies for monomer pairs G+L, L+X and G+X as obtained with M05-2X and MP2. Although the MP2 ΔE values are generally 10–15 kJ/mol more negative than the M05-2X values, the relative ordering of pairwise interactions is the same throughout. MP2 energies are known to be dependent on basis set, and overestimation of van der Waals forces between molecules is common especially with small basis sets such as 6-31G(d,p) [56]. One should consider this − 10 to − 15 kJ/mol discrepancy when applying the ΔE values presented in this study.

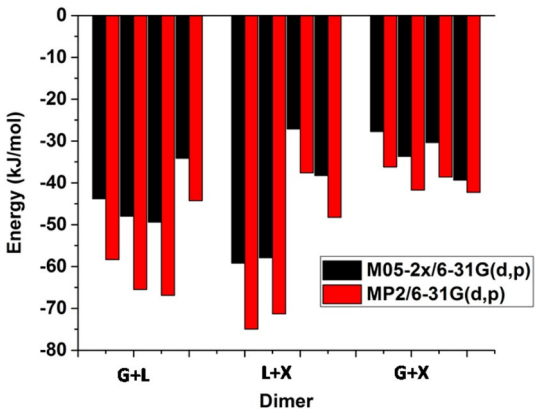
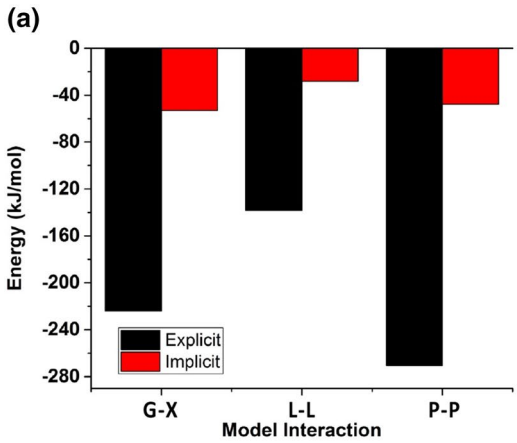


Fig. 3 Comparison of interaction energies for M05-2X/6-31G(d,p)//M05-2X/6-311++G(d,p) (black) and MP2/6-31G(d,p)//M05-2X/6-311++G(d,p) (red) on glucose–lignin (G++L), lignin–xylose (L++X), and glucose–xylose (G++X) pairs of monomers. Each set of interactions (i.e., G+L, L+X and G+X) are ordered as Stacked, Stacked180, Side, and Side180. MP2-based energies are more negative than M05-2X energies, but the relative ordering of interaction energies are the same for all three pairs

Energy minimizations with explicit solvation for tetramer (and larger) models become impractical. For example, a pectin–pectin tetramer with full explicit solvation has a stoichiometry of COH–$n$$H_2O$. Numerous simulation steps on such a model are required to find an energy minimum for such a model requiring months of CPU time. Consequently, implicit solvation is more practical for this size of molecule. To assess the relative accuracy of the IEFPCM against explicit solvation by H_2O molecules, a set of supra-molecular dimers (G–X, L–L, P–P) consisting of various monomeric compounds included in this study were subjected to both explicit and implicit solvation. The G–X, L–L and P–P dimers were selected to represent the common cellulose–hemicellulose interaction as well as the most hydrophobic and hydrophilic compound interactions, respectively.

Figure 4a compares the calculated interaction energies for each pair with both solvation methods. The explicit solvation method values are − 110 to − 220 kJ/mol lower in energy than the implicit solvation values. This is due to the inclusion of the energy of interaction of the H_2O molecules assigned to each monomer as they H-bond in the dimer (Fig. 4b). Furthermore, the implicit solvation method has been formulated to approximate Gibbs free energies of solvation whereas the implicitly solvated models provide potential energy changes. Consequently, the absolute energy differences between explicit and implicit solvation are not directly comparable, and one can focus on the relative energy differences. The L–L interaction has the weakest interaction using either solvation method (not including the entropic component of hydrophobic interactions), but the P–P dimer is the lowest in energy using explicit solvation versus the



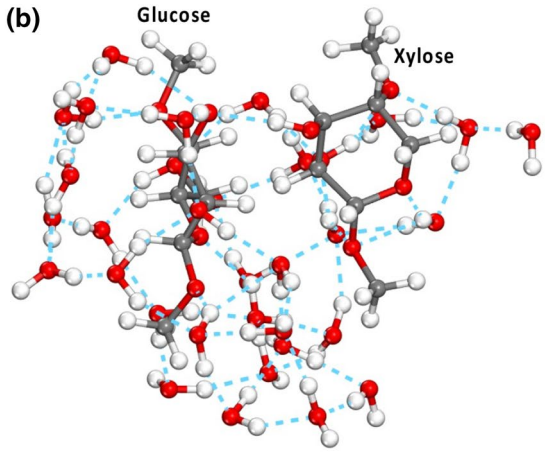


Fig. 4 **a** Comparison of Stacked monomer pair interaction energies as calculated with explicit solvation with H_2O molecules and with the implicit IEFPCM model. **b** G–X explicit solvation models showing H-bonds (blue dashed lines) between H_2O and each monomer in addition to H-bonds between the methylated "Glucose" and "Xylose" models

G–X dimer using implicit solvation. Hence, we must use caution when interpreting our calculated interactions of the tetramers using the implicit solvation model. Energy differences of 20 kJ/mol can be considered significant in light of these test results, but differences on the order of 5 kJ/mol should be considered the same within error.

