2019

# Arabinose substitution effect on xylan rigidity and self-aggregation

Utsab Shrestha
*Oak Ridge National Laboratory*

Sydney Smith
*Oak Ridge National Laboratory*

Sai Venkatesh Pingali
*Oak Ridge National Laboratory*

Hui Yang
*Pennsylvania State University*

Mai Zahran
*CUNY New York City College of Technology*

*See next page for additional authors*

# Arabinose substitution effect on xylan rigidity and self-aggregation

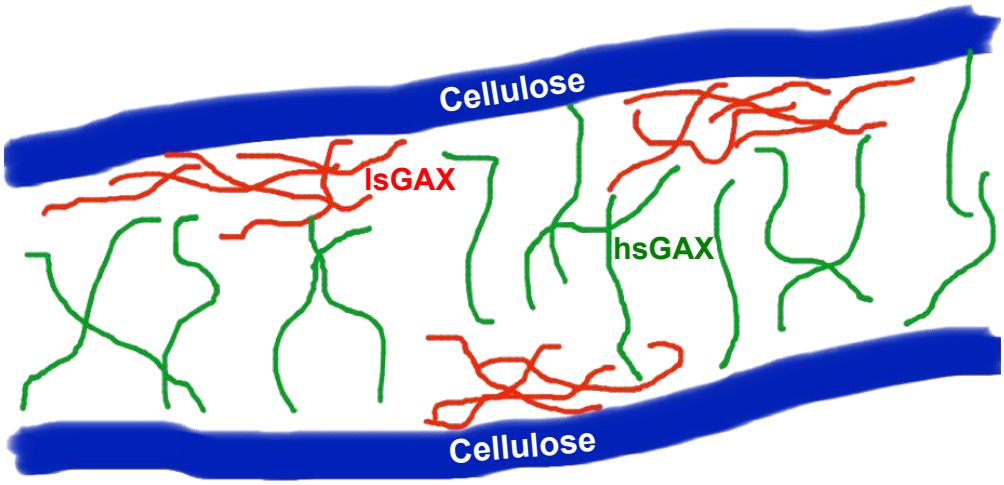
Utsab R. Shrestha,[a] Sydney Smith,[a] Sai Venkatesh Pingali,[b] Hui Yang,[c] Mai Zahran,[d] Lloyd Breunig,[c] Liza A. Wilson,[c] Margaret Kowali,[e] James D. Kubicki,[f] Daniel J. Cosgrove,[c] Hugh M. O'Neill,[b] Loukas Petridis[a,*]

*[a]UT/ORNL Center for Molecular Biophysics, Oak Ridge National Laboratory, Oak Ridge, TN, U.S.A.; [b]Neutron Scattering Division, Oak Ridge National Laboratory, Oak Ridge, TN, U.S.A.; [c]Department of Biology, Pennsylvania State University, University Park, PA, U.S.A.; [d]Department of Biology, New York City College of Technology, New York, NY, U.S.A.; [e]Department of Chemical Engineering, Pennsylvania State University, University Park, PA, U.S.A.; [f]Department of Geological Sciences, University of Texas at El Paso, El Paso, TX, U.S.A.*

**Abstract**

Substituted xylans play an important role in the structure and mechanics of the primary cell wall of plants. Arabinoxylans (AX) consist of a xylose backbone substituted with arabinose, while Glucuronoarabinoxylans (GAX) also contain glucuronic acid substitutions and ferulic acid esters on some of the arabinoses. We provide a molecular-level description on the dependence of xylan conformational, self-aggregation properties and binding to cellulose on the degree of arabinose substitution. Molecular dynamics simulations reveal fully solubilized xylans with a low degree of arabinose substitution (lsAX) to be stiffer than their highly substituted (hsAX) counterparts. Small-angle neutron scattering experiment indicate that both wild-type hsAX and debranched lsAX form macromolecular networks that are penetrated by water. In those networks, lsAX are more folded and entangled than hsAX chains. Increased conformational entropy upon network formation for hsAX contributes to AX loss of solubility upon debranching. Furthermore, simulations show the intermolecular contacts to cellulose are not affected by arabinose substitution (within the margin of error). Ferulic acid is the GAX moiety found here to bind to cellulose most strongly, suggesting it may play an anchoring role to strengthen GAX-cellulose interactions. The above results suggest highly substituted GAX acts as a spacer, keeping cellulose microfibrils apart, whereas low substitution GAX is more localized in plant cell walls and promotes cellulose bundling.

**Graphical Abstract**



## Introduction

Plant cell walls are of fundamental importance because they provide structural integrity and defense against pathogens. They are also used as a raw material by a growing number of industries, such as in food, paper, textiles and more recently in producing lignocellulosic biofuels and bioproducts (Doblin et al. 2014). The structural and mechanical properties of primary cell walls are determined by their mesoscale architecture and the molecular interactions between their constituent polymers. The stiffest component of primary cell wall is cellulose, unbranched chains of glucose that pack into crystalline microfibrils. Cellulose fibers are embedded in a matrix of hemicellulose polysaccharides. In primary walls of monocots, the major hemicellulose is xylan, for example glucuronoarabinoxylan (GAX) in grasses and arabinoxylan in cereal grains. Xylans contain a xylose (Xyl) backbone with arabinose (Ara) and glucuronic acid (GlcA) side chains. In GAX, Ara residues can be further esterified with ferulic acid (FA). Here, we refer to the presence of Ara side-chains as "substitutions" to the Xyl backbone.

There is considerable evidence that GAX plays an important structural role in grasses (Anders et al. 2012; Darvill et al. 1980; Jones et al. 2003; Ochoa-Villarreal et al. 2012). Highly substituted (hs) GAX, in which the majority of Xyl is substituted with Ara, was found to contribute to cell wall strength and mechanics in maize (Tabuchi et al. 2011). Solubilization of hsGAX leads to increased plastic compliance and to a decrease of the force required to break the wall upon extension (Tabuchi et al. 2011).

Due to their importance in grass wall structure, the interactions between GAX and cellulose have been studied extensively. These interactions include both covalent linkages and noncovalent interactions between hemicellulose and cellulose (referred to here as binding). 2D solid-state NMR (ssNMR), a technique that examines cell walls at high resolution and with minimum perturbation (Wang and Hong 2016; Wang et al. 2014; Wang et al. 2016b; White et al. 2014) of Brachypodium primary cell walls revealed that more rigid GAX approaches cellulose fibers at ~1 nm distances, with contacts involving both the backbone (Xyl) and side chains (Ara and FA) of GAX, while more

mobile GAX fills the interfibrillar space (Wang et al. 2014). Characterizing the binding of cell wall loosening β-expansins has led to a conceptual account of GAX proximities to other grass cell wall polymers. Low substitution lsGAX binds to cellulose, whereas hsGAX does not. Instead hsGAX is assumed to bind to lsGAX, xyloglucan (XyG) and mixed-linkage glucan (MLG) (Wang et al. 2016a). The load bearing role of hsGAX is thus attributed to its binding to matrix polysaccharides (hsGAX, XyG, MLG) that interact with cellulose (Wang et al. 2016a).

