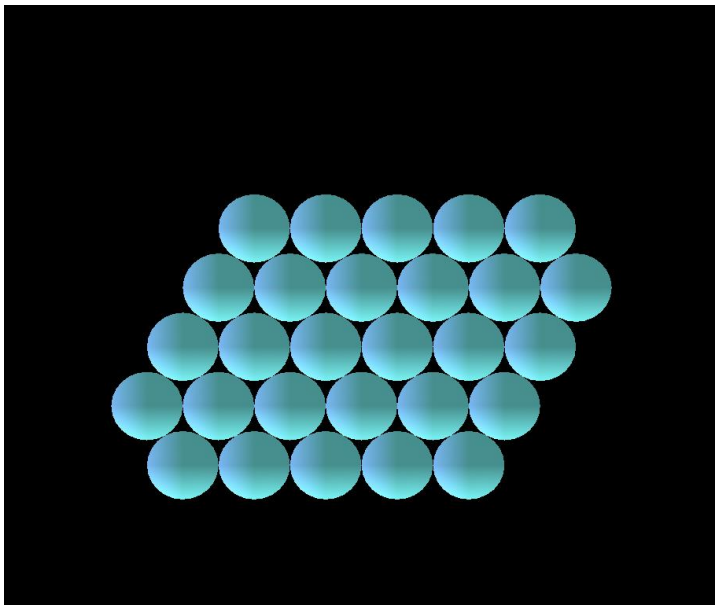
Supplementary Information

1. Classes for Polysaccharides and Enzymes

1.1. Cellulose Class

We used a model predicted by Smith *et al*. [1], based on their finding that 43% of cellulose molecules in Italian ryegrass (*Lolium multiflorum*) primary walls is crystal interior. This model consists of 28 parallel cellulose molecules aligned in a near-rhomboid arrangement, as can be seen from Supplementary Figure 1.

Supplementary Figure 1. End view of cellulose microfibril (CMF) consisting of a bundle of 28 parallel cellulose molecules. Each sphere represents a glucose molecule. 43% of cellulose molecules are interior to the crystal.



An object of the Cellulose class is created with a call specifying its start point (at the centre of the near-rhomboid cross-section), direction, length (in number of glucoses), and the radius of a glucose molecule (in this case set to unity). The direction of the CMF defines the direction for all glucoses, and thus the glucoses in each glucan are aligned head to tail (reducing end to non-reducing end). Boolean variables specify whether adjacent glucoses in the same glucan are covalently bonded (true) or not (false). When a Cellulose object is created, these flags are set to true. By assumption glucoses which are adjacent but in different glucans are connected by hydrogen bonding.

1.2. Xyloglucan Class

The glucoses forming the xyloglucan backbone contact each other head to tail (giving reducing and non-reducing ends) and Boolean variables specify whether adjacent glucoses in the same backbone are covalently bonded (true) or not (false). These flags are set true when a Xyloglucan object is created. Unlike in the Cellulose class, the directions of glucoses are not constrained to be collinear, and two covalently bonded glucose molecules contact each other so that their respective direction vectors form an angle of 0 ± 30 °. This gives flexibility to the xyloglucan, allowing it to curve. In grass (family

Poaceae) xyloglucans, glucose molecules with xylose side branches occur in pairs, with a varying number of glucose units between branched glucose pairs [2]. In our simulation we specified that at each end the last two glucosyl residues would be unbranched (facilitating their attachment to CMFs) with a sequence along the chain of two branched and then two unbranced glucosyl residues. This required that each xyloglucan had a chain length that was an even number not divisible by 4. Each xylose branch contacted its glucose such that their respective direction vectors are oriented to form an angle of 90 ± 30 °, and a Boolean flag is set true to indicate that they are covalently bonded.

When a Xyloglucan object is created, the call specified the number of glucoses in the chain, radii of glucose and xylose (in this case set to unity) and the positions of the first three and last three glucoses in the chain (positions of hydrogen bonding to CMFs). The intermediate glucoses are positioned to ensure a smooth transition without violating constraints on bond angles.