The G and C tetramers represent extreme cases of the varying degree of cellulose crystallinity found in PCWs. If a section of cellulose is tightly constrained and well ordered within a CMF, then the C tetramer will be a better approximation. On the other hand, if a section of a cellulose chain has significant freedom to relax upon interactions with other PCW components, then the G tetramer is closer to reality. The G tetramer will also better reflect interactions that may occur with disordered or "amorphous" cellulose. Figure 5 compares the interaction energies in all four configurations for the G–G, G–X, G–L and G–P dimers of tetramers. The values are reasonable compared to previously reported values for cellulose–hemicellulose interactions in water. CHARMM-based classical MD results produced values of approximately −20 to −40 kJ/mol from simulations [22]; note that those values were based on tetramer units and reported in kcal/mol. The interaction energy results reported here and those previously reported results [22] are much smaller than those reported for gas-phase calculations [57]; this discrepancy can be attributed to the presence of water in the calculations presented here.

Examination of the details of Fig. 6a reveals several important points. First, the G–G stacked interactions are generally stronger than the G–G side-by-side interactions. This implies that in aqueous solution the hydrophobic interactions between layers of cellulose polymers will dominate assembly over the H-bonding interactions found within layers of the CMF, consistent with prior work [58]. This does not consider the entropic contribution, which will also be significant for these hydrophobic interactions [59].

Second, specific G–L and G–P stacked interactions are comparable or greater than the most favorable G–G interaction energy. This prediction has relevance to lignin protecting CMFs from enzymatic degradation in secondary cell walls [60, 61] and the role of pectin as a spacer between CMFs in primary cell walls [62, 63]. Both lignin and pectin interactions with the G tetramer can be stronger than that of the G–X interactions, which may help explain why CMF may have only 20% of their surfaces in contact with hemicellulose in PCWs [10]. This result appears contrary to in vitro adsorption studies of xyloglucan and pectin component onto cellulose that show preference for xyloglucans [8, 64]. The primary source of this difference could be the use of xylans in the tetramer rather than xyloglucans. Another issue is that our results are normalized to a per monomer basis, so direct comparison with experiments where the polymer lengths may be different between adsorbates is not possible. We note