Sequential chemical extraction of maize cell walls, a technique that examines cell wall interactions indirectly, leads to hsGAX being removed more easily than the less substituted lsGAX, where the proportion of Ara to Xyl is reduced (Carpita 1983). This suggests more Ara substitution weakens the association of GAX with cell wall components. The chemical extraction experiments are thus consistent with hsGAX being found in the interstitial matrix, whereas relatively unsubstituted GAX is tightly associated with cellulose microfibrils.

*In-vitro* binding experiments of isolated cell wall components show that GAX lacking FA binds weakly to cellulose, but Ara substitution makes the binding even weaker (Carpita 1983; Köhnke et al. 2011). The weak in vitro binding assays seemingly give different results to ssNMR experiments that show extensive GAX-cellulose contact in cell walls (Wang et al. 2014). Molecular dynamics (MD) simulations, which could be considered as in silico binding "experiments", demonstrate that stabilization of GAX-cellulose binding is determined by the position of the substitution linkage, rather than the chemical nature of the substituent, with α1-2 linked substitutions enhancing the binding most (Pereira et al. 2017).

It has been also shown using NMR diffusometry that the solubility and the hydrodynamic properties of wheat AX depend on both degree of substitution and substitution pattern (Köhnke et al. 2011). Wild type highly substituted arabinoxylan (hsAX), is water-soluble. On the other hand enzymatic hydrolysis by preferential hydrolysis of Ara of monosubstituted Xyl leads to low substituted arabinoxylan (lsAX) that has lower solubility than the wild-type. However, the AX solubility doesn't change when enzymatically releasing the Ara of di-substituted Xyl. Solubility of AX was thus found to decrease as the number of unsubstituted Xyl residues increases. However, an

understanding of the effect of Ara substitution on the conformations of arabinoxylans is lacking.

Here MD simulations and small-angle neutron scattering (SANS) are employed to investigate how the degree of Ara substitution affects the 3D structure, stiffness and binding properties of AX and GAX. SANS data revealed the formation of macromolecular inter-chain networks of AX chains, which are exposed to the aqueous solvent. lsAX samples were found to form larger aggregates compared to hsAX. Molecular dynamics simulations of fully solvated lsAX and hsAX single molecules were employed to understand the role of Ara substitution at molecular level. We found that decreasing Ara substitution from 64% to 29% makes AX stiffer. Such increase in stiffness results in aggregation of lsAX, consistent with SANS measurements. Simulations of GAX-cellulose indicate that substitution doesn't affect the non-covalent interactions of GAX with cellulose. FA moieties are found to contribute strongly to cellulose binding in the MD simulations and density functional theory calculations, suggesting FA may play an anchoring role to attach GAX to the microfibrils. The results are consistent with hsGAX playing a spacer role in primary cell walls whereas lsGAX promotes cellulose fiber bundling.

## Methods

### Samples

Three wheat AX of varying Ara substitution and substitution pattern were purchased from Megazyme: a native, high substitution 'hsAX' (P-WAXYM, 61% Ara substitution, containing some doubly substituted D-xylosyl residues), enzymatically debranched 'ls_eAX' (P-EDWAX30, 43% Ara substitution, essentially devoid doubly substituted D-xylosyl residues) and acid debranched 'ls_aAX' (P-ADWAX26, 35% Ara substitution, containing some doubly substituted D-xylosyl residues).

### Small Angle Neutron Scattering

Small-angle neutron scattering data were collected at the Bio-SANS instrument (CG3) situated in the High-Flux Isotope Reactor Facility (HFIR) at Oak Ridge National Laboratory (Oak Ridge, TN) (Heller et al. 2014; Lynn et al. 2006). A single instrument configuration was employed, which consisted of the main detector at 15.5 m from sample, the west wing detector at 1.4° from direct beam, 6 Å neutrons with 13% wavelength spread ($\Delta\lambda/\lambda$) and a sample aperture of 14 mm diameter. SANS data were collected at 25 ºC using 1 mm thick Hellma cylindrical quartz cells spanning a q-range of $0.003< q<0.85$ Å$^{-1}$, where the wavevector, $q = (4\pi/\lambda\sin\theta$, is a function of scattering angle, $2\theta$ and neutron wavelength, $\lambda$). The two-dimensional scattering data were circularly averaged and reduced to one-dimensional scattering profiles using MantidPlot software (Arnold et al. 2014). Buffer containing the same solvent ratio (% $D_2O$) as the sample were similarly measured and subtracted from the sample scattering as part of background correction. SANS intensities were fitted using the Unified Fit model (Eq. 1) (Beaucage 1995) using the Irena package (Ilavsky and Jemian 2009) which runs in IgorPro Software by Wavemetrics Inc.

AX (10 mg) was wetted with 0.08 ml 95% ethanol followed by 0.9 mL 85% $D_20$/15% $H_2O$ and vigorously mixed using an Eppendorf Thermomixer C, at 100°C for approx. 10 min. The solution was cooled to room temperature and the volume was adjusted to 1 ml with water. For SANS measurements, all the samples in 85% of $D_2O$ buffer were enclosed in 1 mm thick quartz cell and measured at room temperature and atmospheric pressure.

## Atomistic Models

### Arabinoxylan models.

The two models of AX were constructed using “doGlycans” package (Danne et al. 2017). Both models have the same backbone of 14 β(1,4)-linked Xyl residues. Their difference lies in the degree of Ara substitution. The low substitution model of AX (lsAX) has 4 Xyl backbone residues covalently linked (at the O3 position) to Ara side-chains, equivalent to a 29% of Ara substitution. The high substitution (hsAX) model has 9 Xyl backbone residues covalently linked (at the O3 position) to Ara side-chains, equivalent to a 64% of

Ara substitution. The AX molecules were solvated in a cubical box of volume ~680 $nm^3$ with ~68,900 TIP3P (Jorgensen et al. 1983) water atoms with counter ions.

Glucuronoarabinoxylan-cellulose models.

Two structural models of GAX polymers were generated using the experimentally-determined average chemical composition of maize GAX as a guide (Tabuchi et al. 2011). Both models have the same backbone of 56 β(1,4)-linked Xyl residues and, similar to AX, their difference lies in the degree of Ara substitution. The high substitution model, called hsGAX, has 38 Xyl backbone residues covalently linked (at the O3 position) to Ara side-chains, equivalent to a 67% degree of Ara substitution. The low substitution model, called lsGAX has 14 Xyl residues bonded to Ara, a 25% degree of Ara substitution. Both models also include one substitution of GlcA to the backbone (at the O2 position) and two FA are bonded to two Ara side-chains. The positions of all substitutions were assigned randomly.