1.3. (1,3;1,4)-β-Glucan Class

The (1,3;1,4)-β-Glucan class was written to model (1,3;1,4)-β-glucans. As with the Xyloglucan class, the glucoses in the (1,3;1,4)-β-Glucan class are arranged to contact each other head to tail and two Boolean variables are set true for each glucose to indicate that it is covalently bonded to its neighbours. (1,3;1,4)-β-Glucans are kinked, rather than straight, because of the (1,3)-linkages which are never found adjacent to each other. There are usually about 70% (1,4)-linkages and 30% (1,3)-linkages. Usually two or three β-glucosyl residues are linked (1,4)-. However, up to ~10% of the polysaccharide may have up to 20 β-glucosyl residues linked (1,4)- [3,4]. In our model each (1,3;1,4)-β-glucan consisted of a randomly chosen number (from 25 to 35) of linked β-glucosyl molecules. To simulate the expected structure, 5 to 14 sequential glucoses were randomly selected from along backbone; these were designated to be linked (1,4) and were arranged linearly along a randomly selected row of available sites on a randomly selected CMF. Counting outwards from either end of the selected run the next bond was designated to be (1,3)-linked, with the remaining bonds forming a repeating sequence of either two or three (1,4)-links (randomly chosen with equal probability) followed by a (1,3)-link. The bond angles of (1,4)-linked glucosyl residues were constrained to vary between 0 and 30 °, while (1,3)-linked bond angles were constrained to vary between 30 and 60 °.

1.4. GAX Class

Adjacent xyloses forming the xylan backbone contact each other head to tail (giving reducing and non-reducing ends) and are covalently bonded (indicated by Boolean variables), with their relative directions allowed to deviate by up to 60 degrees to allow flexibility, as in the case of the xyloglucan backbone. Each xylose is randomly assigned either a glucuronic acid, an arabinose or no side branch, with respective probabilities of 0.08, 0.32 and 0.6 [5]. Each xylose and its side branch contact each other in an orientation that ensures an angle of 90 ± 30 ° between their direction vectors and appropriate flags are set to indicate the two molecules are covalently bonded. In order to facilitate the attachment of the ends of the xylose backbone to CMFs, the last two xylose residues at each end of the arabinoxylan were specified to be unbranched.

Typically arabinoses in a GAX contain a number of arabinoses which are covalently bonded to ferulates. The ratio of ferulates to arabinoses varies, but for simplicity, we set the probability that an arabinose was covalently bonded to a ferulate to be 0.4 for all GAXs [5]. A ferulate covalently bonds to its arabinose with the head of the acid contacting the tail of the arabinose so that the angle between their direction vectors is 0 ± 30 °, and a Boolean flag is set true to indicate the bond.

GAX objects are created with a call which specifies radii for xylose, arabinose and glucuronic acid (in this case all set to unity), and the length of the xylan backbone. If the GAX is to be semi-tethered, the positions of the first three xyloses of the GAX are also specified to be hydrogen bonded to attachment sites on a CMF. If the GAX is tethered to two CMFs, then the positions of the last three xyloses are specified as attachment sites on a second CMF.

When a GAX object is created and placed in the simulation environment, it attempts to form diferulate crosslinks with other GAXs by attempting to form bonds between its ferulates and any proximal ferulates on neighbouring GAXs. The positioning of the GAX may be adjusted to facilitate this (without violating physical constraints). For two ferulates to bond covalently they must contact each other tail to tail so that their respective direction vectors have an orientation of 180 ± 30 °; a Boolean flag is set to indicate the bond.

