Cellulose Hydrolysis with Evolving Substrate Morphologies

by

Zhiqian Hao

(Under the direction of Heinz-Bernd Schüttler)

Abstract

A systematic modeling and simulation of cellulose hydrolysis with non-complexed cellulase is presented here. Based on previous work, full chain fragmentation model is further sophisticated with time-evolving substrate morphology, which is a direct result of continuous defragmentation and solubilization. This modification not only accounts for actual prolonged hydrolysis timeframe, but also provides an innovative approach explaining the drop in initial hydrolysis rate, which is widely observed in industrial manufacturing. In addition, we present a novel site-number formalism, which keeps track of time evolution of accessible $\beta(1,4)$ glucosidic bonds of different site types. Site-number formalism, unlike its predecessors, is presented in a considerably reduced system size, i.e. for merely a dozen ordinary differential equations (ODE) regardless of system size. The underlying local Poisson (LP) assumption is found to agree very well with exact full chain fragmentation model, under realistic experimental parameters. Furthermore, we discovered two distinctive time-frames for complete hydrolysis, associated with initial outer layer hydrolysis and more profound complete substrate hydrolysis. Future possible work and the model's potential applications are discussed.

Index words: cellulose hydrolysis; substrate morphology; hydrolysis modeling; site number formalism

Cellulose Hydrolysis with Evolving Substrate Morphologies

by

Zhiqian Hao

B.A., Shanghai Jiaotong University, China, 2005

A Dissertation Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2010

Cellulose Hydrolysis with Evolving Substrate Morphologies

by

Zhiqian Hao

Approved:

Major Professor: Heinz-Bernd Schüttler

Committee: Steven P. Lewis
Uwe Happek

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
July 2010

Dedication

This dissertation is dedicated to:

The memory of my dearest grand-aunt, Hao Shujuan,

for your unconditional 7,347 days of love.

My beloved parents,

for your constant caring and love,

for sacrifices you have made throughout the years,

that guided me every single step of my journey.

Acknowledgments

First and foremost, I would like to express my sincere appreciation to my advisor Dr. Heinz-Bernd Schüttler. He has been constantly helping me during the whole process of our research, from big pictures to small details. Without his incisive thoughts, wise directions, crucial comments and kind encouragements, the completion of this dissertation would not have been possible. His patience, optimism and confidence will keep guiding me in my future professional life. It is a truly privilege and pleasure to work with him.

I would also like to give my thanks to Dr. Steven P. Lewis and Dr. Uwe Happek for enhancing this dissertation, offering valuable suggestions, and serving on my committee. Special thanks to Dr. Loris Magnani and Linda Lee, who have always been kind, generous, and supportive.

I would like to extend my gratitude to Dr. Wen Zhou, for his studious work and dedication into my related research. His work and results have been truly an inspiration for me.

Many thanks to my parents HAO, Jianguang and CHEN, Jiemian for offering me the best they can; thanks to my grandparents for your pampering and care; special thanks to my American family, Ron, Dale, Blair and Ben Terry, for your generous love and support.

Moreover, I would like to thank my dearest friends Dr-to-be Yue Fu, Dr. Bing Xu, Dr-to-be Leilei Guo, Dr-to-be Xuehai Liang, Dr-to-be Yanping Huang and Dr. Jing Xu. With them, I shared my 5 years of life; with them, we laughed and teared; with them, we explored many facets of our world; with them, we grew into best friends.

Last, but not least, I am deeply grateful for being given the opportunity to pursue my Ph.D degree in Physics and Astronomy Department here at UGA. Also, computing resources provided by the UGA Research Computing Center are hereby acknowledged.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Cellulose is the structural compound of many green plants, and thus considered as the most abundant renewable biological resource (Zhang and Lynd, 2006). The abundance of cellulose and its potential in environmental-friendly energy production attract many scientists to study the biochemistry of cellulase enzyme systems, and many others formulate modeling frameworks to quantitatively incorporate our understandings/speculations for engineering or manufacture purposes.

Enzymatic hydrolysis breaks down cellulose molecules into short-chain oligomers, which are further metabolized into biofuels. This conversion is achieved by various cellulase systems, which all contain three major enzyme types: endoglucanases, exoglucanases and β-glucosidases. Endoglucanases cut at any internal $\beta(1,4)$ glucosidic bonds in insoluble glucose chains; exoglucanases, however, only act on two ends of insoluble chains; β-glucosidases have been found to hydrolyze preferably soluble oligomers into glucose (Henrissat et al., 1998; Wood, 1975), the final product of hydrolysis. Thus endo- and exo-glucanases depolymerize long chain cellulose molecules and feed their soluble product to β-glucosidases for final touch-up. The depolymerization is believed to be the rate-limiting step (Lynd et al., 2002).

There has been extensive work done on numerical simulations of hydrolysis. Zhang and Lynd categorized available models into four groups: nonmechanistic, semimechanistic, functionally based and structurally based. Nonmechanistic models mainly serve only as data correlations to a very limited set of experiment conditions. Semimechanistic models feature only one substrate state variable or single enzyme activity. Structurally based models pose

a challenge to develop meaningful kinetic models based on our current limited knowledge. Among all, functionally based models lend insight into an adsorption model with multiple substrate variables and multiple enzyme solubilizing activities (Zhang and Lynd, 2004).

However, current functionally based models have fundamental limitations inherited from the assumption of missing spatial interactions between different molecules on solid substrate. This is as if all cellulose chains are treated effectively as isolated, decoupled cellulose chains, readily to be hydrolyzed by enzymes. In real solid substrate, cellulose chains are assembled into random shapes that only expose a fraction of its content to enzyme adsorption. Steric obstructions among surrounding cellulose chains impose further limitations to cellulose availability. As hydrolysis progresses, enzymatic erosion exposes cellulose chains previously buried under the suface. Consequently, the *substrate morphology*, that is, the spatial organization of cellulose chains into a solid material, along with enzymatic fragmentation hydrolysis are mutually dependent and profoundly affect each other. This interaction has been largely neglected.

Due to a lack of knowledge concerning cellulose spatial arrangements, some semi-mechanistic models employed a pre-determined change in substrate morphology (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) The respective functions are highly empirical and non-universal.

In our proposed model, we incorporate both enzymatic fragmentation kinetics and its coupling to concurrent hydrolysis-driven evolution of substrate morphology. For illustration purposes, we define smallest accessible compartments (SACs) 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 shrinkage of SAC units are illustrated in surface ablation formalism which correlates time evolution of morphology with enzymatic hydrolysis of cellulose on the surface. This approach presents a better replication of real-world hydrolysis and thus allows us to simulate to near-completion cellulose conversion.

In this dissertation, the general site representation formalism of enzyme hydrolytic fragmentation coupled with morphology evolution is introduced. Its numerical simulation results are then compared to its corresponding exact chain number formalism (Zhang and Lynd, 2006). Furthermore, we will investigate into model predictions regarding hydrolytically evolving substrate morphologies, their effects on the hydrolysis kinetics and enzyme synergism. Lastly, quasi-steady state analysis is given to provide an alternative approach to large time scale hydrolysis simulation, which is also a determining factor that sets apart our morphological model from aforementioned non-morphological models. The distinctive short time scale behavior vs. long time behavior lends fresh insight into a frequently observed phenomenon in industry: initial hydrolysis rate drop (Lynd et al., 2002). Potential future work will be briefly discussed at the end.

Chapter 2

Fundamentals about Cellulose and Cellulases [1]

2.1 Cellulose Molecules and Hydrolytic Enzymes

Cellulose is a linear condensation of D-anhydroglucopyranose ($C_6H_{10}O_5$) monomers that are interconnected by β(1,4)-glucosidic bonds. Such bonds create alternating directionality between neighboring monomers, leaving even number of hydroxyl groups on each side of the chain. Each chain has two chemically distinctive chain ends, *non-reducing* end (L-end) and *reducing* end (R-end). Due to hydrogen bonds and van de Waal's forces, coupled cellulose molecules form a sheet of parallel-aligned crystalline structure, (ab plane) with multiple sheets stacked perpendicularly, creating a 3-D lattice structure. (Mosier, 1999; Zhang and Lynd, 2004) Therefore there are distinctive faces in cellulose molecules and are believed to be directly related to enzymatic hydrolysis rate. This is due to the fact that *endo*-acting hydrolytic enzymes preferentially attack from a direction perpendicular to ab plane.

Solid cellulose substrates are solubilized by hydrolytic enzymes cutting at various β(1,4)-glucosidic bonds exposed both internally and externally. For illustrative purposes, each D-anhydroglucopyranose monomer is denoted as G_1, and with the polymer of length ℓ conveniently referred as G_ℓ. The enzymatic hydrolysis then produces small fragments of G_ℓ, some dissolves into ambient solution and some remain on the substrate.

We adopt several simplifying assumptions for this hydrolysis kinetics (Okazaki and Moo-Young, 1978; Zhang and Lynd, 2006). For fragments produced by enzymatic hydrolysis with length ℓ smaller than a certain cut-off value, denoted as ℓ_s, those are treated as instantly

[1] Contain contents published in Zhou W, Schüttler HB, Hao Z, Xu Y. 2009. Cellulose hydrolysis in evolving substrate morphologies I: A general modeling formalism. *Biotechnol Bioeng* **104**:261-274. Permission acquired from *Biotechnol Bioeng* journal.

detached from the surface and solubilized. Otherwise, for fragments longer than ℓ_s, they remain on the substrate surface with its own new chain ends. Typical industrial ℓ_s value ranges $\sim 4-7$. Further depolymerization and reattachment of soluble fragments are neglected as insoluble hydrolytic kinetics is the dominating effect here.

In this study, we consider a system with three glucanase enzymes produced by the *Trichoderma* species: cellobiohydrolase I (CBH1, Ce17A), cellobiohydrolase II (CBH2, Ce16B) and endoglucanase I (EG1 or Ce17B). Cellobiohydrolase I/II are believed to "*exo*-cut" at 2nd $\beta(1,4)$-glucosidic bond from both chain ends, producing G_2 cellobiose that is immediately dissolved into the solution. Endoglucanase I, on the other hand, randomly selects a $\beta(1,4)$-glucosidic site within G_ℓ, resulting in G_{ℓ_1} and G_{ℓ_2}, with $\ell_1 + \ell_2 = \ell$. Apparently, if either ℓ_1 or ℓ_2 appears to be smaller than ℓ_s, this *endo*-activity also results in soluble fragments. For generality, in our modeling formalism, the foregoing three enzyme activities will be referred as *exo-L*, *exo-R*, and *endo*, respectively, without reference to the specific microbial source organism.

2.2 Substrate Morphology and Enzyme Accessibility

As-grown substrates are usually subject to pre-hydrolysis treatments, such as mechanical grinding or thermo-chemical procedures. With the ultimate purpose of increasing hydrolysis rate, mechanical and thermo-chemical treatments both serve to increase enzyme accessible surfaces. For mechanical grinding, physical particle sizes are being reduced to increase surface-to-volume ratio; whereas for thermo-chemical pretreatment, it mainly creates additional enzyme-accessible *internal* surfaces by weakening linkage, and by infusion of water, between neighboring cellulosic fibrils units (Himmel et al., 2007; Zhang and Lynd, 2004). "Swollen" internal voids are thus available to hydrolytic enzymes attacks, essentially increasing the number of accessible $\beta(1,4)$-glucosidic bonds. Additionally, the removal and/or spatial-redistribution of non-cellulosic components from thermo-chemical treatments are proved to be beneficial to increase hydrolysis rate as these components (e.g. lignin,

hemicellulose and pectin) can obstruct *endo/exo*-glucanase enzyme adsorption or cause inhibitory competition. (Converse, 1993)

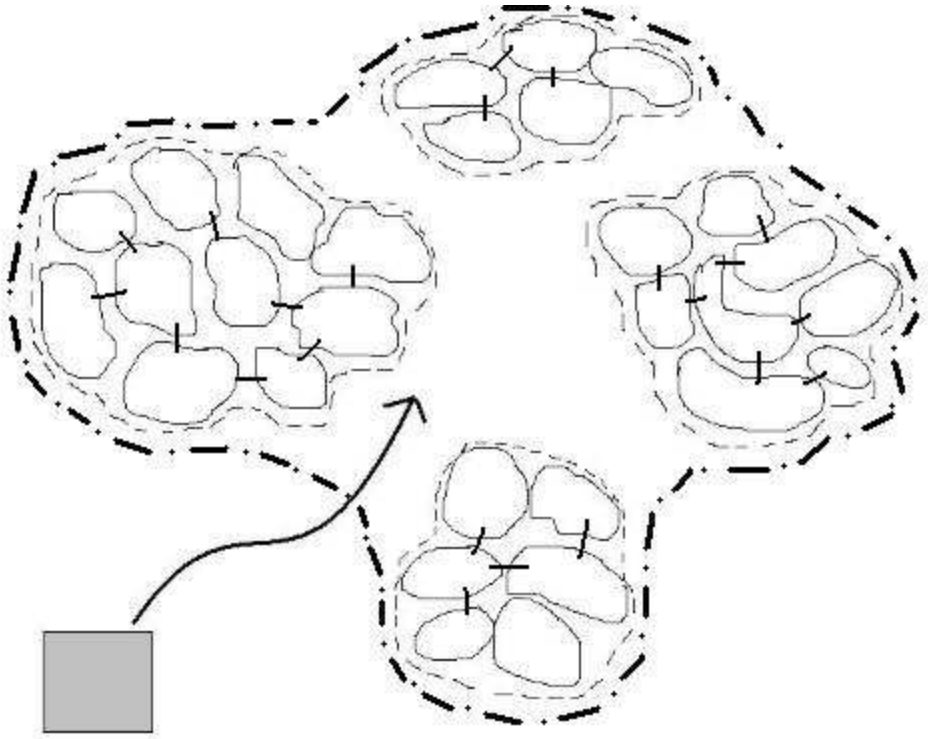

Figure 2.1: Schematic illustration of SAC partitioning by SAV. Schematic illustration of the subdivision of a single contiguous cellulosic substrate particle into SACs by SAVs. The particle's external surface is represented by the dot-dash line. The particle shown comprises four SACs. SAC surfaces, comprising 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.

Previous studies (Chang et al., 1981; Weimer et al., 1990) indicate that enzyme accessible surface area consists mostly of internal hydrated surface area. Therefore it is safe to infer a positive relationship between hydrolysis rate and the availability of internal surface area. Sub-micron imaging (Himmel et al., 2007) provides a direct visual confirmation of the existence

of a system of hydrated internal voids. Enzyme adsorption area only further justifies our assumption by exhibiting a much larger total adsorbing surface area than that of particle's external geometric shape. (Bothwell et al., 1997; Gilkes et alk., 1992; Marshall and Sixsmith, 1974; Zhang and Lynd, 2004) Therefore, most hydrolytic enzyme activity must happen on internal voids that are sufficiently hydrated and of sufficient size, in order to permit enzyme to invade and to access β(1,4)-glucosidic bonds internally.

There are several important parameters describing the properties of different substrates: (i) the Crystallinity Index (CrI) (ii) Degree of Polymerization (DP) and (iii) the fraction of enzyme-accessible β(1,4)-glucosidic bonds (F_{a}). Zhang and Lynd's study showed both the type of substrate and hydrolytic pretreatments have significant effects on the aforementioned structural features. (Zhang and Lynd, 2004). Moreover, it is experimentally impossible to isolate the effects of each parameter on hydrolysis rate.

2.2.1 Crystallinity

Crystallinity (CrI) provides valuable insight of substrate's structural periodicity, and is often correlated with substrate reactivity. Previous study revealed a negative correlation between hydrolysis rate and crystallinity (Lynd et al., 2002), in other words, amorphous substrate with low crystallinity undergoes hydrolytic reaction at a faster rate. Consequently, it is natural to draw the conclusion that crystallinity increases as hydrolysis undertakes (Ooshima et al., 1983). However, contradicting discoveries (Ohmine et al., 1983; Puls and Wood, 1991) render crystallinity as an ineffective indication of hydrolysis rate.

2.2.2 Degree of Polymerization (DP)

Degree of polymerization reflects the average length of cellulose molecules in substrate, counting by unit of anhydroglucose. In other words, degree of polymerization also denotes the

relative abundance of terminal β(1,4)-glucosidic bonds available for exo-glucanase adsorption. Therefore, it is reasonable to predict with lower DP value there should be a higher hydrolysis rate, which has been reported in previous studies. (Wood, 1975)

2.2.3 Accessibility Fraction F_{a} and Accessible Compartments

F_{a}, on the other hand, describes the fraction of β(1,4)-glucosidic bonds exposed to enzymes across the whole substrate. It is thus a geometrical quantity that depends on both the cellulosic substrate morphology *and* enzyme shape. Each cellulosic substrate consists of not only cellulose molecules but inevitably also hydrated voids that likely to span several orders of magnitude (Grethlein, 1985; Marshall and Sixsmith, 1974). Smallest accessible void (SAV), defined based on enzyme size, along with external surfaces comprise of enzyme accessible surfaces. In addition, these SAVs practically delimit the entire substrate into several subdivision, known as aforementioned smallest accessible compartments (SAC), as illustrated schematically in Figure(2.1). That is, by definition, an SAC is a minimal volume of substrate material that is bounded by, but not further divisible by, enzyme accessible surfaces. For the purpose of enzyme hydrolysis, only voids larger than SAV become relevant in our modeling since smaller voids are inaccessible to enzymes. Therefore, it is equivalent of saying all enzyme-accessible glucosidic bonds are exposed on the surface of SAC and SAC surfaces comprise only enzyme-accessible bonds. Overall, F_{a} can be viewed as a dimensionless SAC surface-to-volume ratio, with "*surface area*" defined as accessible glucosic bonds and "*volume*" as total amount of glucosidic bonds contained within each SAC. Similar to DP, consistent evidence suggests an increasing hydrolysis rate is correlated with higher F_{a} value. This is pretty straightforward as the more glucosidic bonds are exposed on SAC surface, the readier and faster substrate will undergo hydrolytic conversion.

In principle, each type of enzyme defines its own SAV and in turn SAC subdivision, since each enzyme can adopt different sizes. Enzyme-type-dependent SAC subdivision can

be readily incorporated into our model. However, due to a current lack of a detailed understanding of substrate morphologies, and because of the similar sizes of the *exo-L*, *exo-R*, and *endo* enzymes of the *Trichoderma* species (Grethlein, 1985), we will assume a single common SAV and in turn SAC subdivision for all three enzymes.

2.3 Hydrolytic Evolution of Substrate Morphology

The term "substrate morphology" specifically refers to the random geometry of the substrate on mesoscopic (>10-100nm) length scales. That is, morphology refers to how cellulose chain molecules are connected into larger units on different length scales larger than the molecular scale (*e.g.* molecules form fibers, fibers form larger structures, such as the SACs defined above and SACs form bigger particles, etc.); and how these larger units are sized, shaped, evolved as quantified *e.g.* by random size and/or random shape distributions.

2.3.1 Elementary Layer

During the process of hydrolysis, the substrate morphology undergoes substantial changes, and such change in turn affects enzymatic hydrolysis rate. Thus, substrate morphology and enzymatic hydrolysis are inter-dependent. In mesoscopic level, enzymes attack β(1,4)-glucosidic bonds on SAC surfaces and gradually solubilizing cellulose molecules, resulting in exposure of new molecules and intact bonds underneath; in macroscopic level, hydrolysis leads to the shrinkage of SAC units that as a consequence increases its surface-to-volume ratio.