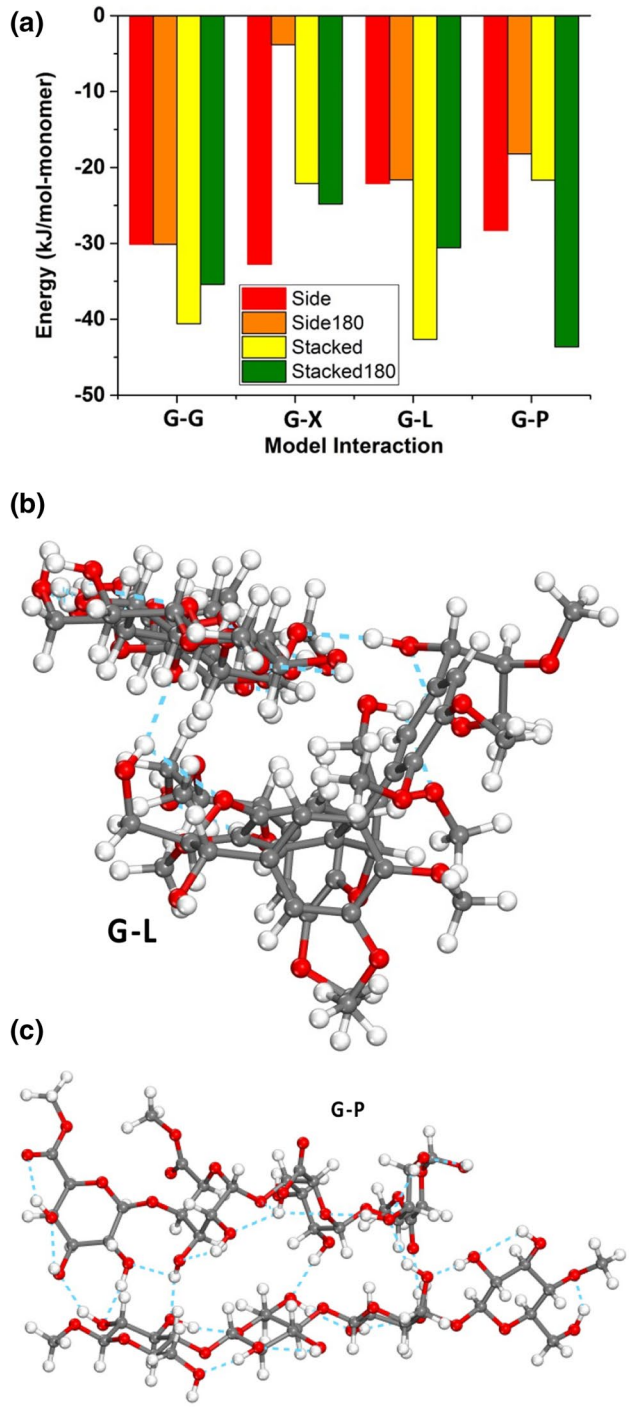


Fig. 5 **a** Interaction energies (ΔE) as calculated using Eq. (1) for the G, X, L and P tetramers interacting with the G tetramer (i.e., unconstrained cellotetraose). **b** G–L Stacked model shows minimal H-bonding with the dimer formation dominated by hydrophobic interactions. **c** Glucose–pectin Stacked180 tetramer interactions show significant H-bonding (dashed blue lines)

that prior results do demonstrate that certain components of pectin (i.e., the arabinan and galactan side chains) do adsorb significantly to cellulose when xyloglucan is not present or

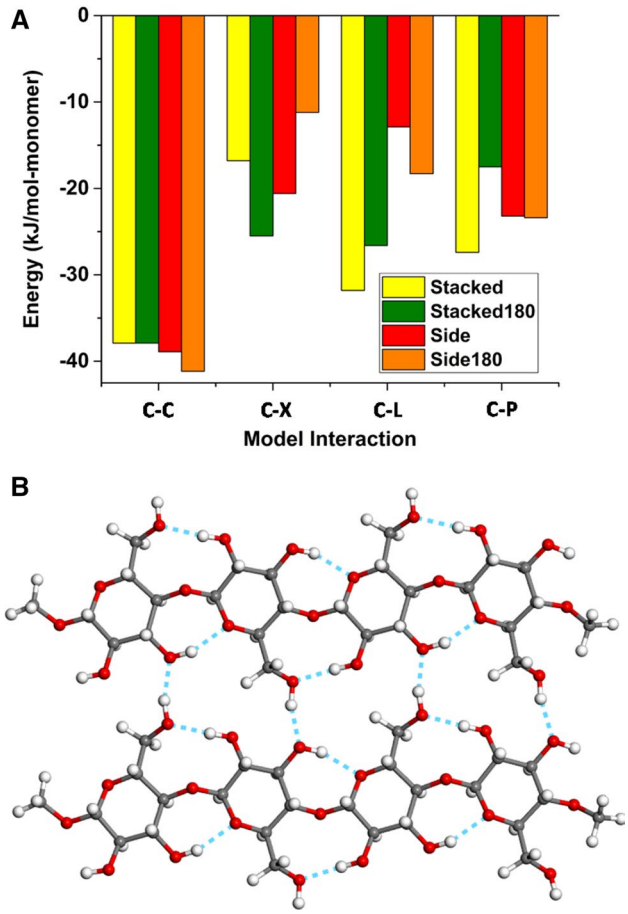


Fig. 6 **a** Energy differences (kJ/mol normalized to a per monomer basis) of constrained (see "Methods" section) cellotetraose ("C" tetramer in the cellulose structure) used to mimc cellulose with other PCW components. **b** Side C–C configuration showing H-bonds (dashed blue lines) that dominate side-by-side interactions as well as influence intramolecular interactions

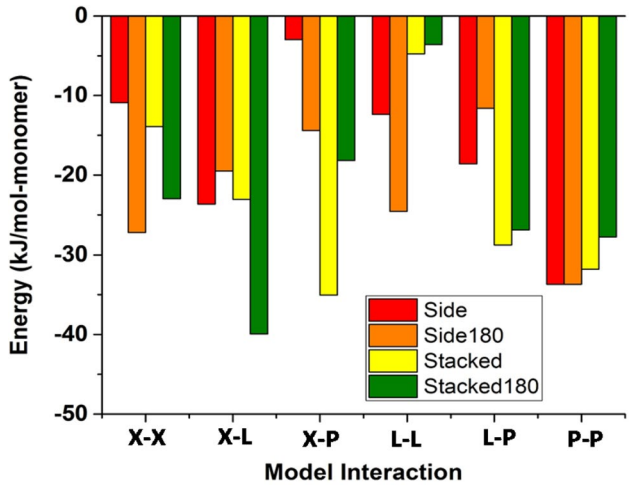


Fig. 7 Energies of X, L and P interactions with each other show the relative preference of each component for other components as well as the more favorable configurations for each type of interaction

is in low concentration [8, 64]. Xyloglucan coverage of cellulose microfibrils in plant cell walls may be 20% or less [10], so there could be opportunity for pectin–cellulose binding without competition from xyloglucan in plant cell walls. Significant pectin–cellulose interactions are also seen in ss-NMR spectra of primary cell walls in *Arabidopsis* and *Brachypodium* [65].

