ARTICLE

BIOTECHNOLOGY and BIOENGINEERING

Cellulose Hydrolysis in Evolving Substrate Morphologies I: A General Modeling Formalism

Wen Zhou,[1,2] Heinz-Bernd Schüttler,[3] Zhiqian Hao,[3] Ying Xu[1,2]

[1]Department of Biochemistry and Molecular Biology, Institute of Bioinformatics, University of Georgia, Athens, Georgia; telephone: 706-542-9779; fax: 706-542-9751; e-mail: xyn@bmb.uga.edu

[2]BioEnergy Science Center (BESC), Oak Ridge, Tennassee

[3]Department of Physics and Astronomy, University of Georgia, Athens, Georgia; telephone: 706-542-3886; fax: 706-542-9751; e-mail: hbs@physast.uga.edu

Received 16 December 2008; revision received 13 March 2009; accepted 27 April 2009

Published online 8 May 2009 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bit.22389

ABSTRACT: We develop a general framework for a realistic rate equation modeling of cellulose hydrolysis using non-complexed cellulase. Our proposed formalism, for the first time, takes into account explicitly the time evolution of the random substrate morphology resulting from the hydrolytic cellulose chain fragmentation and solubilization. This is achieved by integrating novel geometrical concepts to quantitatively capture the time-dependent random morphology, together with the enzymatic chain fragmentation, into a coupled morphology-plus-kinetics rate equation approach. In addition, an innovative site number representation, based on tracking available numbers of β(1,4) glucosidic bonds, of different "site" types, exposed to attacks by different enzyme types, is presented. This site number representation results in an ordinary differential equation (ODE) system, with a substantially reduced ODE system size, compared to earlier chain fragmentation kinetics approaches. This formalism enables us to quantitatively simulate both the hydrolytically evolving random substrate morphology *and* the profound, and heretofore neglected, morphology effects on the hydrolysis kinetics. By incorporating the evolving morphology on an equal footing with the hydrolytic chain fragmentation, our formalism provides a framework for the realistic modeling of the entire solubilization process, beyond the short-time limit and through near-complete hydrolytic conversion. As part I of two companion papers, the present paper focuses on the development of the general modelling formalism. Results and testable experimental predictions from detailed numerical simulations are presented in part II.

Biotechnol. Bioeng. 2009;xxx: xxx–xxx.

KEYWORDS: cellulose hydrolysis; substrate morphology; mathematical model; site number formalism

Introduction

Cellulose has long been recognized as a renewable carbohydrate source for human energy use. It is available in large quantities in plant biomass. In recent years, for both economic and environmental reasons, intense research efforts have been directed at the utilization of cellulose to produce short-chain soluble sugar oligomers, such as glucose or cellobiose, which are subsequently or concurrently metabolized by microorganisms to produce biofuels such as bioethanol. The hydrolytic conversion of cellulose into soluble, fermentable sugars is achieved by means of cellulase systems which contain three major types of enzymatic activities: (1) endoglucanases which cut β(1,4) glucosidic bonds randomly at all internal bond sites of insoluble cellulose chains in solid substrates; (2) exoglucanases which cut bonds only at the ends of insoluble cellulose chains; and (3) β-glucosidases which hydrolyze soluble oligomer sugars into glucose (Henrissat et al., 1998). The endoglucanases and exoglucanases thus depolymerize long solid cellulose chains into short soluble sugars and this is believed to be the rate-limiting step of the cellulose hydrolysis process (Lynd et al., 2002).

Kinetic modeling of cellulose hydrolysis is categorized into four groups (please refer to Zhang and Lynd (2004) for a detailed review). Among the available models, non-mechanistic and semi-mechanistic models are mainly used in data correlation/fitting, while functionally based models describing mechanistic details of enzyme adsorption and hydrolysis by non-complexed cellulase systems often involve more than one substrate variable and more than one

Correspondence to: H.-B. Schüttler and Y. Xu

Contract grant sponsor: U.S. Department of Energy
Contract grant number: 4000063512
Contract grant sponsor: National Science Foundation
Contract grant number: NSF/DBI-0354771; NSF/ITR-IIS-0407204; NSF/DBI-0542119; NSF/CCF0621700

Additional Supporting Information may be found in the online version of this article.

solubilizing enzyme activity, with the goal of probing a detailed mechanistic understanding of enzyme–substrate interactions (Converse and Optekar, 1993; Fenske et al., 1999; Okazaki and Moo-Young, 1978; Suga et al., 1975; Zhang and Lynd, 2006).

Although more advanced than other classes of models, the current functionally based models have some significant limitations. Most importantly, the substrate in these models is treated effectively as a collection of isolated, decoupled cellulose chains, without any explicit reference to the spatial correlations and mutually obstructive interactions between the chains. In real solid substrates, cellulose chains are assembled into dense, spatially correlated structures. Due to steric obstruction by surrounding cellulosic material, most chains in the solid are initially inaccessible to enzymes. At any time during hydrolysis, only a fraction of the substrate's glucosidic bonds admit unobstructed enzyme access, as chains become gradually exposed at external and internal substrate surfaces by hydrolytic removal of overlaying material. Consequently, the *substrate morphology*, that is, the spatial organization of cellulose chains into a solid material, profoundly affects the enzymatic hydrolysis kinetics. At the same time, hydrolytic conversion of substrate into solute continually changes the substrate morphology during hydrolysis. In currently available functionally based models, the effects of substrate morphology on the hydrolysis kinetics, and, vice versa, the effects of hydrolytic conversion on the morphology, are neglected.

To circumvent our lack of understanding of cellulose chain spatial correlations and obstructive interactions, some studies proposed semi-mechanistic models describing a pre-determined change in substrate shape and accessible surface area of substrate along the hydrolysis trajectory (Converse and Grethlein, 1987; Converse et al., 1988; Luo et al., 1997; Movagarnejad et al., 2000; Oh et al., 2001; Philippidis et al., 1992, 1993). In addition, other studies integrated the two-type-substrate model with this surface area approach (Gan et al., 2003; Wald et al., 1984). In reality, the evolution of accessible surface area is likely a complicated function of multiple factors which in turn are driven by the hydrolysis process, and can not be pre-determined.

In the present work, we propose a new modeling approach which explicitly incorporates *both* the hydrolytic cellulose chain fragmentation kinetics *and* its coupling to the concurrent hydrolysis-driven evolution of the substrate morphology. Our treatment of the time-dependent substrate morphology is based on a geometrical construction which organizes the substrate material into time-evolving structural units, to be referred to as smallest accessible compartments (SACs). We define an SAC unit as a minimal volume that is delimited by external surfaces and by internal surfaces exposed to enzyme-accessible hydrated interior voids of the solid substrate material. The hydrolytic time evolution of these SAC structural units is modelled by a surface layer ablation formalism which couples the hydrolytic shrinkage of SAC units to the enzymatic fragmentation and solubilization of surface-exposed cellulose chains. The chain fragmentation kinetics is then modelled by a surface *bond site* formalism which describes the state of surface-exposed chain fragmentation in terms of a small set of glucosidic bond site types and a corresponding set of site number variables.

The present paper focuses entirely on the description of the coupled substrate evolution and chain fragmentation kinetics formalism. This formalism is developed in full generality for any number of enzymatic species, chain site types and substrate morphologies. In a companion article (Zhou et al., 2009), we will present numerical simulation results, comparisons with results of previous models and experimental observations, and predictions for the substrate morphology evolution that could be tested by future experiments.

Materials and methods

Cellulose Chains and Hydrolytic Enzymes

Cellulose consists of linear glucan chain molecules that are built, without branching, from anhydro-glucose ($C_6H_{10}O_5$) monomers, also referred to as G_1 units in the following. The G_1 units are connected by β(1,4)-glucosidic bonds. The chain has the shape of a slightly corrugated ribbon, defined by the approximately parallel-aligned G_1 hexagonal faces (Mosier et al., 1999). The two ends of each chain are chemically distinct, with one end having *non-reducing* and the other *reducing* character; they are referred to in the following as *L*-end and *R*-end, respectively.

In crystalline cellulose, the glucan ribbons are stacked in parallel, with the chain direction defining a crystalline axis, and with the ribbon face defining a crystalline plane. In amorphous cellulosic material, both the ribbon facial orientations and/or the chain directions can be disordered. The degree of positional, directional and/or orientational disorder in a cellulosic substrate is likely relevant for the hydrolysis kinetics, since *endo*-acting hydrolytic enzymes preferentially attack β(1,4)-glucosidic bonds from a direction perpendicular to the ribbon facial surface (as opposed to the lateral ribbon edge) (Mosier et al., 1999).

Solid cellulose substrates are enzymatically solubilized by hydrolytic glucanase enzymes cutting β(1,4)-glucosidic bonds that are accessible at both external and internal surfaces of the solid substrate. The enzymatic cuts produce shorter chain fragments. Chains or chain fragments consisting of ℓ G_1 units are denoted by G_ℓ in the following. The kinetics modeling of this process is typically based on several simplifying assumptions (Okazaki and Moo-Young, 1978; Zhang and Lynd, 2006), which we shall adopt here.

If the fragment length ℓ resulting from a cut is less than a certain minimum insoluble chain length, denoted by ℓ_S, the hydrolyzed oligomeric G_ℓ fragment immediately detaches from the solid substrate surface and goes into solution. Else, if $\ell \geq \ell_S$, the fragment G_ℓ remains insoluble and attached to the solid surface. Re-attachment processes of dissolved

oligomers back to the solid surface will be neglected. Typical cellulosic substrates for industrial applications have ℓ_S values in the range $\ell_S \sim 4 - 7$ (Klemm et al., 1998; Pereira et al., 1988; Stalbrand et al., 1998; Zhang and Lynd, 2003, 2005).

Following earlier studies (Okazaki and Moo-Young, 1978; Suga et al., 1975; Zhang and Lynd, 2006), we consider a model including three hydrolytic glucanase enzymes typically produced, for example, by the *Trichoderma* species: cellobiohydrolase I (CBH1, Ce17A), cellobiohydrolase II (CBH2, Ce16B) and endoglucanase I (EG1 or Ce17B). Cellobiohydrolase I and II are assumed to cut only the 2nd β(1,4)-glucosidic bond from the *R*-end and *L*-end of the chain respectively, generating one soluble G_2 oligomer per cut. Endoglucanases are assumed to cut accessible β(1,4)-glucosidic bonds randomly. Such "*endo*-cuts" can thus produce both soluble oligomers with any $\ell < \ell_S$, as well as insoluble chain fragments, G_ℓ with $\ell \geq \ell_S$, which remain attached to the solid substrate. In the modeling formalism development, we will, for purposes of generality, refer to the foregoing three enzyme (activity) types as *exo-L*, *exo-R*, and *endo*, respectively, without reference to the specific microbial source organism.

