12-21-2015

Mechanism of lignin inhibition of enzymatic biomass deconstruction

Josh V. Vermaas

Loukas Petridis

Xianghong Qi
University of Tennessee, Knoxville

Roland Schulz
University of Tennessee, Knoxville

Benjamin Lindner

See next page for additional authors

Recommended Citation

Vermaas *et al.* "Mechanism of lignin inhibition of enzymatic biomass deconstruction." *Biotechnology for Biofuels* 8, no. 217 (2015). DOI: 10.1186/s13068-015-0379-8.

Vermaas *et al. Biotechnol Biofuels (2015) 8:217*
DOI 10.1186/s13068-015-0379-8

Biotechnology for Biofuels

RESEARCH Open Access

Mechanism of lignin inhibition of enzymatic biomass deconstruction

Josh V. Vermaas[1,2], Loukas Petridis[1], Xianghong Qi[1,3], Roland Schulz[1,3], Benjamin Lindner[1] and Jeremy. C. Smith[1,3,4*]

Abstract

Background: The conversion of plant biomass to ethanol via enzymatic cellulose hydrolysis offers a potentially sustainable route to biofuel production. However, the inhibition of enzymatic activity in pretreated biomass by lignin severely limits the efficiency of this process.

Results: By performing atomic-detail molecular dynamics simulation of a biomass model containing cellulose, lignin, and cellulases (*Tr*Cel7A), we elucidate detailed lignin inhibition mechanisms. We find that lignin binds preferentially both to the elements of cellulose to which the cellulases also preferentially bind (the hydrophobic faces) and also to the specific residues on the cellulose-binding module of the cellulase that are critical for cellulose binding of *Tr*Cel7A (Y466, Y492, and Y493).

Conclusions: Lignin thus binds exactly where for industrial purposes it is least desired, providing a simple explanation of why hydrolysis yields increase with lignin removal.

Keywords: Biofuel, Lignin, Cel7A, Cellulose crystallinity

Background

Sustainable global economic growth requires the development of technologies that will reduce the environmental footprint of energy consumption, including the adoption of renewable, energy-dense transportation fuels [1]. The production of biofuels from abundant lignocellulosic biomass is a potential alternative to fossil fuels. However, a significant barrier to cost-effective cellulosic biofuel production is the current inefficient hydrolysis of cellulose glycosidic bonds to fermentable sugars by cellulase enzymes [2–4].

Cellulose hydrolysis by cellulases is typically preceded by thermochemical pretreatment of biomass to increase the accessibility of the cellulose substrate to the enzyme. Dilute acid pretreatment removes almost all biomass components apart from the cellulose itself and lignin [5–7], a poly-aromatic amorphous and hydrophobic plant polymer [8]. However, even after pretreatment, enzymatic cellulose hydrolysis remains incomplete [9]. Overcoming this inefficiency presents one of the most important challenges in biotechnology [2–4, 10–13].

There is considerable evidence implicating lignin as a major culprit in reducing cellulase efficiency in pretreated biomass [3, 14–23], though its mechanism of action has not been definitively elucidated. Various lignin-related inhibitory processes have been proposed, including cellulose association with lignin, blocking enzymatic access to cellulose [15–18], and the unproductive binding of the enzymes to lignin [19–23]. Unproductive binding has been proposed to be non-specific and to occur via hydrophobic [19, 22, 23] or electrostatic interactions [24–26], although no direct evidence has been observed for either hypothesis. It is also suspected that the cellulose-binding module (CBM) of cellulases participates in lignin binding, as enzymes containing a CBM have a higher affinity for lignin than those without one [20, 22]. However, an atomic-detailed characterization of how cellulases become inhibited by lignin is currently lacking.

In order to rationally design improved pretreatment processes which minimize the lignin's adverse effect

*Correspondence: smithjc@ornl.gov
[4] University of Tennessee/Oak Ridge National Laboratory Center for Molecular Biophysics, P.O.Box 2008, Oak Ridge, TN 37831-6309, USA
Full list of author information is available at the end of the article

in biofuel production and guide current developments in lignin bioengineering, it is important to understand mechanistically how lignin interferes with cellulose degradation [27–29]. Here, we report molecular dynamics (MD) simulations of a model of a pretreated multi-component biomass system, containing lignin, cellulose fibers of different degrees of crystallinity, and the industrially important [30–32] *Trichoderma reesei* fungal cellulase (*Tr*Cel7A) enzyme. The simulation system models the crowded lignocellulosic environment in which *Tr*Cel7A operates during industrial biomass hydrolysis. The results indicate that lignin associates preferentially with the hydrophobic surface of cellulose, which is also the preferred substrate of *Tr*Cel7A. Lignin is also found to bind preferentially to the CBM tyrosine residues 466, 492, and 493, which have been identified as being critical to cellulose binding [33–38]. Thus, lignin directly and competitively inhibits the recognition mechanism of the cellulase consistent with a competitive inhibition mechanism previously postulated by mutagenesis work and biochemical assays [9, 25, 39]. These atomistic details of the interaction of a cellulase within a crowded biomass environment, including both substrate interactions and lignin inhibition, explain why lignin is such an effective barrier to efficient enzymatic hydrolysis of post-pretreated biomass.

Results and discussion

The simulation specifically investigates the binding of Cel7A to cellulose prior to the enzyme hydrolyzing a glucan chain, and how this binding is affected by the presence of lignin. The simulation model was devised to represent a pretreated biomass system of cellulose and lignin at room temperature upon the addition of cellulolytic enzyme. Other components of biomass, such as pectins and hemicellulose, are assumed to have been removed [5]. As detailed in Sect. "Methods," a large variety of experimental data was used to construct a realistic model. The simulation system consisted of nine cellulose fibers, of which six were crystalline and the other three non-crystalline [40], 54 glycosylated *Tr*Cel7A enzymes, and 468 lignin molecules in explicit solvent. In the starting structure of the system, i.e., prior to the simulation (Fig. 1), no enzymes are bound to the biomass, but there is extensive cellulose–lignin association derived from previous simulations of pretreated biomass [40] which remained virtually unchanged after the addition of the enzymes in the current study. Three different cellulose fiber–lignin distribution combinations were present in the simulation: CH (crystalline cellulose, high lignin coverage), CL (crystalline cellulose, low lignin coverage), and NonC (non-Crystalline cellulose, low lignin coverage). These combinations are analyzed independently

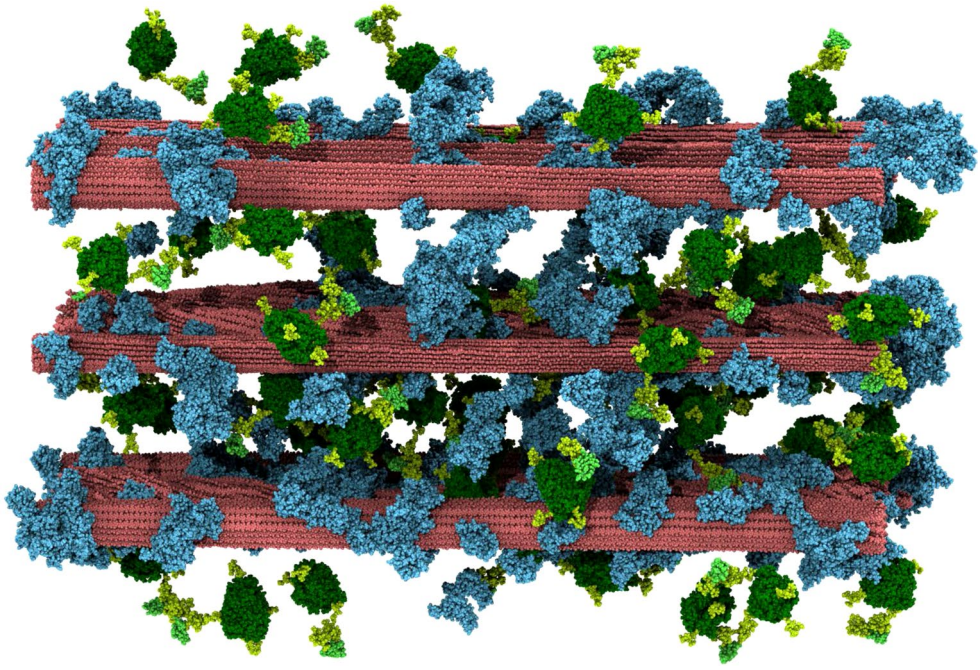

Fig. 1 Side view of the initial state of the lignocellulosic biomass system. Cellulose fibrils are *red*, lignin molecules *blue*, and *Tr*Cel7A enzymes *green*; the CBMs have a lighter color than the CDs, while glycosylations and linker regions are in *pastel green*. An animation from this starting structure is given as an Additional file 1

throughout the text when clear differences were found in the properties observables studied.

Network formation

The intermolecular contacts, a measure of binding thermodynamics and defined in Eq. 1, indicate that during the simulation the degrees of lignin–lignin and lignin–cellulose association do not vary significantly (Fig. 2a), as would be expected for the pre-equilibrated lignocellulose fibrils used here. As the simulation progresses a gradual increase in the number of enzymatic contacts is observed as the enzymes diffuse to the lignocellulose. However, all enzymes are bound to another interaction partner within 600 ns (Fig. 2b), so the growth in the number of enzyme–lignin contacts seen over the second half of the simulation in Fig. 2a arises from enzymes that are already bound optimizing their interfacial area with the lignin.

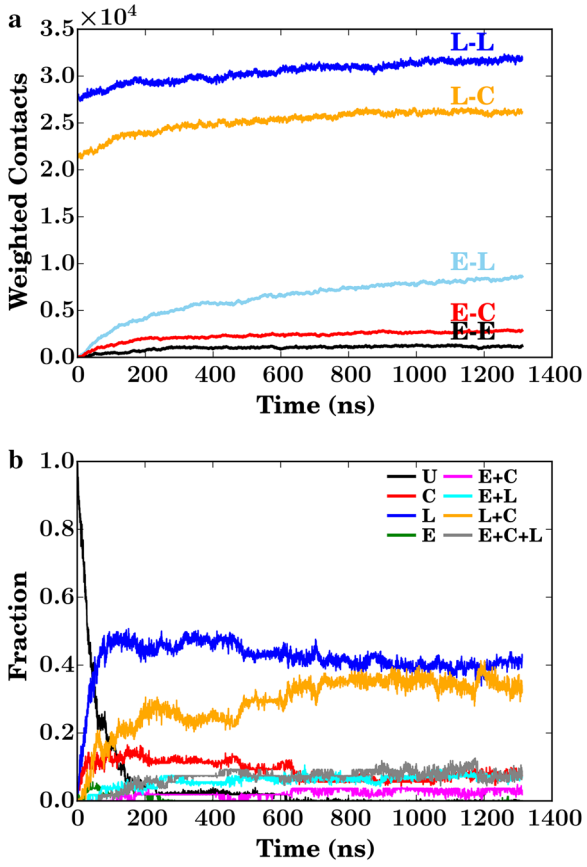

Fig. 2 **a** Contact counts as a function of time between enzyme E, lignin L, and cellulose C molecules. **b** Time traces of the fraction of the 54 enzymes that are unbound, *U*; bound only to cellulose, C; bound only to lignin, L; bound only to other enzymes, E; bound to enzymes and cellulose, *E+C*; bound to enzymes and lignin, *E+L*; bound to lignin and cellulose, *L+C*; bound to other enzymes, lignin and cellulose, *E+C+L*. In this analysis, an enzyme is said to be bound if any of its heavy atoms are within 3.2 Å of a heavy atom in another molecule

The cellulases overwhelmingly interact with either only lignin or both lignin and cellulose. Together, these equally large populations account for approximately 80 % of all enzymes (Fig. 2b). This corresponds to 160 mg of protein bound to 1 g of biomass "solids" (cellulose and lignin), in broad agreement with the experimentally determined cellulase binding capacity of thermochemically pretreated biomass systems, which is 160 mg/g for Douglas-fir softwood [41], 170 mg/g for poplar [42], and 140–150 mg/g for corn stover [7, 43].