Inspection of the G–L and G–P configurations shows that one type of interaction is strongly favored in each case (Fig. 5b "Stacked" and Fig. 5c "Stacked180", respectively). With the alternating nature of glucan orientations within the cellulose polymer [27], these results imply that both pectin and lignin will selectively bind to the CMF while having weaker interactions at adjacent glucan units in primary and secondary cell walls, respectively.

Full relaxation of the "Glucose" tetramer allows the model to deviate significantly from the ordered structure of cellulose microfibrils [27]; to prevent this, we modeled the same tetramer with the C atoms fixed to retain a better approximation of the configuration found in cellulose. The interaction energies for the various components and configurations are shown in Fig. 6a. Compared to the G–G interactions (Fig. 5a), the C–C interactions are slightly stronger with the notable exception of the C–C Side180 configuration. This implies a strong asymmetry for cellulose polymer interactions within a layer and a preference for the alignment shown in Fig. 6b. Both of the "Stacked" interaction energies are as strong as the "Side" interaction, which again is consistent with previous results showing that layer-to-layer interactions dominate cellulose microfibril assembly [58], especially considering that these more hydrophobic interactions would also gain thermodynamic stability from entropy gains due to water release [66, 67] that is not accounted for in this calculation. The interactions of the "C" tetramer with the other model tetramers (Fig. 7a) are similar to those calculated for the "G" tetramer (Fig. 5a) except that the C–P Stacked180 configuration is no longer the strongest. This suggests that pectin interactions with cellulose microfibrils (as opposed to unassembled cellulose polymers) will be similar in a PCW to cellulose–xyloglucan interactions on a per monomer basis. Also notable are the weak interactions of the Side180 C–C and C–X tetramer pairs, which imply that this anti-parallel arrangement is not favorable.

Other PCW interaction energies fall within a similar range as those with the G and C tetramers (Fig. 7). The range of calculated energies is from −10 to −40 kJ/mol-monomer with most values falling in the −20 to 30 kJ/mol-monomer range. The most favorable interaction is the Stacked180 X–L dimer, which is similar to the Stacked G–L dimer (Fig. 5a). Based on these results, one can infer that in the complex assemblage of the PCW, hydrophobic interactions bond the

lignin to the cellulose and hemicellulose at specific points in these polymers. These sites of interaction are likely to exclude water, which will affect the mechanical properties of the polymer mixture by creating points of rigidity and possibly inhibiting enzymatic attack since water is less available. The Stacked X–P and the Side P–P interactions are also relatively strong. Pectin is predicted to be a good binding agent of PCW components even in this aqueous model where the hydrophilic nature of pectin is included in the reacting tetramer molecules [64].

A critical parameter that will influence the thermodynamics of PCW component interactions is the nature of the medium in which the interactions take place (e.g., which solvent). Experiments are often performed in vitro with the assumption that the water within the PCW behaves in a manner similar to bulk water. However, prior work has demonstrated that water within cells can exhibit much slower dynamics than bulk water [68]. Consequently, the decreased dynamic response causes a decrease in the polarizability of the water [69]. On the macroscopic thermodynamic scale, this translates into a lower dielectric constant [69]. Hence, water within the PCW may have a dielectric constant, ε, closer to that of organic solvents (e.g., 10–40) than bulk water at standard temperature and pressure (i.e., 78.4). Such a decrease would enhance the strength of H-bond energies among PCW components relative to the PCW component–water interactions and alter the ΔE of various interactions depending on the hydrophilicity of the components involved. To complicate matters, one would expect ε to vary within PCWs as a function of composition, with primary cell walls exhibiting higher water content and εs than secondary cell walls with their higher hydrophobic lignin content.

In the absence of data on the ε of PCWs, we varied the ε input into Gaussian 09 to test the effect on calculated ΔE values. Dielectric constants representative of methanol and chloroform (32.7 and 4.8) were used to explore the range of variability. (Note: this is not a perfect representation of water with a lower ε, but this method examines the magnitude effect of lowering ε to a first approximation.) Figure 8 shows the results of this computational experiment, and the results vary significantly. For example, the C–C Side interactions are enhanced by a few kJ/monomer in the lowest dielectric constant solvent, chloroform; whereas the C–C Stacked interactions are decreased by approximately one-third. These results make chemical sense because the hydrophilic interactions between PCW components are enhanced in the lower ε solvent, and the hydrophobic interactions are decreased. The models including the lignin tetramer (C–L Stacked, L–L Stacked180 and L–L Side180) do not exhibit any systematic pattern, complicating interpretation. The P–P Side180 interaction is decreased in chloroform when the opposite should be expected for hydrophilic associations, and the P–P Stacked180 resulted in small changes with dielectric