Additionally, β-glucosidase enzymes may be applied in solution to further break down the dissolved oligomers (with $\ell < \ell_S$) into fully hydrolyzed glucose. At sufficiently high dissolved oligomer concentrations, glucose and oligomers may also cause product inhibition of both glucanases and β-glucosidase. All of these "down-stream" effects and enzyme degradation can be straightforwardly incorporated into our modeling approach. However, we will not pursue these model extension in the present paper, since our main focus here is on the hydrolytic solubilization of the solid substrate which is most likely the critical rate-limiting step in industrial applications of enzymatic cellulose hydrolysis.

Substrate Morphology and Enzyme Accessibility

As-grown cellulosic biomass is subjected to mechanical processing (e.g., grinding) and thermo-chemical pre-treatment prior to enzymatic hydrolysis (Hsu, 1996; Jacobsen and Wyman, 2000; McMillan, 1994; Weil et al., 1994). Whereas mechanical processing increases mainly the *external* surface area, by reducing the size of cellulosic substrate particles, one main objective of the thermo-chemical pre-treatment is the creation of additional enzyme-accessible *internal* surface area by weakening the linkage, and by infusion of water, between the cellulosic fibril units (Himmel et al., 2007; Zhang and Lynd, 2004). This has the effect of opening up additional hydrated voids (cracks, channels or pores) in the interior of the substrate which increase enzyme access to β(1,4)-glucosidic bonds thus exposed at the internal surfaces of the hydrated voids. This is illustrated schematically in Figure 1. In typical pre-treated substrates, the enzyme accessible surface area consists predominantly of such internal surface area (Chang et al., 1981; Weimer et al., 1990). Hydrolysis rate and efficiency are therefore largely controlled by the availability of internal surface. A second, closely related objective of the thermo-chemical pre-treatment is the removal and/or spatial re-distribution of non-cellulosic components, such as lignin, hemicellulose and pectin, which provide the linkage between the cellulosic microfibril units and hence are essential for the structural integrity and recalcitrance of as-grown cellulosic biomass (Converse, 1993; Grethlein, 1985). Residual non-cellulosic components (remaining after pre-treatment) can obstruct glucanase enzyme access to substrate cellulose and may also inhibit enzyme activity by competitively adsorbing the enzymes (Converse, 1993).

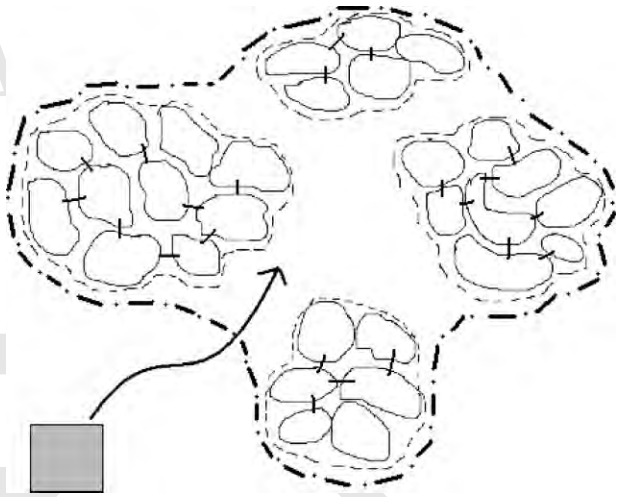

Figure 1. Schematic illustration of the subdivision of a single contiguous cellulosic substrate particle into SACs by SAVs. The particle's external surface is represented by the dot-dash line. The particle shown comprises four SACs. SAC surfaces, comprising both external and internal surface pieces, are indicated by dashed lines. The open void spaces between SACs are SAVs. SAVs are large enough to permit invasion by an enzyme molecule, schematically indicated by the shaded square. Each SAC is shown to consists of smaller irregularly shaped grains that are bounded by full lines and separated by smaller sub-SAV-sized internal voids. These sub-SAV-sized voids that are too small to be invaded by enzymes. Hence, SAC surfaces (dashed lines) comprise the entire enzyme-accessible surface area. Short bridges between grains or between SACs (not shown) represent linking material which provides "solid" structural stability to the substrate particle. Such linking material may consist of non-cellulosic, surface-access-obstructing contaminants or of small bridging cellulosic components. Contaminant-obstructed cellulosic material or surfaces are represented by "*O*-sites" in our model.

Cellulosic substrates are characterized by several important materials parameters: (i) degree of polymerization (DP); (ii) the fraction of enzyme-accessible β(1,4)-glucosidic bonds (F_a); and (iii) the crystallinity index (CrI). All of these parameters are dependent upon the type of as-grown plant material from which the substrate is derived and upon the pre-treatment prior to enzymatic hydrolysis (Zhang and Lynd, 2004). Since different plant materials and pre-treatment conditions usually affect all materials parameters simultaneously, it has been difficult to experimentally isolate the effects of each such parameters on the hydrolysis rate. There does not seem to exist consistent experimental evidence, across different cellulosic substrate materials, for an obvious correlation between CrI and

hydrolysis rate (Mansfield et al., 1999; Ohmine et al., 1983; Puls and Wood, 1991). On the other hand, there does appear to be consistent evidence, across substrate materials, that hydrolysis rates increase with decreasing DP and increasing F_a (Puri, 1984; Sinitsyn et al., 1991; Wood, 1975).

F_a is a geometrical quantity determined by both the cellulosic substrate morphology and by the enzyme size. A contiguous solid cellulose substrate particle presents to the invading enzyme molecule a system of hydrated internal voids of a random size distribution which likely span several orders of magnitude (Grethlein, 1985; Marshall and Sixsmith, 1974). These voids in turn subdivide the cellulose substrate particle into "void-delimited" cellulosic compartments, with a random distribution of compartment sizes that will likely also span several orders of magnitude. The size of the enzyme defines a smallest accessible void (SAV) size, and these SAVs in turn define a subdivision of the substrate into SACs of solid cellulosic material, as schematically illustrated in Figure 1. That is, by definition, an SAC is a minimal volume of substrate material that is bounded by, but not further sub-divided by, enzyme-accessible internal or external surfaces. For purposes of enzyme access, interior surfaces *inside* of an SAC are by definition irrelevant, since they are exposed to sub-SAV-sized voids that are too small to be invaded by enzyme molecules, as illustrated in Figure 1. Thus, by definition, *all enzyme-accessible* glucosidic bonds are exposed at SAC surfaces and SAC surfaces comprise only enzyme-accessible bonds. The accessibility fraction F_a, defined as the ratio of enzyme-accessible bonds to total number of bonds in the substrate, thus has a very simple interpretation in terms of SACs: F_a is the dimensionless SAC surface-to-volume ratio, with "areas" and "volumes" measured in terms of the number of bonds they comprise.

The existence of a system of hydrated internal voids, and corresponding internal surfaces, can be inferred both from sub-micron imaging (Himmel et al., 2007) and from enzyme adsorption data. The latter show that the total adsorbing surface area for most substrates by far exceeds the particles' external particle surface area (Bothwell et al., 1997; Gilkes et al., 1992; Marshall and Sixsmith, 1974; Zhang and Lynd, 2004;). Thus, most of the adsorbing surface area must be internal surface in such substrates; and therefore there must be internal voids that are (i) delimited by such internal surfaces, (ii) sufficiently hydrated and (iii) of sufficient size, in order to permit enzyme to invade and to access to the internal surfaces. Our definition of an SAC also encompasses the special case of highly compact, low-accessibility substrates without enzyme-accessible internal surfaces: the SAC comprises (i.e., is) the entire contiguous particle in that case. Hence, internal surfaces and voids can be conveniently accounted for by the SAC construction, but they are not a precondition for our general SAC-based modeling approach.

SACs are not only a useful geometrical modeling construct, but their surface structures may also be directly observable, in principle, by appropriate surface decoration and imaging methods. For example, one may be able to engineer a gold-conjugated cellulase enzyme (Goldberg, 2008; Lucocq, 2008) or a fluorescently tagged cellulose-binding protein (Hong et al., 2007) of molecular size comparable to cellulases. By adsorbing this protein probe to saturation into a cellulosic substrate particle one could then cover all accessible surfaces (i.e., all SAC surfaces!) by probe molecules and thereby make the SAC surfaces visible, for example, by electron or, respectively, fluorescent confocal, microscopy imaging techniques. Alternatively, cellulase-sized fluorescent nanodots (Wang et al., 2006) might be used as negative staining probes to image the SAC-bounding SAVs. We cannot assess to what extent such experiments would be feasible at the present time.

In principle, each type of enzyme geometrically defines its own distinct SAC subdivision of the substrate, since each enzyme type can have a different molecular size. Enzyme-type-dependent SAC subdivisions could be readily incorporated into our modeling approach. Given the current lack of detailed morphological information, and because of the similarity of sizes for *exo-L*, *exo-R*, and *endo* enzymes of the *Trichoderma* species (Grethlein, 1985), we will assume a single, common SAC subdivision for all three enzyme types. Also, the limiting effects on the enzyme accessibility caused by obstruction of residual non-cellulosic components and adsorbed catalytically inactive enzymes, in addition to the cellulose packing structure, may affect the hydrolytic reaction rates, and need to be taken into account if possible.

Hydrolytic Evolution of Substrate Morphology

The morphology of the solid substrate changes substantially over time as hydrolytic conversion progresses. As solid material is ablated from its surface, at the mesoscopic level, the size of each SAC shrinks, along with this shrinkage, its surface-to-volume ratio increases; at the molecular level, new β(1,4)-glucosidic bonds, previously obscured below the surface by overlaying material, become exposed. To quantitatively model the time evolution of the substrate morphology and its effect on the hydrolysis kinetics, we introduce an SAC "elementary layer" partitioning, as illustrated in Figure 2. This layer partitioning is constructed such that the sequence of elementary layers tracks the inward progression of the SAC surface with time, as the SAC size shrinks during hydrolytic dissolution. Hence, for a given SAC surface, an elementary layer is defined as that fraction of material which will be solubilized and removed from the surface if all outer glucosidic bonds, exposed to enzyme access at *that* surface, and only those, are cut. Removal of said material will then, by definition, generate the outer surface of the next layer, going inward into the material. Note that this definition of an elementary layer does not imply any kind of orderly layered structural organization of the material itself at the molecular or mesoscopic level. It is simply a convenient "accounting" device to introduce a layer number variable λ (see Fig. 2)

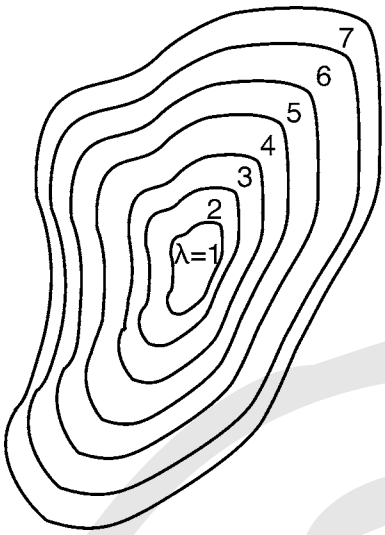

Figure 2. Schematic illustration of the partitioning of an SAC into elementary layers. The layers must be labelled by the layer number λ such that the layer with the highest λ-value is the first one to be removed due to solubilization by the attacking enzymes during hydrolysis.

and to keep track of how many G_1–G_1 bonds are exposed at the SAC surface *vs*. the amount of material enclosed within its bulk at any given time during hydrolytic conversion. Partially dissolved layers are then modeled by treating λ as a continuous variable.