A hexagonal 24-chain (degree of polymerization 70) elementary cellulose fiber (Wang and Hong 2016) was constructed using the cellulose Iβ crystalline structure (Nishiyama et al. 2002). The fiber has two types of faces exposed to the solvent: (200) crystallographic planes, commonly referred to as hydrophobic because non-polar aliphatic hydrogen atoms are exposed, and (110) and (1-10) crystallographic planes, called hydrophilic as they expose the more polar hydroxyl groups.

The GAX-cellulose model has two identical GAX molecules (either lsGAX or hsGAX) initially placed near a cellulose fiber, one GAX molecule close to the hydrophobic face and the other close to the hydrophilic (Figure 1). To check if the results depend on the initial conditions, a "far" model, in which the minimum distance between GAX and cellulose was 0.8 nm and a "near" model, in which the distance was 0.5 nm (data from the two sets of simulations are shown in Figure S4).

All GAX models were hydrated with TIP3P (Jorgensen et al. 1983) water molecules and two sodium ions (Beglov and Roux 1994) per GAX molecule were added to neutralize the system, as FA with pKa = 4.1 is deprotonated (Mota et al. 2008).

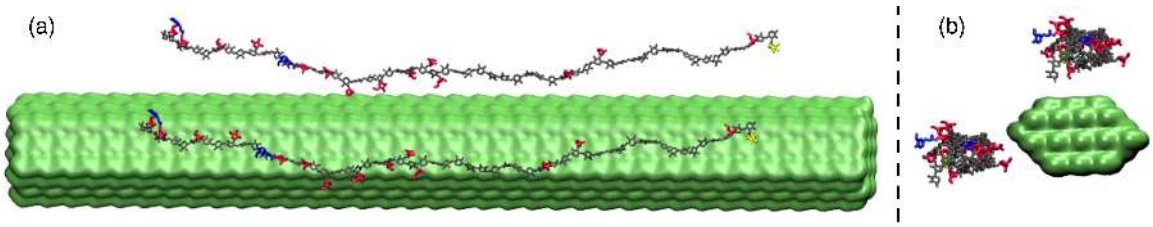



**Fig. 1** Side (a) and axial (b) view of the initial models: cellulose shown in green, Xyl in grey, Ara in red, GlcA in yellow and FA in blue. (c) Side-view of the cellulose fiber with hydrophobic surface chains in red, hydrophilic in blue and core chains in brown.

## Molecular Simulations

### Hamiltonian replica-exchange MD simulations of AX:

CHARMM-GUI (http://www.charmm-gui.org), a web-based graphical user interface was utilized to prepare the simulation systems of AX in explicit water and input files for GROMACS engine. Hamiltonian replica-exchange MD simulations (HREMD) (Bussi 2013; Wang et al. 2011) of lsAX and hsAX were conducted using GROMACS 5.1.4 (Berendsen et al. 1995; Pronk et al. 2013) patched with PLUMED 2.3.4 (Bonomi et al. 2009; Bussi 2013). Specifically, replica-exchange with solute tempering 2 (REST2) was employed (Wang et al. 2011), in which the solute-solute interaction was scaled by a factor $\lambda$, and solute-solvent interaction by $\sqrt{\lambda}$, whereas solvent-solvent interaction was unaltered. Here, $\lambda$ is defined by $T_{eff,0}/T_{eff,i}$, where $i$ = 1, 2,….,n, is the effective temperature of $i^{th}$ highest replica and $T_{eff,0}$ is the effective temperature of lowest replica (i.e., replica 0). This approach allows only the solute molecule to effectively heat up while the solvent remains cold at higher order replicas, such that the number of replicas required to enhance the sampling is greatly reduced compared to the commonly-used temperature replica-exchange MD simulation (Wang et al. 2011). 8 replicas with effective temperature range of 300 to 500 K were used.

The CHARMM36 force field for carbohydrates was employed (Guvench et al. 2008; Guvench et al. 2009). The non-bonded parameters have been shown to reproduce monosaccharide-water interaction energies and distances obtained from Quantum Chemical HF/6-31G(d) calculations. Condensed phase properties were used to validate

the force field: crystal lattice unit cell parameters, aqueous-phase densities, and aqueous NMR ring pucker. The force field was further validated by the calculated free energies of aqueous solvation being in good agreement with experiments. The accurate interaction energies and solvation free energies obtained with this force field provide evidence that the relative free energetics of different monomers are captured accurately in the present simulations.

All bonds involving hydrogen atoms were constrained using LINCS algorithm (Hess et al. 1997). The Verlet leapfrog algorithm was used to numerically integrate the equation of motions with a time step of 2 fs . A cut off of 1.2 nm was used for short-range electrostatics and Lenard-Jones interactions. Long-range electrostatic interactions were calculated by particle-mesh Ewald summation with a fourth order interpolation and a grid spacing of 0.16 nm (Darden et al. 1993). The solute and solvent were coupled separately to a temperature bath of 300 K using modified Berendsen thermostat with a relaxation time of 0.1 ps. The pressure coupling was fixed at 1 bar using Parrinello-Rahman algorithm (Parrinello and Rahman 1981) with a relaxation time of 2 ps and isothermal compressibility of $4.5*10^{-5}$ $bar^{-1}$. The energy of each system was minimized using 1000 steepest decent steps.

All 8 replicas were equilibrated at NVT for 1 ns and then for 5 ns at NPT before attempting an exchange of Hamiltonian every 1 ps based on the imposed detailed balance conditions for the acceptance probability of exchange between neighboring replicas discussed elsewhere (Bussi 2013; Wang et al. 2011). Each of the simulation achieved an average exchange probability of $0.3<p<0.4$, which confirms the sufficient number of replicas used for an efficient sampling by HREMD. Nearly 200 ns long trajectories of lowest replica (i.e., 300 K) of hsAX and lsAX were used in analyses. An aggregate of 3.2 microseconds of simulations were ran. The efficiency of sampling by HREMD is shown in Figure S3, where each replica visits all other replicas frequently over the simulation steps.