The cellulase interactions do not take place in isolation, but rather are part of a crowded mesh formed by the superstructure formed by the biomass constituents (Fig. 3). This shows that lignin mediates the formation of a fully interconnected network of cellulose, lignin, and *Tr*Cel7A, with each molecule linked to all others directly or indirectly. These networks arise spontaneously in the simulations, and are only possible due to the simulation incorporating multiple cellulose fibrils. Within the network, cellulose fibrils act as hubs, i.e., have numerous connections to other molecules. *Tr*Cel7A and lignin

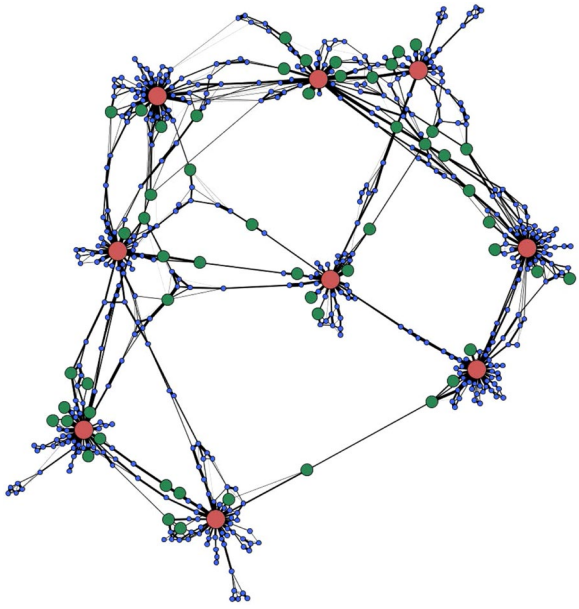

Fig. 3 Schematic representation of the network formed by the individual biomass components at the end of the simulation. Each *circle* represents one element of the system: the large *red circles* are for cellulose fibrils, the *small blue circles* are for lignin molecules, and the intermediate *green circles* are for *Tr*Cel7A enzymes. The *black lines* connecting the components indicate a contact between two components, and the thickness represents the degree of contact (the contact number). The position of the individual particles is arbitrary, with the position determined using the ForceAtlas algorithm of Gephi [45], which treats the connection as springs connecting the elements. An animation of the time-evolution of this representation is given as an Additional file 3

acts as a "glue" connecting these hubs. Within the network, lignin's role depends on its morphology. We identify three types of lignin aggregates (Additional file 2: Figure S1): "sheets," in which lignin monolayers bind to a single cellulose fiber; "piles," in which the lignin aggregates onto a single cellulose fibril but not as a monolayer; and "linkages," in which the lignin aggregates connect cellulose fibrils. If lignin adopts an extended morphology (a sheet or linkage), more surface is exposed, and lignin's propensity to bind to enzymes is increased (Table S1). Therefore, piles are the least effective at trapping enzymes and hence the least inhibitory to cellulase action. It has been shown that increasing the hydrophobicity of lignin reduces its radius of gyration thereby making it more compact [44], which may favor pile formation over other lignin morphologies.

An implication of the existence of lignin-mediated networks is the retardation of enzyme diffusion due to confinement. Indeed, binding to cellulose or lignin leads to a three orders of magnitude slowdown in enzyme translational diffusion, decreasing from an initial $\sim 10^{-6}$ cm^2/s to a final $\sim 10^{-9}$ cm^2/s, and one order in rotational diffusion, from $\sim 10^{6}$ to $\sim 10^{5}$ rad^2/s (Additional file 2: Figure S2). In comparison, the translational diffusion coefficient of proteins in living cells is $\sim 10^{-7}$ cm^2/s [46] and that of bound cellulases processing on a cellulose surface is $\sim 10^{-10} - 10^{-11}$ cm^2/s [47].

Cellulase binding to cellulose in the presence of lignin

Cellulase binding to cellulose is the first step of the mechanism of enzymatic deconstruction. *Tr*Cel7A possesses a typical cellulase multidomain organization, with a large catalytic domain (CD) connected to a CBM via a flexible linker. The enzyme possesses posttranslational modifications, in which the linker is highly *O*-glycosylated and the CD *N*-glycosylated [32, 48]. To obtain a molecular-level description of this binding in the presence of lignin we determined the propensity of the individual enzyme residues to participate in cellulose-*Tr*Cel7A binding and mapped them onto the *Tr*Cel7A structure (Fig. 4a; Additional file 4: Video S1; Additional file 5: Video S2; Additional file 6: Video S3; Table 1). From Fig. 4a, two regions stand out as forming the most contacts to cellulose: three Tyr CBM residues and the linker glycosylation sugars. The linker glycosylations have been previously demonstrated to interact with cellulose [32], although their physiological role has not been fully elucidated. The linker has been suggested to convey resistance to proteolysis [49], increase protein solubility [50], minimize contact between the CD and the CBM [51], and promote binding to cellulose [32]. Here, the glycosylations are found to participate significantly in *Tr*Cel7A binding not only to cellulose, but also to lignin and other *Tr*Cel7A molecules (Table 1).

The flat hydrophobic surface on the CBM formed by three tyrosine residues (Y466, Y492 and Y493) promotes binding to the hydrophobic surfaces of cellulose fibers [33–38]. In the present simulations, the tyrosine residues form extensive contacts with the lignin. Indeed lignin outcompetes cellulose in terms of interacting with these residues (Table 1; Additional file 10).

However, in over half of the trajectories individual enzymes form interactions with the cellulose substrate. Among the 30 enzymes that bind to cellulose within our simulation, there are many that have their substrate tunnel aligned perpendicular to the fibril axis, some of which are only loosely connected via glycosylations to the fibril. A full gallery of all of these interactions is available as an Additional file 11. From our sampling, there are more cases where the substrate tunnel is aligned parallel to

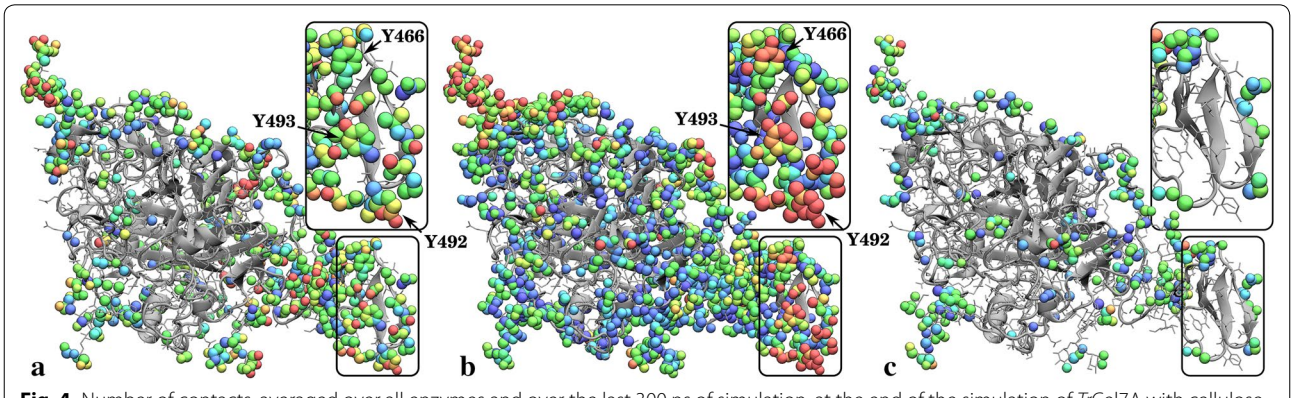

Fig. 4 Number of contacts, averaged over all enzymes and over the last 300 ns of simulation, at the end of the simulation of *Tr*Cel7A with cellulose (**a**), lignin (**b**), and other enzymes (**c**) mapped onto a model of *Tr*Cel7A. Cooler (*blue*) colors indicate fewer contacts, while warmer (*red*) indicate more. These figures are also available as Additional file 4: Video S1; Additional file 5: Video S2; Additional file 6: Video S3 as well as downloadable pdb files where the contact number is in the beta column (Additional files 7, 8 and 9)

Table 1 40 residues of Cel7A interacting most frequently with other enzymes, lignin and cellulose