1.5. Pectin Class

The complexity of polysaccharides comprising pectins, rules out a detailed representation of pectin polymers in our early development of the cell-wall model. Thus the pectin class represented pectins as chains of unbranched galacturonic acid residues (which effectively depicting them as homogalacturonans). The galacturonic acid residues were covalently bonded head to tail, with the direction vector of two adjacent galacturonic acid residues forming an angle of 0 ± 30 °. No side branches were defined for the galacturonic acid residues.The Pectin constructor is called with a specification of start and end positions, the chain length and the radius of a galacturonic acid residue.

1.6. Enzyme Class

All enzymes are represented using one Enzyme class. When an enzyme object is created, the call specifies an enzyme name which is used to assign other properties, such as size, shape, colour, the type of bonds it can cleave and its manner of action when cleaving bonds. The default enzyme is a sphere of radius 8 glucose radii with no activity; a method in the enzyme class allows activities to be assigned after the enzyme is created. Additional methods govern the operation of the enzymes, and these contain both generic algorithms for default enzymes, and specific algorithms for “real” characterised enzymes which are being modelled.

2. Enzyme Motion, and the Detection and Cleavage of Bonds

2.1. Collision Detection

At each iteration, an enzyme will normally move in a random direction with a random step size, with the maximum step size limited to 0.5 glucose radii. As described in the main text, an octagonal tree is then used to identify a list of neighbours (both glycosyl residues or other enzymes) with which

the enzyme may have collided. A collision detection algorithm creates a new list by determining which of these neighbours the enzyme has collided with. Two molecules are considered to have collided if the separation of their centres is less than the sum of radii of their bounding spheres.

2.2. Identification of Cleavable Bonds

At each iteration, after collision detection has occurred, pairs of bonded glycosyl residues are identified from the list of neighbours each enzyme has collided with. These bonded glycosyl residues are ordered by the degree that they overlap the enzyme (separation of centres minus sums of radii). Starting with the most overlapping, an algorithm checks each pair in term to ascertain whether or not the enzyme can cleave the bond. These checks will take into account the type of glycosidic link, molecules involved, DP of the polysaccharide or oligosaccharide, and the presence of branches. If the enzyme is capable of breaking any of the bonds between these pairs of glycoside residues, then it is assigned (with 100% probability) to the first pair found. Setting the assignment probability to 100% was arbitrary and corresponds to a very high affinity of the enzyme to the bond. However, the assignment probability can be specified as a function of enzyme type, glycosidic linkage and polysaccharide type to reflect real affinities. We did not do this, because parameters obtained from applying Michaelis-Menten kinetics to *in vitro* data are not necessarily applicable to modelling discrete systems.

2.3. Modelling Steric Effects

The steric hindrance on both enzyme diffusion and access to bonds is modelled by reflecting colliding objects away from each other. The collision of two objects results in displacements which are governed by the directions and sizes of their steps, the degree to which their bounding spheres overlapped, the separation of the centres of the bounding spheres. At each iteration, after the identification of cleavable bonds, the displacement of an enzyme is calculated for each object that it has collided with (except for a glycosyl residue pair assigned for bond cleavage by that enzyme). These displacements are added as a vector sum to get a nett displacement (whose magnitude truncated at the maximum step size). The enzyme is moved by this displacement, and the presence of other molecules which might cause further collision-reflections is taken into account.

However, if one of the molecules involved in the collision is “immovable” because it forms part of the polysaccharide network or is an enzyme temporarily bound to the polysaccharide network then that molecule will not be reflected but rather remain stationary.

In order that the algorithm work properly and not allow molecules to step (or be reflected) through each other, the maximum step size must be less than the radius of the smallest molecule in the simulation. For safety, we used the step stated above.

2.4. Bond Cleavage

After the effect of collisions have been considered, the next stage of the iterations is to test whether or not each enzyme that has encountered a cleavable bond has been displaced from the bonded glycosyl residue pair involved by collision with other objects. If not, the enzyme is deemed to have

attached to the glycosyl residue pair. The time taken for an enzyme to bind and cleave a bond is simulated by keeping an enzyme stationary in its attached position for a specified number of iterations, which can vary with enzyme and bond type, and the polysaccharide involved. After the number of stationary iterations has elapsed, the enzyme cleaves the bond by changing the states of the glycosyl residues to reflect they are not bonded. It is then able to move.