Standard MD simulations of GAX-cellulose:

The standard MD simulations of GAX-cellulose systems were performed employing the NAMD software (Phillips et al. 2005) and the CHARMM force fields for carbohydrates (Guvench et al. 2008; Guvench et al. 2009) and lignin (Petridis and Smith 2009) (to model FA). Periodic boundary conditions were applied. The Particle Mesh Ewald electrostatics method (Darden et al. 1993; Essmann et al. 1995) with a grid spacing of 0.11 nm were employed for the treatment of Coulomb interactions beyond 1.1 nm. A force switching function smoothly transitioned the Lennard-Jones forces to zero over the range of 1.0-1.1 nm. Multiple timestep integration was used, with timesteps of 2 fs for bonded and short-range non-bonded forces, and 4 fs for the long-range electrostatic forces. The neighbor list was updated every 10 steps with a pair-list distance of 1.25 nm. The SHAKE algorithm (Ryckaert et al. 1977) was used to constrain all covalent bonds involving hydrogen atoms to their equilibrium values. The simulations were performed in the NPT ensemble, and the temperature was kept constant using the Langevin dynamics algorithm implemented in X-PLOR with a damping coefficient of 5 $ps^{-1}$. The pressure was held constant at 1 atm using the Nose-Hoover Langevin piston algorithm (Feller et al. 1995; Martyna et al. 1994).

DFT calculations

A snapshot (at 150 ns) of the hsGAX-cellulose MD simulation was employed to calculate intermolecular interaction energies between cellulose and GAX by ab initio electronic structure calculations. Two GAX fragments were selected, one with a FA (5 residues in total) and the other without (3 residues in total). The positions of all hydrogen atoms were optimized at the HF/6-31+G* (Ditchfield et al. 1971; Petersson et al. 1988; Petersson et al. 1991; Roothaan 1951) level of theory, while the positions of all the non-hydrogen atoms were fixed. The interaction energy, defined as the energy of the complex minus the energy of the GAX and cellulose, between GAX and all cellulose monomers at a distance less than 0.6 nm from the GAX was calculated at the M052X/6-311++G (Krishnan et al. 1980; McLean and Chandler 1980; Zhao et al. 2006) level of theory.

The following high performance computing facilities were used: CADES at Oak Ridge National Laboratory, and EDISON and CORI at National Energy Research Scientific Computing Center.

## Results

### Small Angle Neutron Scattering

SANS experiments probed the structure of wheat arabinoxylan (AX). Three types of AX were studied: wild type AX (hsAX) that has 61% of Ara substitution, enzymatically debranched AX (ls_eAX) that has 43% of Ara substitution, and acid debranched AX (ls_aAX) with 35% of Ara substitution. hsAX and ls_aAX both contain some doubly substituted Xyl residues, whereas ls_eAX is devoid of doubly substituted Xyl residues.

For AX in solution the scattering intensities *I(q)* were fitted using the Unified Fit model (Figure S2, and Table S1 in SI) (Beaucage 1995) :

$$I(q) = G\, e^{\frac{-q^2 R_g^2}{3}} + B \left\{ \frac{\left(\mathrm{erf}\left(\frac{qR_g}{\sqrt{6}}\right)\right)^3}{q} \right\}^P + I_{bkg} \qquad (1)$$

where $q$ is the wavevector, $G$ and $B$ are $q$-independent constants and $I_{bkg}$ is the constant incoherent background. We focus below on parameters $R_g$ and $P$ that characterize the solution structure of AX (Figure 2, Table 1).

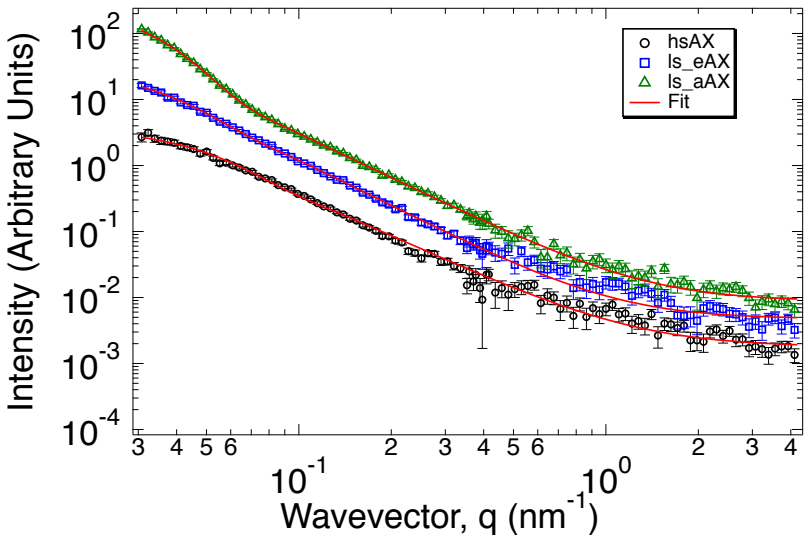



**Fig. 2** Small angle neutron scattering of hsAX (open black circles), ls_eAX (open blue squares) and ls_aAX (open green triangles) samples in 85% $D_2O$ solution, with fits in solid red lines. The ls_eAX and ls_aAX curves were offset by a factor of 2.5 and 5 respectively for the clarity.

$R_g$ is a measure of the characteristic size of the scattering "particles" (here AX). *I(q)* is not flat at the very low $q$ ($q < 0.05$ nm$^{-1}$), which suggests the parameter $R_g$ in Eq. 1 should be interpreted as a lower bound of the radius of gyration of the AX structures. The $R_g$ of hsAX (35.0 ± 1.2 nm) is found to be smaller than that of both ls_eAX and ls_aAX (Table 1). We interpret the trend in $R_g$ as the low-substituted AX having a higher propensity to form polydisperse macromolecular networks/aggregates, which has been reported previously (Köhnke et al. 2011). The absence of a low-$q$ plateau, indicating AX aggregation, was found in the two concentrations we studied (10 and 20 mg/ml, see Figure S1).

At intermediate $q$, the SANS intensities are found to display a power-law dependence on $q$. The power-law or Porod exponent ($P$) describes the degree of entanglement of AX chains, larger $P$ values indicating a more entangled network. $P$ is found here to be smaller than 3 for all the samples (Table 1), indicating that the AX chains form a network that is partially penetrated by the solvent (a "mass fractal"). We find that $P$ decreases slightly, *i.e.* the AX network becomes less entangled, with increasing substitution and with the removal of doubly-substituted Xyl residues.

**Table 1** The radius of gyration ($R_g$) and Porod exponent ($P$) obtained from the Guinier-Porod model fits to the SANS data. The uncertainties in the $R_g$ and $P$ values are calculated from the error evaluation as discussed in Figure S2.

| Samples | $R_g$ (nm) | $P$ |
|---|---|---|
| hsAX | 35 ± 1 | 2.11 ± 0.01 |
| ls_eAX | 49 ± 6 | 2.32 ± 0.07 |
| ls_aAX | 60 ± 1 | 2.23 ± 0.03 |

Molecular Dynamics Simulations

**Conformational Properties of Solvated AX.** We next examine the conformational properties of unbound AX in solution and their dependence on Ara substitution. The bending rigidity/stiffness, i.e. the resistance to bending, of AX is a mechanical property critical to its structural role in plant cell walls. The rigidity of a thermally fluctuating polymer is quantified here by calculating the orientational correlations of unit vectors tangential to Xyl backbone chain:

$$C(s) = \langle \boldsymbol{n_i} \cdot \boldsymbol{n_{i+s}} \rangle = exp\left(-\frac{s}{l_p}\right), \quad (2)$$

where $\boldsymbol{n_i}$ is a unit vector connecting atoms C1 and C4 of Xyl monomer $i$, $s$ is the contour length separating the two residues and <…> indicate time and ensemble averaging. The steeper the decay in $C(s)$, the more flexible the AX backbone is (capable of bending significantly).