Contacts with Enzyme		Contacts with Lignin		Contacts with Cellulose	
Residue	Count	Residue	Count	Residue	Count
Man 522	0.7 ± 0.2	ASN 45	6.2 ± 0.6	Man 524	1.5 ± 0.3
ASN 45	0.5 ± 0.3	TYR 492	3.8 ± 0.4	GLY 22	1.1 ± 0.1
Man 523	0.5 ± 0.2	GluNAc 519	3.4 ± 0.3	Man 518	1.0 ± 0.3
Man 534	0.5 ± 0.2	Man 522	3.3 ± 0.5	Man 522	1.0 ± 0.3
Man 524	0.5 ± 0.2	Man 524	2.9 ± 0.5	Man 499	1.0 ± 0.2
Man 513	0.5 ± 0.1	TYR 466	2.4 ± 0.3	SER 21	0.9 ± 0.1
Man 502	0.4 ± 0.1	LEU 6	2.4 ± 0.3	Man 502	0.9 ± 0.3
Man 505	0.4 ± 0.1	Man 520	2.3 ± 0.3	Man 513	0.9 ± 0.2
Man 530	0.4 ± 0.2	ILE 472	2.2 ± 0.2	Man 529	0.9 ± 0.3
Man 536	0.3 ± 0.1	TYR 493	2.1 ± 0.3	Man 530	0.9 ± 0.3
Man 520	0.3 ± 0.1	Man 523	2.1 ± 0.4	ASN 45	0.8 ± 0.1
GLN 410	0.3 ± 0.1	Man 521	1.8 ± 0.3	TYR 492	0.8 ± 0.2
SER 409	0.3 ± 0.1	ASN 490	1.8 ± 0.2	GluNAc 519	0.7 ± 0.2
Man 512	0.3 ± 0.1	Man 506	1.7 ± 0.2	Man 508	0.7 ± 0.1
TYR 252	0.3 ± 0.1	ARG 450	1.6 ± 0.2	Man 503	0.7 ± 0.1
Man 500	0.3 ± 0.1	GLY 473	1.6 ± 0.2	Man 505	0.7 ± 0.2
Man 532	0.3 ± 0.1	Man 528	1.5 ± 0.4	Man 520	0.7 ± 0.1
Man 533	0.3 ± 0.1	ASN 384	1.5 ± 0.4	Man 501	0.6 ± 0.2
Man 535	0.3 ± 0.2	HSP 465	1.5 ± 0.3	Man 523	0.6 ± 0.2
Man 518	0.2 ± 0.1	LEU 489	1.5 ± 0.2	Man 528	0.6 ± 0.3
ASN 441	0.2 ± 0.1	Man 505	1.3 ± 0.3	ASN 436	0.6 ± 0.1
ASN 70	0.2 ± 0.1	GLY 439	1.3 ± 0.2	TYR 466	0.6 ± 0.1
Man 517	0.2 ± 0.1	PRO 438	1.3 ± 0.2	Man 536	0.6 ± 0.3
HSP 465	0.2 ± 0.1	THR 44	1.3 ± 0.2	Man 507	0.5 ± 0.1
GLN 7	0.2 ± 0.1	THR 5	1.2 ± 0.2	SER 156	0.5 ± 0.1
GLU 1	0.2 ± 0.1	Man 499	1.2 ± 0.2	Man 521	0.5 ± 0.1
Man 509	0.2 ± 0.1	Man 513	1.1 ± 0.3	GLY 473	0.5 ± 0.1
THR 44	0.2 ± 0.1	Man 510	1.1 ± 0.3	THR 296	0.5 ± 0.1
LYS 69	0.2 ± 0.1	GLN 495	1.1 ± 0.2	TYR 493	0.5 ± 0.1
VAL 488	0.2 ± 0.1	Man 516	1.1 ± 0.3	THR 310	0.5 ± 0.1
GLN 487	0.2 ± 0.1	Man 508	1.0 ± 0.2	Man 504	0.5 ± 0.2
GluNAc 531	0.2 ± 0.1	ASN 436	1.0 ± 0.2	Man 506	0.5 ± 0.2
GluNAc 519	0.2 ± 0.2	SER 47	1.0 ± 0.3	Man 510	0.5 ± 0.2
Man 521	0.2 ± 0.1	GLN 7	1.0 ± 0.2	Man 514	0.5 ± 0.1
GLU 190	0.2 ± 0.1	PRO 451	1.0 ± 0.1	Man 534	0.4 ± 0.3
LEU 6	0.2 ± 0.1	SER 320	1.0 ± 0.2	Man 500	0.4 ± 0.2
Man 529	0.2 ± 0.1	Man 500	1.0 ± 0.3	SER 482	0.4 ± 0.1
SER 87	0.2 ± 0.0	LYS 69	1.0 ± 0.2	ILE 472	0.4 ± 0.1
GLY 439	0.2 ± 0.1	ALA 100	1.0 ± 0.2	ILE 426	0.4 ± 0.1
ASN 200	0.2 ± 0.1	ASN 270	0.9 ± 0.1	LYS 354	0.4 ± 0.1
Cum. Sum	11.4		69.6		26.6
Tot. Sum	21.8		151.9		50.6

The values are the average number of contacts a cellulase residue makes with other cellulases (enzyme), lignin and cellulose for simulation times $t > 1000$ ns

The residues are color-coded based on their location within Cel7A

Blue text indicates a residue that is part of the CD, red of the CBM and black of the linker

Lighter text in lower case indicates a glycosylation (sugar) monomer, while bold upper case is an amino acid residue

The cumulative sum of the top 40 contacts and the total sum of all contacts are reported in the final two rows

the cellulose fibril than where it is anti-parallel (Additional file 2: Figure S4). The observed preference toward a parallel orientation would facilitate processive binding, although we can identify no clear mechanism as to the origins of the preferential parallel orientation. It is possible that this orientation is enforced by the directionality of the CBM, as has been previously postulated [38, 52]. However, given how few CBMs are actually bound to cellulose (see the gallery available online provided as Additional file 11), this cannot be determined based on our simulations.

Cellulose association with lignin

The cellulose surface is crowded. Nearly a quarter of the total cellulose surface area is consistently covered by lignin, significantly reducing the area accessible to the enzymes (Fig. 5a). In addition, the presence of lignin molecules on the cellulose surface is likely to interfere with the processive mechanism of cellulose hydrolysis [31], reducing the distance an enzyme bound to cellulose can travel before its path is blocked by a lignin molecule (Fig. 5b).

Non-crystalline cellulose was engaged in twice as many contacts with the enzyme per fibril than does the crystalline polymer (Fig. 6a), which may be due in part to a reduced affinity of non-crystalline cellulose for lignin [40]. The reduced affinity in turn increases the surface area available for enzymatic binding, and in fact the non-crystalline cellulose surface has comparatively little lignin coverage (Table 2). A second factor favoring enzymatic binding to non-crystalline cellulose is the accessibility of surface cellulose hydroxyl groups, which account for more than half of the cellulose–enzyme contacts (Fig. 6b); a larger fraction of these is buried in crystalline cellulose than in the non-crystalline form. Due to the lower lignin coverage of non-crystalline cellulose, enzymes can, in principle, process this form for a larger distance before being blocked by lignin (Fig. 5b).

Chains of crystalline cellulose on hydrophobic surfaces can be more readily decrystallized than those on hydrophilic surfaces [53]. The present simulations reveal a preferential association of both lignin and the enzymes with the hydrophobic face of the cellulose fibers (for a chain-by-chain analysis see Additional file 2: Figure S5). Lignin contacts lead to the hydrophobic chains of crystalline cellulose being only poorly accessible, with 30–40 % of their total surface area covered by lignin and only ~3 % covered by enzymes (Table 2). In contrast, in the non-crystalline fibers, the lignin contact area with the "hydrophobic" face is reduced by about half to ~18 %, while the proportion in contact with cellulases nearly doubles (Table 2). Moreover, the trend line between lignin and enzyme coverage of cellulose for the hydrophobic faces (Fig. 7a) has a negative slope, confirming competitive binding.

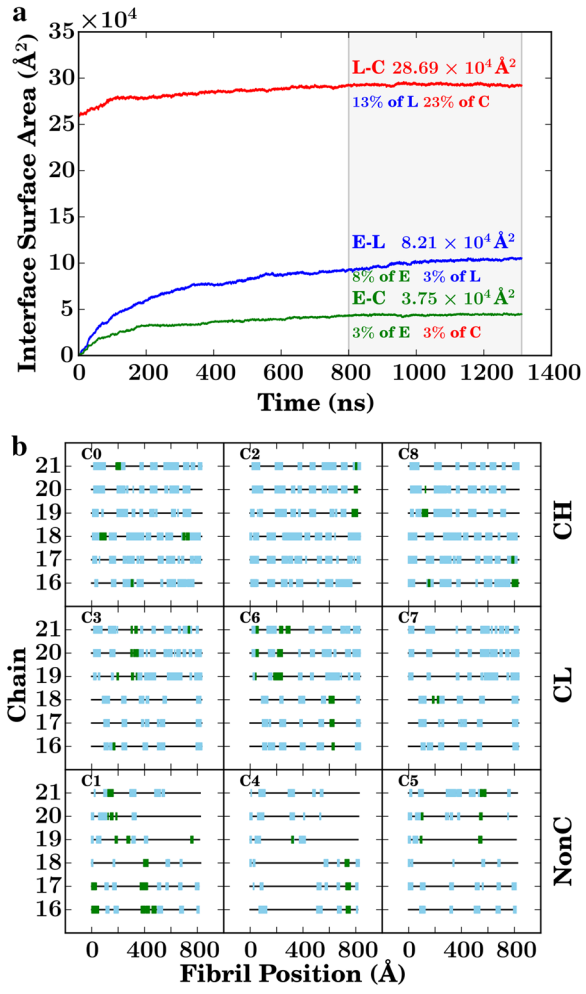

Fig. 5 **a** Interface surface area for cellulose (*C*), lignin (*L*), and enzymes (*E*), their means values (for $t > 800$ ns) labeled above the *curves*. The % fraction of interface area over the total surface area of a species is also labeled below the *curves*. **b** Pictorial representation of the final configuration of the simulation, showing the positions of lignins (*blue*) and enzymes (*green*) on the hydrophobic surface of the nine cellulose fibrils (*black line*). The average "procession length" (distance along the fibril between two lignin clusters) depends on the type of fibril. CH fibrils have the shortest procession lengths (3.5 nm), CL fibrils intermediate (5.5 nm), and NonC the longest (9.2 nm)

Unproductive binding of enzyme to lignin

Enzymes that bind irreversibly to lignin are prevented from binding to their cellulose substrate, such as the example configuration shown in (Fig. 8). The most probable lignin–enzyme contacts involve either CBM residues or glycosylation sugars on the CD (Fig. 4b and Table 1). Three CBM tyrosine residues (Y466, Y492, Y493) that are known to recognize and bind to cellulose [33–38] play an outsized role in the lignin–enzyme association process. In the simulations, the probability of these residues binding to lignin is approximately five times higher than

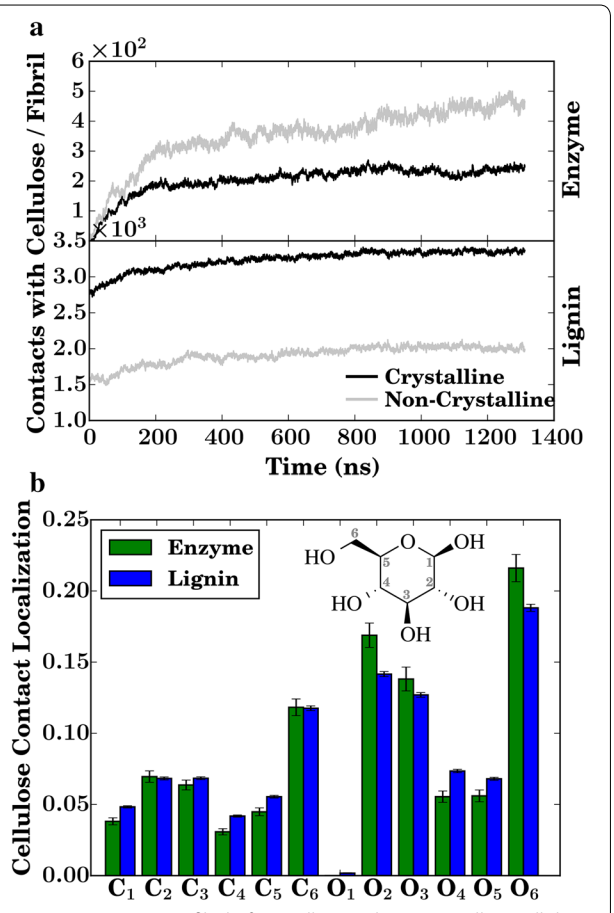

Fig. 6 **a** Contacts per fibril of crystalline and non-crystalline cellulose with the enzyme and with lignin. **b** Normalized number of contacts between any specific cellulose heavy atom and lignin and enzymes

Table 2 Total fibril cellulose surface area (A_T), cellulose–enzyme contact area (A_E), cellulose–lignin contact area (A_L), and their corresponding ratios (A_E/A_T) and (A_L/A_T) for the three initial cellulose–lignin fibril combinations: CH (crystalline cellulose, high lignin coverage), CL (crystalline cellulose, low lignin coverage), and NonC (non-crystalline cellulose, high lignin coverage)

	A_T (10^4 Å^2)	A_E (10^4 Å^2)	A_L (10^4 Å^2)	A_E/A_T	A_L/A_T
CH	***6.51***	***0.18***	***3.02***	***0.03***	***0.46***
	6.74	*0.15*	*1.33*	*0.02*	*0.20*
CL	***6.51***	***0.23***	***2.11***	***0.04***	***0.32***
	6.74	*0.24*	*0.97*	*0.04*	*0.14*
NonC	***6.57***	***0.30***	***1.16***	***0.05***	***0.18***
	7.76	*0.36*	*1.17*	*0.05*	*0.15*

Rows with "bold italic" background correspond to hydrophobic surfaces, while those with "italic" background correspond to the hydrophilic cellulose surfaces

The quantities reported here are the averages over the last 500 ns

their binding to cellulose (Fig. 9). Figure 7b also indicates that, for the most part, the CBM Y466 and Y493 residues interact exclusively with either lignin or cellulose due to geometrical constraints, further suggesting that binding to lignin indeed impedes binding to cellulose. This is shown in another way in Fig. 7, which demonstrates that an individual residue is only rarely in contact with both lignin and cellulose. Taken together, these findings imply a competitive inhibition mechanism of *Tr*Cel7A, in which the binding of lignin to the CBM Tyr residues prevents cellulose recognition.