2.5. Modelling Processive Behaviour

A processive enzyme is not automatically positioned at the next target position on a polysaccharide or oligosaccharide because its access may be impeded by the presence of other molecules. Rather, at the end of the sequence of stationary iterations, a preferred direction is assigned to a processive enzyme, for (the arbitrary number of) 20 iterations. The preferred direction directs the enzyme along the polysaccharide it has just cleaved, towards the next bond that it can cleave. For each of these iterations, the step that the enzyme makes is a weighted sum of the preferred direction and a random direction. The preferred direction is initially favoured, but the weight on it decreases with each iteration.

3. Enzymes Modelled

Working on the basis than an amino acid has roughly the same mass as a glucose molecule (and on the basis of representing enzymes as spherical), we assumed that the volume of an enzyme would be proportional to the volume of a glucose molecule by the number of amino acids in the enzyme. Thus taking the cube root of the number of its enzymes would give the radius of the enzyme in terms of glucose radii.

Cel9D contains 697 amino acids and has a theoretical mass of 77 kDa [6] and this equivalence was used to calculate the radius of an enzyme where its mass was known but the number of amino acids was not.

Since we have not calibrated each iteration of the model to a period of time, modelling of different catalytic rates can only be relative. The fastest activity of the enzymes we modelled was 330 s^{-1} for Endoxylanase 1 activity on xylosidic bonds, which we modelled as a stationary period of three iterations (number of iterations required to cleave a bond). All other catalytic rates were modelled with respect to this calibration.

4. Cel9D

The exoglucanase Cel9D was thus specified a sphere with radius 8.87 glucose radii. Cel9D can attack cellooligosaccharides with DP $\geq$ 3 from the non-reducing end to produce glucose but does not degrade cellobiose [6]. Thus in the model it was restricted to attacking 1,4-β glycosidic links in oligosaccharides with DP > 2, with no upper limit and irrespective of the type of polysaccharide. This allowed it to attack cellulose, xyloglucans and 1,3;1,4-β-glucans as well as oligosaccharides of these, but cannot remove substituted glucosyl residues.

The algorithm to test whether cel9D could cleave a bond that it had encountered first checked that the two glycosyl residues in question were the last two glucosyl residues from the non-reducing end of

a backbone of DP ≥ 3, that neither had side branches, and that they and the third glycosyl residue from the non-reducing end were all 1,4-β-linked.

The average catalytic rate (k_{cat}) of Cel9D when acting on oligosaccharides of DP 3–6 is 14.4 s^{-1} [6].This value was applied to assumed for all polysaccharides and oligosaccharides, irrespective of DP, and was modelled by a stationary period of 69 iterations.

Since Cel9D is a processive enzyme, a preferred direction is set, being the direction vector of that glucose.

5. Cel51A

The endoglucanase Cel51A (formerly CelF) has mass of 118.3kDa [7] and was specified with a radius of 10.2 glucose radii. It attacks oligosaccharides and polysaccharides of DP ≥ 5 with backbones of 1,4-linked β-glucosyl residues and was assumed to attack cellulose, xyloglucans and 1,3,1,4-β-glucans as well as oligosaccharides.

The algorithm to determine if Cel51A could cleave a bond linking two β-glucosyl residues returns a value of true if those residues are part of a backbone of at least five, all of which were 1,4-linked, and that there are at least three 1,4-linked β-glucosyl residues on the non-reducing side of the bond to be cleaved, and at least two 1,4-linked β-glucosyl residues on the other side, and that one of the glucosyl residues linked by the bond to be cleaved is unsubstitued.

No value is given for the catalytic rate of Cel51A, and so a value of k_{cat} = 39.8 s^{-1} given for Cel9B by Qi, *et al.* [8] was used, and modelled as a stationary period of 25 iterations.