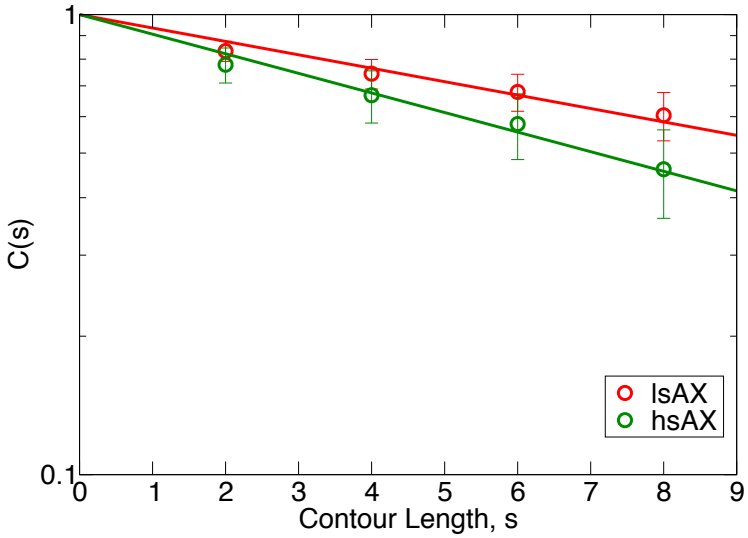



**Fig. 3** Log-linear plot of the tangent correlation function *C(s)* as a function of contour length, *s*, calculated from HREMD simulations of AX molecules in solution. The steeper the decay in *C(s)* the more flexible GAX is. Data are averaged over ~200 ns of HREMD simulation of lowest replica (i.e., $T$ = 300 K). Lines show fits to Eq. 2.

We simulated two types of AX molecules in solution: lsAX and hsAX with 29% and 64% of Ara substitutions respectively. Decreasing the degree of Ara substitution is found to stiffen AX, as evident by the sharper decay in the tangent correlation function, *C(s)* for hsAX compared to lsAX (Figure 3). The persistence length ($l_p$) of an AX chain can be estimated by fitting Eq. 2 to *C(s)*. We find the persistence length of lsAX correspond to 14 Xyl residues and that of hsAX to 10 Xyl residues.

**GAX-cellulose binding.** In standard MD simulations of GAX-cellulose (Figure 1), binding of GAX to cellulose is quantified by calculating the total number of intermolecular GAX-cellulose contacts, NCG (Figure 4). Two simulations were performed with different initial distance between GAX and cellulose. In both simulations, GAX is initially not bound to cellulose (NCG = 0). A gradual increase in NCG is observed as the simulations progress, with NCG reaching a plateau after ~75 ns. The hsGAX and lsGAX25 are found to form statistically similar number of contacts with

cellulose, indicating that increased arabinose substitution does not significantly affect the GAX-cellulose interactions.

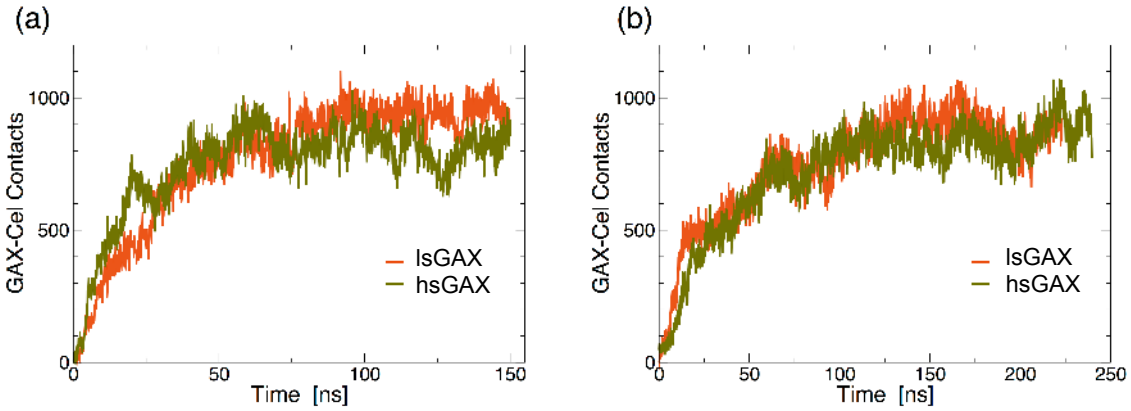
**Table 2** Number of contacts to cellulose per GAX residue, averaged over the last 75 ns of the MD simulations, see Figure S3 (Ferulic acid: FA; glucuronic acid: GlcA; arabinose: Ara; xylose: Xyl).

| | **lsGAX** | **hsGAX** |
|---|---|---|
| Xyl | 5.2±0.6 | 3.6±0.6 |
| Ara | 6.0±0.6 | 4.9±0.4 |
| FA | 10.0±3.8 | 10.3±1.7 |
| GlcA | 8.0±3.8 | 3.2±1.7 |

The backbone (Xyl) and side chain (Ara) per residue contacts to cellulose are not significantly influenced by the degree of Ara substitution (Table 2 and Figure S4). Due to the small number of FA and GlcA residues per GAX molecule (two and one, respectively), the corresponding contacts to cellulose are noisy. Nonetheless, FA is found to bind stronger to cellulose than Xyl and Ara. Previous MD simulations showed $\alpha$1-2 linked (Ara or GlcA) substitutions stabilize the GAX-cellulose binding on the hydrophilic surface (Pereira et al. 2017). Here, only one $\alpha$(1-2) GlcA substitution was simulated and, although GlcA formed interim binding to cellulose in agreement with the previous MD study, the statistics are not adequate to draw firm conclusions.

Density Functional Theory Calculations

The interaction energy between GAX and cellulose was also obtained from single-point DFT calculations, using geometries from the MD simulations at 150 ns (see Methods for details). The interaction energy between cellulose and GAX fragments that include FA are more favorable (-10.6 kJ/mol per GAX monomer) than those of GAX fragments that do not contain FA (-9.1 kJ/mol per GAX monomer). Thus the DFT calculations indicate that (enthalpic) interaction energies contribute to the strong binding of FA to cellulose observed in the MD (Table 2). However, we stress that binding is ultimately determined by free energy that includes entropic contributions.