To obtain further information on the Tyr-lignin binding we examined the stacking interactions of the aromatic side chains of the Tyr residues as determined by the angle γ between the planes of the tyrosine and the lignin/cellulose rings [54]. For the Tyr-cellulose stacking, the two rings are almost parallel, with a relatively narrow distribution peaked at $\gamma \simeq 30°$ that deviates from that that would be obtained in the absence of an angular energetic preference (Fig. 9b) [54]. However, for the interaction of the Tyr residues with the phenolic rings of lignin γ has a broader distribution, which is more similar to what would be expected if there were no intrinsic angular energetic preference. This suggests enthalpy plays a more significant role in determining the orientation preferences of Tyr–cellulose than Tyr–lignin interactions.

It has been suggested that enzymes may become denatured on the lignin surface [9]. However, in the ∼μs timescales examined here, no clear trend was observed between the average residue root mean square fluctuation, an approximate measure of the propensity to denature, and the number of residue-lignin contacts (Additional file 2: Figure S6). Rather than denaturing, the enzymes compact to a mean radius of gyration of 24.8 ± 1.0 Å (Additional file 2: Figure S7) over the course of the simulation, in line with experimentally determined radius of gyration for Cel7A in solution of 26.1 ± 2.1 Å [55].

We find that the interactions lignin makes with other lignin molecules, cellulose, and cellulases are qualitatively different. Although lignin is hydrophobic overall due to its phenolic rings, monolignols also contain a flexible three-carbon (C_7–C_9) chain with hydroxyl groups (Additional file 2: Figure S8). Inter-lignin association is dominated by interactions between the rings, defined here as involving atoms C_1–C_6, O_3, O_4 , and C_{10} (Fig. 9c; Additional file 2: Figure S8). In contrast, enzyme association with the lignin flexible chains (C_7–C_9 and O_7–O_9) is as frequent as with the lignin rings. Finally, when associating with cellulose lignin interacts mostly via its flexible chain atoms. Thus it is not simply a matter that either

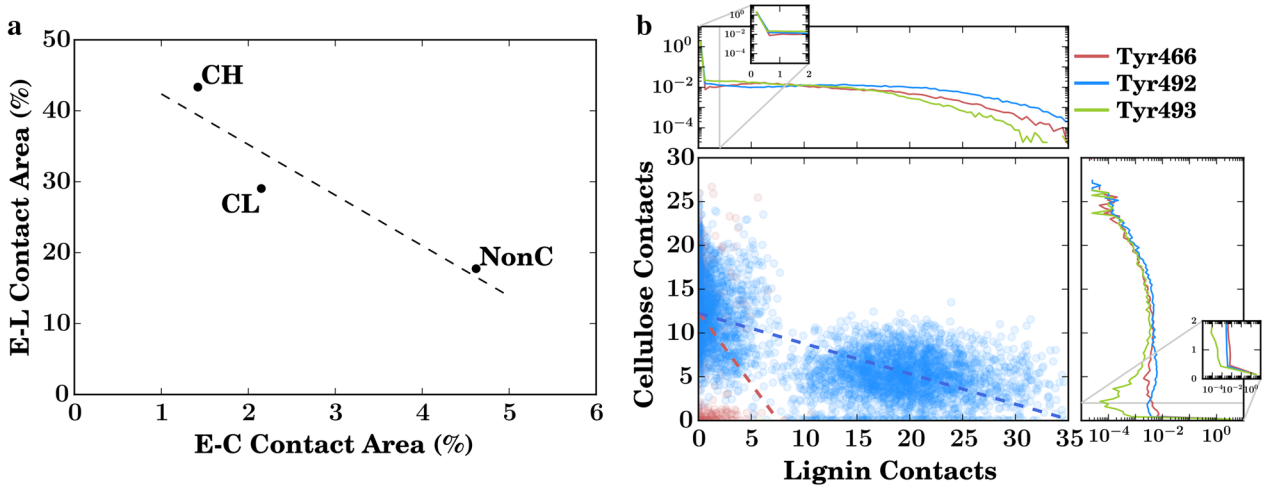

Fig. 7 **a** Fraction of hydrophobic cellulose covered by lignin and enzymes per cellulose fibril type. Individual fibril types are labeled. The *dotted line* is a linear regression to the data. This contains the same information as Table 2. **b** Comparison of the number of simultaneous contacts between the specific CBM tyrosine residues, with a scatterplot in the main panel, and log-probability distributions for direct comparisons along each axis

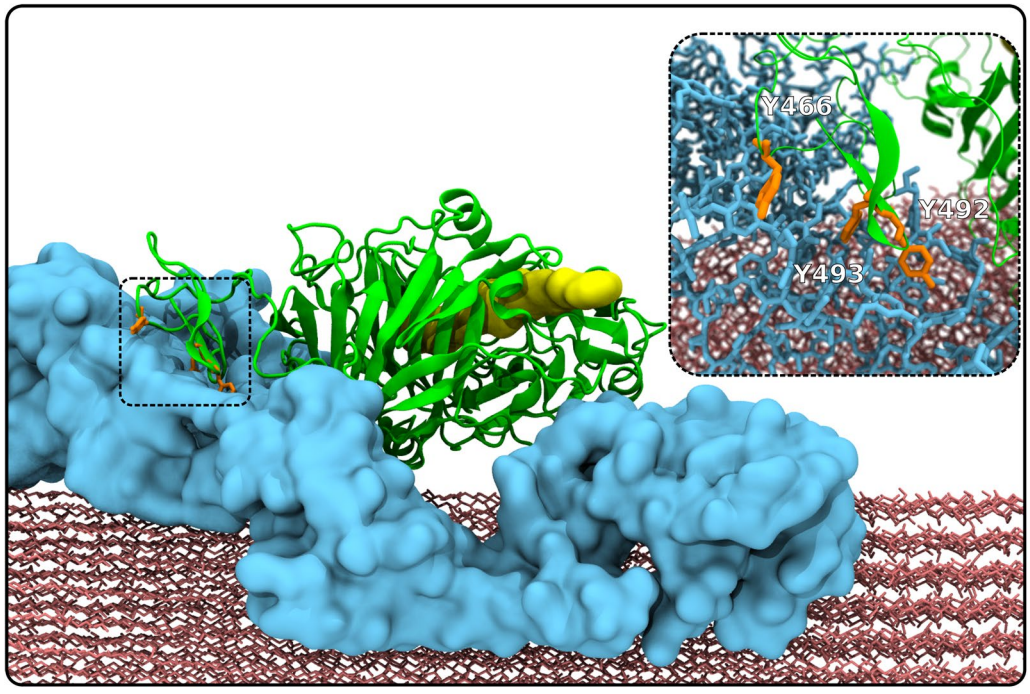

Fig. 8 Snapshot of the simulation in which *Tr*Cel7A (*green cartoon*) is bound unproductively to a lignin cluster (*blue surface*) on a cellulose fiber (*red*). The CBM residues Y466, Y492, and Y493 are *orange*. The location CD catalytic tunnel is shown by a *yellow spacefilling* representation, and is provided for reference. No cellulose was within the tunnel at any point during the simulation, as the complete fibrils did not decrystallize. The *inset* is an enlarged image delineated by the *dotted rectangle*, which highlights the Tyr (*orange*)–lignin (*blue*) interactions. A gallery of images showing the cases where *Tr*Cel7A enzymes interact with cellulose are provided in the supplementary information†

ring-mediated hydrophobic [19, 22, 23] or hydroxyl-mediated electrostatic interactions[24–26] that drive unproductive binding to lignin, but rather both elements contribute to the overall binding.

Conclusions

Atomistic MD simulations of a multi-component system of cellulose, lignin, and an industrially important cellulase, *Tr*Cel7A, described here have led to a mechanistic