An essential input to model the evolving SAC geometry is then the average number of insoluble G_1 monomers per SAC, contained within all layers enclosed by, and including, layer number λ, as a function of λ, to be denoted by $n_V(\lambda)$. Given our current lack of detailed mesoscopic morphological information, we will assume that $n_V(\lambda)$ has a simple power law dependence, as illustrated by the three prototypical SAC geometries shown in Figure 3, The SAC geometry function $n_V(\lambda)$ is then parameterized by just two variables: an ablation dimension d_A and a "volume" prefactor c_V, so that

$$n_V(\lambda) = c_V \lambda^{d_A} \tag{1}$$

This allows us to keep track of the SAC's shrinkage and remaining monomer content during hydrolytic dissolution in terms of the single geometry variable λ. All other SAC geometry information, such as the overall shape of the SAC, its linear size(s) normal to the ablation direction(s), and its average G_1 monomer density, is absorbed into the volume prefactor c_V. The assumed power law behavior $n_V(\lambda) \propto \lambda^{d_A}$ follows from the simplifying assumption that the shape of the SAC, that is, shape similarity, is preserved during hydrolytic shrinkage of the SAC volume. In real random substrate morphologies, more realistic, irregular SAC shapes, as illustrated in Figure 2, may require non-power-law $n_V(\lambda)$ volume functions which can be readily included in our general formalism described below.

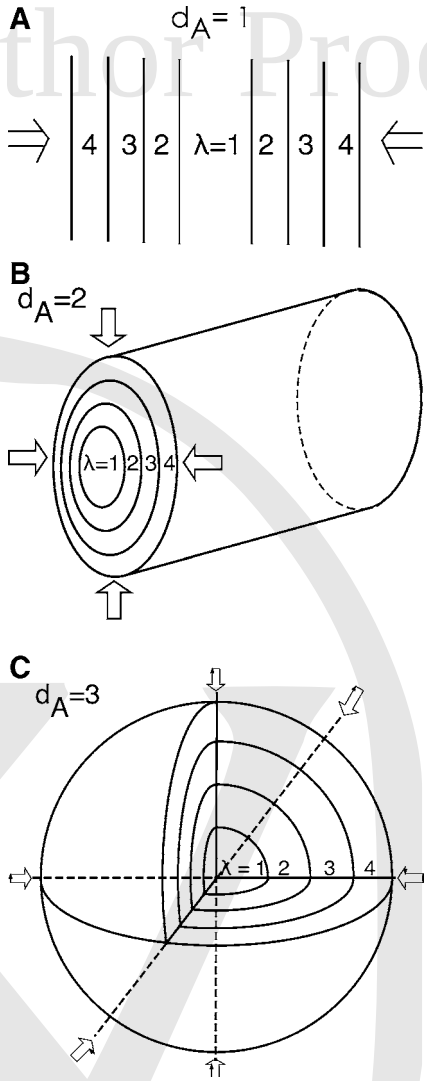

Figure 3. Prototypical SAC layer geometries illustrating the concept of an "ablation dimension" d_A. Prototypes with ablation dimensions $d_A = 1$, 2, and 3 are shown in panels (**A**), (**B**), and (**C**), respectively. Arrows indicate possible directions of enzyme attack during hydrolysis. (A) Dimension $d_A = 1$ is realized if the glucan chains within the SAC exhibit "orientational" order with all glucan ribbon faces oriented approximately parallel to the layer surfaces. This would occur, for example, in a highly crystalline substrate. Directional order is not required for $d_A = 1$. (B) Dimension $d_A = 2$ is realized if the glucan chains within the SAC are orientationally disordered, but *do* exhibit "directional" order, with all glucan chain directions aligned approximately parallel to a common axis, corresponding to the cylinder axis in the drawing. This would likely occur in a substrate consisting of highly aligned fibers of random glucan chain ribbon facial orientations. (C) Dimension $d_A = 3$ is realized if the glucan chains within the SAC are highly disordered, both orientationally and directionally. This would occur, for example, in highly amorphous substrates.

The number of exposed monomers n_M in layer λ is expressed generally in terms of n_V by:

$$n_M(\lambda) = n_V(\lambda) - \Theta(\lambda - 1) n_V(\lambda - 1),$$
$$\Theta(\Delta\lambda) \equiv \begin{cases} 1 & \text{if } \Delta\lambda > 0 \\ 0 & \text{if } \Delta\lambda \le 0 \end{cases} \tag{2}$$

As hydrolysis progresses, outer SAC layers dissolve gradually, while at the same time new underlying layers become gradually exposed at the surface. To model the presence and time evolution of an SAC exhibiting such fractional layers at its surface, the outermost layer number λ will be treated as a continuous dynamical (time-dependent) variable.

A real cellulosic substrate is composed of a wide variety of SACs representing a random distribution of shapes and sizes. We will model this in terms of a finite population of SAC "geometry classes," labeled by an index σ and represented by a population of volume functions, $n_{V,\sigma}(\lambda_\sigma)$ with corresponding $n_{M,\sigma}(\lambda_\sigma)$ from Equation (2), where $\sigma = 1, \ldots M_{MD}$ and M_{MD} is the population size. Adopting the simple power law geometries of Equation (1), we thus introduce a population $(\lambda_\sigma, d_{A,\sigma}, c_{V,\sigma})$ of geometry parameter variables. Let then C_σ be the number of all SACs, in moles of SAC per reactor volume, that have approximately the same geometry $(\lambda_\sigma, d_{A,\sigma}, c_{V,\sigma})$. We can write $x_{M,\sigma}$, the concentration of G_1-monomers (in moles G_1 per reactor volume) exposed at the surfaces of all SACs in class σ, and $x_{V,\sigma}$, the molar concentration of all G_1-monomers contained in all SACs of class σ, as

$$x_{M,\sigma} = C_\sigma n_{M,\sigma}(\lambda_\sigma), \qquad x_{V,\sigma} = C_\sigma n_{V,\sigma}(\lambda_\sigma) \tag{3}$$

$F_{a,\sigma}$, the "partial" fraction of sterically accessible G_1 monomers within a single geometry class σ, and $\overline{F}_a$, the overall macroscopically observable steric accessibility fraction of the entire substrate, are given by

$$F_{a,\sigma}(\lambda_\sigma) = \frac{x_{M,\sigma}}{x_{V,\sigma}} = \frac{n_{M,\sigma}(\lambda_\sigma)}{n_{V,\sigma}(\lambda_\sigma)}, \qquad \overline{F}_a = \frac{x_M}{x_V} = \sum_\sigma \xi_\sigma F_{a,\sigma}(\lambda_\sigma) \tag{4}$$

where $x_M = \sum_\sigma x_{M,\sigma}$, $x_V = \sum_\sigma x_{V,\sigma}$, and $\xi_\sigma = x_{V,\sigma}/x_V$. With this relation, we will be able to make experimentally testable model predictions for the evolution of the F_a-parameter during hydrolysis. We will also use this relation to constrain our morphology models for the simulation of real substrates by the experimentally observed initial (pre-hydrolysis) F_a-values. Inserting Equation (1) and observed values of $F_a \sim 0.1 - 0.001$ (Zhang and Lynd, 2004) into Equation (4), we can also obtain a rough estimate for the typical SAC sizes $\lambda \sim d_A/F_a \sim 20 - 2,000$ or, equivalently, 20–2,000 nm, assuming $d_A = 2$ and a typical thickness of order ~1 nm per elementary layer. (The layer thickness is estimated here to be of the same order as the typical cellulose chain thickness, that is, of order of the ~1 nm glucose molecular size, since, as defined above, removing the outermost layer requires cutting of only the outermost glucosidic bonds, exposed at the SAC surface, and results in the removal of only the outermost chains.) We therefore expect SAC sizes to exceed both typical enzyme molecular sizes (Zhang and Lynd, 2004) and typical cellulosic micro- or elementary fibril sizes (Himmel et al., 2007) by at least an order of magnitude in all but the most highly accessible substrates.

Model Development

Surface Site Ablation Rate Equations

Our hydrolytic surface ablation kinetics formalism keeps track of the availability of hydrolyzable β(1,4)-glucosidic bonds on the substrate surface and of their positioning within their respective cellulose chains. This is achieved by treating the cellulosic substrate material as a composite of several types of "sites." To be specific, we propose a model which distinguishes specifically seven different site types, referred to as *N*-, *O*-, *X*-, *Y*-, *Z*-, *L*-, and *R*-sites and labeled by a site-type index ν, as illustrated in Figures 4 and 5. We then define $n_{\nu,\sigma}(t)$ as the t-dependent average number of accessible surface sites, of each type ν, per SAC, on SACs of geometry class σ.

The first five site types, *N*, *O*, *X*, *Y*, and *Z*, represent intact β(1,4)-glucosidic bonds. Specifically, an *X*-site (*Y*-site) can adsorb, and be cut by, either an *exo-L*- (*exo-R*-) or an *endo*-acting enzyme molecule and it is thus located a distance of k_X (k_Y) G_1-monomers from the *L*-end (R-end) of the chain, where $k_X = 2$ ($k_Y = 2$) for cellohydrobiolase. A *Z*-site can adsorb and be cut by all three enzyme types. It can thus be thought of as an *X*- and *Y*-site coinciding on the same bond. An *N*-site can only adsorb, and only be cut by, an *endo*-acting enzyme molecule. An *O*-site represents an intact bond that may not adsorb any enzyme (e.g., obstruction), or may adsorb but cannot be cut by the *endo*-acting enzyme molecule (e.g., inactive adsorption). The last two, L and R, represent broken bonds or, equivalently, the non-reducing and reducing chain ends respectively. Neither *L*- nor *R*-sites adsorb any enzyme. The type-ν surface site numbers $n_{\nu,\sigma}$ must add up to the total surface monomer number $n_{M,\sigma}$, that is,

$$\sum_\nu w_\nu n_{\nu,\sigma}(t) = n_{M,\sigma}(\lambda_\sigma(t)) \tag{5}$$

where the weight factor w_ν represents the fraction of a surface-exposed monomer that is associated, on average, with one site of type ν. Hence, $w_\nu = 1$ for $\nu =$ N, O, X, Y, Z; and $w_\nu = 1/2$ for $\nu =$ L, R, as illustrated by Figures 4 and 5.