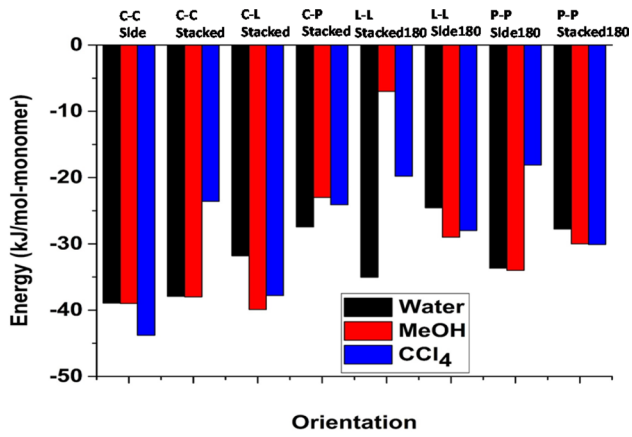


Fig. 8 Interaction energies of various model components in water, methanol and chloroform are shown to highlight the effective of dielectric constant on the calculated ΔE values. The dielectric constant within a PCW is likely to be lower than that of bulk water, so changing ε value illustrates how various types of interactions will be affected differently depending on the ε present

constant (Fig. 8). Due to the simple nature of this test, one cannot derive quantitative predictions on PCW component interactions from these results. However, one can conclude that the ε effect is likely to be significant. Based on this conclusion, obtaining experimentally based estimates of the variations in ε in PCWs is imperative to improve our understanding of PCW architecture and mechanical properties.

To move forward with larger scale, models of PCW component interactions that are more complex, classical molecular simulations will be necessary because the size of the model system quickly becomes too large for quantum calculations to be practical at this point. The CHARMM force field has been used extensively to model certain PCW components [60, 70], so we start with a comparison of DFT versus CHARMM interaction energies for our model tetramers (Fig. 9). The ability to accurately predict ΔE of interaction among these components is a critical test for models that attempt to reproduce PCW structures and behavior because the configurations calculated are based on the energy–structure relationship. (Note: The DFT results themselves will contain significant errors, so calorimetric measurement of these interaction energies is desirable to create a benchmark for both computational methods.)

The results for the G–L and L–L tetramer models are mixed. For example, good agreement was found for the L–L Side, G–L Side180, G–L Stacked and L–L Stacked180 configurations. On the other hand, significant (i.e., 10 kJ/monomer or more) deviation can be seen for the G–L Side, L–L Side180, L–L Stacked, and L–L Stacked180 (Fig. 9). Much of this discrepancy may be due to inadequacies in the DFT method in calculating the van der Waals forces for these interactions. The DFT results may be 10–15 kJ/mol

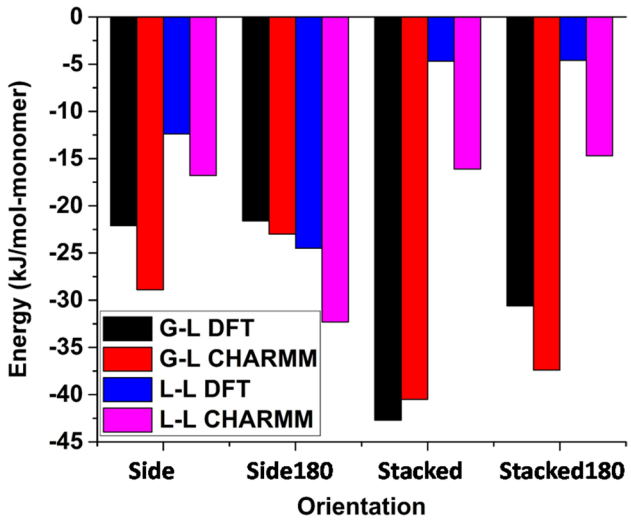


Fig. 9 Comparison of selected model tetramer interaction energies as calculated with DFT versus CHARMM in gas-phase reveal similar relative values although discrepancies on the order or 10 kJ/mol/monomer are present

lower than actual values based on our comparison to MP2 calculations (Fig. 3). We conclude that the CHARMM force field is behaving reasonably well for the G and L component interactions, but further testing against experimental thermodynamic is required to test the force field and guide future parameter refinement.

4 Summary

In this case, we have shown that the M05-2X method tracks the trend of interaction energies calculated with the more computationally intensive MP2 method, but the interaction energies are systematically less negative (i.e., thermodynamically favorable) by 10–15 kJ/mol/monomer. Thus, relative interaction energies can be assessed with M05-2X, but when modeling plant cell wall components with classical or coarse-grained methods, this systematic difference should be accounted for.

The results also indicate that the commonly used implicit solvation method IEFPCM does not replicate interaction energies calculated with explicit solvation by H_2O molecules. This is to be expected because polarized continuum models are not designed to account for stronger, short-range H-bonding. In addition, energy minimization with explicit solvation is more complicated and will result in more local potential energy minima. This discrepancy can be addressed in the future with extensive MD or Monte Carlo simulation of the explicitly solvated interactions and calculating single-point energies with quantum mechanical methods.