In order to quantitatively capture the essence of time evolution of substrate morphology as well as its effects on hydrolysis in return, we introduce the variable that indicates the numbering of layered structure of substrate, denoted as λ. (See Figure. 2.2) With the outermost layer carries the largest λ value, a decreasing λ kinetic corresponds with the hydrolytic inward evolution of SAC morphology. The definition of an elementary layer is therefore that the fraction of material which will be solubilized and removed from SAC surface if all outer

glucosidic bonds, and only those, are hydrolyzed and removed from the substrate, generating, by definition, a new surface layer. Notice, this definition does not require any actual physical structural layer of substrate, and only serves as an "accounting" device to keep track of how many glucosidic bonds are exposed on the surface from the variable λ.

Moreover, λ should not be limited to integers. Obviously, during hydrolysis, partial solubilization of the outermost layer dominate most of the time, until a new layer is completely exposed to enzyme attack. To model such state, λ should be treated as a continuous dynamical (time-dependent) real variable.

2.3.2 Power Law Dependence

An essential variable in our model is the average number of G_1 monomers contained within each SAC unit, enclosed within all layers enclosed by, and including outermost layer λ, to be denoted by $n_{\mathrm{V}}(\lambda)$. Due to a lack of knowledge in mesoscopic morphological information during hydrolysis, we assume substrate shape similarity before and during hydrolysis shrinkage, thus $n_{\mathrm{V}}(\lambda)$ obeys power law dependence parameterized by the volume prefactor c_{V} and the ablation dimension variable d_{A}. This is illustrated by the prototypical layer geometries shown in Fig. 2.3. Thus $n_{\mathrm{V}}(\lambda)$ has the form:

$$n_{\mathrm{V}}(\lambda) = c_{\mathrm{V}} \lambda^{d_{\mathrm{A}}} \tag{2.1}$$

The simple formulae grants us the ability to keep track of the amount of G_1 left within the substrate with only a single geometry variable λ. All other information, such as overall shape of SAC, it's linear size(s) normal to ablation direction(s), and its average G_1 monomer density, are all compacted into the prefactor c_{V}.

As simple as power law gets, the model can be readily equipped with other n_{V} modeling once more detailed morphology information become available.

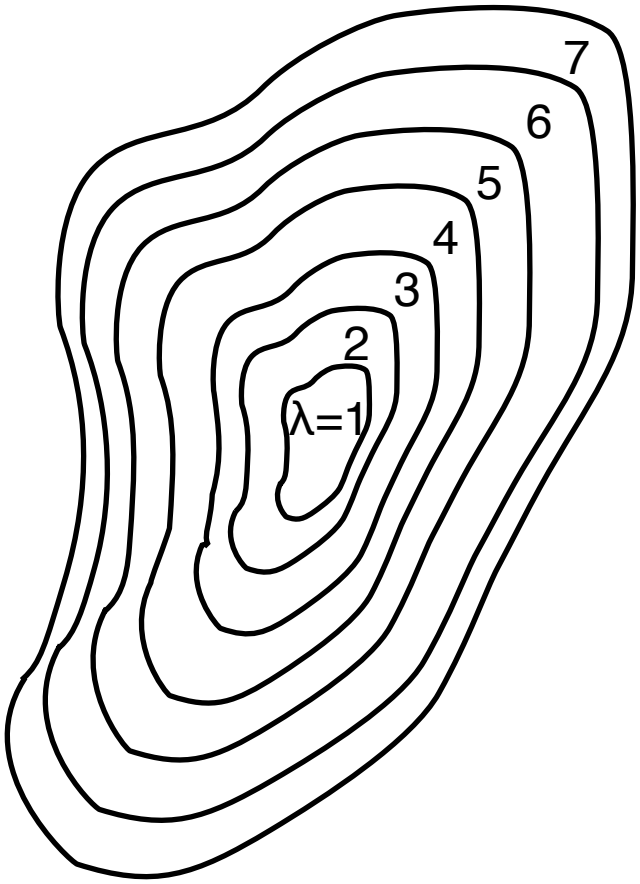

Figure 2.2: SAC elementary layer partitioning. Schematic illustration of the partitioning of a SAC into elementary layers. The layers must be labelled by the layer number λ such that the layer with the highest λ-value is the first one to be removed due to solubilization by the attacking enzymes during hydrolysis.

The number of G_1 monomers contained in each layer λ, denoted as n_{M}, can then be readily expressed in terms of the difference between $n_{\mathrm{V}}(\lambda)$ and $n_{\mathrm{V}}(\lambda-1)$:

$$n_{\mathrm{M}}(\lambda) = n_{\mathrm{V}}(\lambda) - \Theta(\lambda-1)n_{\mathrm{V}}(\lambda-1), \qquad \Theta(\Delta\lambda) \equiv \begin{cases} 1 \text{ if } \Delta\lambda > 0 \\ 0 \text{ if } \Delta\lambda \leq 0 \end{cases} \tag{2.2}$$

2.3.3 Geometry Classes σ

Admittedly, any real substrate will adopt various types of SAC units, whose dimensions will likely span several orders of magnitudes. Impossible to model every single one of them, it is wise to group them into a finite population of several SAC "geometry classes," labeled by subindex σ ($\sigma = 1, ...M_{\mathrm{MD}}$ and M_{MD} is the population size). Each geometry class is modeled by its own λ-variable and volume function $n_{\mathrm{V}}(\lambda)$. Both are then labeled by their class index σ *i.e.*, as λ_σ and $n_{\mathrm{V},\sigma}(\lambda_\sigma)$. In the simple power-law model Eq. (2.1), the classes are parameterized by σ-dependent volume prefactors $c_{\mathrm{V},\sigma}$. The ablation dimension could also be made class-dependent as $d_{\mathrm{A},\sigma}$. In our model calculations, we will not consider this case and use the same d_{A} for all classes.

Adopting power law dependence of Eq. (2.1), we introduce a population $(\lambda_\sigma, d_{\mathrm{A},\sigma}, c_{\mathrm{V},\sigma})$ of geometry parameter variables. Within each class-σ, where all SAC units share similar geometry parameter variables, let C_σ denote the concentration of class-σ SAC units, in units of moles of SACs per reactor volume. Then we can rewrite the concentration of total G_1-monomers, $x_{\mathrm{V},\sigma}$, and the concentration of exposed G_1-monomers on SAC surface, $x_{\mathrm{M},\sigma}$, for each SAC class-σ:

$$x_{\mathrm{M},\sigma} = C_\sigma n_{\mathrm{M},\sigma}(\lambda_\sigma); \;\; x_{\mathrm{V},\sigma} = C_\sigma n_{\mathrm{V},\sigma}(\lambda_\sigma) \tag{2.3}$$

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

$$F_{\mathrm{a},\sigma}(\lambda_\sigma) = \frac{x_{\mathrm{M},\sigma}}{x_{\mathrm{V},\sigma}} = \frac{n_{\mathrm{M},\sigma}(\lambda_\sigma)}{n_{\mathrm{V},\sigma}(\lambda_\sigma)}; \;\; \bar{F}_{\mathrm{a}} = \frac{x_{\mathrm{M}}}{x_{\mathrm{V}}} = \sum_\sigma \xi_\sigma F_{\mathrm{a},\sigma}(\lambda_\sigma) \tag{2.4}$$

where $x_{\mathrm{M}} = \sum_\sigma x_{\mathrm{M},\sigma}$, $x_{\mathrm{V}} = \sum_\sigma x_{\mathrm{V},\sigma}$ and $\xi_\sigma = x_{\mathrm{V},\sigma}/x_{\mathrm{V}}$.

2.3.4 Observable Simulation Parameters

$\bar{F}_{\mathrm{a}}$ serves as a bridge between modeling and real world experiments as it provides a testable predictions for the evolution of F_{a} parameter during hydrolysis. On the other hand, experi-

mentally determined F_{a} values as a fraction of accessible $\beta(1,4)$ glucosidic bonds can be fed back into our simulation as a constraint variable. For the purpose of estimating SAC size by our definition, we can substitute $n_{\mathrm{V}}(\lambda)$ in Eq. ((2.4) with Eq. (2.1), realizing observed F_{a} values are typically $\sim 0.1 - 0.001$ (Zhang and Lynd, 2004), we can approximate SAC sizes $\lambda \sim d_{\mathrm{A}}/F_{\mathrm{a}} \sim 20 - 2000$, assuming $d_{\mathrm{A}} = 2$. According to our previous assumption with each elementary layer being the fraction of material covered only by the outermost glucosidic bonds, and exposed at the SAC surface as a result of removal of only the outermost bonds, it is reasonable to assume the typical thickness of an elementary layer to be the same order of the chain order, that is, of order of the $\sim$ 1nm glucose molecular size. Therefore, a typical SAC sizes λ are 20-2000nm, which is expected to exceed both typical enzyme molecular sizes (Zhang and Lynd, 2004) and typical cellulosic fibril sizes (Himmel et al., 2007) by at least an order of magnitude in all but the most highly accessible substrates.

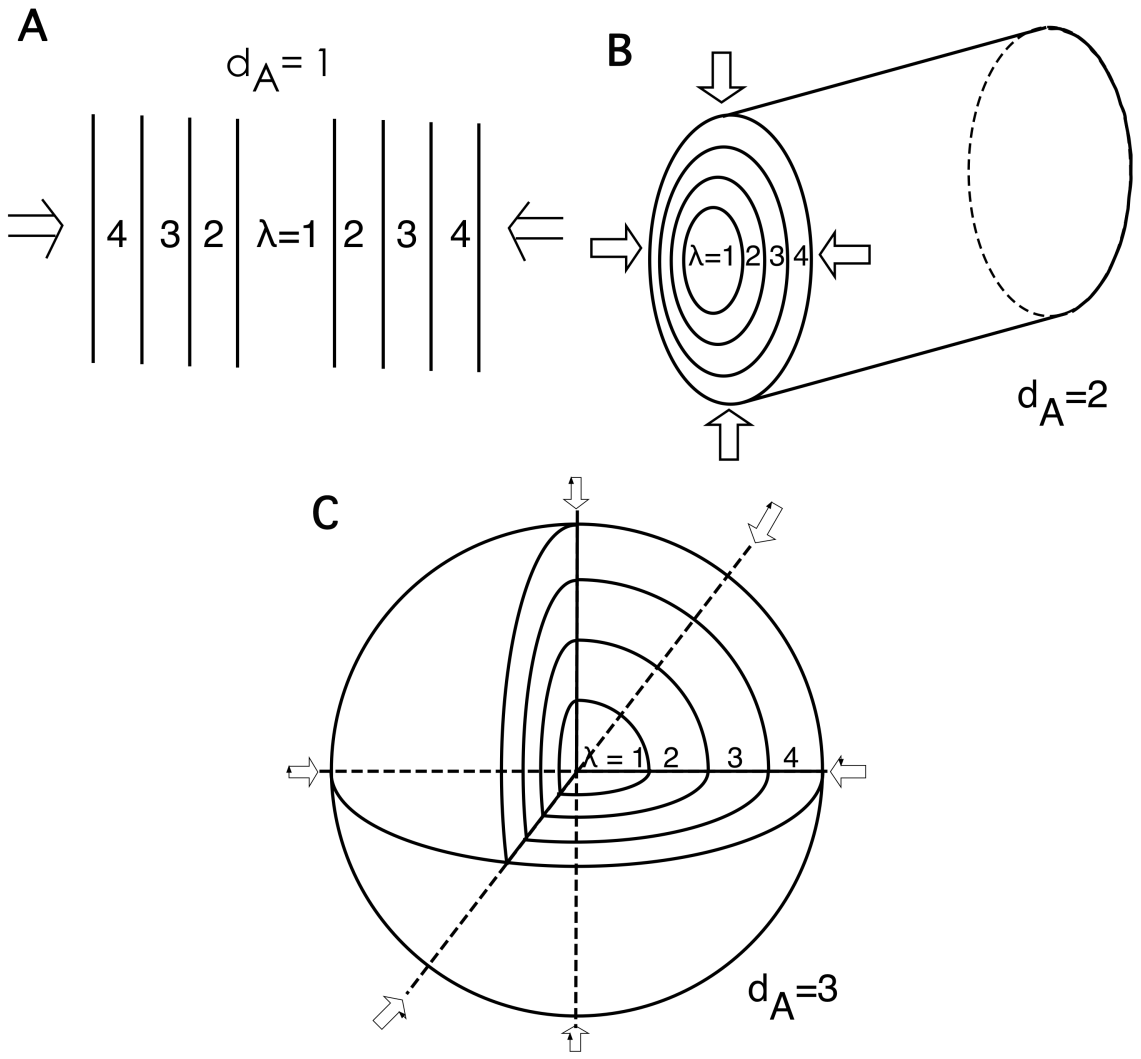

Figure 2.3: Prototypical SAC layer geometries. Prototypical SAC layer geometries illustrating the concept of an "ablation dimension" $d_{\rm A}$. Prototypes with ablation dimensions $d_{\rm A} = 1, 2$ and 3 are shown in panel (A), (B) and (C), respectively. Arrows indicate possible directions of enzyme attack during hydrolysis. (A) Dimension $d_{\rm A} = 1$ 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, *e.g.*, in a highly crystalline substrate. Directional order is not required for $d_{\rm A} = 1$. (B) Dimension $d_{\rm A} = 2$ 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 $d_{\rm A} = 3$ 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.

Chapter 3

General Site Ablation Model Development [1]

3.1 Surface Site Ablation Rate Equations

3.1.1 Site Type Classification

Our predecessors established model in which they kept track of concentrations of chains with different monomer lengths (Zhang and Lynd 2004). However, their model was describing only the process of enzymatic fragmentation of individual chains treated *in isolation*. Their model, in other words, did not account for the effects of substrate solidity and the resulting obstruction of enzyme access to all chains hidden below the substrate surface. It therefore did also not capture the substrate morphology evolution and the surface chain exposure kinetics. The latter must necessarily occur simultaneously with the enzymatic surface chain fragmentation kinetics during a near-complete solubilization of the substrate.

Our hydrolytic surface ablation kinetics formaslim circumvented the problem by keeping track of the availability of different $\beta(1,4)$ glucosidic bonds exposed on substrate surfaces, categorized according to their respective positioning within each chain. Thus with a number of pre-defined site types, we have successfully reduced the amount of variables to a fixed amount regardless of system size. Each cellulose chain can be treated as a "train" of different types of "compartments", and the substrate is therefore an even bigger composite of these basic "compartment" types. In our formalism, for convenience, we construct 7 basic site types, namely N-, O-, X-, Y-, Z-, L- and R-sites and labelled by a site type subindex ν, as

[1] Contain contents published in Zhou W, Schüttler HB, Hao Z, Xu Y. 2009. Cellulose hydrolysis in evolving substrate morphologies I: A general modeling formalism. *Biotechnol Bioeng* **104**:261-274. Permission acquired from *Biotechnol Bioeng* journal.

illustrated in Figs. 3.1 and 3.2. Following the same notation as above, we then define $n_{\nu,\sigma}(t)$ as the time-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-, Z-, represent intact $\beta(1,4)$ glucosidic bonds. Type X- (Y-) is the site where either *exo*-L- (*exo*-R-) or *endo*-acting enzyme can adsorb and hydrolyze, and is thus located k_{X} (k_{Y}) G_1-monomers from the L-end R-end) of the chain, where $k_{\mathrm{X}} = 2$ ($k_{\mathrm{Y}} = 2$) for cellohydrobiolase. Type N-site, on the other hand, only accepts adsorption from and therefore only be cut by *endo*-acting enzymes. Type Z-site can be treated as one where X-, Y- and N- coincide on the same site, and is therefore only found exactly in the midpoint of a chain with length $\ell = k_{\mathrm{X}} + k_{\mathrm{Y}} - 1$. Lastly, a type O-site is one that cannot be absorbed with any type of enzyme due to steric obstruction or cannot be hydrolyzed because of inactive adsorption. Type L- and R-sites are the already broken bonds residing on non-reducing and reducing chain ends respectively, and neither can further absorb any enzyme, let alone undergo hydrolysis.

Straightforwardly, the sum of $n_{\nu,\sigma}(t)$ for all types should, at any time t, obey conversation law: the weighted summation of $n_{\nu,\sigma}(t)$ equals the average total monomer concentration on each SAC, of geometry class-σ, $n_{\mathrm{M},\sigma}$, *i.e.*,

$$\sum_{\nu} w_\nu n_{\nu,\sigma}(t) = n_{\mathrm{M},\sigma}(\lambda_\sigma(t)) \tag{3.1}$$

where the weight factor w_ν represents the fraction of $\beta(1,4)$ glucosidic bonds on average associated with each site type, and is therefore 1 for an intact bond (N-, O-, X-, Y-, Z-) and 0.5 otherwise (L-,R-).

3.1.2 Uniform Segment Exposure

As hydrolysis progresses on real cellulosic substrate, enzyme activities solubilize the target segment from the surface, leaving the underlying glucan chains only partially exposed with the remaining part still subducted. Under the absence of experimental data elaborating the degree of partial exposure, it is reasonable to reckon that any particular segment with length

k on G_ℓ is as likely to be completely exposed/subducted, or partially exposed as any other segment with the same length k on any other glucan chain G_ℓ, regardless of its relative position with respect to each chain ends.

Therefore, each fully-exposed, complete chain under "Uniform Segment Exposure" assumption can be perceived as a reassembly of different pieces: For site counting purposes, upon averaging over all SAC surfaces of the same class-σ, we can construct a fictitious yet sterically unobstructed chain of length ℓ, G_ℓ, based on pieces taken from several real, partially exposed chains. This mathematically constructed chain, G_ℓ acts equivalently as a real, fully-exposed chain with length ℓ on SAC surface. 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*. This should be kept in mind for all following discussions as any "chain whole exposed at surface" is referred to such assembled "whole chain *on average*".

Applying this assumption here, chain ends should obey L-R-symmetry. The number of left chain ends must, on average, equal the number of right chain ends, *i.e.* $n_{L,\sigma} = n_{R,\sigma}$. Notice, this symmetry is based on previous "Uniform Segment Exposure" assumption. Substrates without L-R-symmetry are in principle possible and can be readily treated in our formalism by addition of site types marking L- and R-directed chain subduction loci.

3.1.3 Site Fragmentation Coupled Rate Functions

We are now in a position to derive a set of coupled rate equations governing hydrolytic ablations on SAC surfaces. Let us first define $V_{\nu,\sigma}$ to be the net rate of production of type-ν sites at the class-σ SAC surfaces, due to all types of enzymatic chain fragmentation processes, which can be further categorized into two cases:

1. A site of type ν can be gained or lost as a result of hydrolytic enzyme cuts that produce at least one insoluble substring. For ones residing on the insoluble substring, site of type ν can be created (or annihilated) due to conversion to (or from) another site of

type ν'. For example, additional site type X-, Y-, L-, R- and (possibly) Z-sites are gained, and N-, as well as possibly O-site(s), are lost due to an *endo*-activity at a site at least ℓ_{S} monomers from either one of both pre-existing chain ends.

2. A site of type ν can be lost along with a soluble chain fragment due to enzymatic cut at site less than ℓ_{S} monomers away from either chain ends. Such cut can be generated by either *endo*-activity or *exo*-activity, and all site types can be lost in this manner.

Straightforwardly, $-V_{\nu,\sigma}$ would then be the rate of loss of type-ν sites per class-σ SAC during the aforementioned processes. Each multiplied with its associated monomer fraction w_ν, the sum of $-V_{\nu,\sigma}$ over all site type-ν then indicates the total rate of monomer loss from SAC surface, denoted as $-\bar{V}_\sigma$.

$$-\bar{V}_\sigma = -\sum_\nu w_\nu V_{\nu,\sigma} \tag{3.2}$$

On the other hand, the total rate of monomer loss from SAC class-σ can also be expressed as $\frac{d}{dt} n_{\mathrm{V},\sigma}(\lambda_\sigma(t))$. Thus, setting this expression equal to $\bar{V}_\sigma$, and from chain rule, we arrive at:

$$\dot{\lambda}_\sigma = \frac{\bar{V}_\sigma}{\partial_\lambda n_{\mathrm{V},\sigma}(\lambda_\sigma)} \tag{3.3}$$

where $\partial_\lambda ...$ is shorthand for the λ-derivative $\frac{\partial}{\partial\lambda}$.