As hydrolysis proceeds in a real cellulosic substrate, glucan chains will typically become partially exposed at the ablated SAC surface while at the same time still remaining partially subducted below the surface. Barring experimental morphology data to the contrary, there is no reason to assume that any particular segment of a G_ℓ-chain is more likely to be exposed sooner, or remain subducted longer, than any other segment of the chain of the same segment length. In other words, any contiguous chain segment of given segment length k in a G_ℓ-chain is *as* likely to be fully exposed, partially exposed, or fully subducted, as any other

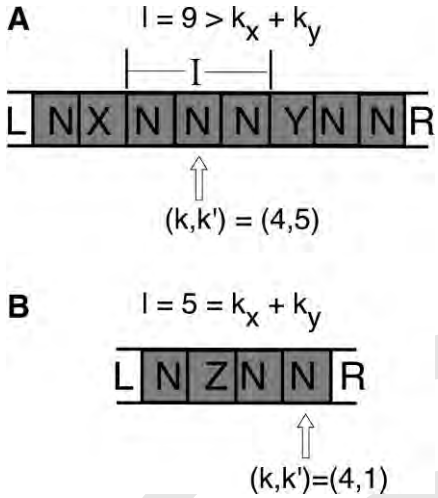

Figure 4. Distribution of the seven site types *N, X, Y, Z, L, R, O* along glucan chains of monomer lengths $\ell = 9$ (**A**) and $\ell = 5 = k_X + k_Y$ (**B**) in the homogeneously dirty chain (HDC) model. Both examples (A) and (B) are for a hypothetical system of *endo-*, *exo-L-* and *exo-R*-acting enzymes with *exo-L-* and *exo-R*-cuts to produce soluble oligomers of lengths $k_X = 2$ and $k_Y = 3$ from the *L-* and *R*-end, respectively. Square boxes represent β(1,4) glucosidic bonds between G_1 monomers; vertical lines separating boxes represent the G_1 monomers themselves. Bonds labelled with letters *N, X, Y,* or *Z* in shaded boxes, are either of the site type indicated by the letter, with probability $1 - \phi_{O,\sigma}$; or they are of site type *O*, with probability $\phi_{O,\sigma}$. Only chains of lengths $\ell \geq k_X + k_Y + 2$ have an "interior segment," indicated in (A) by the horizontal bar labelled "I." Vertical arrows are fragmentation examples: the (*k, k'*) labels indicate the monomer lengths *k* and *k'* of the resulting *L-* and *R*-end fragments if the bonds pointed to were cut.

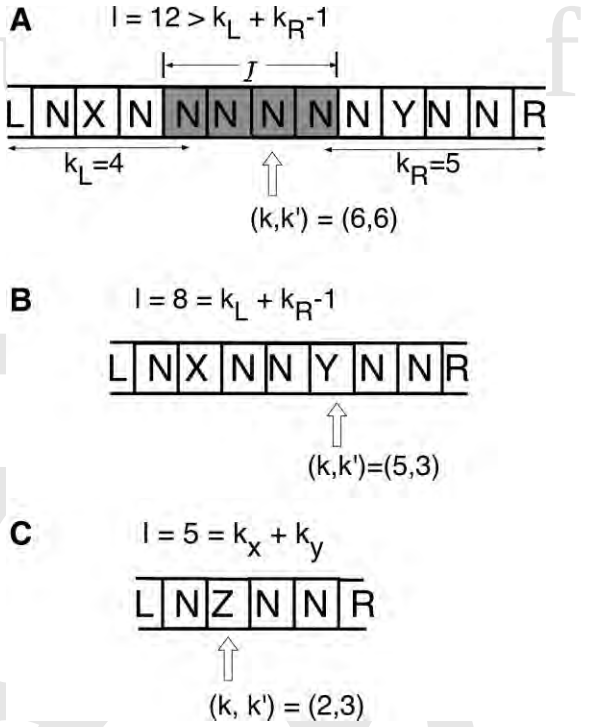

Figure 5. Distribution of the seven site types *N, X, Y, Z, L, R, O* along glucan chains of monomer lengths $\ell = 12$ (**A**), $\ell = 8 = k_L + k_R - 1$ (**B**), and $\ell = 5 = k_X + k_Y$ (**C**), in the clean chain ends (CCE) model. Notation and graphical representation of bonds, monomers, interior chain segment, and possible fragmentation examples are the same as in Figure 4, with the same hypothetical values of $k_X = 2$ and $k_Y = 3$, respectively, in all three examples (A), (B), and (C). The assumed clean chain end segments (kept free of *O*-sites) have monomer lengths $k_L = 4$ and $k_R = 5$, counted from the *L-* and *R*-end, respectively. Bonds labeled with letters *N* in shaded (gray) boxes, are either of the site type *N*, with probability $1 - \phi_{O,\sigma}$; or they are of site type *O*, with probability $\phi_{O,\sigma}$. Bonds labeled with letters *N, X, Y,* or *Z* in unshaded (white) boxes, are of the site type indicated by the letter with probability 1. Only chains of lengths $\ell \geq k_L + k_R$ have an "interior segment," indicated in (A) by the horizontal bar labelled "I." Only a bond within an interior chain segment can be of site type *O* (with probability $\phi_{O,\sigma}$).

segment of the same length k in any other chain G_ℓ of the same chain length ℓ, independent of either k-segment's position relative to its respective chain end. Based on this "uniform segment exposure" assumption, one can then, for accounting purposes, think of all surface-exposed chain segments as being reassembled into fully surface-exposed, complete chains. We can therefore assume that, on average, that is, upon averaging over many SACs of the same geometry class σ, all site counting relations of such an assembly of partially subducted chains G_ℓ are then the same as those of an equivalent, but fictitious assembly of fully exposed, sterically unobstructed chains of the same lengths ℓ. This should be kept in mind for all discussion to follow: when we consider a "whole chain exposed at the surface" we are really referring to the mathematical construct of a "whole chain, on average," reassembled from the randomly exposed pieces of real chains which are partially subducted. In other words, we are not really assuming here that all chains are either fully surface-exposed or fully subducted; but for site counting purposes, we can treat them as if they *were*. Applying this here, the number of left chain ends must, on average, equal the number of right chain ends, that is, $n_{L,\sigma} = n_{R,\sigma}$. Note that this *L*–*R*-symmetry is based on the foregoing simplifying assumption of uniform segment exposure. Substrate chain morphologies with asymmetrically subducted chain ends are in principle possible and could be treated in our formalism by way of additional site types marking *L-* and *R*-directed chain subduction loci.

The state of the SAC surfaces during hydrolytic ablation is then characterized in our model by dynamical variables $\lambda_\sigma(t)$ and $n_{\nu,\sigma}(t)$, governed by a system of coupled rate equations, derived in Supporting Information section A, of the following general form:

$$\dot{n}_{\nu,\sigma} = V_{\nu,\sigma} - \overline{V}_\sigma \eta_\sigma(\lambda_\sigma) g_{\nu\sigma}(\lambda_\sigma) \tag{6}$$

$$\dot{\lambda}_\sigma = \frac{\overline{V}_\sigma}{\partial_\lambda n_{\mathrm{V},\sigma}(\lambda_\sigma)} \tag{7}$$

$$\eta_\sigma(\lambda_\sigma) = 1 - \frac{\partial_\lambda n_{\mathrm{M},\sigma}(\lambda_\sigma)}{\partial_\lambda n_{\mathrm{V},\sigma}(\lambda_\sigma)}, \qquad \overline{V}_\sigma = \sum_\nu w_\nu V_{\nu,\sigma} \tag{8}$$

where $\partial_\lambda \ldots$ is shorthand for the λ-derivative $\partial/\partial\lambda$, the "site fragmentation rate function" $V_{\nu,\sigma}$ denotes the net rate of

production of type-ν sites at the class-σ SAC surfaces, and $\overline{V}_\sigma$ denotes net negative rate of monomer loss ($\overline{V}_\sigma < 0$) from the SAC due to ablation.

The first term in Equation (6), the "site fragmentation rate function" $V_{\nu,\sigma}$ is due to the following two types of site gain or site loss events resulting from enzymatic bond cuts:

1. A site of type ν can be gained or lost at the SAC surface by conversion from or to another SAC surface site type ν', *without* generation of any soluble chain fragment. Specifically, an *X*-, *Y*-, *L*-, *R*- and (possibly) *Z*-site is gained, and *N*-, as well as possibly *O*-site(s), are lost, in this manner when an *endo*-cut occurs at a bond located at least ℓ_S monomers from both pre-existing chain ends.
2. A site of type ν can be lost from the SAC surface by being part of a soluble chain fragment which has been generated by the cut. Specifically, sites of all type can be lost in this manner when either an *endo*- or an *exo*-cut occurs at a bond located less than ℓ_S monomers from (at least) one of the original chain ends.

The detailed functional dependence of $V_{\nu,\sigma}$ on the dynamical variables $n_{\nu,\sigma}$ and λ_σ will be described in the following sections.

The second term in Equation (6), the native surface exposure term $-\overline{V}_\sigma \eta_\sigma g_{\nu,\sigma}$, represents the rate of production of type-ν SAC surface sites due to the fact that new, underlying native substrate material becomes exposed at the surface as a result of hydrolytic removal of overlaying material in the outermost layer(s). The geometrical factor η_σ accounts for the surface curvature effect, that is, for the fact the removal of surface layers shrinks not only the volume, but also shrinks the surface of the SAC (when $d_{A,\sigma} > 1$). Consequently, removal of one monomer from the surface of a convex SAC in general exposes less than one entire monomer in the layer underneath. The native site fraction functions $g_{\nu,\sigma}(\lambda)$ represent the fraction of type-ν sites in the interior of the SAC prior to hydrolysis as long as $\lambda < \lambda_\sigma(t)$. The $g_{\nu,\sigma}$ must obey the normalization condition: $\sum_\nu w_\nu g_{\nu,\sigma}(\lambda_\sigma) = 1$.

The rate of production of dissolved monomers, contained in soluble oligomers G_k of any length $k < \ell_S$, per class-σ SAC, is given by

$$\dot{n}_{S,\sigma} = \sum_{k=1}^{\ell_S - 1} k V_{S,\sigma}(k) \tag{9}$$

where $V_{S,\sigma}(k)$ is the production rate of soluble oligomers G_k, per class-σ SAC. The construction of $V_{S,\sigma}$ and $V_{\nu,\sigma}$ requires the enzymatic bond cutting reaction rate coefficients, the cellulose chain fragmentation probabilities, and the concomitant solutions of the enzyme–substrate adsorption equilibria. This will be discussed in detail in the next section.