For G–G, G–L and G–P interactions, the stacked configurations tend to have the largest interaction energies. Combined with the entropic factor that will favor this type of hydrophobic interaction, the Gibbs free energy of this type of interaction energy is likely to be even more favorable than the side-by-side interactions. Most of the interactions with the "C" (cellulose-like) tetramer are similar considering the computational error. A notable exception is the Side180 interaction that is not found in crystalline cellulose, which suggests the models are reproducing energetics observed in the larger experimental systems. The range of interaction energies is typically −5 to −20 kJ/mol/monomer, but notable exceptions such as the X–L, X–P and P–P interactions can be −30 to −40 kJ/mol/monomer. This could indicate that pectin interactions via H-bonding more strongly influence the mechanical properties of the plant cell wall. These interactions will be a function of hydration with H_2O competing with other PCW components for interaction with pectin, so these results could be re-examined with explicitly solvated models in the future.

The variations in model interaction energies with solvent dielectric in the polarized continuum model are significant and non-systematic. This result points to a knowledge gap in our understanding of plant cell wall behavior, i.e., the state of water within the matrix of the PCW. This work indicates that addressing this factor will be important in generating improved classical and coarse-grained models of PCWs. The CHARMM results are reasonably accurate for the models tested here, but they also have room for improvement by further parameterization. This is particularly problematic for secondary PCWs where lignin is a significant component.

Acknowledgements This work was supported as part of The Center for LignoCellulose Structure and Formation, an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, Basic Energy Sciences under Award # DE-SC0001090. This research was conducted with Advanced CyberInfrastructure computational resources provided by NERSC, The Institute for CyberScience at The Pennsylvania State University (http://ics.psu.edu), and the Texas Advanced Computing Center (TACC). Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the US Department of Energy under Contract DE-AC05-00OR22725.

Compliance with ethical standards

Conflict of interest The authors declare no conflicts of interest.

References

1. Cosgrove DJ, Jarvis MC (2012) Comparative structure and biomechanics of plant primary and secondary cell walls. Front Plant Sci 3:204
2. Cosgrove D (2010) Cell walls: structure, biogenesis, and expansion. In: Taiz L, Zeiger E (eds) Plant physiology. Sinauer, Sunderland, pp 425–452