406

407 **Figure 4.** Total number of GAX-cellulose contacts, defined here as the number of GAX atoms at distance
408 less than 0.3 nm from cellulose, as a function of simulation time. The initial GAX-cellulose distance is 0.5
409 nm in (a), and 0.8 nm in (b).

410

411 **Discussion**

412 Important properties of plant primary cell walls depend on their hierarchical
413 microarchitecture. On the mesoscale and above, structure is controlled by spatial and
414 temporal coordination and localization of biosynthesis of the component biopolymers. At
415 molecular level, the nano-architecture is influenced by interactions determined by the
416 chemical properties of the individual polymers. Here, we have conducted neutron
417 scattering experiments, MD simulations with enhanced sampling and quantum chemical
418 calculations that provide a detailed understanding of how the degree of Ara substitution
419 affects the molecular-level properties of substituted xylans, GAX and AX.

420 The strength of plant cell wall is commonly assessed by biomechanical experiments,
421 during which the wall is extended rapidly, yielding first an irreversible plastic
422 deformation, followed by reversible elastic extension (Tabuchi et al. 2011). GAX is
423 known to contribute to cell wall mechanics: solubilization of hsGAX decreased the
424 plastic compliance, defined as the ratio of strain over stress (the inverse of stiffness), of
425 the cell wall, but it did not change the elastic modulus (Tabuchi et al. 2011).

426 AX structural role in cell walls is determined, in part, by AX binding to other cell
427 wall polymers. AX association with other polymers and with itself is enhanced if its
428 solubility is decreased. Solubility is determined by the free energy difference between

fully solvated single molecules and their aggregated states. SANS experiments (Figure 2) revealed debranching of the Xyl backbone leads to larger AX aggregates that contain more entangled chains. The more entangled lsAX are more flexible and thus have larger entropy than the hsAX. We probed the fully solvated single molecules by enhanced sampling MD simulations (Figure 3) and discovered removal of Ara side-chains increased the intrinsic stiffness and rigidity of AX and thus decreases its entropy. The combination of SANS and MD provides an entropic explanation of the previously-reported insolubility of wheat AX upon enzymatic debranching (Köhnke et al. 2011). lsAX aggregation is favored entropically because the molecules transition from stiff to more flexible conformations.

Previous *in vitro* binding experiments (Carpita 1983) correlate lower Ara substitution with stronger GAX-cellulose binding (Kabel et al. 2007; Selig et al. 2015). Binding is determined by the free energy difference between the bound and unbound states of GAX. The present simulations, which probe the bound cellulose-bound state, show that Ara substitution does not affect significantly the interaction of GAX with cellulose. The present SANS experiments, probing the unbound state, show decreasing Ara substitution increases the self-aggregation of unbound AX, thus lsGAX has lower solubility than hsGAX consistent with previous findings (Bosmans et al. 2014; Köhnke et al. 2011; Pitkanen et al. 2009). Based on the combination of the simulations and SANS results, we attribute the lower binding affinity to cellulose with increasing Ara substitution to an increase of the free energy of the unbound state.

We also found, from both MD simulations and DFT calculations, FA to contribute the most, per residue, to GAX-cellulose binding. This result is in agreement with ssNMR studies of cell walls that found FA and cellulose to be in spatial proximity (<1 nm apart) (Wang et al. 2014). FA may thus play an anchoring function in cell walls, stabilizing GAX-cellulose interactions. Our results may also reconcile ssNMR in vivo experiments (Wang et al. 2014) and in vivo binding assay (Carpita 1983) that seemingly disagree on the extend of GAX-cellulose binding. The *in vivo* experiments, performed on cell walls whose GAX contains FA, show GAX-cellulose are in close spatial proximity (Wang et al. 2014), whereas the *in vitro* experiments, using isolated GAX that does not containing FA, shows GAX only weakly binds to cellulose (Carpita 1983). Considering the strong

binding propensity to cellulose of FA observed in the simulations, the above discrepancy could be explained by the difference in FA incorporation in the samples.

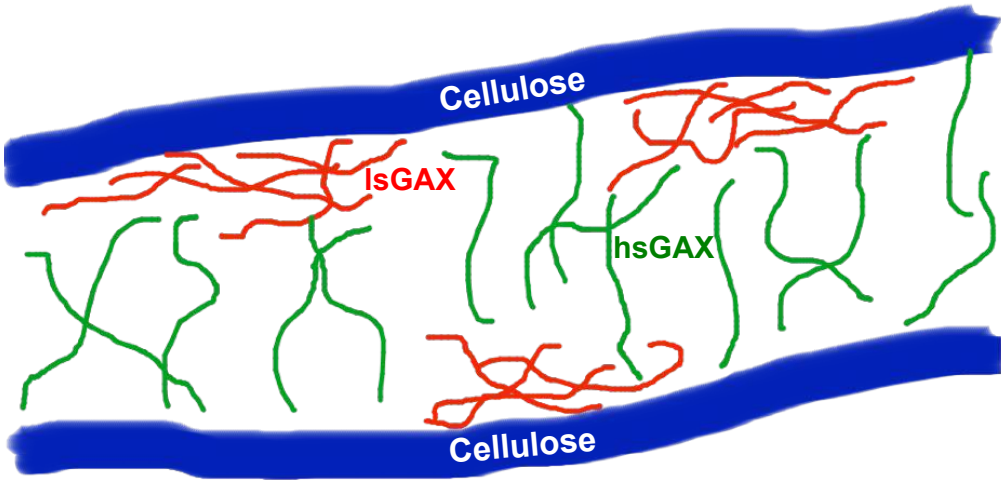



**Figure 5.** Hypothetical arrangement of GAX and cellulose (blue) in grass cell walls. Low substituted lsGAX (red) self-aggregates, and binds more to cellulose (blue) compared to high substituted hsGAX (green).

In addition to being a dominant hemicellulose of the primary walls of grasses, xylan is also a major component of the secondary cell walls in vascular plants. In the latter, substitutions (acetyl, Ara, GlcA) to the xylan backbone are found in evenly spaced Xyl residues (Busse-Wicher et al. 2014; Busse-Wicher et al. 2016; Martinez-Abad et al. 2017), which enables strong binding between cellulose and xylan (Busse-Wicher et al. 2014; Busse-Wicher et al. 2016; Grantham et al. 2017; Pereira et al. 2017; Simmons et al. 2016). Xylan adopts in solution a 3-fold screw conformation, but transitions to a flat ribbon 2-fold screw conformation when binding to cellulose, which facilitates binding to cellulose (Busse-Wicher et al. 2014; Busse-Wicher et al. 2016; Martinez-Sanz et al. 2017). Our study focuses on xylans found in the primary wall of grasses, whose precise substitution pattern in unknown.