Ablation and Oligomer Rate Functions

For an *endo*- or *exo*-acting enzyme to cut an intact $G_1 - G_1$ bond, the enzyme must first form an enzyme–substrate (ES) complex, by adsorption to the bond site exposed on the SAC surface. The cutting rate for intact bonds of site type μ, having adsorbed an enzyme of type κ on the surface of an SAC of class σ, is thus given by the product of rate coefficient and ES complex concentration, written in the form $\gamma_{\kappa,\mu} m_{\kappa,\mu,\sigma}$, where, in our three enzyme model, the values $\kappa = 1$, 2, or 3 represent the *endo*-, *exo*-*L*- and *exo*-*R*-acting glucanase, respectively; and $m_{\kappa,\mu,\sigma}$ is the number of (κ,μ) ES complexes per class-σ SAC; and $\gamma_{\kappa,\mu}$ is the cutting rate coefficient (in units of cuts per second per ES complex). As described in Supporting Information section B, the $m_{\kappa,\nu,\sigma}$ are calculated as functions of the $n_{\mu,\sigma}$ by solving the enzyme adsorption equilibrium equations. We assume $\gamma_{\kappa,\mu}$ to be the same for all geometry classes σ. If warranted by experimental evidence, $\gamma_{\kappa,\mu}$ could easily be made dependent on σ and/or λ_σ in our formalism.

Cutting a bond at a site of site type μ will then in general change the number of sites for several or all site types ν at the SAC surface, on average, by an increment $\Delta\overline{N}_{\nu,\mu,\sigma}$ per μ-bond being cut. The net rate of production of type-ν sites resulting from cuts of all bond site types μ subject to all enzyme types κ, per SAC of class σ, can thus be written as

$$V_{\nu,\sigma} = \sum_{\kappa,\mu} \gamma_{\kappa,\mu} m_{\kappa,\mu,\sigma} \Delta\overline{N}_{\nu,\mu,\sigma} \tag{10}$$

$$\Delta\overline{N}_{\nu,\mu,\sigma} = \sum_{k=1}^{\infty} \sum_{k'=1}^{\infty} P_\sigma(k,k'|\mu,+1) \Delta N_{\nu,\sigma}(k,k') \tag{11}$$

In Equation (11), $P_\sigma(k,k'|\mu,+1)$ denotes the probability for a randomly selected intact bond of given site type μ to be located k monomers from the *L*- and k' monomers from the *R*-end of a surface exposed chain on a class-σ SAC. $P_\sigma(k,k'|\mu,+1)$ is therefore also the probability for the cut of a randomly selected type-μ bond to generate two chain fragments consisting of k monomers at the original *L*-end and k' monomers at the original *R*-end, for the surface-exposed chain G_ℓ of length $\ell = k + k'$. Also, $\Delta N_{\nu,\sigma}(k,k')$ denotes the increment of type-ν sites that is produced by a bond cut generating a $(G_k, G_{k'})$ chain fragment pair. As detailed in Supporting Information section C, $\Delta N_{\nu,\sigma}(k,k')$ depends on the distribution of site types along the chain which must be provided as a model input. Two specific examples of such chain site distribution models, the homogeneously dirty chain (HDC) model and clean chain ends (CCE) model, are illustrated in Figures 4 and 5, respectively, and described in Supporting Information section C.

The production rate of soluble oligomers G_k, $V_{S,\sigma}(k)$, is likewise be expressed as:

$$V_{S,\sigma}(k) = \sum_{\kappa,\mu} \gamma_{\kappa,\mu} m_{\kappa,\mu,\sigma} \sum_{k'=\ell_S-k}^{\infty} [P_\sigma(k,k'|\mu,+1) + P_\sigma(k',k|\mu,+1)] \quad (12)$$

The resulting oligomer production rates $\dot{n}_{S,\sigma}$ from Equation (9) can then be shown to obey general monomer conservation laws, derived in Supporting Information section D.

In Supporting Information section E, $P_\sigma(k,k'|\mu,+1)$ is expressed in terms of the chain length probability of surface-exposed chains, denoted by $P_\sigma(\ell)$; and in terms of a conditional probability, $P_\sigma(\mu|k,k',+1)$, for a cut to occur at site of type μ, given the resulting fragment lengths k and k'. As discussed in Supporting Information section E, $P_\sigma(\mu|k,k',+1)$ is a model input determined by the chain site distribution model. Using Equation (50) of Supporting Information section E, the site rate functions $V_{\nu,\sigma}$ and the soluble oligomer production rate functions $V_{S,\sigma}(k)$ for $k < \ell_S$ then become:

$$V_{\nu,\sigma} = \sum\nolimits_{\kappa,\mu} \gamma_{\kappa,\mu} m_{\kappa,\mu,\sigma} \left(\frac{n_{L,\sigma}}{n_{\mu,\sigma}}\right) \times \sum\nolimits_{k,k'=1}^{\infty} \Delta N_{\nu,\sigma}(k,k') P_\sigma(\mu|k,k',+1) P_\sigma(k+k') \quad (13)$$

$$V_{S,\sigma}(k) = \sum\nolimits_{\kappa,\mu} \gamma_{\kappa,\mu} m_{\kappa,\mu,\sigma} \left(\frac{n_{L,\sigma}}{n_{\mu,\sigma}}\right) \times \sum\nolimits_{k'=\ell_S-k}^{\infty} [P_\sigma(\mu|k,k',+1) + P_\sigma(\mu|k',k,+1)] P_\sigma(k+k') \quad (14)$$

We thus need, in principle, the complete surface chain length distribution $P_\sigma(\ell)$, including its full time evolution during hydrolysis, as an input into Equation (13) in order to evaluate $V_{\nu,\sigma}$ and $V_{S,\sigma}(k)$. However, as we will show in the next section, within the framework of the site number formalism developed here, one can devise approximate treatments where, in fact, one does not have to solve the corresponding rate equations for the full chain length distribution.

Chain End Decomposition

In this section, we will show that the set of chain number probability variables, $P_\sigma(\ell)$ entering into the rate functions $V_{\nu,\sigma}$ and $V_{S,\sigma}$ can effectively be reduced to only a few "short" chain lengths ℓ up to some short-chain cut-off, ℓ_C, which is typically of the same order as ℓ_S. To do so, we rewrite both $P_\sigma(\mu|k,k',+1)$ and $N_{\nu,\sigma}(\ell)$, so as to separate the effects of the near-chain-end sites, where both *exo* and *endo*-activity may occur, from the chain interior sites, where only *endo*-activity occurs. Underlying this decomposition is the assumption of "chain homogeneity" and "chain end locality," that is, the basic idea that the interior of a sufficiently long chain is essentially homogeneous and unaffected by chain ends.

Formally, this is expressed by assuming that the conditional site type probability can be decomposed into chain end (L,R), chain interior (I) and short-chain (S) contributions of the following functional form:

$$P_\sigma(\mu|k,k',+1) = p^{(I)}_{\mu,\sigma} + \Theta_L(k) p^{(L)}_{\mu,\sigma}(k) + \Theta_R(k') p^{(R)}_{\mu,\sigma}(k') + \Theta_L(k)\Theta_R(k') p^{(S)}_{\mu,\sigma}(k,k') \quad (15)$$

The cut-off factors, $\Theta_L(k) \equiv \Theta(\ell_L - k)$ and $\Theta_R(k') \equiv \Theta(\ell_R - k')$, subdivide a chain G_ℓ of length $\ell = k + k'$ with $\ell > \ell_{LR} \equiv \ell_L + \ell_R - 1$ into three segments, as illustrated in Figures 4A and 5B: an interior (I-) segment of $\ell - \ell_{LR}$ (intact bond) sites, flanked by a terminal L-segment containing the ℓ_L left-most monomers, and by a terminal R-segment containing the ℓ_R right-most monomers. The terminal segment lengths ℓ_L and ℓ_R signify the maximum "range" of the chain end effects on the physical properties of a site inside the chain: the L-end (R-end) can affect only bond sites within ℓ_L (ℓ_R) monomers from the L-end (R-end).

The four chain segment terms $p^{(I)}_{\mu,\sigma}$, $p^{(L)}_{\mu,\sigma}(k)$, $p^{(R)}_{\mu,\sigma}(k')$, and $p^{(S)}_{\mu,\sigma}(k,k')$, denote contributions from the interior, L-terminal and R-terminal segments, and from short chains with $\ell \le \ell_{LR} \equiv \ell_L + \ell_R - 1$, respectively. The crucial point to note is that, for chain lengths $\ell = k + k' > \ell_{LR}$, $P_\sigma(\mu|k,k',+1) = p^{(I)}_{\mu,\sigma}$ is independent of k and k' at interior site locations, that is, when $k \ge \ell_L$ and $k' \ge \ell_R$. Likewise, for L-terminal sites with $k < \ell_L$ and $k' \ge \ell_R$, $P_\sigma(\mu|k,k',+1) = p^{(I)}_{\mu,\sigma} + p^{(L)}_{\mu,\sigma}(k)$ is independent of k'; and for R-terminal sites with $k' < \ell_R$ and $k \ge \ell_L$, $P_\sigma(\mu|k,k',+1) = p^{(I)}_{\mu,\sigma} + p^{(R)}_{\mu,\sigma}(k')$ is independent of k.

These four chain segment terms can be straightforwardly extracted from the underlying chain site distribution model, such as the HDC model with $\ell_L = k_X + 1$ and $\ell_R = k_Y + 1$; and the CCE model with $\ell_L = k_L$ and $\ell_R = k_R$. By way of Equation (47), Equation (15) implies that the chain site counting functions $N_{\nu,\sigma}$ can be decomposed in the form

$$N_{\nu,\sigma}(\ell) = b_{\nu,\sigma}\ell + a_{\nu,\sigma} + \Theta_T(\ell) d_{\nu,\sigma}(\ell) \quad (16)$$

with ℓ-independent $a_{\nu,\sigma}$ and $b_{\nu,\sigma}$, and $d_{\nu,\sigma}(\ell) \equiv N_{\nu,\sigma}(\ell) - b_{\nu,\sigma}\ell - a_{\nu,\sigma}$. The cut-off factor is $\Theta_T(\ell) = \Theta(\ell_T - \ell)$ where $\ell_T \equiv \max(\ell_S, \ell_L + \ell_R - 1)$. Further details concerning $a_{\nu,\sigma}$, $b_{\nu,\sigma}$ and $d_{\nu,\sigma}(\ell)$ are given in Supporting Information section F.