Now let's derive the native surface exposure term. Consider the removal of a small fraction of a layer, $d\lambda_\sigma < 0$, during a short time interval dt, resulting in a total monomer loss of $-dn^{(\mathrm{fra})}_{\mathrm{V},\sigma} = -\partial_\lambda n_{\mathrm{V},\sigma} d\lambda_\sigma > 0$ on SAC surface; and meanwhile, such ablation will cause a mean shrinkage of SAC surface by $-dn^{(\mathrm{fra})}_{\mathrm{M},\sigma} = -\partial_\lambda n_{\mathrm{M},\sigma} d\lambda_\sigma$ monomers. Taking consideration of both exposing underlying monomers from outermost layer ablation *and* geometrical shrinkage, the net number of newly exposed monomers is thus

$$dn^{(\mathrm{exp})}_{\mathrm{M},\sigma} = -dn^{(\mathrm{fra})}_{\mathrm{V},\sigma} + dn^{(\mathrm{fra})}_{\mathrm{M},\sigma} = -(\partial_\lambda n_{\mathrm{V},\sigma} - \partial_\lambda n_{\mathrm{M},\sigma}) d\lambda_\sigma \tag{3.4}$$

Of these newly exposed monomers, only a fraction $g_{\nu,\sigma}$ consists of sites of type ν, contributing a positive increment to the type-ν surface site count due to surface exposure *alone*. The newly

exposed sites from surface exposure contribution is thus

$$dn_{\nu,\sigma}^{(\mathrm{exp})} = dn_{\mathrm{M},\sigma}^{(\mathrm{exp})} g_{\nu,\sigma} = -\bar{V}_\sigma (1 - \partial_\lambda n_{\mathrm{M},\sigma}/\partial_\lambda n_{\mathrm{V},\sigma}) g_{\nu,\sigma} dt \tag{3.5}$$

where we substituted $d\lambda_\sigma$ with $dt\bar{V}_\sigma/\partial_\lambda n_{\mathrm{V},\sigma}$ from Eq.(3.3)

Therefore, the net increment of type-ν surface sites, combining effects from both surface exposure term and fragmentation term, becomes $dn_{\nu,\sigma} = dn_{\nu,\sigma}^{(\mathrm{fra})} + dn_{\nu,\sigma}^{(\mathrm{exp})}$. Noticing $dn_{\nu,\sigma}^{(\mathrm{fra})} = V_{\nu,\sigma} dt$, Eq.(3.5) and $\dot{n}_{\nu,\sigma} \equiv dn_{\nu,\sigma}/dt$, we finally arrive at

$$\dot{n}_{\nu,\sigma} = V_{\nu,\sigma} - \bar{V}_\sigma \cdot \eta_\sigma(\lambda_\sigma) \cdot g_{\nu\sigma}(\lambda_\sigma) \tag{3.6}$$

$$\eta_\sigma(\lambda_\sigma) = 1 - \frac{\partial_\lambda n_{\mathrm{M},\sigma}(\lambda_\sigma)}{\partial_\lambda n_{\mathrm{V},\sigma}(\lambda_\sigma)} \tag{3.7}$$

So far, we have derived a system of coupled rate equations (Eq. (3.3) and Eq. (3.6)), describing the state of SAC surfaces during hydrolytic ablation, characterized only by two dynamical variables $\lambda_\sigma(t)$ and $n_{\nu,\sigma}(t)$.

The second term in Eq. (3.6), $\bar{V}_\sigma \eta_\sigma(\lambda_\sigma) g_{\nu\sigma}(\lambda_\sigma)$, denoted as surface exposure term, governs the amount of site of type ν being exposed at the surface due to removal of covering outermost layer(s). $\bar{V}_\sigma$ denotes net rate of monomer loss from the SAC due to ablation. Geometric parameter $\eta_\sigma(\lambda_\sigma)$ accounts for the shape curvature effect. As hydrolysis progresses inward on SAC unit, not only volume but also surface area decreases when $d_{\mathrm{A},\sigma} > 1$, and as a result, with each unit surface area being removed from the outermost layer, *less* than unit surface area will be exposed underneath. $g_{\nu\sigma}(\lambda_\sigma)$, the native site fraction function, describes the fraction of type-ν sites enclosed in SAC unit surface prior to hydrolysis as long as $\lambda < \lambda_\sigma(t)$. Apparently, the weighted sum of site fraction across all type ν must be 1: $\sum_\nu w_\nu g_{\nu,\sigma}(\lambda_\sigma) = 1$.

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

$$\dot{n}_{\mathrm{S},\sigma} = \sum_{k=1}^{\ell_{\mathrm{S}}-1} k V_{\mathrm{S},\sigma}(k) \tag{3.8}$$

where $V_{\mathrm{S},\sigma}(k)$ is the production rate of soluble oligomers G_k, per class-σ SAC. The construction of $V_{\mathrm{S},\sigma}$ 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.

3.2 Ablation and Oligomer Rate Functions

For each hydrolytic cut, *endo*- or *exo*-acting enzymes must first bond with a target site and successfully form an Enzyme-Substrate (ES) complex. Previous studies (Lynd *et al.*, 2002; Zhang and Lynd, 2004) assumes that the complex is formed at a much faster rate than the actual bond cutting kinetics, and therefore an enzyme adsorption quasi-steady state is maintained at the SAC surfaces during hydrolysis, governed by the law of mass action:

$$z_{\kappa,\mu,\sigma} = L_{\kappa,\mu} v_\kappa y_{\mu,\sigma} \tag{3.9}$$

where $z_{\kappa,\mu,\sigma}$ is the molar ES complex concentrations, $y_{\mu,\sigma}$ the molar concentration of free type-μ sites exposed on class-σ SAC surfaces, and v_κ the free type-κ enzyme concentrations, $L_{\kappa,\mu}$ the adsorption coefficient which is the inverse of the conventional *de*sorption equilibrium coefficient (Zhang and Lynd, 2004). The number of ES complexes per SAC, $m_{\kappa,\mu,\sigma}$, is related to $z_{\kappa,\mu,\sigma}$ by a factor of C_σ, the SAC molar concentrations:

$$z_{\kappa,\mu,\sigma} \equiv C_\sigma m_{\kappa,\mu,\sigma} \tag{3.10}$$

The *free* enzyme and surface site concentrations, v_κ and $y_{\mu,\sigma}$ are related to the corresponding *total* concentrations, u_κ and $x_{\mu,\sigma} \equiv C_\sigma n_{\mu,\sigma}$, respectively, by way of the total enzyme and total site balance relations:

$$u_\kappa = v_\kappa + \sum_{\mu,\sigma} z_{\kappa,\mu,\sigma} \tag{3.11}$$

$$x_{\mu,\sigma} = y_{\mu,\sigma} + \sum_{\kappa} z_{\kappa,\mu,\sigma} + \sum_{\kappa,\mu'} f_{\mu,\sigma}\beta_\kappa z_{\kappa,\mu',\sigma} \tag{3.12}$$

where $f_{\mu,\sigma} \equiv n_{\mu,\sigma}/n_{\mathrm{M},\sigma} = x_{\mu,\sigma}/x_{\mathrm{M},\sigma}$ and $x_{\mathrm{M},\sigma} \equiv C_\sigma n_{\mathrm{M},\sigma}$. The last term in Eq. (3.12) arises due to the face that the dimension of enzyme is greater than that of a single $\beta(1,4)$ glucosidic bond site (Zhang and Lynd, 2006). Hence, a type-κ enzyme molecule, bound to a type-μ'

surface site, will prevent further adsorption from other enzymes to sites that are located in spatial proximity to the target type-μ' binding site and effectively creating obstruction to some number, β_κ, of other "collateral" surface sites. Eqs. (3.9), (3.11) and (3.12) can then be solved simply by iteration, for v_κ, $y_{\mu,\sigma}$ and $z_{\kappa,\mu,\sigma}$.

However, in the low-free-enzyme limit, defined by the condition

$$\sum_\kappa \beta_\kappa L_{\kappa,\mu} v_\kappa \ll 1 \quad \text{for all} \ \ \mu \ , \tag{3.13}$$

we can approximate $y_{\mu,\sigma} \cong x_{\mu,\sigma}$ and evaluate v_κ without iteration directly from the following equation, which can be easily derived from mass action and balance relations presented above.

$$v_\kappa = \frac{u_\kappa}{1 + \sum_{\mu,\sigma} L_{\kappa,\mu} y_{\mu,\sigma}} \tag{3.14}$$

Following earlier nomenclature, set $V_{\nu,\sigma}$ as 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. The complex formation and hydrolytic cut process can be mathematically formulated as

$$V_{\nu,\sigma} = \sum_{\kappa,\mu} \gamma_{\kappa,\mu} m_{\kappa,\mu,\sigma} \Delta \bar{N}_{\nu,\mu,\sigma} \tag{3.15}$$

$$\Delta \bar{N}_{\nu,\mu,\sigma} = \sum_{k=1}^{\infty} \sum_{k'=1}^{\infty} P_\sigma(k, k' | \mu, +1) \Delta N_{\nu,\sigma}(k, k') \tag{3.16}$$

In Eq. (3.15), $\gamma_{\kappa,\mu}$ is the cutting rate coefficient for an ES complex formed by enzyme type κ and substrate site type μ, in units of cuts per second per ES complex. In our formalism, $\kappa = 1, 2$ or 3, representing the *endo*-, *exo-L*- and *exo-R*-acting glucanase, respectively. The product, $\gamma_{\kappa,\mu} m_{\kappa,\mu,\sigma}$, then illustrates the cutting rate for intact bonds of site type ν, having absorbed an enzyme of type κ on the surface of an SAC of class-σ. $\Delta \bar{N}_{\nu,\mu,\sigma}$ contains the information of an average change in site type-ν concentration at surfaces of SAC class-σ, as a result of cutting a bond at site type μ. As mentioned before, such cut may affect concentrations of several, if not all, site types.

In Eq. (3.16), $P_\sigma(k,k'|\mu,+1)$ denotes the probability of finding an intact bond of site type μ on the surface of SAC class-σ, to be located k monomers away from the L-nonreducing end and k' monomers away from the other R-reducing end. Equivalently, it is the same probability for a cut of a randomly selected type-μ bond to generate two segments length k and k', from the L-end and R-end, respectively, for the surface-exposed chain G_ℓ of length $\ell = k + k'$. $\Delta N_{\nu,\sigma}(k,k')$ denotes the increment of type-ν sites that is produced by a bond cut generating a $(G_k, G_{k'})$ chain fragment pair. We will elaborate further its dependence and propose two corresponding chain site distribution models.

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

$$V_{\mathrm{S},\sigma}(k) = \sum_{\kappa,\mu} \gamma_{\kappa,\mu} m_{\kappa,\mu,\sigma} \sum_{k'=\ell_\mathrm{S}-k}^{\infty} [P_\sigma(k,k'|\mu,+1) + P_\sigma(k',k|\mu,+1)] \tag{3.17}$$

The resulting oligomer production rates $\dot{n}_{\mathrm{S},\sigma}$ from Eq. (3.8) can then be shown to obey general monomer conservation laws.

3.3 Chain Site Distribution Models

The change in type-ν site number, $\Delta N_{\nu,\sigma}(k,k')$, originated from enzymatic cut at site that generates two fragments G_k and $G_{k'}$ can be expressed in terms of $N_{\nu,\sigma}(k)$, which denotes for average number of type-ν sites per insoluble cellulose chain G_k. Naturally, for any soluble chain, $k < \ell_\mathrm{S}$, $N_{\nu,\sigma}(k)$ should by definition set to 0.

$$\Delta N_{\nu,\sigma}(k,k') = N_{\nu,\sigma}(k) + N_{\nu,\sigma}(k') - N_{\nu,\sigma}(k+k') \ . \tag{3.18}$$

This is easy to understand, as the increment in site type-ν is simply the difference between the number of site type-ν on two segments (G_k, $G_{k'}$) after the cut and the number of same site type-ν on the original chain $G_{k+k'}$.

The specific information contained $N_{\nu,\sigma}(k)$ must be provided as model input, and depends on the site distribution along the chain. Here we propose two simplified models and future calculation results will be presented based on these two models. Again, more realistic or

complicated models can be readily built into our model upon the availability of experimental knowledge over cellulosic chain structures.

Fig. 3.1 illustrates a simple site distribution called "Homogeneously Dirty Chain" (HDC model). The model is constructed on the assumption that O-sites are randomly distributed over $(\ell - 1)$ intact bonds on chain G_ℓ at a fixed probability $\phi_{O,\sigma}$. The site distribution for the rest of site types can be done through straightforward site counting, and $N_{\nu,\sigma}(k)$ is thus calculated as:

$$N_{\nu,\sigma}^{(\mathrm{HDC})}(\ell) = \begin{cases} (1-\phi_{O,\sigma})(\ell - 3 + \delta_{\ell,k_\mathrm{X}+k_\mathrm{Y}}) \ , & \nu = N; \\ (1-\phi_{O,\sigma})(1 - \delta_{\ell,k_\mathrm{X}+k_\mathrm{Y}}) \ , & \nu = X \text{ or } Y; \\ (1-\phi_{O,\sigma})\delta_{\ell,k_\mathrm{X}+k_\mathrm{Y}} \ , & \nu = Z; \\ 1 \ , & \nu = L \text{ or } R; \\ \phi_{O,\sigma}(\ell - 1) \ , & \nu = O \ . \end{cases} \qquad \ell \geq \ell_\mathrm{S} \geq 3 \tag{3.19}$$

Another more sophisticated model, known as "Clean Chain Ends" (CCE) model is illustrated in Fig. 3.2. As opposed to HDC model where O-sites are randomly distributed among all sites over the chain, leaving possibility of obstructing X-,Y-, or Z-sites, hence blocking *exo*-access from enzymes, CCE restrains possible O-sites only within an interior segment, leaving two short terminal chain segments unobstructed. These two terminal chain segments are designed to be of monomer lengths k_L and k_R away from L-end and R-end respectively. With $k_\mathrm{L} \geq k_\mathrm{X} + 1$ and $k_\mathrm{R} \geq k_\mathrm{Y} + 1$, we assured that no obstruction will prevent *exo*-access to X-,Y-sites to occur due to near-end O-sites. As far as interior segment goes, it only exists for chain lengths $\ell \geq k_\mathrm{L} + k_\mathrm{R}$, and contains $\ell - k_\mathrm{L} - k_\mathrm{R} + 1$ bonds. And in CCE model, $\phi_{O,\sigma}$ is the fraction of interior chain segment sites that are O-sites, with the complement being N-sites. Based on these model assumptions, we get $N_{\nu,\sigma}(\ell)$ in the CCE model for insoluble

chains G_ℓ with $\ell \geq \ell_{\rm S} > 3$, $k_{\rm L} > k_{\rm X}$ and $k_{\rm R} > k_{\rm Y}$:

$$N^{(\rm CCE)}_{\nu,\sigma}(\ell) = \begin{cases} \ell - 3 + \delta_{\ell,k_{\rm X}+k_{\rm Y}} - \phi_{O,\sigma}\max(\ell - k_{\rm T}, 0)\ , & \nu = N; \\ 1 - \delta_{\ell,k_{\rm X}+k_{\rm Y}}\ , & \nu = X \text{ or } Y; \\ \delta_{\ell,k_{\rm X}+k_{\rm Y}}\ , & \nu = Z; \\ 1 & \nu = L \text{ or } R \\ \phi_{O,\sigma}\max(\ell - k_{\rm T}, 0)\ , & \nu = O; \end{cases} \tag{3.20}$$

where $k_{\rm T} = k_{\rm L} + k_{\rm R} - 1$.

The chain site number functions $N_{\nu,\sigma}(\ell)$ are actually closely related with the conditional site type probabilities $P_\sigma(\mu|k, k', \zeta)$, which will be introduced in the next section.

3.4 Super Chain and Chain Fragmentation Probability

3.4.1 Super Chain

We are first going to construct purely fictitious Super Chain for the purpose of a mathematically accounting device and later aid in the process of $P_\sigma(k, k'|\mu, +1)$ derivation from chain length distribution $P_\sigma(\ell)$.

Now assume we have a sufficiently large collection of chains, with random chain lengths ℓ drawn according to $P_\sigma(\ell)$ and sample size $\mathcal{N}_{\rm L} \to \infty$. Have these $\mathcal{N}_{\rm L} \to \infty$ chains to be concatenated, in random order, into a "superchain" in a way that the L-end of one chain is bonded with R-end of another to form an imaginary bond, referred to as a "-1-bond", while all intact, real internal bonds between monomers inside each chain are referred to as "$+1$-bonds". According to this definition, we therefore assign each bond a "bond integrity" variable ζ with $\zeta = +1$ for bonds inside chains (*i.e.* for N, X, Y, Z, or O-sites) and $\zeta = -1$ for the bonds between adjacent chain ends (*i.e.* for a pair of adjacent L, R-sites).

3.4.2 Derivation of Chain Fragmentation Probability $P_\sigma(k, k'|\mu, +1)$

Both average increment in site type-ν, $\Delta\bar{N}_{\nu,\mu,\sigma}$, and site number rate functions, $V_{\nu,\sigma}$, among many others are expressed in terms of chain fragmentation probabilities $P_\sigma(k, k'|\mu, +1)$.

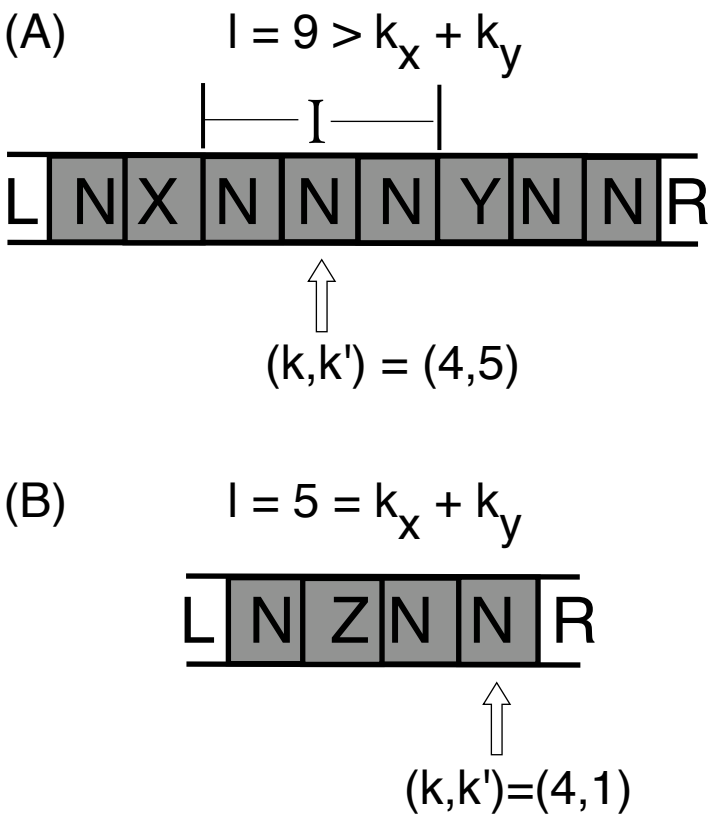

Figure 3.1: Homomgeneously Dirty Chain. Distribution of the seven site types N, X, Y, Z, L, R, O along glucan chains of monomer lengths $\ell = 9$ (A) and $\ell = 5 = k_{\rm X} + k_{\rm Y}$ (B) in the "homogeneously dirty chain" (HDC) model. Both examples (A) and (B) are for a hypothetical system of *endo*-, *exo*-L- and *exo*-R-acting enzymes with *exo*-L- and *exo*-R-cuts to produce soluble oligomers of lengths $k_{\rm X} = 2$ and $k_{\rm Y} = 3$ from the L- and R-end, respectively. Square boxes represent $\beta(1,4)$ 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 $\ell \geq k_{\rm X} + k_{\rm Y} + 2$ 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.