3. Park YB, Cosgrove DJ (2012) A revised architecture of primary cell walls based on biomechanical changes induced by substrate-specific endoglucanases. Plant Physiol 158:1933–1943
4. Pauly M, Albersheim P, Darvill A, York WS (1999) Molecular domains of the cellulose/xyloglucan network in the cell walls of higher plants. Plant J 20:629–639
5. Dick-Perez M, Wang T, Salazar A, Zabotina OA, Hong M (2012) Multidimensional solid-state NMR studies of the structure and dynamics of pectic polysaccharides in uniformly 13C-labeled Arabidopsis primary cell walls. Magn Reson Chem 50:539–550
6. Wang T, Park YB, Cosgrove DJ, Hong M (2015) Cellulose–pectin spatial contacts are inherent to never-dried arabidopsis primary cell walls: evidence from solid-state nuclear magnetic resonance. Plant Physiol 168:871–884
7. White PB, Wang T, Park YB, Cosgrove DJ, Hong M (2014) Water–polysaccharide interactions in the primary cell wall of arabidopsis thaliana from polarization transfer solid-state NMR. J Am Chem Soc 136:10399–10409
8. Zykwinska A, Gaillard C, Buléon A, Pontoire B, Garnier C, Thibault JF, Ralet MC (2007) Assessment of in vitro binding of isolated pectic domains to cellulose by adsorption isotherms, electron microscopy, and X-ray diffraction methods. Biomacromol 8:223–232
9. Zykwinska A, Thibault JF, Ralet MC (2008) Modelling of xyloglucan, pectins and pectic side chains binding onto cellulose microfibrils. Carbohydr Polym 74:23–30
10. Dick-Pérez M, Zhang Y, Hayes J, Salazar A, Zabotina OA, Hong M (2011) Structure and interactions of plant cell-wall polysaccharides by two- and three-dimensional magic-angle-spinning solid-state NMR. Biochemistry 50:989–1000
11. Busse-Wicher M, Gomes TCF, Tryfona T, Nikolovski N, Stott K, Grantham NJ, Bolam DN, Skaf MS, Dupree P (2014) The pattern of xylan acetylation suggests xylan may interact with cellulose microfibrils as a twofold helical screw in the secondary plant cell wall of *Arabidopsis thaliana*. Plant J 79:492–506
12. Langan P, Petridis L, O'Neill HM, Pingali SV, Foston M, Nishiyama Y, Schulz R, Lindner B, Hanson BL, Harton S, Heller WT, Urban V, Evans BR, Gnanakaran S, Ragauskas AJ, Smith JC, Davison BH (2014) Common processes drive the thermochemical pretreatment of lignocellulosic biomass. Green Chem 16:63–68
13. Hill SJ, Franich RA, Callaghan PT, Newman RH (2009) Nature's nanocomposites: a new look at molecular architecture in wood cell walls. N Z J For Sci 39:251–257
14. Parthasarathi R, Bellesia G, Chundawat SPS, Dale BE, Langan P, Gnanakaran S (2011) Insights into hydrogen bonding and stacking interactions in cellulose. J Phys Chem A 115:14191–14202
15. Kubicki JD, Mohamed MN-A, Watts HD (2013) Quantum mechanical modeling of the structures, energetics and spectral properties of Iα and Iβ cellulose. Cellulose 20:9–23
16. Kubicki JD, Watts HD, Zhao Z, Zhong L (2014) Quantum mechanical calculations on cellulose–water interactions: structures, energetics, vibrational frequencies and NMR chemical shifts for surfaces of Iα and Iβ cellulose. Cellulose 21:909–926
17. Watts HD, Mohamed MNA, Kubicki JD (2014) A DFT study of vibrational frequencies and ^{13}C NMR chemical shifts of model cellulosic fragments as a function of size. Cellulose 21:53–70
18. Watts H, Mohamed M, Kubicki J (2011) Comparison of multistandard and TMS-standard calculated NMR shifts for coniferyl alcohol and application of the multistandard method to lignin dimers. J Phys Chem B 115:1958–1970
19. Parthasarathi R, Romero RA, Redondo A, Gnanakaran S (2011) Theoretical study of the remarkably diverse linkages in lignin. J Phys Chem Lett 2:2660–2666
20. Zhao Z, Shklyaev OE, Nili A, Mohamed MNA, Kubicki JD, Crespi VH, Zhong L (2013) Cellulose microfibril twist, mechanics, and implication for cellulose biosynthesis. J Phys Chem A 117:2580–2589
21. Diehl BG, Watts HD, Kubicki JD, Regner MR, Ralph J, Brown NR (2014) Towards lignin–protein crosslinking: amino acid adducts of a lignin model quinone methide. Cellulose 21:1395–1407
22. Zhao Z, Crespi VH, Kubicki JD, Cosgrove DJ, Zhong L (2014) Molecular dynamics simulation study of xyloglucan adsorption on cellulose surfaces: effects of surface hydrophobicity and side-chain variation. Cellulose 21:1025–1039
23. Hohenberg P, Kohn W (1964) Inhomogeneous. Electron Gas Phys Rev 136:B864–B871
24. Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140:A1133–A1138
25. Guvench O, Hatcher ER, Venable RM, Pastor RW, MacKerell, AD Jr (2009) CHARMM additive all-atom force field for glycosidic linkages. J Chem Theory Comput 5:2353–2370
26. Materials Studio (2007) Release 7.0, Accelrys Software Inc., San Diego
27. Nishiyama Y, Langan P, Chanzy H (2002) Crystal structure and hydrogen-bonding system in cellulose Iβ from synchrotron X-ray and neutron fiber diffraction. J Am Chem Soc 124:9074–9082
28. Sun HCOMPASS (1998) An AB initio force-field optimized for condensed-phase applicationsoverview with details on alkane and benzene compounds. J Phys Chem B 102:7338–7364
29. Zhao Y, Schultz NE, Truhlar DG (2006) Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions. J Chem Theory Comput 2:364–382
30. Rassolov VA, Ratner MA, Pople JA, Redfern PC, Curtiss LA (2001) 6-31G* basis set for third-row atoms. J Comput Chem 22:976–984
31. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2009) Gaussian 09 Revision E.01. Wallingford, CT
32. Raju RK, Ramraj A, Hillier IH, Vincent MA, Burton NA (2009) Carbohydrate–aromatic π interactions: a test of density functionals and the DFT-D method. Phys Chem Chem Phys 11:3411
33. Weigend F, Häser M (1997) RI-MP2: first derivatives and global consistency. Theor Chem Acc Theory Comput Model (Theoretica Chim Acta) 97:331–340
34. Weigend F, Häser M, Patzelt H, Ahlrichs R (1998) RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. Chem Phys Lett 294:143–152
35. Hellweg A, Hättig C, Höfener S, Klopper W (2007) Optimized accurate auxiliary basis sets for RI-MP2 and RI-CC2 calculations for the atoms Rb to Rn. Theor Chem Acc 117:587–597
36. Hehre WJ, Ditchfield K, Pople JA (1972) Self-consistent molecular orbital methods. XII. Further extensions of gaussian-type basis sets for use in molecular orbital studies of organic molecules. J Chem Phys 56:2257–2261
37. Ditchfield R, Hehre WJ, Pople JA (1971) Self-consistent molecular-orbital methods. IX. An extended Gaussian-type basis for