Using Equations (15) and (16) via Equation (39) in Supporting Information section C, the average increment

functions $\Delta\overline{N}_{\nu,\mu,\sigma}$ are decomposed into

$$\Delta\overline{N}_{\nu,\mu,\sigma} = \left(\frac{1}{n_{\mu,\sigma}}\right)[A_{\nu,\mu,\sigma} n_{\mathrm{M},\sigma}(\lambda_\sigma) + (B_{\nu,\mu,\sigma} + D_{\nu,\mu,\sigma}) n_{\mathrm{L},\sigma}] \tag{17}$$

By way of Equation (10), Equation (17) generates a chain-end decomposition for $V_{\nu,\sigma}$. An analogous decomposition of $V_{\mathrm{S},\sigma}(k)$ for $k < \ell_{\mathrm{S}}$ is provided in Supporting Information section F. As detailed in Supporting Information section F, the $A_{\nu,\mu,\sigma}$- and $B_{\nu,\mu,\sigma}$-terms contain contributions from cutting chains of any length, whereas the $D_{\nu,\mu,\sigma}$-terms contain contributions from cutting only short chains G_ℓ with lengths $\ell \le \ell_{\mathrm{C}}$ where $\ell_{\mathrm{C}} \equiv \ell_{\mathrm{T}} + \max(\ell_{\mathrm{L}}, \ell_{\mathrm{R}}) - 2$. The $A_{\nu,\mu,\sigma}$ and $B_{\nu,\mu,\sigma}$ are therefore constant coefficients which do not depend on any $P_\sigma(\ell)$ or other dynamical variables, whereas the $D_{\nu,\mu,\sigma}$ depend explicitly *only* on the $P_\sigma(\ell)$ with $\ell \le \ell_{\mathrm{C}}$. As a result, $\Delta\overline{N}_{\nu,\mu,\sigma}$, thus $V_{\nu,\sigma}$ and likewise $V_{\mathrm{S},\sigma}$ (see Supporting Information section F), depend explicitly *only* on chain number variables $P_\sigma(\ell)$ with $\ell \le \ell_{\mathrm{C}}$, while their dependence on the longer-chain $P_\sigma(\ell)$-variables (with $\ell > \ell_{\mathrm{C}}$) is absorbed entirely into the site number variables $n_{\nu,\sigma}$ and $n_{\mathrm{M},\sigma}(\lambda_\sigma)$ entering in Equation (17). We therefore need to find a solution *only* for the short-chain part of $P_\sigma(\ell)$, with chain lengths $\ell \le \ell_{\mathrm{C}}$, serving as input into the $D_{\nu,\mu,\sigma}$-contributions, to generate a complete closed system of rate equations for the site number variables $n_{\nu,\sigma}$.

Rate Equation Closure in the Long-Chain Limit

In order to find a solution for $P_\sigma(\ell)$ with $\ell \le \ell_{\mathrm{C}}$, we first develop a rate equation system, equivalent to the site number formalism, for chain number variables $H_\sigma(\ell) \equiv P_\sigma(\ell) n_{L,\sigma}$; and then apply the chain-end decomposition to the resulting chain number rate functions. This is described in Supporting Information section G. To solve these chain number rate equations for $H_\sigma(\ell)$ with lengths ℓ up to some short-chain cut-off $\ell_{\mathrm{D}} \ge \ell_{\mathrm{C}}$, only short-chain values of $P_\sigma(j)$, for chain lengths $j \le \ell_{\mathrm{D}} + \ell_{\mathrm{E}}$, are required as input into the chain number rate functions: just as in the case of the site number rate functions, the dependence of the chain number rate functions on the longer-chain $P_\sigma(j)$-variables (with $j > \ell_{\mathrm{D}} + \ell_{\mathrm{E}}$) can be absorbed entirely into site number variables $n_{\nu,\sigma}$ and $n_{\mathrm{M},\sigma}(\lambda_\sigma)$.

Our goal is thus to construct a closed system of coupled rate equations involving *only* short-chain number variables $H_\sigma(\ell)$ with lengths $\ell \le \ell_{\mathrm{D}}$, combined with the site number rate equations (6) and (7), but *without* having to explicitly solve any chain number rate equations for longer chains having $\ell > \ell_{\mathrm{D}}$. We propose the local Poisson (LP) approximation scheme, described below, to achieve this closure in the long-chain limit (LCL). The LCL is defined by the condition that, at least at the start of hydrolysis, at time $t^{(\mathrm{o})}$, the vast majority of all cellulose chains have lengths that are much greater than the relevant "short-chain" length parameter ℓ_{C}. In other words, in the LCL, there is negligible probability weight for finding any "short chains," of order $\ell \sim \ell_{\mathrm{C}}$.

Furthermore, under LCL conditions, the vast majority of short insoluble chains are generated by *endo*-cuts of long chains G_j with $j \gg \ell_{\mathrm{C}}$. Also, all short chains G_ℓ with lengths of the order $\ell \sim \ell_{\mathrm{C}}$, but $\ell \ge \ell_{\mathrm{C}}$, are being produced at similar rate from the long G_j-chains, since the relevant bond cuts occur in the interior segments of the G_j-chains. Consequently, for short chain lengths with $\ell \ge \ell_{\mathrm{C}}$, we can expect $H_\sigma(\ell)$ to be a slowly varying function of ℓ. Therefore, given the values of, say, $H_\sigma(\ell_{\mathrm{D}})$ and $H_\sigma(\ell_{\mathrm{D}} - 1)$ for some short $\ell_{\mathrm{D}} \sim \ell_{\mathrm{C}}$, but $\ell_{\mathrm{D}} \ge \ell_{\mathrm{C}}$, we should be able to approximate locally, for example by extrapolation, the values of $H_\sigma(j)$ at larger, "nearby" chain lengths $j = \ell_{\mathrm{D}} + 1, \ldots, \ell_{\mathrm{D}} + \ell_{\mathrm{E}}$. Our proposed LP approximation scheme uses a linear extrapolation of $\log P_\sigma(j)$ which can be expressed as

$$P_\sigma(j) \simeq P_\sigma(\ell_{\mathrm{D}}) \left[\frac{P_\sigma(\ell_{\mathrm{D}})}{P_\sigma(\ell_{\mathrm{D}} - 1)}\right]^{j-\ell_{\mathrm{D}}} \quad \text{for } j = \ell_{\mathrm{D}} + 1, \ldots, \ell_{\mathrm{D}} + \ell_{\mathrm{E}} \tag{18}$$

with $P_\sigma(\ell) \equiv H_\sigma(\ell)/n_{L,\sigma}$ for $\ell \le \ell_{\mathrm{D}}$.

Consequently, the LP approximation results in a closed ODE system, for the site number formalism, consisting of Equations (6), (7), (8), (10), (17), and (18); of Equations (60) and (65) for $\ell = \ell_{\mathrm{S}}, \ldots, \ell_{\mathrm{D}}$, and (66) in Supporting Information section G, of Equations (52) and (53) in Supporting Information section F; and of the enzyme adsorption equilibrium equations in Supporting Information section B. The independent dynamical variables of this ODE system are the $n_{\nu,\sigma}$, λ_σ and short-chain numbers $H_\sigma(\ell)$, $\ell = \ell_{\mathrm{S}}, \ldots, \ell_{\mathrm{D}}$. Intermediate variables $m_{\kappa,\mu,\sigma}$ are evaluated from enzyme adsorption equations (see Supporting Information section B). Other intermediate variables are evaluated directly from the expressions given before. In addition, one can then use Equations (9) and (55), (56), (57) in Supporting Information section F to obtain the concentrations of dissolved oligomers G_k, $k = 1, \ldots, \ell_{\mathrm{S}}$ and total dissolved monomers $n_{\mathrm{S},\sigma}$

The foregoing ODE system is subject to the initial conditions, at starting time $t^{(\mathrm{o})}$:

$$n_{\nu,\sigma}(t^{(\mathrm{o})}) = g_{\nu,\sigma}\left(\lambda_\sigma{}^{(\mathrm{o})}\right) n_{\mathrm{M},\sigma}\left(\lambda_\sigma{}^{(\mathrm{o})}\right) \tag{19}$$

$$\lambda_\sigma(t^{(\mathrm{o})}) = \lambda_\sigma{}^{(\mathrm{o})} \tag{20}$$

$$H_\sigma(\ell, t^{(\mathrm{o})}) = \epsilon_Q\, n_{L,\sigma}(t^{(\mathrm{o})}) \quad \text{for } \ell = \ell_{\mathrm{S}}, \ldots, \ell_{\mathrm{D}} \tag{21}$$

We thus restrict ourselves to native chain length distributions without short chains, that is, $Q_\sigma(\ell, \lambda) = \varepsilon_Q$ for $\ell \le \ell_{\mathrm{D}}$, with near-zero ε_{Q} (e.g., 10^{-20}, but cannot be exact zero). Numerically, this initialization is found to produce $P_\sigma(j, t)$, for $j = \ell_{\mathrm{D}} - 1, \ldots, \ell_{\mathrm{D}} + \ell_{\mathrm{E}}$, that are

approximately exponentially *decreasing* with j. This distribution shape is also supported by an exact analytical result, showing that the Poisson shape is indeed an exactly preserved shape of $H_\sigma(\ell)$, and thus of $P_\sigma(\ell)$, under time evolution by the full-chain rate equations in the absence of the new surface exposure term, as proven in Supporting Information section H. In our companion article II (Zhou et al., 2009), an approximate Poisson distribution shape at short chain lengths is also consistently observed, from early through intermediate to late stages of hydrolytic conversion, in numerical solutions of the full chain number rate equations. In that paper, we will demonstrate that the LP approximation is indeed highly accurate for typical substrates with typical DP-values exceeding chain lengths of 20 monomers, using naturally occurring enzyme compositions.

Discussion

A chain-number-based fragmentation kinetics model was recently applied to the three-enzyme cellulose hydrolysis system by Zhang and Lynd (2006), following earlier work by Okazaki and Moo-Young (1978). Our general chain number surface ablation formalism, developed in Supporting Information section G, reduces mathematically to their model if we allow only for the five basic site types $\nu = N$, X, Y, L and R; include only a single geometry class $\sigma = 1$; adopt the low-enzyme limit; and discard the surface exposure term, which means that all glucan chains are assumed to be already fully exposed at enzyme-accessible surfaces at the start of hydrolysis. To correct for this substrate "over-exposure," the effective chain concentrations available for enzyme adsorption were reduced by a time-independent factor equal to the initial accessibility fraction $\overline{F}_a(t^{(o)})$ (Okazaki and Moo-Young, 1978; Zhang and Lynd, 2006). This correction may be adequate in the limit of very short hydrolysis times and in the presence of a very small amount of enzymes, so that only a very small fraction of substrate is converted to solute; and thus the accessibility fraction F_a can be assumed to remain unchanged. This was, appropriately, the limit actually studied by Zhang and Lynd.