Based on the results provided here we provide below a hypothetical arrangement of GAX and cellulose polymers in grass walls (Figure 5). hsGAX forms weak interactions with cellulose and adopts more mobile conformations in solution. hsGAX can therefore act as

a spacing agent that prevents cellulose microfibrils from coming together. Conversely lsGAX is stiffer in solution, interacts more strongly with cellulose and self-aggregates. lsGAX may therefore be more localized in plant cell walls and promote cellulose bundling. The close association of rigid GAX with cellulose obtained by ssNMR (Wang et al. 2014) is consistent with the above conceptual model of grass cell walls.

## Conclusions

Substituted xylans, major hemicelluloses in the primary cell walls of grasses, contribute to cell wall strength and mechanics. They consist of a Xyl backbone substituted with Ara side-chains (the latter sometimes also contain FA). To rationalize the function of AXs, we determine how their chemical compositions, specifically the degree of Ara substitution, affect their structural and binding properties. In solution, AX was found by SANS to form a macromolecular network penetrated by solvent. SANS showed lower substitution AX to form larger networks, whose AX molecules are more entangled, than high substitution AX. HREMD simulations show that individually solvated AX molecules are more flexible when highly substituted. Thus entropic considerations, that the self-aggregated state of low-substituted has favorable configurational entropy, may explain the decreased solubility of lsAX as well as its strong binding to cellulose. MD simulations and density functional theory calculations further show that ferulic acid, an aromatic moiety found in some of the Ara residues, makes strong non-covalent interactions with cellulose fibers. The above results suggest ferulic acid potentially playing an anchoring role to facilitate binding to cellulose.

## Notes

Correspondence should be addressed to L.P. E-mail: petridisl@ornl.gov

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

514 This research was supported by the Center for Lignocellulose Structure and Formation, an Energy
515 Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic
516 Energy Sciences under Award DE-SC0001090. This research used resources of two DOE Office of
517 Science User Facilities: the National Energy Research Scientific Computing Center, a supported
518 under Contract No. DE-AC02-05CH11231, and the High Flux Isotope Reactor at Oak Ridge
519 National Laboratory. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.
520 S. Department of Energy under Contract DE-AC05-00OR22725.

521 **References**

522
523 Anders N et al. (2012) Glycosyl transferases in family 61 mediate arabinofuranosyl transfer onto
524 xylan in grasses. Proc Natl Acad Sci USA 109:989-993
525 Arnold O et al. (2014) Mantid—Data analysis and visualization package for neutron scattering and
526 μ SR experiments. Nucl Instr Meth Phys Res Section A: Accelerators, Spectrometers,
527 Detectors and Associated Equipment 764:156-166
528 Beaucage G (1995) Approximations leading to a unified exponential/power-law approach to small-
529 angle scattering. J Appl Crystallogr 28:717-728
530 Beglov D, Roux B (1994) Finite representation of an infinite bulk system: Solvent boundary
531 potential for computer simulations. J Chem Phys 100:9050-9063
532 Berendsen HJC, van der Spoel D, van Drunen R (1995) GROMACS: A message-passing parallel
533 molecular dynamics implementation. Comput Phys Commun 91:43-56
534 Bonomi M et al. (2009) PLUMED: A portable plugin for free-energy calculations with molecular
535 dynamics. Comput Phys Commun 180:1961-1972
536 Bosmans TJ, Stepan AM, Toriz G, Renneckar S, Karabulut E, Wagberg L, Gatenholm P (2014)
537 Assembly of debranched xylan from solution and on nanocellulosic surfaces.
538 Biomacromolecules 15:924-930
539 Busse-Wicher M et al. (2014) The pattern of xylan acetylation suggests xylan may interact with
540 cellulose microfibrils as a twofold helical screw in the secondary plant cell wall of
541 Arabidopsis thaliana. Plant J 79:492-506
542 Busse-Wicher M et al. (2016) Evolution of xylan substitution patterns in gymnosperms and
543 angiosperms: Implications for xylan interaction with cellulose. Plant Physiol 171:2418-
544 2431
545 Bussi G (2013) Hamiltonian replica exchange in GROMACS: a flexible implementation. Mol Phys
546 112:379-384
547 Carpita NC (1983) Hemicellulosic polymers of cell walls of zea coleoptiles. Plant Physiol 72:515
548 Danne R, Poojari C, Martinez-Seara H, Rissanen S, Lolicato F, Rog T, Vattulainen I (2017)
549 doGlycans-Tools for preparing carbohydrate structures for atomistic simulations of
550 glycoproteins, glycolipids, and carbohydrate polymers for GROMACS. J Chem Inf Model
551 57:2401-2406
552 Darden T, York D, Pedersen L (1993a) Particle mesh Ewald: An N·log(N) method for Ewald sums
553 in large systems. J Chem Phys 98:10089-10092
554 Darvill JE, McNeil M, Darvill AG, Albersheim P (1980) Structure of plant cell walls. Plant Physiol
555 66:1135
556 Ditchfield R, Hehre WJ, Pople JA (1971) Self-consistent molecular-orbital methods. IX. An
557 extended Gaussian-type basis for molecular-orbital studies of organic molecules. J Chem
558 Phys 54:724-728
559 Doblin MS, Johnson KL, Humphries J, Newbigin EJ, Bacic A (2014) Are designer plant cell walls
560 a realistic aspiration or will the plasticity of the plant's metabolism win out? Curr Opin
561 Biotechnol 26:108-114
562 Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle
563 mesh Ewald method. J Chem Phys 103:8577-8593
564 Feller SE, Zhang Y, Pastor RW, Brooks BR (1995) Constant pressure molecular dynamics
565 simulation: The Langevin piston method. J Chem Phys 103:4613-4621