molecular-orbital studies of organic molecules. J Chem Phys 54:724–728
38. Becke AD (1993) Becke's three parameter hybrid method using the LYP correlation functional. J Chem Phys 98:5648–5652
39. Lee C, Yang W, Parr RG (1988) Development of the Colle–Salvetti correlation–energy formula into a functional of the electron density. Phys Rev B 37:785–789
40. Hariharan PC, Pople JA (1974) Accuracy of AHn equilibrium geometries by single determinant molecular orbital theory. Mol Phys 27:209–214
41. McLean AD, Chandler GS (1980) Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, $Z = 11$–18. J Chem Phys 72:5639–5648
42. Clark T, Chandrasekhar J, Spitznagel GW, Schleyer PVR (1983) Efficient diffuse function-augmented basis sets for anion calculations. III. The 3–21 + G basis set for first-row elements, Li-F. J Comput Chem 4:294–301
43. Frisch MJ, Pople JA, Binkley JS (1984) Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. J Chem Phys 80:3265–3269
44. Neese F (2012) The ORCA program system. Wiley Interdiscip Rev Comput Mol Sci 2:73–78
45. Weigend F, Ahlrichs R (2005) Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: design and assessment of accuracy. Phys Chem Chem Phys 7:3297
46. Weigend F (2008) Hartree-fock exchange fitting basis sets for H to Rn. J Comput Chem 29:167–175
47. Marenich AV, Cramer CJ, Truhlar DG (2009) Unviersal solvation modle based on solute electron density and a contiuum model of the solvent defind by the bulk dielectric constant and atomic surface tensions. J Phys Chem B 113:6378–6396
48. Cancès E, Mennucci B, Tomasi J (1997) A new integral equation formalism for the polarizable continuum model: theoretical background and applications to isotropic and anisotropic dielectrics. J Chem Phys 107:3032–3041
49. Yang H (2014) Schrödinger Maestro, Release 2014-1. Schrödinger, LLC, New York
50. Bowers K, Chow E, Xu H, Dror R, Eastwood M, Gregersen B, Klepeis J, Kolossvary I, Moraes M, Sacerdoti F, Salmon J, Shan Y, Shaw D (2006) Scalable algorithms for molecular dynamics simulations on commodity clusters. In: ACM/IEEE SC 2006 conference (SC'06); IEEE, pp 43–43
51. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26:1781–1802
52. Guvench O, Hatcher E, Venable RM, Pastor RW, MacKerell AD (2009) CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. J Chem Theory Comput 5:2353–2370
53. Guvench O, Greenr SN, Kamath G, Brady JW, Venable RM, Pastor RW, Mackerell AD (2008) Additive empirical force field for hexopyranose monosaccharides. J Comput Chem 29:2543–2564
54. Petridis L, Smith JC (2009) A molecular mechanics force field for lignin. J Comput Chem 30:457–467
55. Møller C, Plesset MS (1934) Note on an approximation treatment for many-electron systems. Phys Rev 46:618–622
56. Haldar S, Gnanasekaran R, Hobza P (2015) c A comparison of ab initio quantum-mechanical and experimental D0 binding energies of eleven H-bonded and eleven dispersion–bound complexes. Phys Chem Chem Phys 17:26645–26652
57. Hanus J, Mazeau K (2006) The xyloglucan–cellulose assembly at the atomic scale. Biopolymers 82:59–73
58. Cousins SK, Brown RM (1995) Cellulose I microfibril assembly: computational molecular mechanics energy analysis favours bonding by van der Waals forces as the initial step in crystallization. Polymer 36:3885–3888
59. Hummer G, Garde S, García AE, Paulaitis ME, Pratt LR (1998) Hydrophobic effects on a molecular scale. J Phys Chem B 102:10469–10482
60. Lindner B, Petridis L, Schulz R, Smith JC (2013) Solvent-driven preferential association of lignin with regions of crystalline cellulose in molecular dynamics simulation. Biomacromol 14:3390–3398
61. Vermaas JV, Petridis L, Qi X, Schulz R, Lindner B, Smith JC (2015) Mechanism of lignin inhibition of enzymatic biomass deconstruction. Biotechnol Biofuels 8:217
62. Cosgrove DJ (2014) Re-constructing our models of cellulose and primary cell wall assembly. Curr Opin Plant Biol 22:122–131
63. Jarvis MC (1992) Control of thickness of collenchyma cell walls by pectins. Planta 187:218–220
64. Zykwinska A, Thibault JF, Ralet MC (2008) Competitive binding of pectin and xyloglucan with primary cell wall cellulose. Carbohydr Polym 74:957–961
65. Wang T, Hong M (2016) Solid-state NMR investigations of cellulose structure and interactions with matrix polysaccharides in plant primary cell walls. J Exp Bot 67:503–514
66. Abraham MH (1982) Free energies, enthalpies, and entropies of solution of gaseous nonpolar nonelectrolytes in water and nonaqueous solvents. The hydrophobic effect. J Am Chem Soc 104:2085–2094
67. Sedov IA, Magsumov TI, Solomonov BN (2016) Solvation of hydrocarbons in aqueous–organic mixtures. J Chem Thermodyn 96:153–160
68. Bakulin AA, Cringus D, Pieniazek PA, Skinner JL, Jansen TLC, Pshenichnikov MS (2013) Dynamics of water confined in reversed micelles: multidimensional vibrational spectroscopy study. J Phys Chem B 117:15545–15558
69. Huang YR, Liu KH, Mou CY, Sun CK (2015) Relaxation dynamics of surface-adsorbed water molecules in nanoporous silica probed by terahertz spectroscopy. Appl Phys Lett 107:081607
70. Nimlos MR, Beckham GT, Matthews JF, Bu L, Himmel ME, Crowley MF (2012) Binding preferences, surface attachment, diffusivity, and orientation of a family 1 carbohydrate-binding module on cellulose. J Biol Chem 287:20603–20612