By contrast, when a sizeable fraction of the substrate is hydrolyzed, the surface morphology itself, and hence the accessible substrate surfaces, will be modified substantially in the process. Our surface ablation approach is specifically designed to explicitly model the hydrolytic time evolution of the substrate morphology, coupled to the hydrolysis kinetics. Both the effects of morphology on the hydrolysis kinetics, and the effects of hydrolytic conversion on the morphology are taken into account on an equal footing. Our approach should thus be applicable to the entire hydrolysis process, up to the point of complete hydrolytic conversion. In our companion article II (Zhou et al., 2009), we will demonstrate this capability by a series of numerical simulations based on the general formalism developed here.

One significant computational advantage of the site number formalism, compared to conventional chain number approaches, is a substantial reduction in the size of the resulting ODE system that needs to solve. The number of independent variables $H_\sigma(\ell)$ for conventional chain number approaches are of the order of the initial substrate DP, for each geometry class σ. By contrast, in our site number approach, the number of independent variables per geometry class σ is independent of the substrate DP and no more than the number of site types (e.g., 7). Thus, the site number approach could easily realize a saving of at least 1–2 orders of magnitude in ODE system size and computational effort, relative to the corresponding full chain number formalism.

To capture the time-evolving substrate morphology, two essential geometrical concepts have been introduced: the subdivision of the substrate, on the basis of the *enzyme accessibility* of the substrate internal surfaces, into SACs; and the subdivision of each such SAC unit into elementary layers. These two geometrical constructs are of very general applicability, relying only on the notion of steric enzyme accessibility, but not on any specific assumptions about the crystallinity, chain ordering or other spatial organization. The specific very simple prototypical power-law SAC geometry model for $n_V(\lambda)$ assumed here is of course an over-simplification which may need to be replaced by more sophisticated models that are informed by detailed mesoscopic morphology data. Our modeling formalism developed here is of sufficient generality and flexibility to accommodate such model refinements. Lastly, although SACs represent physically well-defined particulate subunits of the substrate, it is important to realize that an SAC cannot necessarily be identified with a conventional "particle" as observed, for example, by conventional microscopy: the SAC construct depends not only on substrate morphology, but also on enzyme molecular size and shape. Special surface decoration experiments would therefore be required to achieve a direct imaging of the SAC surfaces. Order-of-magnitude estimates suggest that the linear sizes of our SAC units in typical real substrates can be significantly larger, by factors of up to 10^2, than either the relevant enzyme molecular sizes or the sizes of, say, micro- or elementary fibril units identified in as-grown cellulosic materials (Himmel et al., 2007).

Conclusions

We have developed a general theoretical framework for the kinetic rate equation modeling of the enzymatic hydrolysis of solid cellulose substrates in the presence of hydrolytic evolution of a random substrate morphology. This is achieved by explicitly coupling the enzymatic fragmentation kinetics of surface-exposed, enzyme-accessible cellulosic glucan chains to a geometric rate equation description of the time-evolving substrate morphology, resulting from the hydrolytic loss of solid material from substrate external and internal surfaces. To account for the fact that real substrates will typically exhibit a random distribution of surface

geometries, we then represent the macroscopic substrate as a population of SAC geometry classes. The foregoing morphological constructs are then integrated with a general chain fragmentation kinetics approach. The resulting surface ablation formalism describes the enzymatic hydrolysis of β(1,4) glucosidic bonds that are exposed at enzyme accessible SAC surfaces with time-dependent SAC geometries. The SAC geometries are described by time-dependent SAC layer number variables λ_σ which evolve, for each geometry class σ, according to geometry rate equations.

To treat the chain fragmentation kinetics, we have developed a novel site number representation which results in a substantially reduced ODE system size, compared to the earlier conventional chain fragmentation kinetics approaches based on a chain number representation. Our new integrated morphology-plus-kinetics rate equation formalism thus opens up a new modeling approach for detailed quantitative simulations of both the hydrolytically evolving random substrate morphology *and* of the profound, and heretofore neglected, morphology effects on the hydrolysis kinetics. By incorporating the morphology evolution on an equal footing with the hydrolytic chain fragmentation, our formalism provides a framework for the realistic modeling of the entire solubilization process, beyond the short-time limit and through near-complete hydrolytic conversion. This numerical modeling capability will be illustrated by detailed simulation results and testable experimental predictions presented in our companion paper II (Zhou et al., 2009).

Nomenclature

$a_{\nu,\sigma}$, $A_{\nu,\mu,\sigma}$	decomposition parameters used in Equations (16) and (17)
$b_{\nu,\sigma}$, $B_{\nu,\mu,\sigma}$	decomposition parameters used in Equations (16) and (17)
C_σ	class-σ SAC concentration, mM (moles of SACs in class-σ per reactor volume)
$c_{V,\sigma}$	volume prefactor to calculate $n_{V,\sigma}$
$d_{A,\sigma}$	ablation dimension for class-σ SACs
$d_{\nu,\sigma}$, $D_{\nu,\mu,\sigma}$	decomposition parameters used in Equations (16) and (17)
$F_{a,\sigma}$	fraction of accessible G_1 for class-σ SACs, $\equiv n_{M,\sigma}/n_{V,\sigma}$
$\overline{F}_a$	overall accessibility fraction of accessible G_1, $\equiv n_M/n_V$
$f_{\nu,\sigma}$	type-ν site fraction on class-σ SAC surfaces, $\equiv n_{\nu,\sigma}/n_M$
$g_{\nu,\sigma}$	native type-ν site fraction in class-σ SACs
G_1	anhydro-glucose ($C_6H_{10}O_5$) monomers
G_ℓ	glucan chain consisting of ℓ G_1 units
$H_\sigma(\ell)$	number of surface-exposed G_ℓ per class-σ SAC
k, k'	number of G_1-monomers in a glucan chain or chain fragment G_k or $G_{k'}$, respectively
k_L, (k_R)	length of terminal chain segments at the L-end (R-end), which are devoid of O-sites, in the CCE model
k_X, (k_Y)	site position from L-end (R-end) where exo-L (*exo*-R) act at
ℓ	number of G_1-monomers in an insoluble glucan chain G_ℓ
ℓ_S	minimum insoluble glucan chain length
$\langle\ell\rangle_\sigma$	average chain length for chains exposed on class-σ SAC surfaces
ℓ_L, (ℓ_R)	length of terminal segment that L-end (R-end) can affect
ℓ_{LR}	$\ell_{LR} \equiv \ell_L + \ell_R - 1$
ℓ_T	$\ell_T \equiv \max(\ell_S, \ell_L + \ell_R - 1)$
ℓ_C	$\ell_C \equiv \max(\ell_S, \ell_L + \ell_R - 1) + \max(\ell_L, \ell_R) - 2$
ℓ_E	$\ell_E \equiv \max(\ell_L, \ell_R) - 1$
ℓ_D	$\ell_D \sim \ell_C$, but $\ell_D \geq \ell_C$
$L_{\kappa,\mu}$	adsorption coefficient for (κ, μ) ES complex (1/mM)
$m_{\kappa,\mu,\sigma}$	number of (κ, μ) ES complexes per class-σ SAC
M_{MD}	population size of SAC geometries
$n_{M,\sigma}$	number of G_1 exposed at the surface per class-σ SAC
$n_{\nu,\sigma}$	number of type-ν sites at the surface per class-σ SAC
$n_{S,\sigma}$	number of dissolved G_1 produced per class-σ SAC
$n_{V,\sigma}$	total number of G_1 contained per class-σ SAC
$N_{\nu,\sigma}(k)$	average number of type-ν sites per glucan chain G_k, in class-σ SACs
$\Delta\overline{N}_{\nu,\mu,\sigma}$	mean increment of type-ν sites on class-σ SAC surfaces per μ-bond being cut
$\Delta N_{\nu,\sigma}(k,k')$	increment of type-ν sites produced by a bond cut generating a $(G_k, G_{k'})$ chain fragment pair
$p^{(I)}_{\mu,\sigma}$, $p^{(L)}_{\mu,\sigma}$, $p^{(R)}_{\mu,\sigma}$, $p^{(S)}_{\mu,\sigma}$	contributions to probability for finding type-μ, on class-σ SAC surfaces from the interior, L-terminal and R-terminal segments, and from short chains with $\ell \leq \ell_{LR}$, respectively
$P_\sigma(\ell)$	probability of a randomly selected insoluble glucan chain, exposed on a class-σ SAC surface, to contain ℓ G_1 monomers; $\equiv H_\sigma(\ell)/n_{L,\sigma}$
$P_\sigma(k,k',\zeta)$	probability that a bond randomly selected from the superchain is a ζ-bond, *and* that this bond be located $k \geq 1$ monomers *and* $k' \geq 1$ monomers from its nearest L-end and R-end, respectively
$P_\sigma(k,k'\|\mu,+1)$	probability for a randomly selected intact bond of given site type μ to be located k monomers from the L- and k' monomers from the R-end of a surface exposed chain on a class-σ SAC
$P_\sigma(\mu\|k,k',\zeta)$	probability for a randomly selected superchain bond to be of site type μ, *given that* the bond *is* a ζ-bond; and given that it *is* located k and k' monomers from its nearest L-end and R-end, respectively
$Q_\sigma(\ell)$	native chain length distribution in class-σ SACs
μ_κ	total type-κ enzyme concentration (mM)
ν_κ	free type-κ enzyme concentration (mM)
$V_{H,\sigma}(\ell)$	production rate of G_ℓ at the surface per class-σ SAC, (number of G_ℓ per min per SAC unit)
$V_{\nu,\sigma}$	production rate of type-ν site at the surface per class-σ SAC, (number of type-ν site per min per SAC unit)
$V_{S,\sigma}(k)$	production rate of soluble oligomer G_k per class-σ SAC
$V_\sigma(\ell \to k,k')$	rate at which chains G_ℓ, exposed on class-σ SAC surfaces, are being cut into fragments G_k and $G_{k'}$, from the original chain L- and R-end, respectively
$\overline{V}_\sigma$	negative rate of monomer loss into solution per class-σ SAC
w_ν	number of monomers represented by one type-ν site

$x_{\nu,\sigma}$	concentration of type-ν site exposed on class-σ SAC surfaces, mM, $\equiv C_\sigma n_{\nu,\sigma}$
x_{M}	total concentration of G_1 exposed on surfaces (mM)
$x_{\mathrm{M},\sigma}$	concentration of G_1 exposed on class-σ SAC surfaces (mM)
x_{V}	total concentration of G_1 in solid substrate (mM)
$x_{\mathrm{V},\sigma}$	concentration of G_1 contained in class-σ SACs (mM)
$y_{\mu,\sigma}$	concentration of free type-μ sites on class-σ SAC surfaces (mM)
$z_{\kappa,\mu,\sigma}$	concentration of (κ, μ) ES complex on class-σ SAC surfaces, mM, $\equiv C_\sigma m_{\kappa,\mu,\sigma}$