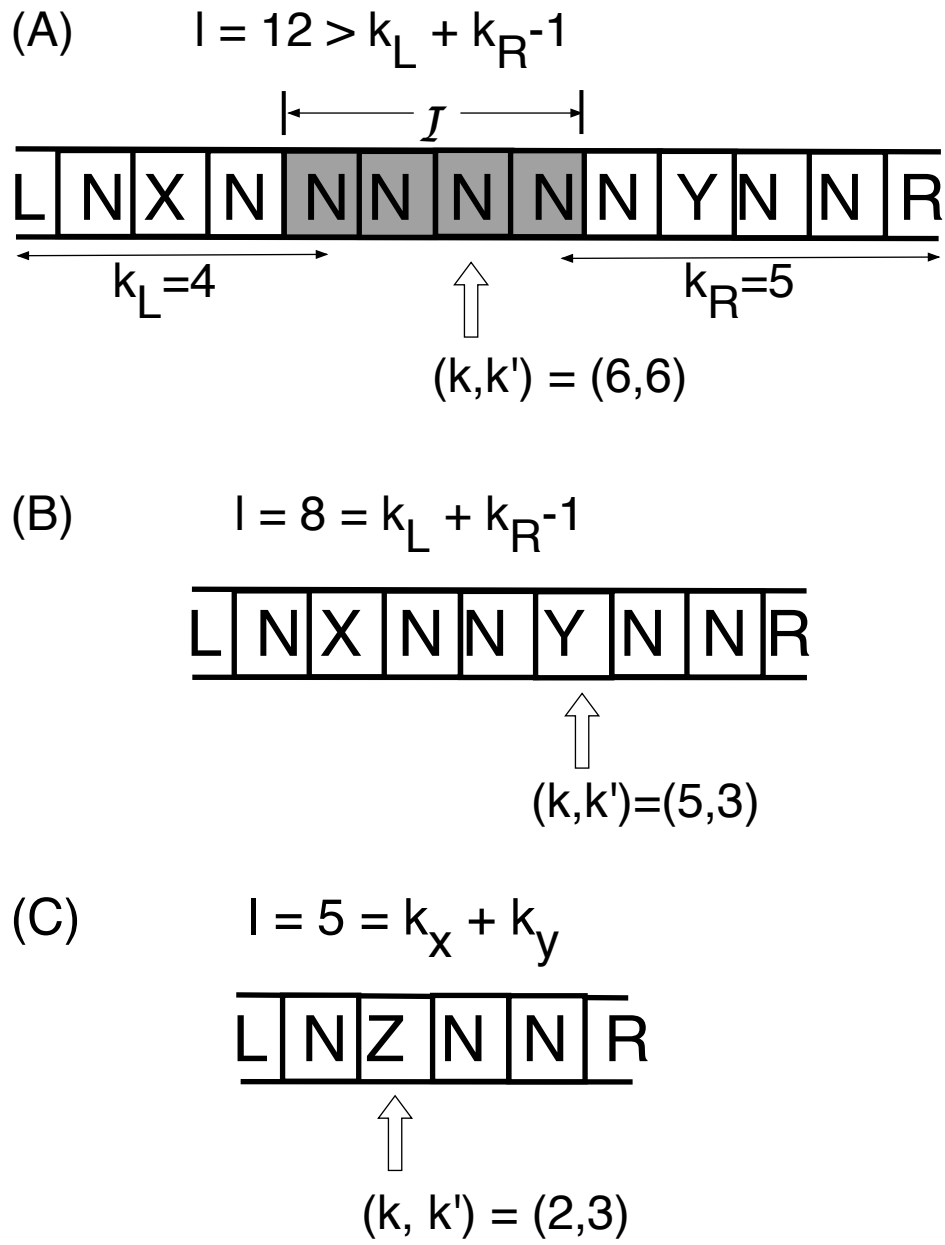

Figure 3.2: Clean Chain Ends. Distribution of the seven site types N, X, Y, Z, L, R, O along glucan chains of monomer lengths $\ell = 12$ (A), $\ell = 8 = k_{\rm L} + k_{\rm R} - 1$ (B), and $\ell = 5 = k_{\rm X} + k_{\rm Y}$ (C), in the "clean chain ends" (CCE) model. Notation and graphical representation of bonds, monomers, interior chain segment, and possible fragmentation examples are the same as in Figure 4, with the same hypothetical values of $k_{\rm X} = 2$ and $k_{\rm Y} = 3$, respectively, in all three examples (A), (B) and (C). The assumed clean chain end segments (kept free of O-sites) have monomer lengths $k_{\rm L} = 4$ and $k_{\rm R} = 5$, counted from the L- and R-end respectively. Bonds labelled with letters N in shaded (grey) 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 labelled 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 $\ell \geq k_{\rm L} + k_{\rm R}$ have an "interior segment", indicated in (A) by the horizontal bar labelled "I". Only a bond within an interior chain segment can of site type O (with probability $\phi_{O,\sigma}$).

In this section, we will construct chain fragmentation probabilities through surface chain length probabilities $P_\sigma(\ell)$ and conditional site type probabilities $P_\sigma(\mu|k,k',+1)$, using Bayes' theorem.

$P_\sigma(\ell)$ is a time-dependent variable that describes the probability of randomly selected insoluble glucan chain, exposed on surface of class-σ SAC, to be comprised of ℓ G_1 monomers. $P_\sigma(\mu|k,k',+1)$, on the other hand, is the probability for a selected *intact* bond ($\mu = N$,X,Y,Z,O), at a position k monomers away from the L-end and k' monomers away from the R-end, to be of type μ. Thus, $P_\sigma(\mu|k,k',+1)$ contains information about the site type distribution along the glucan chain, and similar to $N_{\nu,\sigma}(\ell)$, must be provided as model input.

Consider first the site type fractions, $f_{\nu,\sigma}$, defined as the ratio between number of site type-ν on all SAC surfaces of class-σ over number of all site types on these same SAC surfaces.

$$f_{\nu,\sigma} = \frac{n_{\nu,\sigma}}{n_{\mathrm{M},\sigma}} \tag{3.21}$$

Next, define $P_\sigma(k,k',\zeta)$ to be the probability that a randomly selected bond on a superchain to be a ζ bond ($\zeta = \pm 1$), located at k monomers away from its nearest L-end (*i.e.* nearest $\zeta = -1$ bond to its left) and k' monomers away from its nearest R-end (*i.e.* nearest $\zeta = -1$ bond to its right). Therefore, we are in a position to write out $P_\sigma(k,k',\zeta)$ in terms of site type fractions and surface chain length probabilities:

$$P_\sigma(k,k',\zeta) = \delta_{\zeta,+1} f_{L,\sigma} P_\sigma(k+k') + \delta_{\zeta,-1} f_{L,\sigma} P_\sigma(k) P_\sigma(k') \tag{3.22}$$

This equation can be conceived in two scenarios. First, if $\zeta = +1$, $P_\sigma(k,k',\zeta)$ is equivalent of a probability finding intact bond (bond type $\mu = N,X,Y,Z,O$) on a chain whose total length is $k+k'$. This probability is then simply the product of probability finding the chain L-end, $f_{L,\sigma}$, and probability of that chain to have total length $k+k'$, $P_\sigma(k+k')$. Similarly, if $\zeta = -1$ (bond type $\mu = L,R$), $P_\sigma(k,k',\zeta)$ is the probability to find specifically L- or R-end flanked in between two chains with length equal to k and k', respectively. Notice, we used the

L-R-symmetry adopted earlier where it gives us $f_{L,\sigma} = f_{R,\sigma}$. Also, it can be easily verified that, $\sum_{k,k'\geq 1}\sum_{\zeta=\pm 1} P_\sigma(k,k',\zeta) = 1$.

Next we are in a position to introduce "conditional site type probability, given fragments," denoted by $P_\sigma(\mu|k,k',\zeta)$. This is a probability for a randomly selected site type to be μ, given that its bond integrity is ζ; located k monomers away from nearest L-end to its left and k' monomers away from nearest R-end to its right. Apparently $P_\sigma(\mu|k,k',\zeta)$ also depends on chain site distribution probabilities, and we are now going to give its format according to models we discussed earlier (HDC and CCE models). As we defined earlier, $\zeta = +1$ bonds on superchain corresponds to intact glucan bond types (N,X,Y,Z or O) while $\zeta = -1$ bonds on superchain maps into L-,R-ends with probability 1/2 ($\mu = L$ or R). The latter is consistent with the assignment of monomer weights $w_\mu = 1/2$ for $\mu = L$ or R introduced in Eq. (3.1).

Following the same HDC and CCE model descriptions illustrated in Fig. 3.1 and 3.2, $P_\sigma(\mu|k,k',\zeta)$ is given as:

$$P_\sigma^{(\mathrm{HDC})}(\mu|k,k',\zeta) = \begin{cases} \delta_{\zeta,+1}(1-\phi_{O,\sigma})(1-\delta_{k,k_\mathrm{X}})(1-\delta_{k',k_\mathrm{Y}}) \ , & \mu = N; \\ \delta_{\zeta,+1}(1-\phi_{O,\sigma})\delta_{k,k_\mathrm{X}}(1-\delta_{k',k_\mathrm{Y}}) \ , & \mu = X; \\ \delta_{\zeta,+1}(1-\phi_{O,\sigma})(1-\delta_{k,k_\mathrm{X}})\delta_{k',k_\mathrm{Y}} \ , & \mu = Y; \\ \delta_{\zeta,+1}(1-\phi_{O,\sigma})\delta_{k,k_\mathrm{X}}\delta_{k',k_\mathrm{Y}} \ , & \mu = Z; \\ \delta_{\zeta,+1}\phi_{O,\sigma} \ , & \mu = O; \\ \delta_{\zeta,-1}/2 \ , & \mu = L,R; \end{cases} \tag{3.23}$$

with $\phi_{O,\sigma}$ defined as in section 3.3, Eq. (3.19); and

$$P_{\sigma}^{(\mathrm{CCE})}(\mu|k,k',\zeta) = \begin{cases} \delta_{\zeta,+1}(1-\delta_{k,k_{\mathrm{X}}})(1-\delta_{k',k_{\mathrm{Y}}}) \\ \quad [1-\phi_{O,\sigma}\Theta(k-k_{\mathrm{L}}+1) \\ \qquad \Theta(k'-k_{\mathrm{R}}+1)] \ , & \mu = N; \\ \delta_{\zeta,+1}\delta_{k,k_{\mathrm{X}}}(1-\delta_{k',k_{\mathrm{Y}}}) \ , & \mu = X; \\ \delta_{\zeta,+1}(1-\delta_{k,k_{\mathrm{X}}})\delta_{k',k_{\mathrm{Y}}} \ , & \mu = Y; \\ \delta_{\zeta,+1}\delta_{k,k_{\mathrm{X}}}\delta_{k',k_{\mathrm{Y}}} \ , & \mu = Z; \\ \delta_{\zeta,+1}\phi_{O,\sigma}\Theta(k-k_{\mathrm{L}}+1) \\ \qquad \Theta(k'-k_{\mathrm{R}}+1) \ , & \mu = O; \\ \delta_{\zeta,-1}/2 \ , & \mu = L,R; \end{cases} \tag{3.24}$$

with $\Theta(\Delta k) \equiv 0$ ($\equiv 1$) for $\Delta k \leq 0$ (> 0), and with $\phi_{O,\sigma}$ defined as in section section 3.3, Eq. (3.20). Note that in the case of a "clean" cellulosic substrate, *i.e.*, in the absence of O-sites ($\phi_{O,\sigma} \equiv 0$), the HDC and CCE models become identical. Therefore, for the purposes of fragmentation kinetics, $P_{\sigma}(\mu|k,k',\zeta)$ comprises the complete mathematical description of the chain site distribution model in the superchain language. From their shared dependence on site distribution along the chain, $P_{\sigma}(\mu|k,k',\zeta)$ and $N_{\mu,\sigma}(k)$ are related mathematical language from different perspective and are thus *NOT* totally independent of each other. Straightforwardly, $P_{\sigma}(\mu|k,k',\zeta)$ and $N_{\mu,\sigma}(k)$ must adhere to the following site number counting relations:

$$N_{\mu,\sigma}(\ell) = \Theta(\ell-\ell_{\mathrm{S}}+1)\sum_{k=1}^{\ell-1} P_{\sigma}(\mu|k,\ell-k,+1) \quad \text{for} \ \ \mu = N,X,Y,Z,O \tag{3.25}$$

with $\Theta(\Delta\ell) \equiv 0$ ($\equiv 1$) for $\Delta\ell \leq 0$ (> 0). Thus, $P_{\sigma}(\mu|k,k',\zeta)$ completely determines $N_{\mu,\sigma}(k)$ for intact bond types μ. In addition, of course, $P_{\sigma}(\mu|k,k',\zeta)$ must be normalized according to $\sum_{\mu} P_{\sigma}(\mu|k,k',\zeta) = 1$.

Finally, we are equipped with everything we need to construct "conditional fragmentation probability, given the site type", $P_{\sigma}(k,k'|\mu,\zeta)$. It describes that given the bond is a ζ-bond *and* of site type-μ, what is the probability a hydrolytic enzyme cut on that bond will generate k and k' monomers from its nearest -1-bond to the left and to the right, respectively.

According to Bayes' Theorem, we have the following:

$$P_\sigma(k,k'|\mu,\zeta) = \frac{P_\sigma(k,k',\mu,\zeta)}{P_\sigma(\mu,\zeta)} = \frac{P_\sigma(\mu|k,k',\zeta)\times P_\sigma(k,k',\zeta)}{P_\sigma(\mu,\zeta)} \tag{3.26}$$

where the unconditional site type probability $P_\sigma(\mu,\zeta)$ is given by:

$$P_\sigma(\mu,\zeta) = \sum_{k,k'\geq 1} P_\sigma(k,k',\mu,\zeta) = \begin{cases} \delta_{\zeta,+1} f_{\mu,\sigma} & \text{for } \mu = N,X,Y,Z,O \\ \delta_{\zeta,-1} f_{\mu,\sigma}/2 & \text{for } \mu = L,R \end{cases} \tag{3.27}$$

Of course, $P_\sigma(k,k'|\mu,\zeta)$ has to be normalized to $\sum_{k=1}^{\infty}\sum_{k'=1}^{\infty} P_\sigma(k,k'|\mu,\zeta) = 1$.

Inserting Eqs. (3.26) and (3.27) with $\zeta = +1$ into Eqs. (3.16) and (3.17), immediately yields the following two equations: [*via* Eq. (3.15)]

$$\begin{aligned} V_{\nu,\sigma} &= \sum_{\kappa,\mu} \gamma_{\kappa,\mu} m_{\kappa,\mu,\sigma} (n_{L,\sigma}/n_{\mu,\sigma}) \\ &\quad \times \sum_{k,k'=1}^{\infty} \Delta N_{\nu,\sigma}(k,k')\ P_\sigma(\mu|k,k',+1)\ P_\sigma(k+k') \end{aligned} \tag{3.28}$$

$$\begin{aligned} V_{\mathrm{S},\sigma}(k) &= \sum_{\kappa,\mu} \gamma_{\kappa,\mu} m_{\kappa,\mu,\sigma} (n_{L,\sigma}/n_{\mu,\sigma}) \\ &\quad \times \sum_{k'=\ell_{\mathrm{S}}-k}^{\infty} \left[P_\sigma(\mu|k,k',+1) + P_\sigma(\mu|k',k,+1)\right]\ P_\sigma(k+k')\ . \end{aligned} \tag{3.29}$$

Eq. (3.28) and Eq. (3.29) provide a gateway for us to assess $V_{\nu,\sigma}$ and $V_{\mathrm{S},\sigma}(k)$ from a complete profile of surface chain length distribution as well as their time evolution trajectory.

3.5 Chain End Decomposition

Given the realistic variables of cellulosic substrate, the longest chain length ℓ_{max} could easily go up to hundreds, if not thousands, rendering the task of evaluating a complete profile of chain length probabilities too time-consuming. In this section, however, we will introduce an approximation where the chain number probability variables, $P_\sigma(\ell)$ does *NOT* need to be fully determined for the purpose of calculating rate functions $V_{\nu,\sigma}$ and $V_{\mathrm{S},\sigma}(k)$. In order to do so, we first need to dissect the chain into fictitious functional segments to isolate the effects of near-chain-end sites, where both *exo* and *endo*-activity may occur, from the chain

interior sites, where only *endo*-activity occurs. The underlying assumption justifying such decomposition is "chain homogeneity" and "chain end locality," that the chain is sufficiently long with respect to the length of chain ends, so that its interior is practically homogeneous and unaffected by chain end effects.

Formally the decomposition can be expressed as:

$$P_\sigma(\mu|k,k',+1) = p^{(\mathrm{I})}_{\mu,\sigma} + \Theta_{\mathrm{L}}(k)p^{(\mathrm{L})}_{\mu,\sigma}(k) + \Theta_{\mathrm{R}}(k')p^{(\mathrm{R})}_{\mu,\sigma}(k') + \Theta_{\mathrm{L}}(k)\Theta_{\mathrm{R}}(k')p^{(\mathrm{S})}_{\mu,\sigma}(k,k') \quad (3.30)$$

where $p^{(\mathrm{I})}_{\mu,\sigma}$, $p^{(\mathrm{L})}_{\mu,\sigma}(k)$, $p^{(\mathrm{R})}_{\mu,\sigma}(k')$ and $p^{(\mathrm{S})}_{\mu,\sigma}(k,k')$ stand for contributions from chain interior (I), chain end (L, R) and short-chain (S) ($\ell \le \ell_{\mathrm{LR}} \equiv \ell_{\mathrm{L}} + \ell_{\mathrm{R}} - 1$) respectively; $\Theta_{\mathrm{L}}(k) \equiv \Theta(\ell_{\mathrm{L}} - k)$ and $\Theta_{\mathrm{R}}(k') \equiv \Theta(\ell_{\mathrm{R}} - k')$ are the cut-off factors. ℓ_{L} (ℓ_{R}) signifies the maximum range of chain-end effects from L-(R-) ends: it is thus to say, for sites positioned $\ell > \ell_{\mathrm{L}}$ ($> \ell_{\mathrm{R}}$) monomers away from L-end (R-end), terminal effects would no longer be relevant to their physical properties. For chains of length $\ell \ge \ell_{\mathrm{LR}}$, this decomposition essentially dissects each one into three segments, illustrated in Figures 3.1A and 3.2A: an interior (I-) segment with $\ell - \ell_{\mathrm{LR}}$ intact bonds, flanked by two terminal segments. (L-segment with ℓ_{L} monomers and R-segment with ℓ_{R} monomers). These four chain segments can be straightforwardly evaluated from the underlying chain site distribution model (HDC, CCE)discussed earlier, with $\ell_{\mathrm{L}} = k_{\mathrm{X}} + 1$ and $\ell_{\mathrm{R}} = k_{\mathrm{Y}} + 1$ for HDC model ; and $\ell_{\mathrm{L}} = k_{\mathrm{L}}$ and $\ell_{\mathrm{R}} = k_{\mathrm{R}}$ for CCE model.