566 Grantham NJ et al. (2017) An even pattern of xylan substitution is critical for interaction with
567 cellulose in plant cell walls. Nat Plants 3:859-865
568 Guvench O, Greene SN, Kamath G, Brady JW, Venable RM, Pastor RW, Mackerell AD (2008)
569 Additive empirical force field for hexopyranose monosaccharides. J Comput Chem
570 29:2543-2564
571 Guvench O, Hatcher E, Venable RM, Pastor RW, MacKerell AD (2009) CHARMM Additive all-
572 atom force field for glycosidic linkages between hexopyranoses. J Chem Theory Comput
573 5:2353-2370
574 Heller WT et al. (2014) The Bio-SANS instrument at the high flux isotope reactor of Oak Ridge
575 National Laboratory. J Appl Crystallogr 47:1238-1246
576 Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for
577 molecular simulations. J Comput Chem 18:1463-1472
578 Ilavsky J, Jemian PR (2009) Irena: tool suite for modeling and analysis of small-angle scattering. J
579 Appl Crystallogr 42:347-353
580 Jones L, Milne JL, Ashford D, McQueen-Mason SJ (2003) Cell wall arabinan is essential for guard
581 cell function. Proc Natl Acad Sci USA 100:11783-11788
582 Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple
583 potential functions for simulating liquid water. J Chem Phys 79:926-935
584 Kabel MA, van den Borne H, Vincken J-P, Voragen AGJ, Schols HA (2007) Structural differences
585 of xylans affect their interaction with cellulose. Carbohydr Polym 69:94-105
586 Köhnke T, Östlund Å, Brelid H (2011) Adsorption of arabinoxylan on cellulosic surfaces: Influence
587 of degree of substitution and substitution pattern on adsorption characteristics.
588 Biomacromolecules 12:2633-2641
589 Krishnan R, Binkley JS, Seeger R, Pople JA (1980) Self-consistent molecular orbital methods. XX.
590 A basis set for correlated wave functions. J Chem Phys 72:650-654
591 Lynn GW, Heller W, Urban V, Wignall GD, Weiss K, Myles DAA (2006) Bio-SANS—A
592 dedicated facility for neutron structural biology at Oak Ridge National Laboratory. Physica
593 B: Condensed Matter 385-386:880-882
594 Martinez-Abad A et al. (2017) Regular motifs in xylan modulate molecular flexibility and
595 interactions with cellulose surfaces plant physiology 175:1579-1592
596 Martinez-Sanz M, Mikkelsen D, Flanagan BM, Gidley MJ, Gilbert EP (2017) Multi-scale
597 characterisation of deuterated cellulose composite hydrogels reveals evidence for different
598 interaction mechanisms with arabinoxylan, mixed-linkage glucan and xyloglucan. Polymer
599 124:1-11
600 Martyna GJ, Tobias DJ, Klein ML (1994) Constant pressure molecular dynamics algorithms. J
601 Chem Phys 101:4177-4189
602 McLean AD, Chandler GS (1980) Contracted Gaussian basis sets for molecular calculations. I.
603 Second row atoms, Z=11–18. J Chem Phys 72:5639-5648
604 Mota FL, Queimada AJ, Pinho SP, Macedo EA (2008) Aqueous solubility of some natural phenolic
605 compounds. Ind Eng Chem Res 47:5182-5189
606 Nishiyama Y, Langan P, Chanzy H (2002) Crystal structure and hydrogen-bonding system in
607 cellulose Iβ from synchrotron x-ray and neutron fiber diffraction. J Am Chem Soc
608 124:9074-9082
609 Ochoa-Villarreal M, Aispuro-Hernández E, Vargas-Arispuro I, Martínez-Téllez MÁ (2012) Plant
610 cell wall polymers: Function, structure and biological activity of their derivatives. In:
611 Gomes ADS (ed) Polymerization. InTech, Rijeka, p Ch. 04
612 Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: A new molecular
613 dynamics method. J Appl Phys 52:7182-7190
614 Pereira CS, Silveira RL, Dupree P, Skaf MS (2017) Effects of xylan side-chain substitutions on
615 xylan–cellulose interactions and implications for thermal pretreatment of cellulosic
616 biomass. Biomacromolecules 18:1311-1321
617 Petersson GA, Bennett A, Tensfeldt TG, Al-Laham MA, Shirley WA, Mantzaris J (1988) A
618 complete basis set model chemistry. I. The total energies of closed-shell atoms and
619 hydrides of the first-row elements. J Chem Phys 89:2193-2218

620 Petersson GA, Tensfeldt TG, Montgomery JA (1991) A complete basis set model chemistry. III.
621 The complete basis set-quadratic configuration interaction family of methods. J Chem Phys
622 94:6091-6101
623 Petridis L, Smith JC (2009) A molecular mechanics force field for lignin. J Comput Chem 30:457-
624 467
625 Phillips JC et al. (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26:1781-1802
626 Pitkanen L, Virkki L, Tenkanen M, Tuomainen P (2009) Comprehensive multidetector HPSEC
627 study on solution properties of cereal arabinoxylans in aqueous and DMSO solutions.
628 Biomacromolecules 10:1962-1969
629 Pronk S et al. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular
630 simulation toolkit. Bioinformatics 29:845-854
631 Roothaan CCJ (1951) New developments in molecular orbital theory. Rev Mod Phys 23:69-89
632 Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of
633 motion of a system with constraints: Molecular dynamics of n-alkanes. J Comput Phys
634 23:327-341
635 Selig MJ, Thygesen LG, Felby C, Master ER (2015) Debranching of soluble wheat arabinoxylan
636 dramatically enhances recalcitrant binding to cellulose. Biotechnol Lett 37:633-641
637 Simmons TJ et al. (2016) Folding of xylan onto cellulose fibrils in plant cell walls revealed by
638 solid-state NMR. Nat Commun 7:13902

639 Tabuchi A, Li L-C, Cosgrove DJ (2011) Matrix solubilization and cell wall weakening by β-
640 expansin (group-1 allergen) from maize pollen. Plant J 68:546-559

641 Wang L, Friesner RA, Berne BJ (2011) Replica exchange with solute scaling: a more efficient
642 version of replica exchange with solute tempering (REST2). J Phys Chem B 115:9431-
643 9438
644 Wang T, Chen Y, Tabuchi A, Hong M, Cosgrove DJ (2016a) The Target of β-Expansin EXPB1 in
645 Maize Cell Walls from Binding and Solid-State NMR Studies. Plant Physiol 172:2107-
646 2119
647 Wang T, Hong M (2016) Solid-state NMR investigations of cellulose structure and interactions
648 with matrix polysaccharides in plant primary cell walls. J Exp Bot 67:503-514
649 Wang T, Salazar A, Zabotina OA, Hong M (2014) Structure and dynamics of brachypodium
650 primary cell wall polysaccharides from two-dimensional 13C solid-state nuclear magnetic
651 resonance spectroscopy. Biochemistry 53:2840-2854
652 Wang T, Yang H, Kubicki JD, Hong M (2016b) Cellulose structural polymorphism in plant
653 primary cell walls investigated by high-field 2D solid-state NMR spectroscopy and density
654 functional theory calculations. Biomacromolecules 17:2210-2222
655 White PB, Wang T, Park YB, Cosgrove DJ, Hong M (2014) Water–polysaccharide interactions in
656 the primary cell wall of arabidopsis thaliana from polarization transfer solid-state NMR. J
657 Am Chem Soc 136:10399-10409
658 Zhao Y, Schultz NE, Truhlar DG (2006) Design of density functionals by combining the method of
659 constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics,
660 and noncovalent interactions. J Chem Theory Comput 2:364-382

661