6. Endoxylanase 1

Endoxylanase 1 has mass of 53.7 kDa [9] and was specified with a radius of 7.86 glucose radii. In addition to endoxylanase activity on xylans and xylooligosaccharides of DP ≥ 4, endoxylanase also exhibits α-arabinofuranosidase activity, apparently before xylanase action [9].

The algorithm to test whether Endoxylanse 1 could cleave a bond that it encounters returns true if the bond links an α-arabinosyl residue to a xylosyl residue. It also returns true if the bond links two 1,4-linked β-xylosyl residues, each of which is 1,4-linked to another β-xylosyl residue, and all four xylosyl residues are unsubstituted.

In the absence of information on catalytic rates for Endoxylanase 1, we used the k_{cat} value for Xyn10B of 330 s^{-1} [10] for Endoxylanase 1 activity on xylosidic bonds and modelled this with a stationary period of three iterations. We used a k_{cat} value for general α-arabinofuranosidase activity of of 240 s^{-1} [10] for the α-arabinofuranosidase activity of Endoxylanase 1, and modelled this with a stationary period of four iterations.

References

1. Smith, B.G.; Harris, P.J.; Melton, L.D.; Newman, R.H. Crystalline cellulose in hydrated primary cell walls of three monocotyledons and one dicotyledon. *Plant Cell Physiol.* **1998**, *39*, 711–720.
2. Hsieh, Y.S.Y.; Harris, P.J. Xyloglucans of monocotyledans have diverse structures. *Mol. Plant* **2009**, *2*, 943–965.

3. Fincher, G.B. Exploring the evolution of (1,3;1,4)-β-D-glucans in plant cell walls: Comparative genomics can help! *Curr. Opin. Plant Biol.* **2009**, *12*, 140–147.
4. Harris, P.J.; Fincher, G.B. Distribution, fine structure and function of (1,3;1,4)-β-glucans in the grasses and other taxa. In *Chemistry, Biochemistry, and Biology of 1–3 Beta Glucans and Related Polysaccharides*; Bacic, A., Fincher, G.B., Stone, B.A., Eds.; Academic Press: San Diego, CA, USA, 2009; pp. 621–654.
5. Chesson, A.; Gordon, A.H.; Lomax, J.A. Methylation analysis of mesophyll, epidermis, and fibre cell-walls isolated from the leaves of perennial and Italian ryegrass. *Carbohydr. Res.* **1985**, *141*, 137–147.
6. Qi, M.; Jun, H.S.; Forsberg, C.W. Cel9d, an atypical 1, 4-β-D-glucan glucohydrolase from *Fibrobacter succinogenes*: Characteristics, catalytic residues, and synergistic interactions with other cellulases. *J. Bacteriol.* **2008**, *190*, 1976–1984.
7. Malburg, S.R.; Malburg, L.M., Jr.; Liu, T.; Iyo, A.H.; Forsberg, C.W. Catalytic properties of the cellulose-binding endoglucanase F from *Fibrobacter succinogenes* S85. *App. Environ. Microbiol.* **1997**, *63*, 2449–2453.
8. Qi, M.; Jun, H.S.; Forsberg, C.W. Characterization and synergistic interactions of *Fibrobacter succinogenes* glycoside hydrolases. *App. Environ. Microbiol.* **2007**, *73*, 6098–6105.
9. Matte, A.; Forsberg, C.W., Purification, characterization, and mode of action of endoxylanases 1 and 2 from *Fibrobacter succinogenes* S85. *App. Environ. Microbiol.* **1992**, *58*, 157–168.
10. Jordan, D.B.; Bowman, M.J.; Braker, J.D.; Dien, B.S.; Hector, R.E.; Lee, C.C.; Mertens, J.A.; Wagschal, K. Plant cell walls to ethanol. *Biochem. J.* **2012**, *442*, 241–252.