One critical point following such formality: for chain length $\ell = k + k' > \ell_{\mathrm{LR}}$ ($k \ge \ell_{\mathrm{L}}$ and $k' \ge \ell_{\mathrm{R}}$), $P_\sigma(\mu|k,k',+1)$ collapses into one single term $p^{(\mathrm{I})}_{\mu,\sigma}$ and is thus independent of k or k'. Similarly, for L-terminal sites with $k < \ell_{\mathrm{L}}$ and $k' \ge \ell_{\mathrm{R}}$, $P_\sigma(\mu|k,k',+1) = p^{(\mathrm{I})}_{\mu,\sigma} + p^{(\mathrm{L})}_{\mu,\sigma}(k)$ is independent of k'; and for R-terminal sites with $k' < \ell_{\mathrm{R}}$ and $k \ge \ell_{\mathrm{L}}$, $P_\sigma(\mu|k,k',+1) = p^{(\mathrm{I})}_{\mu,\sigma} + p^{(\mathrm{R})}_{\mu,\sigma}(k')$ is independent of k.

Substitute k' with $(\ell - k)$ in Eq. (3.30), and through Eq. (3.25) we defined earlier, formally the chain site counting functions $N_{\nu,\sigma}$ for intact bond sites (μ=N,X,Y,Z,O) can be

decomposed into:

$$
\begin{aligned}
N_{\mu,\sigma}(\ell) &= b_{\mu,\sigma}\ell + a_{\mu,\sigma} + \Theta_{\rm T}(\ell) d_{\mu,\sigma}(\ell) && (3.31)\\
b_{\mu,\sigma} &= p^{(\rm I)}_{\mu,\sigma} && (3.32)\\
a_{\mu,\sigma} &= -p^{(I)}_{\mu,\sigma} + \sum_{k\geq 1}[\Theta_{\rm L}(k) p^{(L)}_{\mu,\sigma}(k) + \Theta_{\rm R}(k) p^{(R)}_{\mu,\sigma}(k)] && (3.33)
\end{aligned}
$$

with $d_{\nu,\sigma}(\ell) \equiv N_{\nu,\sigma}(\ell) - b_{\nu,\sigma}\ell - a_{\nu,\sigma}$; and $\Theta_{\rm T}(\ell) = \Theta(\ell_{\rm T} - \ell)$ where $\ell_{\rm T} \equiv \max(\ell_{\rm S}, \ell_{\rm L} + \ell_{\rm R} - 1)$. Eq. (3.31) also applies to partial site type $\mu = L$ and R, with $a_{\mu,\sigma} = 1$ and $b_{\mu,\sigma} = 0$, since $N_{L,\sigma}(\ell) = N_{R,\sigma}(\ell) \equiv 1$ for all $\ell \geq \ell_{\rm S}$. In the HDC and CCE chain site distribution models discussed above, only N- and O-sites can have a non-zero ℓ-linear contribution, namely, $b_{N,\sigma} = 1 - \phi_{O,\sigma}$ and $b_{O,\sigma} = \phi_{O,\sigma}$, respectively, while $b_{\mu,\sigma} = 0$ for all site types which are restricted to near-chain-end locations, *i.e.*, for $\mu = X, Y, Z, L, R$.

Then from Eqs. (3.30) and (3.31) *via* Eq. (3.18), the average increment functions $\Delta\bar{N}_{\nu,\mu,\sigma}$ are decomposed into

$$
\Delta\bar{N}_{\nu,\mu,\sigma} = (1/n_{\mu,\sigma})[A_{\nu,\mu,\sigma} n_{{\rm M},\sigma}(\lambda_\sigma) + (B_{\nu,\mu,\sigma} + D_{\nu,\mu,\sigma}) n_{{\rm L},\sigma}] \tag{3.34}
$$

with $A_{\nu,\mu,\sigma}$, $B_{\nu,\mu,\sigma}$ and $D_{\nu,\mu,\sigma}$ explicitly expressed as:

$$
\begin{aligned}
A_{\nu,\mu,\sigma} &= a_{\nu,\sigma} p^{(\rm I)}_{\mu,\sigma}\\
B_{\nu,\mu,\sigma} &= -a_{\nu,\sigma} p^{(\rm I)}_{\mu,\sigma} + 2\textstyle\sum_{k=1}^{\ell_{\rm T}-1} d_{\nu,\sigma}(k) p^{(\rm I)}_{\mu,\sigma}\\
&\quad + \textstyle\sum_{k=1}^{\ell_{\rm L}-1}[a_{\nu,\sigma} + d_{\nu,\sigma}(k)] p^{(\rm L)}_{\mu,\sigma}(k) + \sum_{k'=1}^{\ell_{\rm R}-1}[a_{\nu,\sigma} + d_{\nu,\sigma}(k')] p^{(\rm R)}_{\mu,\sigma}(k')\\
D_{\nu,\mu,\sigma} &= -\textstyle\sum_{k\geq 1}[\Theta_{\rm L}(k) p^{(\rm L)}_{\mu,\sigma}(k)(a_{\nu,\sigma} + d_{\nu,\sigma}(k))\\
&\qquad + \Theta_{\rm R}(k) p^{(\rm R)}_{\mu,\sigma}(k)(a_{\nu,\sigma} + d_{\nu,\sigma}(k))\\
&\qquad + 2\Theta_{\rm T}(k) p^{(\rm I)}_{\mu,\sigma} d_{\nu,\sigma}(k)] \textstyle\sum_{\ell=1}^{k} P_\sigma(\ell)\\
&\quad + \textstyle\sum_{k,k'\geq 1}[\Theta_{\rm T}(k)\Theta_{\rm R}(k') d_{\nu,\sigma}(k) p^{(\rm R)}_{\mu,\sigma}(k')\\
&\qquad + \Theta_{\rm T}(k')\Theta_{\rm L}(k) d_{\nu,\sigma}(k') p^{(\rm L)}_{\mu,\sigma}(k)\\
&\qquad + \Theta_{\rm L}(k)\Theta_{\rm R}(k')(a_{\nu,\sigma} + d_{\nu,\sigma}(k) + d_{\nu,\sigma}(k')) p^{(\rm S)}_{\mu,\sigma}(k,k')\\
&\qquad - \Theta_{\rm T}(k+k') d_{\nu,\sigma}(k+k') P_\sigma(\mu|k,k',\zeta)] P_\sigma(k+k')
\end{aligned} \tag{3.35}
$$

It is critical to notice that $A_{\nu,\mu,\sigma}$ and $B_{\nu,\mu,\sigma}$ are contributions from cutting chains of any length, and are thus constants, *i.e.* independent of any chain length probabilities $P_\sigma(\ell)$ or

other dynamical variables. $D_{\nu,\mu,\sigma}$, on the other hand, is constructed based on chain length probabilities, but *only* those $P_\sigma(\ell)$ with $\ell \leq \ell_\mathrm{C}$, where $\ell_\mathrm{C} \equiv \ell_\mathrm{T} + \max(\ell_\mathrm{L}, \ell_\mathrm{R}) - 2$.

By way of Eq. (3.15), we are in a position to write out chain-end decomposition for the rate function $V_{\nu,\sigma}$ with the same variables we constructed above:

$$V_{\nu,\sigma} = \sum_{\kappa,\mu} \gamma_{\kappa,\mu}(m_{\kappa,\mu,\sigma}/n_{\mu,\sigma})[A_{\nu,\mu,\sigma} n_{\mathrm{M},\sigma}(\lambda_\sigma) + (B_{\nu,\mu,\sigma} + D_{\nu,\mu,\sigma}) n_{\mathrm{L},\sigma}] \tag{3.36}$$

Analogously but without detailed derivations, rate equations for soluble oligomers for $k < \ell_\mathrm{S}$ are given as:

$$V_{\mathrm{S},\sigma}(k) = \textstyle\sum_{\kappa,\mu} \gamma_{\kappa,\mu} m_{\kappa,\mu,\sigma}(n_{\mathrm{L},\sigma}/n_{\mu,\sigma})[B_{\mathrm{S},\mu,\sigma}(k) + D_{\mathrm{S},\mu,\sigma}(k)] \tag{3.37}$$

The $B_{\mathrm{S},\mu,\sigma}(k)$ again denote constant coefficients which do not depend on any $P_\sigma(\ell)$ or other dynamical variables, whereas the $D_{\mathrm{S},\mu,\sigma}(k)$-contributions depend explicitly *only* on short-chain number variables $P_\sigma(\ell)$ with $\ell \leq \max(\ell_\mathrm{L}, \ell_\mathrm{R}) - 1 \leq \ell_\mathrm{C}$, as follows:

$$\begin{aligned} B_{\mathrm{S},\mu,\sigma}(k) &= 2p^{(\mathrm{I})}_{\mu,\sigma} + \Theta_\mathrm{L}(k) p^{(\mathrm{L})}_{\mu,\sigma}(k) + \Theta_\mathrm{R}(k) p^{(\mathrm{R})}_{\mu,\sigma}(k) \\ D_{\mathrm{S},\mu,\sigma}(k) &= \textstyle\sum_{k' \geq \ell_\mathrm{S} - k} [\Theta_\mathrm{L}(k') p^{(\mathrm{L})}_{\mu,\sigma}(k') + \Theta_\mathrm{R}(k') p^{(\mathrm{R})}_{\mu,\sigma}(k') \\ &\quad + \Theta_\mathrm{L}(k)\Theta_\mathrm{R}(k') p^{(\mathrm{S})}_{\mu,\sigma}(k, k') \\ &\quad + \Theta_\mathrm{R}(k)\Theta_\mathrm{L}(k') p^{(\mathrm{S})}_{\mu,\sigma}(k', k)] P_\sigma(k + k') \end{aligned} \tag{3.38}$$

Overall, it is clearly shown in aforementioned derivations that under chain-end decomposition, $\Delta\bar{N}_{\nu,\mu,\sigma}$ and thus rate equations $V_{\nu,\sigma}$, $V_{\mathrm{S},\sigma}$ become dependent *only* on short chain length probabilities $P_\sigma(\ell)$ with $\ell \leq \ell_\mathrm{C}$ via the D-contributions while contributions from all other longer-chain probabilities have been completely absorbed into site number variables $n_{\nu,\sigma}$ and $n_{\mathrm{M},\sigma}(\lambda_\sigma)$.

3.6 Rate Equation Closure in the Long-Chain Limit

3.6.1 Rate Equations for Chain Number Variables $H_\sigma(\ell)$ with Chain End Decomposition

Before we set out to solve rate equations for $P_\sigma(\ell)$ for short chains, we first develop a set of rate equations similar to rate equations we developed for site number formalism before.

Chain number variable is defined as $H_\sigma(\ell) \equiv P_\sigma(\ell) n_{L,\sigma}$, and is thus time-dependent ($H_\sigma(\ell) \equiv H_\sigma(\ell, t)$). $H_\sigma(\ell)$ describes the average number of insoluble glucan chains G_ℓ exposed on SAC surface of length $\ell \geq \ell_\mathrm{S}$, per class-σ SAC. Following the same definition, for all $\ell < \ell_\mathrm{S}$, $H_\sigma(\ell) \equiv 0$.

All surface site number variables we developed earlier can now be expressed in terms of $H_\sigma(\ell)$ as well:

$$n_{\mathrm{M},\sigma}(\lambda_\sigma) = \sum_{\ell=\ell_\mathrm{S}}^{\infty} \ell \; H_\sigma(\ell); \qquad n_{\nu,\sigma} = \sum_{\ell=\ell_\mathrm{S}}^{\infty} N_{\nu,\sigma}(\ell) H_\sigma(\ell) \tag{3.39}$$

Thus, with these three equalities, we have equipped ourselves with tools to bridge between site number formalism and chain number formalism.

Analogous to site ablation rate equations (3.3) and (3.6), we can write out rate equations for chain number formalism as:

$$\dot{H}_\sigma(\ell) = V_{\mathrm{H},\sigma}(\ell) - \bar{V}_\sigma \eta_\sigma(\lambda_\sigma) Q_\sigma(\ell, \lambda_\sigma) / \sum_{j=\ell_\mathrm{S}}^{\infty} j Q_\sigma(j, \lambda_\sigma) \tag{3.40}$$

$$V_{\mathrm{H},\sigma}(\ell) = -\sum_{k,k'=1}^{\infty} V_\sigma(\ell \to k, k') + \sum_{k=1}^{\infty} \sum_{j=\ell+1}^{\infty} [V_\sigma(j \to k, \ell) + V_\sigma(j \to \ell, k)] \tag{3.41}$$

$$V_\sigma(\ell \to k, k') = \sum_{\kappa,\mu} \gamma_{\kappa,\mu} m_{\kappa,\mu,\sigma} P_\sigma(k, k' | \mu, +1) \delta_{\ell, k+k'} \tag{3.42}$$

Similar to corresponding terms in Eq. (3.6), the *r.h.s.* of Eq. (3.40) is comprised of both fragmentation term and surface exposure term. $V_{\mathrm{H},\sigma}(\ell)$ gives the rate of production and consumption of chains of length ℓ due to enzymatic bond cutting events (fragmentation); and the second term contributes the rate of exposure of new chains due to the removal of overlaying material (surface exposure). $Q_\sigma(\ell, \lambda)$ is the native (pre-hydrolysis) chain length distribution of substrate material in layer λ, that is time-independent and pre-determined, and should be fed into our model as a morphological model input. $V_\sigma(\ell \to k, k')$ is the rate, per SAC, at which surface-exposed chains G_ℓ on class-σ SACs are being cut into fragments G_k and $G_{k'}$, from the original chain *L*- and *R*-end, respectively. As in the site number formalism, the volume ablation rate $\bar{V}_\sigma$ and $\eta_\sigma(\lambda)$ are again given by Eq. (3.2), but with $V_{\nu,\sigma}$ now being

expressed in terms of $V_{\mathrm{H},\sigma}(\ell)$ by

$$V_{\nu,\sigma} = \sum_{\ell=\ell_{\mathrm{S}}}^{\infty} N_{\nu,\sigma}(\ell) V_{\mathrm{H},\sigma}(\ell) \tag{3.43}$$

So far we have assembled a closed ODE system, consisted of Eqs. (3.40) for all chains $\ell \geq \ell_{\mathrm{S}}$ and λ rate equation given by Eq. (3.3).

Actually if we substitute Eq. (3.39) with the native site-type fractions $g_{\nu,\sigma}(\lambda)$ given by

$$\begin{aligned} g_{L,\sigma}(\lambda) &= 1/[\textstyle\sum_{\ell=\ell_{\mathrm{S}}}^{\infty} \ell \; Q_{\sigma}(\ell,\lambda)] \\ g_{\nu,\sigma}(\lambda) &= g_{L,\sigma}(\lambda) \textstyle\sum_{\ell=\ell_{\mathrm{S}}}^{\infty} N_{\nu,\sigma}(\ell) \; Q_{\sigma}(\ell,\lambda) \; . \end{aligned} \tag{3.44}$$

The chain number rate equation system Eq. (3.40) is mathematically *exactly* equivalent to the site number rate equations Eq. (3.6). Via relationships between site number formalism and chain number formalism given at the beginning of this subsection, the full chain length distribution $P_{\sigma}(\ell)$ can thus be completely determined.

Then we apply chain end decomposition, applying Eq. (3.30) to decompose $V_{\mathrm{H},\sigma}(\ell)$ we arrive at:

$$V_{\mathrm{H},\sigma}(\ell) = \sum_{\kappa,\mu} \gamma_{\kappa,\mu} m_{\kappa,\mu,\sigma} \frac{n_{\mathrm{L},\sigma}}{n_{\mu,\sigma}} [-N_{\mu,\sigma}(\ell) P_{\sigma}(\ell) + B_{\mathrm{H},\mu,\sigma}(\ell) + D_{\mathrm{H},\mu,\sigma}(\ell)] \tag{3.45}$$

where

$$\begin{aligned} B_{\mathrm{H},\mu,\sigma}(\ell) &= [p^{(\mathrm{L})}_{\mu,\sigma}(\ell)\Theta_{\mathrm{L}}(\ell) + p^{(\mathrm{R})}_{\mu,\sigma}(\ell)\Theta_{\mathrm{R}}(\ell) + 2p^{(\mathrm{I})}_{\mu,\sigma}](1 - \textstyle\sum_{j=1}^{\ell} P_{\sigma}(j)) \\ D_{\mathrm{H},\mu,\sigma}(\ell) &= \textstyle\sum_{j=\ell+1}^{\ell+\ell_{\mathrm{E}}} \big\{ [p^{(\mathrm{S})}_{\mu,\sigma}(\ell, j-\ell)\Theta_{\mathrm{L}}(\ell) + p^{(\mathrm{R})}_{\mu,\sigma}(j-\ell)]\Theta_{\mathrm{R}}(j-\ell) \\ &\quad + [p^{(\mathrm{S})}_{\mu,\sigma}(j-\ell, \ell)\Theta_{\mathrm{R}}(\ell) + p^{(\mathrm{L})}_{\mu,\sigma}(j-\ell)]\Theta_{\mathrm{L}}(j-\ell)\big\} P_{\sigma}(j) \end{aligned} \tag{3.46}$$

Here, $\ell_{\mathrm{E}} = \max(\ell_{\mathrm{L}}, \ell_{\mathrm{R}}) - 1 \leq \ell_{\mathrm{C}}$, and $D_{\mathrm{H},\mu,\sigma}(\ell)$ contains information regarding production of G_{ℓ}-chains limited to hydrolysis of "nearby" longer chain lengths j, with $\ell + \ell_{\mathrm{E}} \geq j \geq \ell + 1$. On the other hand, it is equivalent to say if we want to solve for chain number rate equations for $V_{\mathrm{H},\sigma}(\ell)$, (and hence $H_{\sigma}(\ell)$), up to some short chain cut-off ℓ_{D}, only the lower partition of complete $P_{\sigma}(j)$ ($j \leq \ell + \ell_{\mathrm{E}}$), along with site number variables $n_{\nu,\sigma}$ are needed for calculation. All other longer chain length contributions are, again, completely absorbed into $n_{\nu,\sigma}$ and $n_{\mathrm{M},\sigma}(\lambda_{\sigma})$.

3.6.2 Local Poisson Approximation in the Long Chain Limit

With a set of equations governing chain number variables $H_\sigma(\ell)$, we now need to construct a closed system of coupled rate equations involving *only* short-chain number variable contributions for ℓ up to short chain cut-off $\ell_{\rm D}$ mentioned above, along with site number rate equations derived as in Eqs. (3.6) and (3.3). Here we propose Local Poisson (LP) approximation scheme under Long Chain Limit (LCL), defined as follows:

1. At the beginning of hydrolysis, time $t^{\rm o}$, the vast majority of cellulose chains are found to have length much greater than our preassigned short chain cut-off $\ell_{\rm C}$. (The superscript labeling $^{\rm (o)}$ signifies initial values at the start of hydrolysis, here and in the following, for all quantities so labeled.) In other words, the probabilities of finding short chain lengths on SAC surfaces are negligible.

2. The short chains are mainly being produced due to *endo*-cuts from chains with much longer lengths, G_j with $j \gg \ell_{\rm C}$. In addition, chain length ($\ell > \ell_{\rm C}$) within the same magnitude as $\ell_{\rm C}$ are being generated in the same fashion, i.e. from *endo*-cuts on much longer chains.

Consequently for that pool of short chains, one should expect the chain number variables $H_\sigma(\ell)$ to be a slow varying function of ℓ. That is to say, given two neighboring chain number variables, *e.g.* $H_\sigma(\ell_{\rm D})$ and $H_\sigma(\ell_{\rm D}-1)$, according to a preset extrapolation scheme, one can estimate chain number variable at its near neighborhood, say in this case $H_\sigma(\ell_{\rm D}+1)$. This provides an estimation method that can extrapolate chain number variables for longer chain lengths ($\ell > \ell_{\rm D}$) from those for short chains $\ell_{\rm S} < \ell \le \ell_{\rm D}$.