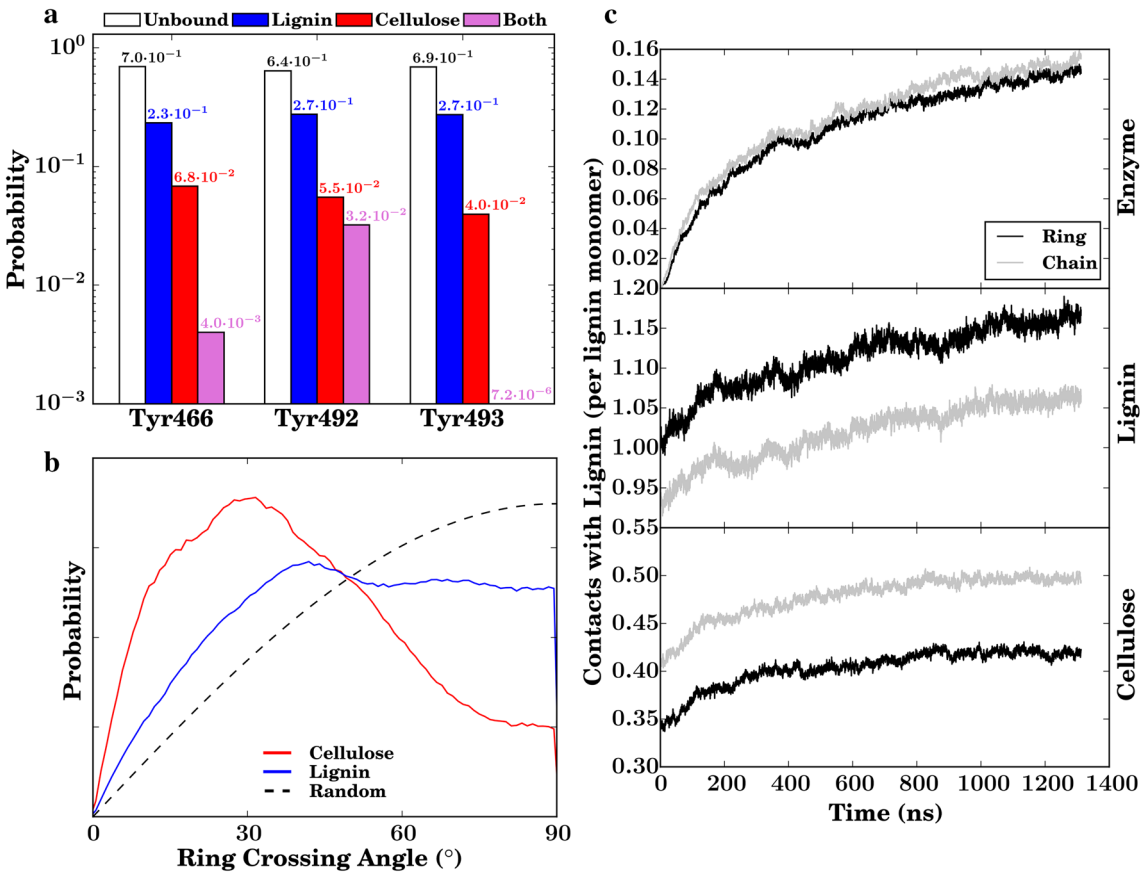

Fig. 9 **a** Probabilities of the three CBM Tyr residues (466, 492, and 493) being contact in contact to only lignin, only cellulose, both lignin and not bound to either (unbound). **b** The crossing angle between the ring normals of the three CBM Tyr residues (466, 492, and 493) and the closest (within 5 Å) biomass ring (the glucose ring of cellulose or the phenolic ring of lignin). The *dotted lines* are distributions that would be obtained without an angular energetic preference from a random distribution. **c** Number of contacts per lignin residue with the enzyme (*top*), other lignins (*middle*), or cellulose (*bottom*). Contacts are labeled as "ring" when involving the lignin atoms C_1–C_6, O_3, O_4, and C_{10}, while "chain" involves atoms C_7–C_9, O_7–O_9

understanding of how lignin in biomass systems impedes binding of cellulase enzymes to cellulose, thus hindering hydrolysis. Lignin is known to directly associate with cellulose and restrict its hydrolysis by cellulases [15, 16, 18]. The present simulations confirm the binding of lignin to cellulose, which decreases both the surface area available for enzymatic binding (Figs. 5a, 6a) and the length of the cellulose chain that can be processed before a lignin blocks its path (Fig. 5b) [18, 31]. Furthermore, lignin is found to bind preferentially to the hydrophobic faces of cellulose (Table 2), as does *Tr*Cel7A [36, 56], amplifying the inhibitory effect. Importantly, the relationship between lignin and enzymatic binding (Fig. 7a) indicates a competitive binding mechanism, in which both enzyme and inhibitor (lignin) bind favorably to the substrate (cellulose). The simulations thus establish a link between cellulose accessibility to cellulases, a key physical property influencing pretreated biomass hydrolysis [57], and cellulose–lignin association.

Secondly, *Tr*Cel7A is also known to bind unproductively to lignin, further limiting its ability to hydrolyze cellulose [19–23]. The present simulations confirm this and provide atomic details of the interactions. Lignin forms specific interactions with those Tyr residues (Y466, Y492 and Y493) on the CBM that have been shown to anchor the enzyme to its cellulosic substrate (Fig. 8; Table 1). The relationship between Tyr binding to lignin and cellulose (Fig. 7b) indicates a second mechanism for competitive inhibition, in which specific binding of the inhibitor (lignin) to the recognition site on the enzyme (CBM) blocks the enzyme substrate binding. The Tyr–lignin interactions may be particularly difficult to engineer away in the enzyme, as mutations to the CBM that might disrupt the interaction with lignin will likely also reduce the affinity of the CBM for cellulose. Engineering the lignin within biomass may be a better approach, possibly by making it more hydrophobic such that it compacts [44] and presents a smaller interaction surface area.

In conclusion, the present study furnishes a detailed description of interactions of a cellulase in a model crowded, pretreated, lignocellulosic environment. Lignin impedes enzymatic action by two competitive binding processes, the molecular bases of which are described here: binding to the hydrophobic face of cellulose, the preferred substrate of *Tr*Cel7A; and specific binding to the tyrosine residues of the CBM that recognize and bind cellulose. Lignin thus binds exactly where for industrial purposes it is least desired, providing a simple explanation why hydrolysis yields increase with lignin removal. These findings explain why lignin is so effective at blocking cellulose hydrolysis by *Tr*Cel7A. This molecular-level description may be used to rationally optimize biofuel production processes which minimize lignin interference. This could, for example, be achieved by pretreatments that lead to non-crystalline cellulose, which associates less with lignin than the crystalline form.

Methods

Model

A 23.7-million atom, multi-component simulation model was build to represent a pretreated biomass system of cellulose and lignin at room temperature upon the addition of cellulolytic enzyme. The model consists of cellulose fibers, lignin molecules, and Cel7A cellulases. Other components of biomass, such as pectins and hemicellulose, are assumed to have been removed by dilute acid pretreatment [5].

Hexagonal cellulose fibers were constructed, each containing 36 glucose chains [58] of degree of polymerization (d.p.) 160. Pretreated cellulose has a d.p. $\gtrsim$140 [59]. Since cellulose in pretreated biomass exists in both highly crystalline and more amorphous forms, both types of fibers were modeled: six crystalline fibers, obtained from the crystal structure of cellulose $I\beta$ [60]; and three non-crystalline, obtained by simulating crystalline cellulose at 650 K for 1 ns [40].

468 lignin molecules (52 per cellulose fibril) were included, comprising 18 copies each of 26 distinct lignin molecules obtained from previous studies [61, 62]. All lignin molecules consisted of 61 monolignol monomers, and the lignin molecular weight, degree of branching, monomer, and linkage composition are consistent with those of softwood lignin [61]. Briefly, structural models of the individual lignin molecules were generated by first deriving the bonding topologies of the molecules and subsequently generating the 3D coordinates. To generate the topologies, a variety of experimental data on the bulk chemical composition of softwood lignins was used. Softwood lignins are composed mainly of G units [63–65] and therefore only G units were used here. The average linkage composition used is typical of softwoods [65, 66]: β-*O*-4′ 50 %, 5-5′ 30 %, α-*O*-4′ 10 %, and β-5′ 10 %. The models also contain equal amounts of left- and right-handed β-*O*-4′, α-*O*-4′ and β-5′ linkages, so as to make the molecules optically inactive, in accord with experiment [67]. Each molecule comprised 61 G units leading to a molecular weight of 13 kDa, within the experimentally determined range [68]. Finally, an average crosslink density of 0.052, or 3.2 branch points per 61 monomers, was used, again as has been derived experimentally, for spruce wood [69]. The number of branch points per molecule and their location along the chain were assigned randomly using a computer algorithm: the resulting 26 distinct lignin topologies have varying degrees of branching: one molecule has zero branch points, three have one, four have two, six have three, seven have four, three have five, and one molecule has six.

Subject to the constraints imposed by the above experimental data, random primary structures of lignins were generated, producing 25 molecules that are different from each other but consistent with the average chemical properties of softwood lignin. For example, although for all 26 molecules 50 % of linkages are of the β-*O*-4′ kind, the positions of these linkages varies between molecules, as does the position of the branch points, and the lengths of the branches are different. Relaxed 3D structures for the lignin molecules were obtained from previous simulations [40].

The starting lignin and cellulose coordinates were obtained from the final state of previous MD simulations of pretreated lignocellulose, in which 52 lignin molecules aggregated on the surface of individual cellulose fibers [40]. Three states were used here, obtained from the end states of three prior simulations [40]: crystalline cellulose with high lignin coverage (CH), crystalline cellulose with low lignin coverage (CL), and non-crystalline cellulose with low lignin coverage (NonC). (In our previous work, CH, CL, and NonC were denoted NC, FC, and FN, respectively [40]). Nine cellulose fibers (and the lignin molecules associated with them) were placed parallel to each other, such that all cellulose fibers (three NC, three FC, and three NonC) have the same neighbors when periodic boundary conditions are applied.

54 identical *tr*Cel7A enzymes were constructed using the crystal structure of the catalytic domain [70] and the NMR structure of the CBM [33]. The linker sequence was built as a linear segment connecting the two domains. *N*-glycans were attached to residues 45, 270, and 384 of the catalytic domain, and *O*-glycans were attached to the linker, in a manner consistent with experimental data [48, 71]. This glycosylation pattern is that suggested by mass spectrometric methods [48, 71]. The 54 enzymes were placed in the unoccupied space of the simulation box using a local algorithm that randomly varied their positions and

orientations until placements were achieved without steric clashes with other macromolecules already in the system. The system was solvated by 7.1 M water molecules and was subsequently neutralized using Na ions.

The relative mass ratio $R_{c:l}$ of cellulose to lignin is 1.5 g cellulose per g of lignin, which is typical of thermochemically pretreated biomass: $R_{cl} \approx 1.8 - 1.9$ for pretreated corn stover [7, 72, 73], $R_{cl} \approx 1.7$ for pretreated switchgrass [74], $R_{cl} \approx 1.2$ for pretreated poplar [75], and $R_{cl} \approx 0.9 - 1.2$ for pretreated pine [76, 77]. Overall, the absolute concentration of the solutes was higher than in typical enzyme binding experiments. For example, the cellulose concentration was 60 g/L (6 % w/v), while that commonly employed in enzyme binding is typically ~ 10 g/L (1 % w/v) [7, 42, 77]. The enzyme loading corresponds to 230 mg protein/g of biomass solids (cellulose and lignin), which is within the range typically used in enzyme binding experiments (0–2000 mg/g) [7, 42, 77].

The dimensions of the simulation box are 95 nm × 62.5 nm × 62.5 nm. The overall size of the system is determined by several requirements. The first is to match physical characteristics of the system, i.e., that pretreated cellulose fibers have lengths $\gtrsim$100 nm [59], the lignin-to-cellulose ratio and the typical enzyme loading. The second is to obtain statistically meaningful enzyme binding propensities, which require ~50 *tr*Cel7A molecules to be simulated. Finally, the system consists of highly heterogeneous mesoscale interactions determined by the variety of lignin polymers and association modes.

Molecular dynamics simulations

The simulations were performed with GROMACS 4.6 [78] using the TIP3P water model [79] and the CHARMM36 carbohydrate [80–82], protein [83, 84, 85], and lignin [86] force fields. Fast hydrogen angle vibrations and rotations were removed employing the virtual sites method [87], thus allowing a 4 fs integration time step. The non-bonded electrostatic interactions were calculated using the reaction field zero (RFZ) method [88] with a 12 Å force and 15.68 Å neighbor-list cutoff. It has been shown that RFZ is of accuracy similar to the commonly used Particle Mesh Ewald method for biomass systems while allowing significantly better parallel computational efficiency above 10,000 cores [89]. A shifting function was applied to the entire Van der Waals potential so that the interaction is zero at the cutoff distance of 12 Å. Neighbor searching was performed every 16 time steps. Bonds were constrained using the LINCS algorithm [90] and the water internal dynamics was constrained using the SETTLE routine [91].The system was simulated in the NPT ensemble.