Greek Symbols

β_κ	type-κ enzyme footprint
η_σ	geometrical factor accounting for surface curvature effect
κ	index of enzyme types, $\kappa = 1$, 2 or 3 represent the *endo-*, *exo-L-* and *exo-R*-acting glucanase, respectively
$\gamma_{\kappa,\mu}$	cutting rate coefficient (cuts per time per (κ, μ) ES complex)
λ_σ	layer number variable of class-σ SACs
ν, μ	index of site types, *N*, *L*, *R*, *X*, *Y*, *Z* or *O*
$\phi_{O,\sigma}$	fraction of *O*-sites in class-σ SAC chains
σ	index of SAC classes
$\Theta(\Delta\ell)$	heavyside step function, $= 1$ if $\Delta\ell > 0$, $= 0$ otherwise; for any real or integer $\Delta\ell$
ξ_σ	molar fraction of G_1 contained in class-σ SACs, $\equiv x_{\mathrm{V},\sigma}/x_{\mathrm{V}}$
ζ	integrity variable with $\zeta = +1$ (-1) indicating intact (broken) bond, in the superchain construction of fragmentation probability

Abbreviations

CCE	"clean chain ends," a chain site distribution model
DP	degree of polymerization
endo enzyme	endoglucanase
ES	"enzyme substrate" complex
exo-L enzyme	exoglucanase acting at non-reducing end of a cellulose chain
exo-R enzyme	exoglucanase acting at reducing end of a cellulose chain
GP	global Poisson
HDC	"homogeneously dirty chain," a chain site distribution model
LCL	long chain limit
L-end	non-reducing end of a cellulose chain, also called "left" end
LP	"local Poisson" an approximation scheme
ODE	ordinary differential equation
R-end	reducing end of a cellulose chain, also called "right" end
SAC	smallest accessible compartment
SAV	smallest accessible void

We acknowledge the U.S. Department of Energy (grant # 4000063512) and the National Science Foundation (grant #: NSF/DBI-0354771, NSF/ITR-IIS-0407204, NSF/DBI-0542119, and NSF/CCF0621700) for financial support. We also acknowledge computing support from the University of Georgia Research Computing Center. The BioEnergy Science Center is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

References

Bothwell MK, Daughhetee SD, Chaua GY, Wilson DB, Walker LP. 1997. Binding capacities for *Thermomonospora fusca* E3, E4 and E5, the E3 binding domain, and *Trichoderma reesei* CBHI on Avicel and bacterial microcrystalline cellulose. Biores Technol 60:169–178.

Chang MM, Chou TYC, Tsao GT. 1981. Structure, pretreatment and hydrolysis of cellulose. Adv Biochem Eng 20:15–42.

Converse AO. 1993. Substrate factors limiting enzymatic hydrolysis. In: Saddler JN, editor. Bioconversion of forest and agricultural plant residues. Wallingford: CAB International. pp. 93–106.

Converse AO, Grethlein HE. 1987. On the use of an adsorption model to represent the effect of steam explosion pretreatment on the enzymatic hydrolysis of lignocellulosic substances. Enzyme Microb Technol 9:79–82.

Converse AO, Optekar JD. 1993. A synergistic kinetics model for enzymatic cellulose hydrolysis compared to degree-of-synergism experimental results. Biotechnol Bioeng 42:145–148.

Converse AO, Matsuno R, Tanaka M, Taniguchi M. 1988. A model of enzyme adsorption and hydrolysis of microcrystalline cellulose with slow deactivation of the adsorbed enzyme. Biotechnol Bioeng 32:38–45.

Fenske JJ, Penner MH, Bolte JP. 1999. A simple individual-based model of insoluble polysaccharide hydrolysis: The potential for autosynergism with dual-activity glycosidases. J Theor Biol 199:113–118.

Gan Q, Allen SJ, Taylor G. 2003. Kinetic dynamics in heterogeneous enzymatic hydrolysis of cellulose: An overview, an experimental study and mathematical modeling. Proc Biochem 38:1003–1018.

Gilkes NR, Jervis E, Henrissat B, Tekant B, Miller RC, Jr., Warren RAJ, Kilburn DG. 1992. The adsorption of a bacterial cellulase and its two isolated domains to crystalline cellulose. J Biol Chem 267:6743–6749.

Goldberg MW. 2008. Immunolabeling for scanning electron microscopy (SEM) and field emission SEM. Meth Cell Biol 88:109–130.

Grethlein HE. 1985. The effect of pore size distribution on the rate of enzymatic hydrolysis of cellulose substrates. Bio Technol 3:155–160.

Henrissat B, Teeri TT, Warren RAJ. 1998. A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants. FEBS Lett 425:352–354.

Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD. 2007. Biomass recalcitrance: Engineering plants and enzymes for biofuels production. Science 315:804–807.

Hong J, Ye X, Zhang Y-HP. 2007. Quantitative determination of cellulose accessibility to cellulase based on adsorption of a nonhydrolytic fusion protein containing CBM and GFP with its applications. Langmuir 23:12535–12540.

Hsu T-A. 1996. Pretreatment of biomass. In: Wyman CE, editor. Handbook on biomass ethanol. Washington, DC: Taylor and Francis.

Jacobsen SE, Wyman CE. 2000. Cellulose and hemicellulose hydrolysis models for application to current and novel pretreatment processes. Appl Biochem Biotechnol 84:81–96.

Klemm D, Philipp B, Heinze T, Heinze U, Wagenknecht W. 1998. Comprehensive cellulose chemistry. I. Fundamentals and analytical methods. Wiley-VCH. Weinheim.

Lucocq J. 2008. Quantification of structures and gold labeling in transmission electron microscopy. Meth Cell Biol 88:59–82.

Luo J, Xia L, Lin J, Cen P. 1997. Kinetics of simultaneous saccharification and lactic acid fermentation process. Biotechnol Prog 13:762–767.

Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. 2002. Microbial cellulose utilization: Fundamentals and biotechnology. Microbiol Mol Biol Rev 66:506–577.

Mansfield SD, Mooney C, Saddler JN. 1999. Substrates and enzyme characteristics that limit cellulose hydrolysis. Biotechnol Prog 15:804–816.

Marshall K, Sixsmith D. 1974. Some physical characteristics of microcrystalline cellulose. 1. Powders for pharmaceutical use. Drug Dev Commun 1:57–71.

McMillan JD. 1994. Pretreatment of lignocellulosic biomass. ACS Symp Ser 566:292–324.

Mosier NS, Hall P, Ladisch CM, Ladisch MR. 1999. Reaction kinetics, molecular action, and mechanisms of cellulolytic proteins. Adv Biochem Eng Biotechnol 65:23–40.

Movagarnejad K, Sohrabi M, Kaghazchi T, Vahabzadeh F. 2000. A model for the rate of enzymatic hydrolysis of cellulose in heterogeneous solid-liquid systems. Biochem Eng J 4:197–206.

Oh K-K, Kim S-W, Jeong Y-S, Hong S-I. 2001. Bioconversion of cellulose into ethanol by nonisothermal simultaneous saccharification and fermentation. Appl Biochem Biotechnol 89:15–30.

Ohmine K, Ooshima H, Harano Y. 1983. Kinetic study of enzymatic hydrolysis of cellulose by cellulase from Trichoderma viride. Biotechnol Bioeng 25:2041–2053.

Okazaki M, Moo-Young M. 1978. Kinetics of enzymatic hydrolysis of cellulose: Analytical description of a mechanistic model. Biotechnol Bioeng 20:637–663.

Pereira AN, Mobedshahi M, Ladisch MR. 1988. Preparation of cellodextrins. Methods Enzymol 160:26–43.

Philippidis GP, Spindler DD, Wyman CE. 1992. Mathematical modeling of cellulose conversion to ethanol by the simultaneous saccharification and fermentation process. Appl Biochem Biotechnol 34/35:543–556.

Philippidis GP, Smith TK, Wyman CE. 1993. Study of the enzymatic hydrolysis of cellulose for production of fuel ethanol by the simultaneous saccharification and fermentation process. Biotechnol Bioeng 41:846–853.

Puls J, Wood TM. 1991. The degradation pattern of cellulose by extracellular cellulases of aerobic and anaerobic microorganisms. Biores Technol 36:15–19.

Puri VP. 1984. Effect of crystallinity and degree of polymerization of cellulose on enzymatic saccharification. Biotechnol Bioeng 26:1219–1222.

Sinitsyn AP, Gusakov AV, Vlasenko EY. 1991. Effect of structural and physico-chemical features of cellulosic substrates on the efficiency of enzymatic hydrolysis. Appl Biochem Biotechnol 30:43–59.

Stalbrand H, Mansfield SD, Saddler JN, Kilburn DG, Warren RAJ, Gilkes NR. 1998. Analysis of molecular size distributions of cellulose molecules during hydrolysis of cellulose by recombinant *Cellulomonas fimi* β-1,4-glucanases. Appl Environ Microbiol 64:2374–2379.

Suga K, van Dedem G, Moo-Young M. 1975. Degradation of polysaccharides by endo and exo enzymes: A theoretical analysis. Biotechnol Bioeng 17:433–439.

Wald S, Wilke CR, Blanch HW. 1984. Kinetics of the enzymatic hydrolysis of cellulose. Biotechnol Bioeng 26:221–230.

Wang F, Tan WB, Zhang Y, Fan X, Wang. M. 2006. Luminescent nanomaterials for biological labelling. Nanotechnology 17:R1–R13.

Weil J, Westgate PJ, Kohlmann K, Ladisch MR. 1994. Cellulose pretreatments of lignocellulosic substrates. Enzyme Microb Technol 16:1002–1004.

Weimer PJ, Lopez-Guisa JM, French AD. 1990. Effect of cellulose fine structure on kinetics of its digestion by mixed ruminal microorganisms in vitro. Appl Environ Microbiol 56:2419–2421.

Wood TM. 1975. Properties and Mode of action of cellulases. Biotechnol Bioeng Symp 5:111–137.

Zhang Y-HP, Lynd LR. 2003. Cellodextrin preparation by mixed-acid hydrolysis and chromatographic separation. Anal Biochem 322:225–232.

Zhang Y-HP, Lynd LR. 2004. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems. Biotechnol Bioeng 88:797–824.

Zhang Y-HP, Lynd LR. 2005. Determination of the number-average degree of polymerization of cellodextrins and cellulose with application to enzymatic hydrolysis. Biomacromolecules 6:1510–1515.

Zhang Y-HP, Lynd LR. 2006. A functionally based model for hydrolysis of cellulose by fungal cellulase. Biotechnol Bioeng 94:888–898.

Zhou W, Hao Z, Xu Y, Schüttler HB. 2009. Cellulose hydrolysis in evolving substrate morphologies. II. Numerical results and analysis. Biotechnol Bioeng xxx:xxx–xxx.