Our proposed Local Poisson (LP) approximation utilizes a linear extrapolation of chain number logarithm, $\log P_\sigma(j)$, given as:

$$P_\sigma(j) \simeq P_\sigma(\ell_{\rm D})\Big[\frac{P_\sigma(\ell_{\rm D})}{P_\sigma(\ell_{\rm D}-1)}\Big]^{j-\ell_{\rm D}} \quad \text{for} \;\; j = \ell_{\rm D}+1, ..., \ell_{\rm D}+\ell_{\rm E} \tag{3.47}$$

with $P_\sigma(\ell) \equiv H_\sigma(\ell)/n_{L,\sigma}$ for $\ell \le \ell_{\rm D}$. Again, any reasonably proposed approximation can be readily incorporated into our model.

With LP approximation, we have successfully completed a closed ODE system, for the site number formalism, consisting of Eqs. (3.6), (3.3), (3.2), (3.15), (3.34) (3.35) (3.46) and (3.47); of Eqs. (3.40), and (3.45), for $\ell = \ell_{\rm S}, ..., \ell_{\rm D}$. The independent dynamical variables of this ODE system are the $n_{\nu,\sigma}$, λ_σ and short-chain numbers $H_\sigma(\ell),\ \ell = \ell_{\rm S}, ..., \ell_{\rm D}$.

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

$$n_{\nu,\sigma}(t^{\rm o}) = g_{\nu,\sigma}(\lambda_\sigma^{\rm o}) n_{{\rm M},\sigma}(\lambda_\sigma^{\rm o}) \tag{3.48}$$

$$\lambda_\sigma(t^{\rm o}) = \lambda_\sigma^{\rm o} \tag{3.49}$$

$$H_\sigma(\ell, t^{\rm o}) = \epsilon_{\rm Q}\ n_{L,\sigma}(t^{\rm o}) \quad \text{for} \ \ \ell = \ell_{\rm S}, ..., \ell_{\rm D} \tag{3.50}$$

Following the assumption under Long Chain Limit, we thus restrict ourselves to native chain length distributions *without* short chains, *i.e.*, $Q_\sigma(\ell, \lambda) = \epsilon_{\rm Q}$ for $\ell \leq \ell_{\rm D}$, with near-zero $\epsilon_{\rm Q}$ (e.g. 10^{-20}, but cannot be exact zero for computational purposes).

Chapter 4

Five-Site Ablation Model Applications and Parameterization [1]

4.1 Five-Site Ablation Model in Site Number Formalism

We have so far developed a general cellulose hydrolysis model with site ablation in terms of site number variables $n_{\nu,\sigma}$. For the purpose of direct comparison to experimentally measurable variables, we are going to develop such general framework into one with consideration of only five site types ($\nu = N, X, Y, L$ and R), as well as work under molar site concentrations,

$$x_{\nu,\sigma} \equiv C_\sigma n_{\nu,\sigma} \ . \tag{4.1}$$

Similarly, we introduce $G_\sigma(\ell)$, the molar concentrations of chains with length ℓ that are exposed on the surface of SAC class-σ. Relating to earlier variables, $G_\sigma(\ell)$ satisfies $G_\sigma(\ell) \equiv C_\sigma H_\sigma(\ell) = x_{L,\sigma} P_\sigma(\ell)$. Due to the absence of Z- and O-sites, we have

$$x_{L,\sigma} = x_{R,\sigma} = x_{X,\sigma} = x_{Y,\sigma} \tag{4.2}$$

$$x_{\mathrm{M},\sigma}(\lambda_\sigma) = x_{N,\sigma} + 3x_{L,\sigma} \tag{4.3}$$

$$x_{\mathrm{V},\sigma}(\lambda) \equiv C_\sigma n_{\mathrm{V},\sigma}(\lambda) = B_{\mathrm{V},\sigma}\ \lambda^{d_{\mathrm{A},\sigma}} \tag{4.4}$$

$$x_{\mathrm{M},\sigma}(\lambda) \equiv C_\sigma n_{\mathrm{M},\sigma}(\lambda) = x_{\mathrm{V},\sigma}(\lambda)\ -\ x_{\mathrm{V},\sigma}(\lambda - 1)\ \Theta(\lambda - 1) \tag{4.5}$$

where $B_{\mathrm{V},\sigma} \equiv C_\sigma c_{\mathrm{V},\sigma}$ are the molar volume prefactors and $\Theta(\lambda - 1) \equiv 1\ (\equiv 0)$ if $\lambda \geq 1$ ($\lambda < 1$). As discussed later, $B_{\mathrm{V},\sigma}$ is determined via the initial total molar fraction of substrate contained in class-σ SACs, average layer number for class-σ SACs and ablation dimension

[1] Contain contents published in Zhou W, Hao Z, Xu Y, Schüttler HB. 2009. Cellulose hydrolysis in evolving substrate morphologies II. Numerical results and analysis. *Biotechnol Bioeng* **104**:275-289. Permission acquired from *Biotechnol Bioeng* journal.

for the same class. Note that the $\lambda_\sigma = \lambda_\sigma(t)$ are time-*dependent* variables denoting the outermost (surface) layer number of hydrolytically evolving SACs in class-σ, whereas λ is an *independent* variable denoting any layer number in the interior or surface of a SAC.

Aside from representing the molar concentrations of G_1 monomers exposed on the surface and total amount in substrate of SAC class-σ, $x_{\mathrm{M},\sigma}$ and $x_{\mathrm{V},\sigma}$ can also be conceived as the amount of enzyme-accessible solid substrate surface area and the amount of the total solid substrate volume, measured in units of some appropriate average area or volume per G_1 monomer, respectively, for class-σ SACs. We will hereafter also refer to $x_{\mathrm{M},\sigma}$ and $x_{\mathrm{V},\sigma}$ as the class-σ "SAC surface" and "SAC volume", respectively. The total insoluble substrate monomer content x_{V} and the total surface monomer concentration x_{M}, counting across all SAC classes, are given as follows

$$x_{\mathrm{V}} = \sum_{\sigma=1}^{M_{\mathrm{MD}}} x_{\mathrm{V},\sigma}(\lambda_\sigma), \qquad x_{\mathrm{M}} = \sum_{\sigma=1}^{M_{\mathrm{MD}}} x_{\mathrm{M},\sigma}(\lambda_\sigma) \tag{4.6}$$

Under the foregoing simplifications, we are now able to assemble a set of rate equations for site number formalism that governs site molar concentrations $x_{\nu,\sigma}$, SAC outer layer number λ and for a handful of short-chain concentrations $G_\sigma(\ell)$:

$$\dot{x}_{N,\sigma} = R_{N,\sigma} - \bar{R}_\sigma\ \eta_\sigma(\lambda_\sigma)\ g_{N,\sigma}(\lambda_\sigma) \tag{4.7}$$

$$\dot{x}_{L,\sigma} = R_{L,\sigma} - \bar{R}_\sigma\ \eta_\sigma(\lambda_\sigma)\ g_{L,\sigma}(\lambda_\sigma) \tag{4.8}$$

$$\dot{\lambda}_\sigma = \bar{R}_\sigma / \partial_\lambda x_{\mathrm{V},\sigma}(\lambda_\sigma) = \bar{R}_\sigma / (B_{\mathrm{V},\sigma}\, d_{\mathrm{A},\sigma}\, \lambda_\sigma^{d_{\mathrm{A},\sigma}-1}) \tag{4.9}$$

$$\begin{aligned} \dot{G}_\sigma(\ell) = {} & \Gamma_{\mathrm{N},\sigma}\, x_{L,\sigma} \left[2 - 2\sum_{k=\ell_{\mathrm{S}}}^{\ell} P_\sigma(k) - (\ell-1)\ P_\sigma(\ell)\right] \\ & + \Gamma_{\mathrm{X},\sigma}\, x_{L,\sigma} \left[P_\sigma(\ell + k_{\mathrm{X}}) - P_\sigma(\ell)\right] \\ & + \Gamma_{\mathrm{Y},\sigma}\, x_{L,\sigma} \left[P_\sigma(\ell + k_{\mathrm{Y}}) - P_\sigma(\ell)\right] \qquad \text{for } \ell = \ell_{\mathrm{S}}, ...\ell_{\mathrm{D}} \end{aligned} \tag{4.10}$$

where $\ell_{\mathrm{D}} \geq \ell_{\mathrm{C}}$ with $\ell_{\mathrm{C}} = \ell_{\mathrm{S}} + \max(k_{\mathrm{X}}, k_{\mathrm{Y}}) - 1$ and

$$\eta_\sigma(\lambda_\sigma) \equiv 1 - \partial_\lambda x_{\mathrm{M},\sigma}(\lambda) / \partial_\lambda x_{\mathrm{V},\sigma}(\lambda) = \Theta(\lambda_\sigma - 1)\,(1 - 1/\lambda_\sigma)^{d_{\mathrm{A},\sigma}-1}\ . \tag{4.11}$$

The enzyme chain cutting rate factors are

$$\Gamma_{\mathrm{N},\sigma} = \gamma_{1,\mathrm{N}}\ L_{1,\mathrm{N}}\ v_1\ \frac{y_{N,\sigma}}{x_{N,\sigma}} \tag{4.12}$$

$$\Gamma_{\mathrm{X},\sigma} = \gamma_{2,\mathrm{X}}\ L_{2,\mathrm{X}}\ v_2\ \frac{y_{X,\sigma}}{x_{X,\sigma}} + \gamma_{1,\mathrm{N}}\ L_{1,\mathrm{N}}\ v_1\ (\frac{y_{X,\sigma}}{x_{X,\sigma}} - \frac{y_{N,\sigma}}{x_{N,\sigma}}) \tag{4.13}$$

$$\Gamma_{\mathrm{Y},\sigma} = \gamma_{3,\mathrm{Y}}\ L_{3,\mathrm{Y}}\ v_3\ \frac{y_{Y,\sigma}}{x_{Y,\sigma}}\ + \gamma_{1,\mathrm{N}}\ L_{1,\mathrm{N}}\ v_1\ (\frac{y_{Y,\sigma}}{x_{Y,\sigma}} - \frac{y_{N,\sigma}}{x_{N,\sigma}}) \tag{4.14}$$

where $\gamma_{\kappa,\mu}$ is the cutting rate coefficient defined earlier; $L_{\kappa,\mu}$ is the adsorption coefficient for (κ, μ) ES complex; v_κ is the free type-κ enzyme concentration; and $y_{\mu,\sigma}$, $x_{\mu,\sigma}$ being concentrations of free type-μ sites and all exposed type-μ sites on class-σ SAC surfaces respectively. In addition, we are following the same convention set before that κ=1, 2, or 3 representing the *endo-*, *exo-L-* and *exo-R*-acting glucanase, respectively.

Since *endo-* acting enzyme can attack any sites other than *L-* or *R*-ends, we thus set $\gamma_{1,X} = \gamma_{1,Y} = \gamma_{1,N}$ and likewise $L_{1,X} = L_{1,Y} = L_{1,N}$. We set all $L_{\kappa,\nu}$- and $\gamma_{\kappa,\nu}$-values according to a table listed further below (Table 1), and any variable not listed there will be set to 0. The concentrations of free substrate sites and free enzymes, $y_{\nu,\sigma}$ and v_κ, respectively, are in general obtained as functions of the corresponding total concentrations $x_{\nu,\sigma}$ and u_κ, by iterative solution of the coupled non-linear enzyme adsorption equilibrium equations. However, in the low-enzyme limit, we can approximate $y_{\mu,\sigma} \cong x_{\mu,\sigma}$ and

$$\begin{aligned}
\Gamma_{\mathrm{N},\sigma} &\cong \gamma_{1,\mathrm{N}}\ L_{1,\mathrm{N}}\ u_1\ /\ [1 + L_{1,\mathrm{N}} \sum_{\sigma'} (x_{M,\sigma'} - x_{L,\sigma'})] \\
\Gamma_{\mathrm{X},\sigma} &\cong \gamma_{2,\mathrm{X}}\ L_{2,\mathrm{X}}\ u_2\ /\ (1 + L_{2,\mathrm{X}} \sum_{\sigma'} x_{L,\sigma'}) \\
\Gamma_{\mathrm{Y},\sigma} &\cong \gamma_{3,\mathrm{Y}}\ L_{3,\mathrm{Y}}\ u_3\ /\ (1 + L_{3,\mathrm{Y}} \sum_{\sigma'} x_{L,\sigma'})
\end{aligned} \tag{4.15}$$

Then, the fragmentation rate functions $R_{\nu,\sigma}$ and $\bar{R}_\sigma$ are given in terms of :

$$\begin{aligned} R_{N,\sigma} \equiv C_\sigma V_{N,\sigma} &= \Gamma_{\mathrm{N},\sigma}\big[-3x_{N,\sigma} + (-\ell_{\mathrm{S}}^2 + 7\ell_{\mathrm{S}} - 12)\; x_{L,\sigma}\big] \\ &\quad -\Gamma_{\mathrm{X},\sigma}\; k_{\mathrm{X}}\; x_{L,\sigma} - \Gamma_{\mathrm{Y},\sigma}\; k_{\mathrm{Y}}\; x_{L,\sigma} \\ &\quad +\Gamma_{\mathrm{X},\sigma}\, x_{L,\sigma} \sum_{\ell=\ell_{\mathrm{S}}}^{\ell_{\mathrm{S}}+k_{\mathrm{X}}-1} (3 + k_{\mathrm{X}} - \ell)\; P_\sigma(\ell) \\ &\quad +\Gamma_{\mathrm{Y},\sigma}\, x_{L,\sigma} \sum_{\ell=\ell_{\mathrm{S}}}^{\ell_{\mathrm{S}}+k_{\mathrm{Y}}-1} (3 + k_{\mathrm{Y}} - \ell)\; P_\sigma(\ell) \qquad (4.16) \\ R_{L,\sigma} \equiv C_\sigma V_{L,\sigma} &= \Gamma_{\mathrm{N},\sigma}\big[x_{N,\sigma} - (2\ell_{\mathrm{S}} - 4)\; x_{L,\sigma}\big] \\ &\quad -\Gamma_{\mathrm{X},\sigma}\, x_{L,\sigma} \sum_{\ell=\ell_{\mathrm{S}}}^{\ell_{\mathrm{S}}+k_{\mathrm{X}}-1} P_\sigma(\ell) \\ &\quad -\Gamma_{\mathrm{Y},\sigma}\, x_{L,\sigma} \sum_{\ell=\ell_{\mathrm{S}}}^{\ell_{\mathrm{S}}+k_{\mathrm{Y}}-1} P_\sigma(\ell) \qquad (4.17) \\ \bar{R}_\sigma \equiv C_\sigma \bar{V}_\sigma &= R_{N,\sigma} \;+\; 3R_{L,\sigma} \qquad (4.18) \end{aligned}$$

Here, the required $P_\sigma(\ell)$ can be calculated from $G_\sigma(\ell) = x_{L,\sigma} P_\sigma(\ell)$, via Eqs. (4.10) for all short $G_\sigma(\ell)$ chains, $\ell = \ell_{\mathrm{S}}...\ell_{\mathrm{D}}$. For longer chains, we can estimate through LP extrapolation approximation:

$$P_\sigma(j) \;= P_\sigma(\ell_{\mathrm{D}})\; [P(\ell_{\mathrm{D}})/P(\ell_{\mathrm{D}}-1)]^{j-\ell_{\mathrm{D}}} \text{ for } j = \ell_{\mathrm{D}}+1, ...\ell_{\mathrm{D}}+\ell_{\mathrm{E}} \qquad (4.19)$$

with $\ell_{\mathrm{E}} = \max(k_{\mathrm{X}}, k_{\mathrm{Y}})$. For cellohydrobiolases, we must use $k_{\mathrm{X}} = k_{\mathrm{Y}} = 2$ for the *exo*-cutting sites.

Hence $\ell_{\mathrm{L}} = \ell_{\mathrm{R}} = k_{\mathrm{X}} + 1 = 3$, $\ell_{\mathrm{C}} = \ell_{\mathrm{S}} + k_{\mathrm{X}} - 1 = \ell_{\mathrm{S}} + 1$, and $\ell_{\mathrm{E}} = 2$. By choosing $\ell_{\mathrm{D}} = \ell_{\mathrm{C}}$, we then have only two short-chain molar concentration variables to solve for in Eq. (4.10): $G_\sigma(\ell_{\mathrm{S}}) = x_{L,\sigma}\, P_\sigma(\ell_{\mathrm{S}})$ and $G_\sigma(\ell_{\mathrm{S}}+1) = x_{L,\sigma}\, P_\sigma(\ell_{\mathrm{S}}+1)$, for each SAC class-$\sigma$.

The dynamical variables in the LP site number formalism for our five-site model are subject to the following initial conditions

$$x_{L,\sigma}(t^{(\mathrm{o})}) \;=\; g_{L,\sigma}(\lambda_\sigma^{(\mathrm{o})})\; x_{M,\sigma}(\lambda_\sigma^{(\mathrm{o})}) \tag{4.20}$$

$$x_{N,\sigma}(t^{(\mathrm{o})}) \;= x_{M,\sigma}(\lambda_\sigma^{(\mathrm{o})}) - 3x_{L,\sigma}(t^{(\mathrm{o})}) \tag{4.21}$$

$$\lambda_\sigma(t^{(\mathrm{o})}) \;\equiv\; \lambda_\sigma^{(\mathrm{o})}\;. \tag{4.22}$$

$$G_\sigma(\ell, t^{(\mathrm{o})}) = 10^{-20}\, x_{L,\sigma}(t^{(\mathrm{o})}) \;\text{ for } \ell_\mathrm{S} \le \ell \le \ell_\mathrm{D} \tag{4.23}$$

Under Long Chain Limit (LCL), we assumed vast majority of chains are initially of length much greater than ℓ_C, and thus native substrate should not contain any short chains, with length $\ell \le \ell_\mathrm{C}$. Therefore, the short-chain rate equations (4.10) do not contain any surface exposure term, but only fragmentation contributions. We also assume λ-independent native site type-μ fraction $g_{\nu,\sigma}$, *i.e.*, $g_{\nu,\sigma}(\lambda_\sigma) \equiv g_{\nu,\sigma}(\lambda_\sigma^{(\mathrm{o})})$ and

$$g_{L,\sigma}(\lambda_\sigma^{(\mathrm{o})}) = 1/DP^{(\mathrm{o})} \tag{4.24}$$

where $DP^{(\mathrm{o})}$ is the initial degree of polymerization and can be obtained from experiments.

From above equations, it can be seen that

$$\begin{aligned} \dot{x}_{\mathrm{V},\sigma} &= -\dot{x}_{\mathrm{S},\sigma} = \bar{R}_\sigma \\ \dot{x}_{\mathrm{M},\sigma} &= \bar{R}_\sigma\, [1 - \eta_\sigma(\lambda_\sigma)] \end{aligned} \tag{4.25}$$

Here, $x_{\mathrm{S},\sigma} \equiv \sum_{k=1}^{\ell_\mathrm{S}-1} x_{\mathrm{S},\sigma}(k)$ is the total number of dissolved G_1-monomers in soluble G_k oligomers, generated from class-σ SACs. $-\bar{R}_\sigma > 0$ is the molar rate of total insoluble monomer loss from the substrate, and $-\bar{R}_\sigma(1-\eta_\sigma)$ is net the molar rate of monomer depletion at the substrate's surfaces, for class-σ SACs, considering surface curvature effect. The difference $\dot{x}_{\mathrm{M},\sigma} - \dot{x}_{\mathrm{V},\sigma} = -\bar{R}_\sigma \eta_\sigma$ is the rate at which new monomers are being exposed at the SAC surface due to hydrolytic ablation of chains from the outermost SAC layer.