The equilibration was performed in three steps, during which the temperature was controlled with the Nose–Hoover [92] algorithm (time constant $\tau = 1$ ps) and, apart from the second step, pressure was controlled with the Berendsen algorithm [93] ($\tau = 1$ ps). First, 3000 steps were performed, with pressure coupling, employing an integration time step of 1 fs, no virtual sites and constraining only bonds containing hydrogen atoms. Subsequently, 50,000 steps without pressure coupling were performed, with a time step of 2 fs, no virtual sites and position restraints applied on all solute atoms. Finally, 25,000 steps with pressure coupling were performed, with a 4 fs time step, virtual sites on and bonds containing all atoms constrained.

For production, the temperature and pressure were controlled using the velocity rescale thermostat [94] ($\tau =$ 1 ps) and the Parrinello–Rahman barostat [95] ($\tau = 4$ ps). Virtual sites and a 4 fs time step were used and all bond lengths were constrained. The total simulation time was 1312 ns. The simulations were carried out on the TITAN XC6 Supercomputer at Oak Ridge National Laboratory, using 60,000 cores at a peak performance of 45 ns/day.

Analysis methodology

The analysis of multi-million atom, μs-long MD simulations introduces unique challenges, chief among them being the computational time required to obtain quantities of interest over the entire trajectory using serial approaches. To address this in part, our analysis was was carried out with purpose-build python-based VMD scripts [96] on only the heavy atoms of the solutes (cellulose, lignin, and enzyme), thus reducing the number of atoms to be analyzed by a factor of 20. This reduces the memory requirement of the analysis scripts as well as the time to solution, as the time to execute many basic operations (such as selecting subsets of atoms or loading trajectory files) scales linearly with the number of atoms.

The critical concept underlying most of the analysis is that of contact. Traditionally, a "contact" would use a fixed cutoff distance, and if two atoms were within this cutoff, they would be considered in contact. However, the choice of the cutoff value will impact tremendously the number of contacts found. Short cutoffs favor strong interactions such as hydrogen bonds, while longer cutoffs will begin to capture non-specific hydrophobic interactions. We strike a balance between these two extremes by adopting a weighted contact definition similar to the native contact definition introduced by Sheinerman and Brooks [97]. Specifically, the number of contacts between heavy atom i in interaction group A and all the heavy atoms in interaction group B is defined as

$$C_i = \sum_{j \in B} \frac{1}{1 + \exp\left(5\,\text{Å}^{-1}\left(d_{ij} - 4\,\text{Å}\right)\right)}. \quad (1)$$

Here, groups A and B are subsets of the system (cellulose, lignin, or enzyme), and d_{ij} is the distance between atoms i and j. If groups A and group B are identical (for instance, in the calculation of lignin–lignin contacts), we only count the contacts between unique molecules, neglecting internal molecular contacts. This approach will count both weaker hydrophobic and stronger electrostatic interactions, and will give more weight to the stronger short-range interactions.

Contacts are made and broken repeatedly over the course of the simulation. Indeed, 83–93 % of interactions formed break within 100 ns in our analysis. However, due to some particularly long-lived interactions, on the μs timescale, the mean duration of binding events to cellulases is on the order of tens of nanoseconds (Additional file 2: Figure S3). This may not be representative of the overall binding time in vivo due to limitations in timescale for typical MD simulations. While classical MD now routinely brings to life multi-million atom structures [98], atomistic MD of large complexes remains limited to ns-μs timescales due to the fs-scale timesteps required for accurate integration in time. Therefore, slow (relaxation time > μs) enzyme-biomass dissociation processes and similarly long binding events are not captured here. Explicit rare-event methods or biased sampling may be useful for characterizing such kinetics.

Further analysis was performed to determine the orientation of the bound Cel7A relative to the long axis of the cellulose and the rotational and translational diffusion constants. These analyses were implemented as python-based VMD [96] scripts, stored using numpy [99], and plotted using matplotlib [100]. In addition, the formation and time evolution of the interaction networks present in the simulation were carried out using the NetworkX library [101] and the Gephi program [45].

Surface area computation

Computing surface area for large systems using conventional algorithms, where many random points on a sphere around every atom in the selection are checked for proximity to nearby atoms, was determined to be too inefficient for our purposes, as a single calculation on the complete trajectory was estimated to take a month in a serial process. Instead, we developed a new tool to efficiently calculate interfacial surface area by utilizing methodologies from the computer graphics literature which had already been incorporated into VMD [96]. In brief, we calculate the surface area using the grid-based QuickSurf [102] representation, and combined the surfaces from different groups of atoms to obtain the interfacial surface area between two groups. This approach is ~100 times faster than the conventional solvent accessible surface area (SASA) calculation implemented in VMD. A conventional SASA calculation on 100,000 atoms evaluates 500 points per atom and determines if they are within a cutoff distance (3–5 Å) of other nearby atoms (20–30 atoms) in that selection, which overall requires over 1 billion distance comparisons. In contrast, the QuickSurf surface calculation performed on the same 100,000 atoms evaluates the value of a Gaussian on a grid with a resolution on the order of 1 Å. The Gaussian function is assumed to be negligible 5–7 Å away from its center (depending on the resolution requested), and therefore in total we only evaluate the Gaussian ~100 million times for each atom selection for which the area is computed. Additional computation is required to generate a surface using the marching cubes algorithm [103] and to calculate the surface area from the resulting triangles. All of the aforementioned steps were carried out on a GPU and the net result is a calculation that is 100–300 fold faster (Fig. 10), depending on the size of the

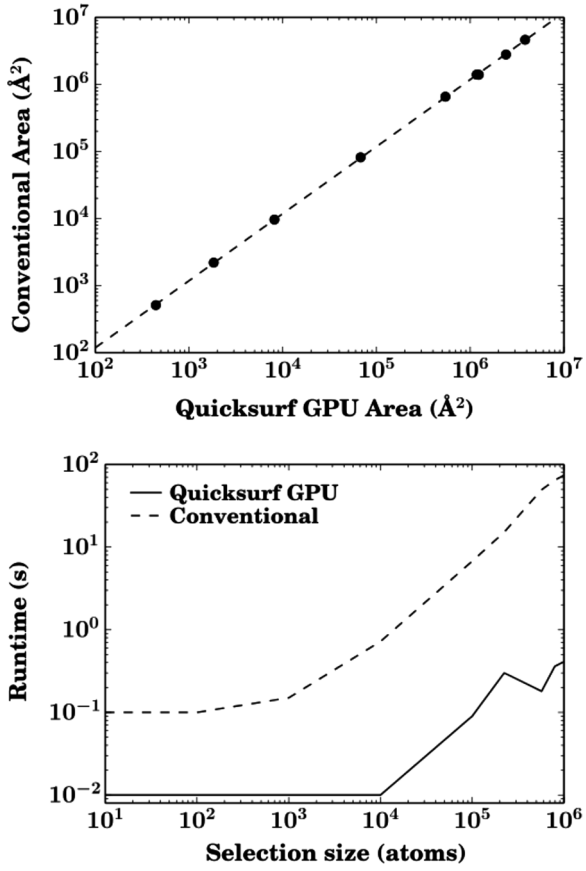

Fig. 10 Accuracy (*top*) and runtime (*bottom*) of a conventional approach vs. our GPU-accelerated surface area calculation for test atom selections of a given size. The r-value for the linear fit between the conventional surface area and the GPU-calculated surface area is 0.99997 with a slope of 0.9997; however the intercept in the plot is not zero, indicating a consistent percentage offset of ~20 %. The runtimes represent the time required to calculate the surface area of a single atom selection once

selection, compared to a conventional SASA calculation performed on one CPU.

To compute the surface, we added 3 Å to the radius of every heavy atom, so as to represent the radii of both the heavy atom and the missing hydrogens, then scaled them by 0.47 when calculating the Gaussian, and use 0.4 as the Gaussian density threshold for computing the surface. These parameters were determined by converting the optimal parameters found by Grant and Pickup [104], with a 1.5 Å grid spacing found through experimentation. Example surfaces and how they compare are shown in Additional file 2: Figure S9.

One particular caution to using the above approach is that the surfaces tend to be 10–20 % smaller than those computed by SASA, due to the smoother Gaussian surfaces that paper over the nooks and crannies between atoms (Additional file 2: Figure S10). However, while the absolute values may be different, the trends and the relative surface areas are consistent between the two methods. In our particular application, where we are interested in the interface area relative to the total surface area, the difference between this method and conventional SASA is expected to be minimal.

Additional files

Additional file 1. This is an animation of the full trajectory, using the same representation as in Fig. 1.

Additional file 2. Supplementary document providing supplementary figures and tables.

Additional file 3. This is an animation of how the contacts change with time, using the same representation as in Fig. 3.

Additional file 4: Video S1. 3-D representation of the contacts points between two TrCel7A enzymes. Heavy atoms that are found to make-contacts are colored based on the contact number, with warmer colors having more contacts relative to coolercolors. For context, the remaining protein structure is also shown, along with the heavy atoms for the remainingresidues. This animation is related to Fig. 4c.

Additional file 5: Video S2. 3-D representation of contacts points between TrCel7A and lignin. Heavy atoms that are found to make contactsare colored based on the contact number, with warmer colors having more contacts relative to cooler colors. Forcontext, the remaining protein structure is also shown, along with the heavy atoms for the remaining residues. Thisanimation is related to Fig. 4b.

Additional file 6: Video S3. 3-D representation of contacts points between TrCel7A copies and cellulose. Heavy atoms that are found to makecontacts are colored based on the contact number, with warmer colors having more contacts relative to coolercolors. For context, the remaining protein structure is also shown, along with the heavy atoms for the remainingresidues. This animation is related to Fig. 4a.

Additional file 7. PDB file of a single copy of the enzyme taken from the trajectory where the average enzyme-enzyme contactnumber is provided in the beta column.

Additional file 8. PDB file of a single copy of the enzyme taken from the trajectory where the average lignin-enzyme contact numberis provided in the beta column.

Additional file 9. PDB file of a single copy of the enzyme taken from the trajectory where the average cellulose-enzyme contactnumber is provided in the beta column.

Additional file 10. A file containing the initial coordinates for the simulation system for independent analysis.

Additional file 11. A zip archive containing a gallery of each of the cellulases that bound to cellulose in the context of their environment. Each image within the gallery is one snapshot taken from the end of the trajectory showing the relative position of each enzyme (green) that makes contact with the cellulose (red). Nearby lignins are shown in blue, and the substrate tunnel is a yellow surface to orient the viewer. The three tyrosine residues are shown in orange. Note that for each protein, there are 4 images, taken from different relative orientations to the cellulose fibril (0, 90, 180, and 270), and are labeled accordingly in their filenames.

Abbreviations

CBM: Cellulose-binding module; TrCel7A: *Trichoderma reesei* fungal cellulase Cel7A; MD: Molecular dynamics; CH: Crystalline cellulose, high lignin coverage; CL: Crystalline cellulose, low lignin coverage; NonC: Non-crystalline cellulose, low lignin coverage; CD: Catalytic domain.

Authors' contributions

JVV analyzed the simulations and wrote the manuscript. LP designed the study, conducted the simulations and wrote the manuscript. XQ analyzed the simulations and helped draft the manuscript. RS conducted the simulations and helped draft the manuscript. BL conducted the simulations and helped draft the manuscript. JCS designed the study and wrote the manuscript. All authors have read and approved of the final version of the manuscript.