We also calculate the concentrations of soluble glucan oligomers, as these are the desired products of the cellulose enzymatic hydrolysis. The molar production rate of soluble oligomers

G_k, having chain lengths $k < \ell_{\mathrm{S}}$, can be written, according to I, as:

$$\begin{aligned}\dot{x}_{\mathrm{S}}(k) = \quad & k \sum_{\sigma} \Big[\left(2\Gamma_{\mathrm{N},\sigma} + \delta_{k,k_{\mathrm{X}}}\ \Gamma_{\mathrm{X},\sigma} + \delta_{k,k_{\mathrm{Y}}}\ \Gamma_{\mathrm{Y},\sigma}\right)\ x_{L,\sigma} \\ & + \Gamma_{\mathrm{X},\sigma}\, x_{L,\sigma}\, P_{\sigma}(k + k_{\mathrm{X}}) + \Gamma_{\mathrm{Y},\sigma}\, x_{L,\sigma}\, P_{\sigma}(k + k_{\mathrm{Y}}) \Big] \end{aligned} \tag{4.26}$$

with $P_{\sigma}(\ell) \equiv 0$ for all $\ell < \ell_{\mathrm{S}}$. Here, $x_{\mathrm{S}}(k)/k$ is the concentration of dissolved oligomers G_k. Hence, $x_{\mathrm{S}}(k)$ itself denotes the corresponding concentration of G_1 monomers contained in such G_k oligomers. The initial values of $x_{\mathrm{S}}(k)$ for the integration of Eq. (4.26) are set to zero.

4.2 Five-Site Ablation Model in Chain Number Formalism

The chain number formalism, by its definition, consists of a huge set of coupled rate equations that governs chain length from ℓ_{S} to $\ell_{\max}$. $\ell_{\max}$ denotes a pre-determined cut-off for maximum chain length exist in substrate, and conveniently consider $G_{\sigma}(\ell) = 0$ for all chains with $\ell > \ell_{\max}$. Following earlier discussions regarding chain hydrolytic dynamics, each rate equation for an arbitrary $G_{\sigma})(\ell)$ should be constructed from enzymatic fragmentation term and surface exposure term. This can be expressed in the most general form under the five-site model as:

$$\dot{G}_{\sigma}(\ell) = \dot{G}_{\sigma}^{(\mathrm{fra})}(\ell) - \bar{R}_{\sigma}\eta_{\sigma}(\lambda_{\sigma})Q_{\sigma}(\ell, \lambda_{\sigma}) / \sum_{j=\ell_{\mathrm{S}}}^{\infty} jQ_{\sigma}(j, \lambda_{\sigma}) \tag{4.27}$$

where the chain fragmentation term $\dot{G}_{\sigma}^{(\mathrm{fra})}(\ell)$ here is exactly the same as *r.h.s.* of Eq. (4.10), but applied for all chain lengths $\ell \geq \ell_{\mathrm{S}}$, and $Q_{\sigma}(\ell, \lambda)$ is the native chain length probability distribution in layer λ, which should be fed into our calculation as a model input. the chain length probabilities $P_{\sigma}(\ell)$ are given by $P_{\sigma}(\ell) = G_{\sigma}(\ell)/x_{L,\sigma}$ for all ℓ. For the purpose of direct comparisons between two formalisms, the site number concentrations expressed in terms of chain number formalism are:

$$x_{L,\sigma} = \sum_{\ell=\ell_{\mathrm{S}}}^{\ell_{\max}} G_{\sigma}(\ell) \tag{4.28}$$

$$x_{N,\sigma} = \sum_{\ell=\ell_{\mathrm{S}}}^{\ell_{\max}} (\ell - 3) G_{\sigma}(\ell) \tag{4.29}$$

So far, Eqs. (4.27) and (4.9) then automatically constitute a closed equation system without further approximation.

Like its counterpart, chain number formalism also takes several initial conditions. Apart from directly sharing Eq. (4.21) with site number formalism, the complete set of $G_\sigma(\ell)$ is initialized by native chain length distribution $Q_\sigma(\ell, \lambda_\sigma)$ by

$$G_\sigma(\ell, t^{(o)}) = x_{\mathrm{L},\sigma}(t^{(o)}) Q_\sigma(\ell, \lambda_\sigma^{(o)}) \ . \tag{4.30}$$

We consider three different λ-independent distribution shapes, for the native chain length probability distribution Q_σ: a Delta shape, a (truncated) Gaussian shape and a Global Poisson (GP) shape where

$$Q_\sigma^{(\mathrm{Gauss})}(\ell) = N_{\mathrm{Q},\sigma}\ \Theta(4\ell_{\mathrm{Wid}} - |\ell - \ell_{\mathrm{Avg}}|)\ \exp\big[-(\ell - \ell_{\mathrm{Avg}})^2/2\ell_{\mathrm{Wid}}^2\big] \tag{4.31}$$

with $\Theta(...)$ denoting the step function; $N_{\mathrm{Q},\sigma}$ being determined by the normalization of $Q_\sigma(\ell)$; and $4\ell_{\mathrm{Wid}} < \ell_{\mathrm{Avg}} - \ell_{\mathrm{D}}$ so that $Q_\sigma(\ell) \equiv 0$ for all $\ell \leq \ell_{\mathrm{D}}$, as assumed in the short-chain rate equations. The substrate's native (=initial) degree of polymerization within SACs of class σ then is $DP_\sigma^{(\mathrm{o})} = \ell_{\mathrm{Avg}}$. The Delta shape can be viewed as the limiting case of zero width Gaussian, *i.e.* $\ell_{\mathrm{Wid}} \to 0^+$ with integer ℓ_{Avg}, thus $Q_\sigma^{(\mathrm{Delta})}(\ell) = \delta_{\ell,\ell_{\mathrm{Avg}}}$.

For testing purposes, we will consider also the case of a "global Poisson" (GP) distribution

$$Q_\sigma^{(\mathrm{GP})}(\ell) = N_{\mathrm{Q},\sigma}\ \Theta(\ell - \ell_{\mathrm{S}} + 1/2)\ \exp\big[-s_\sigma^{(\mathrm{o})}(\ell - \ell_{\mathrm{S}})\big] \tag{4.32}$$

where $N_{\mathrm{Q},\sigma} = 1 - \exp(-s_\sigma^{(\mathrm{o})})$ and $s_\sigma^{(\mathrm{o})}$ is determined by the initial average chain length $DP_\sigma^{(\mathrm{o})}$, *via*

$$DP^{(\mathrm{o})} = \ell_{\mathrm{S}} - 1 + \frac{1}{1 - \exp(-s_\sigma^{(\mathrm{o})})} \tag{4.33}$$

We are able to show that the Global Poisson shape is exactly preserved in the full chain number formalism without surface exposure term. In this case, the local Poisson closure extrapolation in our site number formalism becomes *exactly equivalent* to the solution of the corresponding full chain number formalism with a GP-shaped initial chain length distribution, provided that the initial condition Eq. (4.23) for the short chain lengths $\ell = \ell_{\mathrm{S}}$, ... ℓ_{C} is replaced by Eq. (4.30) with Q_σ from Eq. (4.32).

4.3 Comparisons Between Two Formalisms

Up until now, we have studied Five-Site models under two individual formalisms: site number formalism and chain number formalism.

Before we dive into our two formalisms, it is worthwhile to mention what Zhang and Lynd have achieved in their model, which was essentially build upon chain number variables. They assumed that *all* chains are readily exposed at the enzyme-accessible surfaces for hydrolysis. In order to correct this "over-exposure" of substrate, they introduced $\bar{F}_{\mathrm{a}}^{(\mathrm{o})}$ factor, which denotes the fraction of total substrate that are effectively exposed to enzyme attacks. Limitations behind this correction are due to $\bar{F}_{\mathrm{a}}^{(\mathrm{o})}$ being a time independent variable, and only provides a rough averaged estimation of the accessibility fraction over complete hydrolysis. However under their time scale of study, which constrains themselves strictly to early stage fragmentation, this accessibility fraction can indeed be constant.

For chain number formalism, it is conceptually similar to the model Zhang and Lynd proposed in their paper (Zhang and Lynd, 2006), augmented by surface exposure term which captures the cellulose morphology evolution as well as hydrolysis kinematics. Unlike site number formalism, chain number formalism does not require either short chain cut-off or Local Poisson approximation, which renders this formalism an exact solution to hydrolytic process. This approach also highlights a full evolution profile of complete chain number $G_\sigma(\ell)$, up to some cut-off $\ell_{\max}$, at the price of solving for $\ell_{\max}$ coupled rate equations for $G_\sigma(\ell)$ *and* one for λ_σ via Eq. (4.9). In total, there are $\ell_{\max}+1$ ODE equations for every SAC type-σ, which in realistic could be to the order of hundreds, if not thousands.

By contrast, site number formalism provides an alternative that circumvents such problem. Instead of solving for all chain number, site number formalism demands only a handful of short chain equations, and any other longer chain number is readily approximated under Long-Chain-Limit and Local Poisson approximation. Therefore, the set of independent variables is independent of system size, *i.e.* $\ell_{\max}$, per SAC class-σ, and promises a great potential in its application over large size systems. However, in exchange for this advantage,

one loses detailed information regarding the evolution of chains other than exactly solved short chains, as longer chain contributions have been effectively absorbed into site number variables $x_{\nu,\sigma}$ and can only be *estimated*.

From the foregoing it is easy to see that, in the case of a purely *endo*-acting (EG1) enzyme system, the LP closure approach of the site number formalism is exactly equivalent to the chain number formalism without any approximation. Specifically, if $u_2 = 0$ and $u_3 = 0$, we get $\Gamma_{\mathrm{X},\sigma} = \Gamma_{\mathrm{Y},\sigma} = 0$: the site number rate functions $R_{N,\sigma}$, $R_{L,\sigma}$ and $\bar{R}_\sigma$ can then be evaluated as functions of only the $x_{L,\sigma}$ and $x_{N,\sigma}$, *i.e.*, $R_{N,\sigma}$, $R_{L,\sigma}$ and $\bar{R}_\sigma$ become independent of the short-chain concentrations $G_\sigma(\ell)$. Hence, no short-chain approximation is required in the site number model. Note that this is true even beyond the "Single-layer Single-geometry" (SS) model, that is, in the full surface ablation model with $\eta_\sigma(\lambda) \neq 0$.

4.4 Model Parameterization

We are going to list all simulation parameters used in numerical calculations for our model. We test our model over a pure cellulose substrate (post pre-hydrolysis state where all non-cellulosic substances are disposed of, *i.e.* Avicel) interacting with a system of non-complex cellulase. Their kinetics and concentration parameters are listed in Table 1 from Zhang and Lynd (2006) unless otherwise indicated. We choose the non-complex enzyme system derived from *Trichoderma reesei*, and kept their natural composition ratio among each other. Notice in Table 1, we only listed three enzymes, EG1 (*endo*-), CBH2 (*exo*-X/non-reducing end), CBH1 (*exo*-Y/reducing end) with a total concentration of 27.6 mg/L. Their concentration ratio is 12% EG1, 60% CBH1 and 20% CBH2. The remaining 8% is consisted of other glycoside hydrolases, which are trivial with regards to the purpose of our simulation and hence neglected. Furthermore, we assign $k_{\mathrm{X}} = k_{\mathrm{Y}} = 2$, for the chain-end cutting sites of the cellohydrobiolases CBH2 and CBH1, respectively. We also assume the minimum insoluble chain length $\ell_{\mathrm{S}} = 7 > k_{\mathrm{X}} + k_{\mathrm{Y}} = 4$, thus eliminating Z-sites. With an additional assumption of $\phi_{O,\sigma} = 0$, we arrive at a *pure* five-site cellulose substrate model without site types Z and

Table 4.1: Simulation parameters

Parameter	Unit / Value	Remarks
M_1	55,000 (g/mol)	Molar mass of EG1
M_3	65,000 (g/mol)	Molar mass of CBH1
M_2	58,000 (g/mol)	Molar mass of CBH2
M_{G_1}	162 (g/mol)	Molar mass of anhydroglucose G_1 ($C_6H_{10}O_5$)
$\gamma_{1,N}/M_1$	$0.40(\mu\text{mol bonds/mg} \cdot \text{min})$	Specific enzyme activity (by mass) of EG1 on N, X and Y sites
$\gamma_{3,Y}/M_3$	$0.08(\mu\text{mol bonds/mg} \cdot \text{min})$	Specific enzyme activity (by mass) of CBH1 on Y sites
$\gamma_{2,X}/M_2$	$0.16(\mu\text{mol bonds/mg} \cdot \text{min})$	Specific enzyme activity (by mass) of CBH2 on X sites
$L_{1,N}$	3.0(Liter/mmol)	Adsorption equilibrium coefficient of EG1 to N, X and Y sites
$L_{3,Y}$	4.0(Liter/mmol)	Adsorption equilibrium coefficient of CBH1 to Y sites
$L_{2,X}$	4.0(Liter/mmol)	Adsorption equilibrium coefficient of CBH2 to X sites
M_1u_1	0.0036(g/Liter)	Concentration (by mass) of EG1
M_3u_3	0.0180(g/Liter)	Concentration (by mass) of CBH1
M_2u_2	0.0060(g/Liter)	Concentration (by mass) of CBH2
$M_{G_1}x_V^{(o)}$	10.0(g/Liter)	Concentration (by mass) anhydroglucose G_1 in solid
ℓ_{S}	7	Minimum length ℓ of insoluble chains G_ℓ
k_{X},k_{Y}	2	L-end and R-end exo-cutting lengths (producing cellobiose)

O. As a result of this simplification, we have successfully unified two chain site distribution models, HDC and CCE, and both reduced to the same five-site model that we have already discussed above.

In addition, we use the experimentally observed initial (pre-hydrolysis) values for the degree of polymerizaton, $DP^{(0)} = 300$, and for the enzymatic surface accessibility fraction, $\bar{F}_{\rm a}^{(\rm o)} = 0.006$, from a typical pure cellulosic substrate such as Avicel, and also a realistic value for the initial substrate monomer concentration, $x_{\rm V}^{(\rm o)} \equiv x_{\rm V}(\lambda_\sigma^{(\rm o)}) = 61.73$mM as given in Table 1. $DP^{(0)}$ then in turn determines $g_{L,\sigma}(\lambda_\sigma^{(\rm o)})$ *via* Eq. (4.24).

$\bar{F}_{\rm a}^{(\rm o)}$ can be used to constrain the initial SAC sizes $\lambda_\sigma^{(\rm o)}$ or their corresponding molar fractions $\xi_\sigma^{(\rm o)}$ by:

$$\bar{F}_{\rm a} = \sum_\sigma \xi_\sigma \, F_{{\rm a},\sigma} = \sum_\sigma \xi_\sigma \Big(1 - \Theta(\lambda_\sigma - 1) \Big(1 - \frac{1}{\lambda_\sigma} \Big)^{d_{{\rm A},\sigma}} \Big) \tag{4.34}$$

As introduced in previous sections, the molar fraction of substrate monomers residing in SACs of geometry class-σ can be expressed as $\xi_\sigma \equiv x_{{\rm V},\sigma}/x_{\rm V}$ and likewise the corresponding partial surface accessibility fractions as $F_{{\rm a},\sigma} \equiv x_{{\rm M},\sigma}/x_{{\rm V},\sigma}$, using Eqs. (4.4), (4.5) and (4.6). Of course, as opposed to Zhang and Lynd model, ξ_σ and $\bar{F}_{\rm a}$ are time-dependent in our surface layer ablation models and Eq. (4.34) applies both for the initial ($t^{(\rm o)}$) values and at all later times $t > t^{(\rm o)}$. It will therefore also be used to calculate the time evolution of $\bar{F}_{\rm a}$.

Lastly, given $\xi_\sigma^{(\rm o)}$, $\lambda_\sigma^{(\rm o)}$, and the ablation dimensions $d_{{\rm A},\sigma}$, the initial substrate concentration $x_{\rm V}^{(\rm o)}$ determines the time-independent volume prefactors $B_{{\rm V},\sigma}$ entering into Eq. (4.4) *via*

$$B_{{\rm V},\sigma} = x_{\rm V} \xi_\sigma / \lambda_\sigma^{d_{{\rm A},\sigma}} \tag{4.35}$$

This is the basic approach we have taken to parameterize the "Multiple-Layer, Single-Geometry" (MS) and "Multiple-Layer, Multiple-Geometry" (MM) models.

For the MS model with $M_{\rm MD} = 1$, all substrate masses are concentrated within one geometry, rendering $\xi_\sigma \equiv \xi = 1.0$, so that the prescribed $\bar{F}_{\rm a}^{(\rm o)} = 0.006$ determines the initial

λ_σ-value to $\lambda_\sigma^{(o)} = 333$ via Eq. (4.34). The total initial substrate monomer concentration $x_V^{(o)}$ from Table 1 then determines the prefactor $B_{V,\sigma}$ by Eq. (4.35).

The two MM models, MM82-1 and MM82-2, approximate continuous distributions of substrate morphologies by way of a population of $M_{MD} = 82$ geometry classes, representing SAC geometries with 82 equidistantly spaced initial λ_σ-values, *i.e.* ($\lambda_{\sigma=1}^{(o)} = 20$, $\lambda_{\sigma=2}^{(o)} = 40$, ..., $\lambda_{\sigma=82}^{(o)} = 1640$). Model MM82-1 assumes a uniform distribution of the molar monomer concentration per geometry class, *i.e.*, $\xi_\sigma^{(o)} = 1/82$ for all σ; while MM82-2 assumes a Gaussian distribution:

$$\xi_\sigma^{(o)} = \exp(-(\frac{\lambda_\sigma - 458}{200})^2/2) \; / \sum_{\sigma'} \exp(-(\frac{\lambda_{\sigma'} - 458}{200})^2/2) \tag{4.36}$$

These values are chosen so that again $\bar{F}_a^{(o)} = 0.006$ is obtained in both models MM82-1 and MM82-2. The total initial substrate monomer concentration $x_V^{(o)}$ from Table 1 then again determines the prefactors $B_{V,\sigma}$ *via* Eq. (4.35).

For the discussion of the simulation results obtained with these models, it is important to note that the MS model described above also represents the "zero-width" limit of the Gaussian MM model. Likewise, the uniform distribution model MM82-1 represents the "infinite-width" limit of the Gaussian MM model.

Model SS, as already described above, becomes equivalent to the Zhang-Lynd chain fragmentation model in the low-enzyme limit. Since this model has no morphology and treats all chains in the substrate as being immediately fully accessible, Zhang and Lynd (2006) corrected for partial accessibility by reducing the total chain bonds available for EG1 adsorption by a time-independent *ad hoc* factor, set equal to $\bar{F}_a^{(o)}$. Effectively, their approach amounts to replacing $L_{1,N}$ by $\bar{F}_a^{(o)} L_{1,N}$ in Eq. (4.15). For purposes of comparing the SS model to our morphology-based MS and MM models, we adopt the same approach. Our simulation results of the SS model are therefore, as expected, in excellent numerical agreement with those of Zhang and Lynd (2006). We should emphasize here, and will demonstrate below, that the SS model can really not be used to model the entire hydrolytic conversion process, but only the very early stages of it. Zhang and Lynd (2006) indeed limited its application

to study the short-time behavior, for a duration of only 60min. We are presenting SS results for full hydrolytic conversion here solely for the purpose of comparison and contrast with the morphology-based surface ablation models.