Author details

[1] UT/ORNL Center for Molecular Biophysics, Oak Ridge National Laboratory, 37831 Oak Ridge, TN, USA. [2] Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, 61801 Urbana, IL, USA. [3] Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, 37996 Knoxville, TN, USA. [4] University of Tennessee/Oak Ridge National Laboratory Center for Molecular Biophysics, P.O.Box 2008, Oak Ridge, TN 37831-6309, USA.

Acknowledgements

This research is funded by the Genomic Science Program, Office of Biological and Environmental Research, US Department of Energy. An award of computer time was provided by the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program. This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725. J.V. acknowledges support from the DOE Computational Sciences Graduate Fellowship supported by Grant DE-FG02-97ER25308.

Competing interests

The authors declare that they have no competing interests.

Received: 15 July 2015 Accepted: 9 November 2015
Published online: 21 December 2015

References

1. Chu S, Majumdar A. Opportunities and challenges for a sustainable energy future. Nature. 2012;488(7411):294–303.
2. Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD. Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science. 2007;315(5813):804–7.
3. Jorgensen H, Kristensen JB, Felby C. Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. Biofuels Bioprod Biorefin. 2007;1(2):119–34.
4. Chen F, Dixon RA. Lignin modification improves fermentable sugar yields for biofuel production. Nat Biotechnol. 2007;25(7):759–61.

5. Yang B, Wyman CE. Pretreatment: the key to unlocking low-cost cellulosic ethanol. Biofuel Bioprod Biorefining. 2008;2(1):26–40.
6. Pu Y, Hu F, Huang F, Davison B, Ragauskas A. Assessing the molecular structure basis for biomass recalcitrance during dilute acid and hydrothermal pretreatments. Biotechnol Biofuels. 2013;6(1):15.
7. Gao X, Kumar R, Singh S, Simmons BA, Balan V, Dale BE, Wyman CE. Comparison of enzymatic reactivity of corn stover solids prepared by dilute acid, AFEX, and ionic liquid pretreatments. Biotechnol Biofuels. 2014;7:71.
8. Ragauskas AJ, Beckham GT, Biddy MJ, Chandra R, Chen F, Davis MF, Davison BH, Dixon RA, Gilna P, Keller M, Langan P, Naskar AK, Saddler JN, Tschaplinski TJ, Tuskan GA, Wyman CE. Lignin valorization: improving lignin processing in the biorefinery. Science. 2014;344(6185):709.
9. Rahikainen J, Mikander S, Marjamaa K, Tamminen T, Lappas A, Viikari L, Kruus K. Inhibition of enzymatic hydrolysis by residual lignins from softwood-study of enzyme binding and inactivation on lignin-rich surface. Biotechnol Bioeng. 2011;108(12):2823–34.
10. Somerville C, Youngs H, Taylor C, Davis SC, Long SP. Feedstocks for lignocellulosic biofuels. Science. 2010;329(5993):790–2.
11. Chundawat SPS, Donohoe BS, da Costa Sousa L, Elder T, Agarwal, UP, Lu F, Ralph J, Himmel ME, Balan V, Dale BE. Multi-scale visualization and characterization of lignocellulosic plant cell wall deconstruction during thermochemical pretreatment. Energy Environ. Sci. 2011;4:973–84.
12. Socha AM, Parthasarathi R, Shi J, Pattathil S, Whyte D, Bergeron M, George A, Tran K, Stavila V, Venkatachalam S, Hahn MG, Simmons BA, Singh S. Efficient biomass pretreatment using ionic liquids derived from lignin and hemicellulose. Proc Natl Acad Sci USA. 2014;111(35):3587–95.
13. 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. Common processes drive the thermochemical pretreatment of lignocellulosic biomass. Green Chem. 2014;16(1):63–8.
14. Mansfield SD, Mooney C, Saddler JN. Substrate and enzyme characteristics that limit cellulose hydrolysis. Biotechnol Prog. 1999;15(5):804–16.
15. Donohoe BS, Decker SR, Tucker MP, Himmel ME, Vinzant TB. Visualizing lignin coalescence and migration through maize cell walls following thermochemical pretreatment. Biotechnol Bioeng. 2008;101(5):913–25.
16. Kumar L, Arantes V, Chandra R, Saddler J. The lignin present in steam pretreated softwood binds enzymes and limits cellulose accessibility. Bioresource Technol. 2012;103(1):201–8.
17. Ding S-Y, Liu Y-S, Zeng Y, Himmel ME, Baker JO, Bayer EA. How does plant cell wall nanoscale architecture correlate with enzymatic digestibility? Science. 2012;338(6110):1055–60.
18. Li H, Pu Y, Kumar R, Ragauskas AJ, Wyman CE. Investigation of lignin deposition on cellulose during hydrothermal pretreatment, its effect on cellulose hydrolysis, and underlying mechanisms. Biotechnol Bioeng. 2014;111(3):485–92.
19. Eriksson T, Borjesson J, Tjerneld F. Mechanism of surfactant effect in enzymatic hydrolysis of lignocellulose. Enzyme Microb Technol. 2002;31(3):353–64.
20. Palonen H, Tjerneld F, Zacchi G, Tenkanen M. Adsorption of Trichoderma reesei CBH I and EG II and their catalytic domains on steam pretreated softwood and isolated lignin. J Biotechnol. 2004;107(1):65–72.
21. Berlin A, Balakshin M, Gilkes N, Kadla J, Maximenko V, Kubo S, Saddler J. Inhibition of cellulase, xylanase and beta-glucosidase activities by softwood lignin preparations. J Biotechnol. 2006;125(2):198–209.
22. Borjesson J, Engqvist M, Sipos B, Tjerneld F. Effect of poly(ethylene glycol) on enzymatic hydrolysis and adsorption of cellulase enzymes to pretreated lignocellulose. Enzyme Microb Technol. 2007;41(1–2):186–95.
23. Sammond DW, Yarbrough JM, Mansfield E, Bomble YJ, Hobdey SE, Decker SR, Taylor LE, Resch MG, Bozell JJ, Himmel ME, Vinzant TB, Crowley MF. Predicting enzyme adsorption to lignin films by calculating enzyme surface hydrophobicity. J Biol Chem. 2014;289(30):20960–9.
24. Nakagame S, Chandra RP, Kadla JF, Saddler JN. Enhancing the enzymatic hydrolysis of lignocellulosic biomass by increasing the carboxylic acid content of the associated lignin. Biotechnol Bioeng. 2011;108(3):538–48.
25. Rahikainen JL, Evans JD, Mikander S, Kalliola A, Puranen T, Tamminen T, Marjamaa K, Kruus K. Cellulase-lignin interactions-the role of carbohydrate-binding module and pH in non-productive binding. Enzyme Microbial Technol. 2013;53(5):315–21.
26. Lan TQ, Lou H, Zhu JY. Enzymatic saccharification of lignocelluloses should be conducted at elevated ph 5.2-6.2. BioEnerg Res. 2013;6(2):476–85.
27. Engineering plant cell walls. tuning lignin monomer composition for deconstructable biofuel feed stocks or resilient biomaterials. Green Chem. 2014;16(5):2627.
28. Van Acker R, Vanholme R, Storme V, Mortimer JC, Dupree P, Boerjan W. Lignin biosynthesis perturbations affect secondary cell wall composition and saccharification yield in Arabidopsis thaliana. Biotechnol Biofuels. 2013;6(1):46.
29. Fu C, Mielenz JR, Xiao X, Ge Y, Hamilton CY, Rodriguez M, Chen F, Foston M, Ragauskas A, Bouton J, Dixon RA, Wang Z-Y. Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. Proc Natl Acad Sci USA. 2011;108(9):3803–8.
30. Payne CM, Knott BC, Mayes HB, Hansson H, Himmel ME, Sandgren M, Stahlberg J, Beckham GT. Fungal cellulases. Chem Rev. 2015;115(3):1308–448.
31. Igarashi K, Uchihashi T, Koivula A, Wada M, Kimura S, Okamoto T, Penttila M, Ando T, Samejima M. Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface. Science. 2011;333(6047):1279–82.
32. Payne CM, Resch MG, Chen L, Crowley MF, Himmel ME, Taylor LE, Sandgren M, Ståhlberg J, Stals I, Tan Z, Beckham GT. Glycosylated linkers in multimodular lignocellulose-degrading enzymes dynamically bind to cellulose. Proc Natl Acad Sci USA. 2013;110(36):14646–51.
33. Kraulis J, Clore GM, Nilges M, Jones TA, Pettersson G, Knowles J, Gronenborn AM. Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from Trichoderma reesei. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. Biochemistry. 1989;28(18):7241–57.
34. Reinikainen T, Ruohonen L, Nevanen T, Laaksonen L, Kraulis P, Jones TA, Knowles JKC, Teeri TT. Investigation of the function of mutated cellulose-binding domains of trichoderma reesei cellobiohydrolase i. Proteins Struct Funct Bioinf. 1992;14(4):475–82.
35. Linder M, Mattinen M-L, Kontteli M, Lindeberg G, Stahlberg J, Drakenberg T, Reinikainen T, Pettersson G, Annila A. Identification of functionally important amino acids in the cellulose-binding domain of trichoderma reesei cellobiohydrolase i. Protein Sci. 1995;4(6):1056–64.
36. Lehtio J, Sugiyama J, Gustavsson M, Fransson L, Linder M, Teeri TT. The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. Proc Natl Acad Sci USA. 2003;100(2):484–9.
37. Beckham GT, Matthews JF, Bomble YJ, Bu L, Adney WS, Himmel ME, Nimlos MR, Crowley MF. Identification of amino acids responsible for processivity in a family 1 carbohydrate-binding module from a fungal cellulase. J Phys Chem B. 2010;114(3):1447–53.
38. Nimlos MR, Beckham GT, Matthews JF, Bu L, Himmel ME, Crowley MF. Binding preferences, surface attachment, diffusivity, and orientation of a family 1 carbohydrate-binding module on cellulose. J Biol Chem. 2012;287(24):20603–12.
39. Strobel KL, Pfeiffer KA, Blanch HW, Clark DS. Structural insights into the affinity of cel7a carbohydrate-binding module for lignin. J Biol Chem. 2015;290(37):22818–26.
40. Lindner B, Petridis L, Schulz R, Smith JC. Solvent-driven preferential association of lignin with regions of crystalline cellulose in molecular dynamics simulation. Biomacromolecules. 2013;14(10):3390–8.
41. Lu Y, Yang B, Gregg D, Saddler J, Mansfield S. Cellulase adsorption and an evaluation of enzyme recycle during hydrolysis of steam-exploded softwood residues. Appl Biochem Biotechnol. 2002;98–100(1–9):641–54.
42. Kumar R, Wyman CE. Access of cellulase to cellulose and lignin for poplar solids produced by leading pretreatment technologies. Biotechnol Prog. 2009;25(3):807–19.
43. Gao D, Chundawat SPS, Uppugundla N, Balan V, Dale BE. Binding characteristics of trichoderma reesei cellulases on untreated, ammonia fiber expansion (AFEX), and dilute-acid pretreated lignocellulosic biomass. Biotechnol Bioeng. 2011;108(8):1788–800.
44. Carmona C, Langan P, Smith JC, Petridis L. Why genetic modification of lignin leads to low-recalcitrance biomass. Phys Chem Chem Phys. 2015;17:358–64.

45. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. In: ICWSM, vol. 8. 2009. p. 361–62.
46. Capoulade J, Wachsmuth M, Hufnagel L, Knop M. Quantitative fluorescence imaging of protein diffusion and interaction in living cells. Nat Biotechnol. 2011;29(9):835–42.
47. Jervis EJ, Haynes CA, Kilburn DG. Surface diffusion of cellulases and their isolated binding domains on cellulose. J Biol Chem. 1997;272(38):24016–23.
48. Stals I, Sandra K, Devreese B, Van Beeumen J, Claeyssens M. Factors influencing glycosylation of Trichoderma reesei cellulases. II: *N*-glycosylation of Cel7A core protein isolated from different strains. Glycobiology. 2004;14(8):725–37.
49. Langsford ML, Gilkes NR, Singh B, Moser B, Miller RC, Warren RAJ, Kilburn DG. Glycosylation of bacterial cellulases prevents proteolytic cleavage between functional domains. Letters. 1987;225(1):163–7.
50. Gupta R, Baldock SJ, Fielden PR, Grieve BD. Capillary zone electrophoresis for the analysis of glycoforms of cellobiohydrolase. J Chromatogr A. 2011;1218(31):5362–8.
51. Beckham GT, Dai Z, Matthews JF, Momany M, Payne CM, Adney WS, Baker SE, Himmel ME. Harnessing glycosylation to improve cellulase activity. Curr Opin Biotechnol. 2012;23(3):338–45.
52. Alekozai EM, GhattyVenkataKrishna PK, Uberbacher EC, Crowley MF, Smith JC, Cheng X. Simulation analysis of the cellulase Cel7A carbohydrate binding module on the surface of the cellulose Iβ. Cellulose. 2014;21(2):951–71.
53. Beckham GT, Matthews JF, Peters B, Bomble YJ, Himmel ME, Crowley MF. Molecular-level origins of biomass recalcitrance: decrystallization free energies for four common cellulose polymorphs. J Phys Chem B. 2011;115(14):4118–27.
54. McGaughey GB, Gagne M, Rappe AK. pi-stacking interactions—alive and well in proteins. J Biol Chem. 1998;273(25):15458–63.
55. Pingali SV, O'Neill HM, McGaughey J, Urban VS, Rempe CS, Petridis L, Smith JC, Evans BR, Heller WT. Small angle neutron scattering reveals pH-dependent conformational changes in Trichoderma reesei cellobiohydrolase I: implications for enzymatic activity. J Biol Chem. 2011;286(37):32801–9.
56. Liu Y-S, Baker JO, Zeng Y, Himmel ME, Haas T, Ding S-Y. Cellobiohydrolase hydrolyzes crystalline cellulose on hydrophobic faces. J Biol Chem. 2011;286(13):11195–201.
57. Grethlein HE. The effect of pore size distribution on the rate of enzymatic hydrolysis of cellulosic substrates. Nat Biotechnol. 1985;3:155–60.
58. Ding SY, Himmel ME. The maize primary cell wall microfibril: a new model derived from direct visualization. J Agric Food Chem. 2006;54(3):597–606.
59. Hallac BB, Ragauskas AJ. Analyzing cellulose degree of polymerization and its relevancy to cellulosic ethanol. Biofuels Bioprod Biorefining. 2011;5(2):215–25.
60. Nishiyama Y, Langan P, Chanzy H. Crystal structure and hydrogen-bonding system in cellulose Iβ from synchrotron X-ray and neutron fiber diffraction. J Am Chem Soc. 2002;124(31):9074–82.
61. Petridis L, Pingali SV, Urban V, Heller WT, O'Neill HM, Foston M, Ragauskas A, Smith JC. Self-similar multiscale structure of lignin revealed by neutron scattering and molecular dynamics simulation. Phys Rev E. 2011;83(6):061911.
62. Petridis L, Schulz R, Smith JC. Temperature dependence of lignin structure and dynamics. J Am Chem Soc. 2011;133(50):20277–87.
63. Mooney CA, Mansfield SD, Touhy MG, Saddler JN. The effect of initial pore volume and lignin content on the enzymatic hydrolysis of softwoods. Bioresour Technol. 1998;64(2):113–9.
64. Boerjan W, Ralph J, Baucher M. Lignin biosynthesis. Annu Rev Plant Biol. 2003;54:519–46.
65. Pu Y, Zhang D, Singh PM, Ragauskas AJ. The new forestry biofuels sector. Biofuels Bioprod Biorefining. 2008;2:58–73.
66. Chakar F, Ragauskas A. Review of current and future softwood kraft lignin process chemistry. Ind Crops Prod. 2004;20(2):131–41.
67. Ralph J, Peng JP, Lu FC, Hatfield RD, Helm RF. Are lignins optically active? J Agric Food Chem. 1999;47(8):2991–6.
68. Brunow G, Kilpelainen I, Lapierre C, Lundquist K, Simola LK, Lemmetyinen J. The chemical structure of extracellular lignin released by cultures of picea abies. Phytochemistry. 1993;32(4):845–50.
69. Yan JF, Pla F, Kondo R, Dolk M, McCarthy JL. Lignin. 21. depolymerization by bond-cleavage reactions and degelation. Macromolecules. 1984;17(10):2137–42.
70. Divne C, Stahlberg J, Teeri TT, Jones TA. High-resolution crystal structures reveal how a cellulose chain is bound in the 50 angstrom long tunnel of cellobiohydrolase I from Trichoderma reesei. J Mol Biol. 1998;275(2):309–25.
71. Harrison MJ, Nouwens AS, Jardine DR, Zachara NE, Gooley AA, Nevalainen H, Packer NH. Modified glycosylation of cellobiohydrolase I from a high cellulase-producing mutant strain of Trichoderma reesei. Eur J Biochem. 1998;256(1):119–27.
72. Zhu Z, Sathitsuksanoh N, Vinzant T, Schell DJ, McMillan JD, Zhang Y-HP. Comparative study of corn stover pretreated by dilute acid and cellulose solvent-based lignocellulose fractionation: Enzymatic hydrolysis, supramolecular structure, and substrate accessibility. Biotechnol Bioeng. 2009;103(4):715–24.
73. Ohgren K, Bura R, Saddler J, Zacchi G. Effect of hemicellulose and lignin removal on enzymatic hydrolysis of steam pretreated corn stover. Bioresour Technol. 2007;98(13):2503–10.
74. Li C, Knierim B, Manisseri C, Arora R, Scheller HV, Auer M, Vogel KP, Simmons BA, Singh S. Comparison of dilute acid and ionic liquid pretreatment of switchgrass: Biomass recalcitrance, delignification and enzymatic saccharification. Bioresour Technol. 2010;101(13, SI):4900–6.
75. Kumar R, Mago G, Balan V, Wyman CE. Physical and chemical characterizations of corn stover and poplar solids resulting from leading pretreatment technologies. Bioresour Technol. 2009;100(17):3948–62.
76. Sannigrahi P, Ragauskas AJ, Miller SJ. Effects of two-stage dilute acid pretreatment on the structure and composition of lignin and cellulose in loblolly pine. Bioenerg Res. 2008;1(3–4):205–14.
77. Tu M, Chandra RP, Saddler JN. Recycling cellulases during the hydrolysis of steam exploded and ethanol pretreated Lodgepole pine. Biotechnol Prog. 2007;23(5):1130–7.
78. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B, Lindahl E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics. 2013;29(7):845–54.
79. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983;79(2):926–35.
80. Guvench O, Greene SN, Kamath G, Brady JW, Venable RM, Pastor RW, Mackerell AD. Additive empirical force field for hexopyranose monosaccharides. J Comput Chem. 2008;29(15):2543–64.
81. Guvench O, Hatcher ER, Venable RM, Pastor RW, Mackerell AD. CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. J Chem Theory Comput. 2009;5(9):2353–70.
82. Guvench O, Mallajosyula SS, Raman EP, Hatcher E, Vanommeslaeghe K, Foster TJ, Jamison FW, Mackerell AD. CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling. J Chem Theory Comput. 2011;7(10):3162–80.
83. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, Mackerell AD. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ(1) and χ(2) dihedral angles. J Chem Theory Comput. 2012;8(9):3257–73.
84. MacKerell AD, Feig M, Brooks CL. Improved treatment of the protein backbone in empirical force fields. J Am Chem Soc. 2004;126(3):698–9.
85. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FT, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B. 1998;102(18):3586–616.
86. Petridis L, Smith JC. A molecular mechanics force field for lignin. J Comput Chem. 2009;30(3):457–67.
87. Hess B, Kutzner C, van der Spoel D, Lindahl E. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput. 2008;4(3):435–47.
88. Tironi IG, Sperb R, Smith PE, van Gunsteren WF. A generalized reaction field method for molecular dynamics simulations. J Chem Phys. 1995;102(13):5451.

89. Schulz R, Lindner B, Petridis L, Smith JC. Scaling of multimillion-atom biological molecular dynamics simulation on a petascale supercomputer. J Chem Theory Comput. 2009;5(10):2798–808.
90. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. J Comput Chem. 1997;18(12):1463–72.
91. Miyamoto S, Kollman PA. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. J Comput Chem. 1992;13(8):952–62.
92. Martyna GJ, Tobias DJ, Klein ML. Constant pressure molecular dynamics algorithms. J Chem Phys. 1994;101(5):4177–89.
93. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984;81(8):3684–90.
94. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys. 2007;126(1):014101.
95. Parrinello M, Rahman A. Crystal structure and pair potentials: a molecular-dynamics study. Phys Rev Lett. 1980;45:1196–9.
96. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996;14(1):33–8.
97. Sheinerman FB, Brooks CL III. Calculations on folding of segment B1 of streptococcal protein G. J Mol Biol. 1998;278(2):439–56.
98. Perilla JR, Goh BC, Cassidy CK, Liu B, Bernardi RC, Rudack T, Yu H, Wu Z, Schulten K. Molecular dynamics simulations of large macromolecular complexes. Curr Opin Struct Biol. 2015;31:64–74.
99. van der Walt S, Colbert SC, Varoquaux G. The NumPy array: a structure for efficient numerical computation. Comput Sci Eng. 2011;13(2):22–30.
100. Hunter JD. Matplotlib: A 2D graphics environment. Comput Sci Eng. 2007;9(3):90–5.
101. Hagberg AA, Schult DA, Swart PJ. Exploring network structure, dynamics, and function using NetworkX. In: Varoquaux G, Travis V, Millman J, editors. Proceedings of the 7th Python in Science Conference (SciPy2008). USA: Pasedena; 2008. p. 11–5.
102. Krone M, Stone J, Ertl T, Schulten K. Fast visualization of gaussian density surfaces for molecular dynamics and particle system trajectories. EuroVis Short Papers. 2012;2012:67–71.
103. Lorensen WE, Cline HE. Marching cubes: A high resolution 3D surface construction algorithm. In: Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques—SIGGRAPH '87. ACM Press, New York. 1987; pp. 163–9.
104. Grant JA, Pickup BT. A gaussian description of molecular shape. J Phys Chem. 1995;99(11):3503–10.