For all four models, we have used the kinetics parameters given in Table 1, but with two different sets of mixed enzyme concentrations: enzyme set "E1" consists of the enzyme concentrations as given in Table 1; the other, "E200", has concentrations 200 times the values of E1, for all three enzyme types. For the assumed $\bar{F}_{\mathrm{a}}^{(\mathrm{o})}$ and total initial substrate monomer concentration $x_{\mathrm{V}}^{(\mathrm{o})}$, the E200 enzyme concentrations are close to the high-enzyme limit, whereas E1, as noted before is well within the low-enzyme limit. For all simulations with enzyme set E200, we have therefore used and solved the full coupled non-linear enzyme adsorption equilibrium equation system, to obtain the free enzyme and free substrate site concentrations v_κ and $y_{\nu,\sigma}$ from the corresponding total concentrations u_κ and $x_{\nu,\sigma}$.

The enzyme footprint β_κ entering into these adsorption equilibrium equations has been quantified experimentally in terms of a parameter α, denoting the total number of dimer (G_2) units covered up by the adsorbed enzyme molecule (Zhang and Lynd, 2004). Our β_κ-parameter is related to the α-parameter by $\beta_\kappa = 2\alpha_\kappa - 1$, since β_κ denotes the number of collateral surface bond sites covered up by enzyme κ in addition to the adsorbing site. With an estimated $\alpha_\kappa \sim 15 - 40$ dimer units for a typical *endo*-glucanases (Zhang and Lynd, 2004), we get $\beta_\kappa \sim 30 - 80$ collateral surface bonds. Lacking more detailed experimental information, we will assume the same β_κ-value for the two *exo*-glucanases in our model and use $\alpha_\kappa = 20$, and hence $\beta_\kappa = 39$ for all three enzymes, $\kappa = 1, 2, 3$, in solving the enzyme adsorption equilibrium equations. We also assume β_κ to be the same for all geometry classes σ.

Also, in the three surface ablation models we have assumed substrate morphologies with ablation dimension $d_{\mathrm{A},\sigma} \equiv d_{\mathrm{A}} = 2$ for all σ. The ablation dimension $d_{\mathrm{A}} = 2$ is representative of a substrate where the glucan chains within the SAC, or the microfibril (Himmel et al. 2007) remnants comprising the chains, are orientationally disordered, but do exhibit directional

order, as illustrated and explained in Fig. 2.3. This means that most chains or microfibril remnants are aligned with their chain direction approximately parallel to a common axis, while being rotated at random angles around that axis. This is likely applicable for substrates, like Avicel, where a fibrous alignment structure is maintained up to typical SAC-size length scales, but without any orientational ordering on those length scales.

It is worth pointing out that the parameterization requirements for the surface ablation models MS and MM introduced above do not significantly exceed those of the pure chain fragmentation single-layer model introduced in (Zhang and Lynd, 2006). More realistic morphology modeling efforts will of course require further parameterization, if supported by more detailed microscopic experimental data for the meso-scale structure of real cellulosic substrates.

For sufficiently dense morphology grids, multi-geometry surface ablation models such as MM82-1 and MM82-2 approximate continuous morphology distributions. It is then informative to also study the hydrolytic evolution of certain continuous, weighted SAC size (λ_σ-) density distributions. For SACs having sizes within some infinitesimal interval $[\lambda - d\lambda/2, \lambda + d\lambda/2]$, let $Dx_{\mathrm{V}}(\lambda,t)d\lambda$ and $Px_{\mathrm{M}}(\lambda,t)d\lambda$ denote the molar concentration of all substrate monomers contained in these SACs' volumes, and the molar fraction of all surface monomers residing on these SACs' surfaces, respectively. From the time-evolving, discrete SAC geometry populations in the MM models, we can construct these continuous density distributions $Dx_{\mathrm{V}}(\lambda,t)$ and $Px_{\mathrm{M}}(\lambda,t)$ as follows: for λ-values falling *on* the discrete $\lambda_\sigma(t)$-grid at time t, we set

$$Dx_{\mathrm{V}}(\lambda,t) = \frac{x_{\mathrm{V},\sigma}(\lambda_\sigma(t))}{\Delta\lambda_\sigma(t)}, \quad Px_{\mathrm{M}}(\lambda,t) = \frac{x_{\mathrm{M},\sigma}(\lambda_\sigma(t))}{x_{\mathrm{M}}(t)\,\Delta\lambda_\sigma(t)} \ \text{ if } \lambda = \lambda_\sigma(t) \tag{4.37}$$

where

$$\Delta\lambda_\sigma(t) = \begin{cases} \lambda_2(t) - \lambda_1(t) & \text{for } \sigma = 1\ ; \\ (\lambda_{\sigma+1}(t) - \lambda_{\sigma-1}(t))/2 & \text{for } \sigma = 2, ..., 81\ ; \\ \lambda_{82}(t) - \lambda_{81}(t) & \text{for } \sigma = 82\ . \end{cases} \tag{4.38}$$

For λ-values *off* the discrete $\lambda_\sigma(t)$-grid, but within the "cut-off interval" $[\lambda_1(t) - \Delta\lambda_1(t)/2$, $\lambda_{82}(t) + \Delta\lambda_{82}(t)/2]$, $Dx_{\rm V}(\lambda, t)$ and $Px_{\rm M}(\lambda, t)$ are then defined by linear interpolation or extrapolation from the nearest grid points. Outside of the cut-off interval, we set $Dx_{\rm V}(\lambda, t) = 0$ and $Px_{\rm M}(\lambda, t) = 0$. The density distribution functions then obey the normalization conditions

$$\int_0^\infty d\lambda\ Dx_{\rm V}(\lambda, t) = x_{\rm V}(t), \quad \int_0^\infty d\lambda\ Px_{\rm M}(\lambda, t) = 1\ . \tag{4.39}$$

Chapter 5

Numerical Results and Analysis [1]

5.1 Testing the LP Approximation

First, we are going to test the accuracy of LP approximation in site number formalism against the simulation results from the corresponding full chain number formalism. We primarily want to focus on the case where no-morphology ($\eta_\sigma \equiv 0$) is considered, and thus all glucose chains are fully exposed on the surface, or equivalently, all SACs contain only one single layer ($\sigma = 1 \equiv M_{\mathrm{MD}}$). This is, as previously mentioned, referred to as "Single-layer, Single-geometry" (SS) model and maybe formally regarded, for the purpose of unifying all into one model, as the infinite-dimensional limit, $d_{\mathrm{A}} \to \infty$, of the morphological model ($\eta_\sigma \neq 0$) models. Consequently, $x_{\mathrm{M}} = x_{\mathrm{V}}$, and thus the overall accessibility fraction $\bar{F}_{\mathrm{a}} = x_{\mathrm{M}}/x_{\mathrm{V}} \equiv 1$, the $x_{L,\sigma}$- and $x_{N,\sigma}$-rate equations in the SS model become decoupled from λ_σ, and we can ignore the λ_σ-rate equation altogether. In the site number formalism, we then solve the coupled rate equations Eqs. (4.7) and (4.8) ; and in the corresponding chain number formalism, we solve the rate equation system Eq. (4.27) for *all* chain lengths ℓ, with $\eta_\sigma \equiv 0$ in both. Two enzyme systems, a mixed *endo*-*exo* EG1-CBH1,2 enzyme system with naturally occurring enzyme composition and a purely *exo*-acting CBH1,2 enzyme system, are used in the simulations. As explained in section 4.3, LP approximation in pure *endo*-acting enzyme system becomes exact and does not need further numerical testing. In the chain number formalism, the initial chain length distribution has to be provided as a model input. We consider three

[1] Sec 5.1 - Sec 5.5 contain contents published in Zhou W, Hao Z, Xu Y, Schüttler HB. 2009. Cellulose hydrolysis in evolving substrate morphologies II. Numerical results and analysis. *Biotechnol Bioeng* **104**:275-289. Permission acquired from *Biotechnol Bioeng* journal.

aforementioned initial distributions in section(4.2): a Delta shape, a (truncated) Gaussian shape and a Global Poisson (GP) shape.

5.1.1 Mixed EG1-CBH1,2 Enzyme System

Fig. 5.1 shows the results from model calculations for the full chain formalism from Zhang-Lynd model and our corresponding site number formalism with LP approximation, for a mixture of EG1, CBH1 and CBH2, under different initial DP values. Notice calculations are carried out in SS model, assuming all materials are exposed at the surface of SACs for enzyme attack. It is evident that the results from site number formalism with LP approximation and exact full chain solution are in excellent agreement with each other for native degree of polymerization (DP^{o}) exceeds 20 monomers. We are also showing results for $DP^{(o)} = 10$. Here the LCL condition, *e.g.*, $DP^{(o)} \gg \ell_C$, is not satisfied, since $\ell_C = 8$. As expected, the deviations between site number LP and chain number formalism become quite noticeable here as hydrolysis progresses. Such deviations drastically shrink to less than 1% and become unnoticeable for larger DP values. Although divergence do exist for low DP values, this result is in consistent with long-chain-limit (LCL) we discussed earlier and note that typical cellulosic substrates possesses DP-values well above 20 monomers (Zhang and Lynd, 2004).

In addition, same level of accuracy is achieved by the LP approximation across all observable quantities that are relevant to hydrolysis, including the total remaining solid substrate monomer concentration $x_V(= x_M)$ in Fig. 5.1A; the soluble oligomeric monomer concentrations $x_S(k,t)$, shown in Fig. 5.1B for oligomer length $k = 1$ (=glucose) and in Fig. 5.1C for $k = 2$ (=cellobiose); and the total chain (end) concentration $x_L(t)$ shown in Fig. 5.1D. Note that x_L and x_M determine the hydrolytically evolving DP of surface exposed chains by $DP = x_M/x_L$.

We also tested full chain results with three different chain length distribution shapes mentioned in Sec.4.2. Notice that for $DP^{(o)} \geq 60$, all three models almost give identical results. This is again fully consistent with the general discussion of the LCL : as long as the

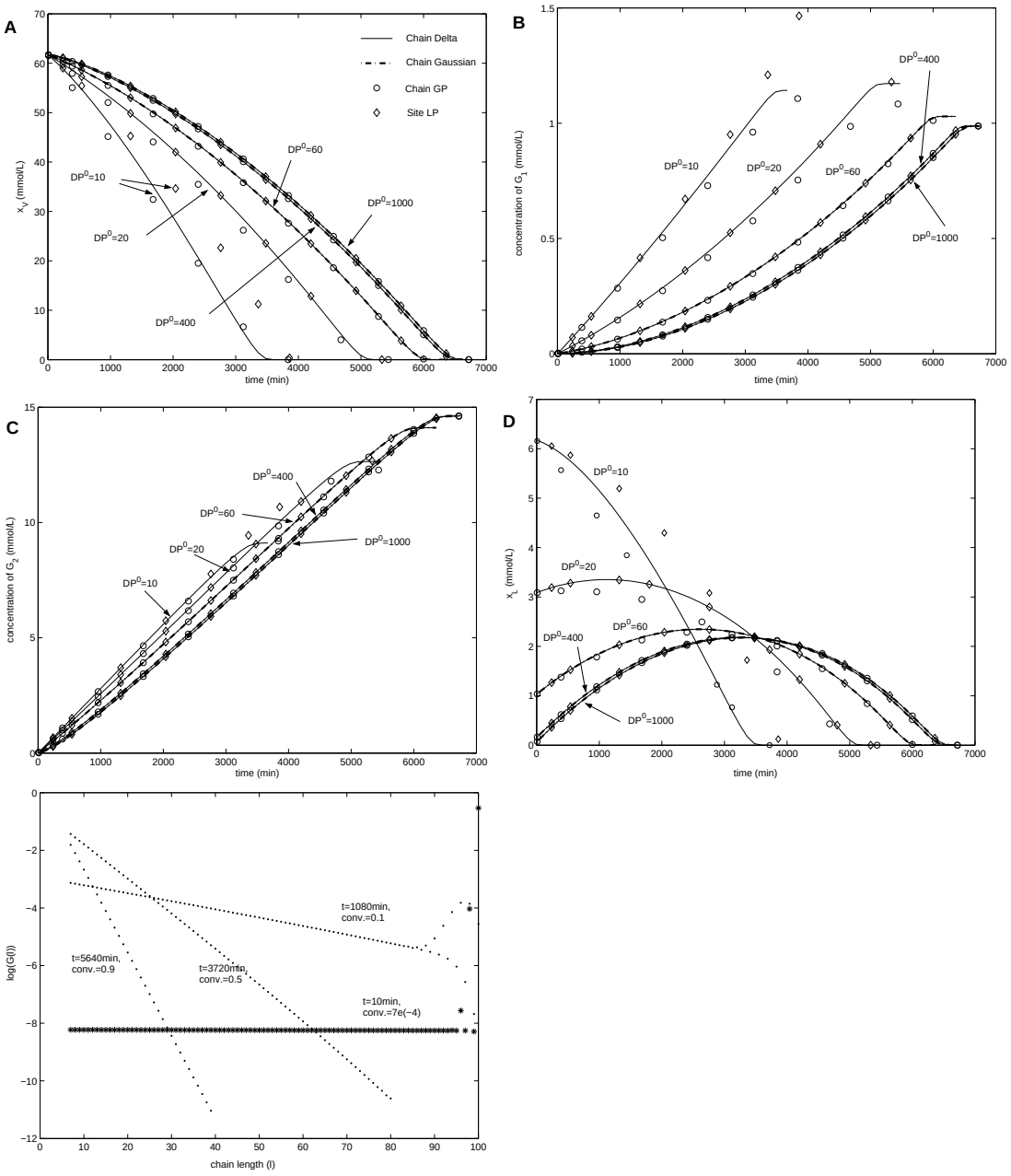

Figure 5.1: Model comparison with mixed enzymes. Comparison of the SS model results from chain number formalism (Zhang-Lynd Model) and site number formalism, with different initial chain length distributions, (as previously discussed, for the case of a GP-shaped initial chain length distribution, the results of site formalism "Site GP" and the chain formalism "Chain GP" are identical.) for the mixed EG1-CBH1,2 enzyme system. In panels A-D, full lines, dot-dashed lines and circles are for the chain number formalism with delta-, Gaussian- and global-Poisson-(GP-) shaped initial chain length distributions; diamonds are for the corresponding local Poisson (LP) approximation in the site number formalism. A: total monomer concentration $x_{\mathrm{V}}(\mathrm{t})$ in solid versus time t; (B) concentration of G_1 in solution, $x_{\mathrm{S}}(1,\mathrm{t})$, versus time t; (C) concentrations of G_2 in solution, $x_{\mathrm{S}}(2,\mathrm{t})/2$, versus time t; (D) concentration of non-reducing chain ends, $x_{\mathrm{L}}(\mathrm{t})$, versus time t; (E) typical log chain length distribution, $\log G(\ell,\mathrm{t})$, versus chain length ℓ at several times t, from SS model chain number formalism with delta-shaped initial distribution from Eq. (4.31) with $\ell_{\mathrm{Wid}} \to 0$ and $\ell_{\mathrm{Avg}} = 100$.

initial chain length distribution satisfies the LCL conditions, the hydrolysis kinetics is very insensitive to the actual initial chain length distribution shape. The only parameter that matters under LCL conditions is the initial average chain length, *i.e.*, the $DP^{(o)}$-value; other details of the distribution shape become essentially irrelevant. Therefore, delta initial chain distribution is highly favored for its simplicity in setting up and smaller $\ell_{\max}$ value compared to other equivalent initial distribution models.

5.1.2 Pure CBH1,2 Enzyme System

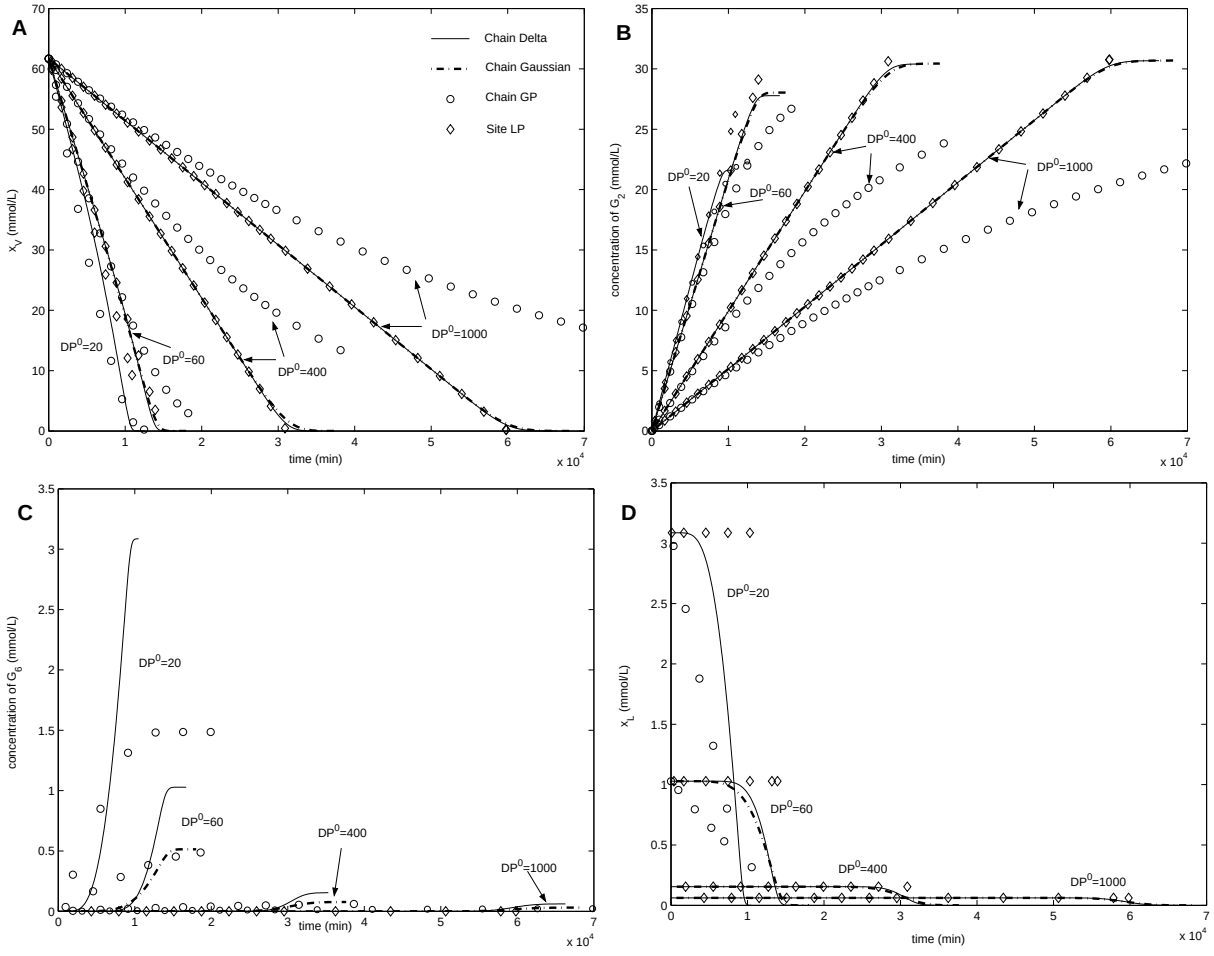

Figure 5.2: Model comparison with pure exo-enzymes. Comparison of the SS model results from chain number formalism (Zhang-Lynd Model) and site number formalism, with different initial chain length distributions, for the pure CBH1,2 enzyme system. In all panels A-D, abbreviations, full lines, dot-dashed lines, circles and diamonds are defined as in Fig.5.1. A and D: see descriptions in Figure.5.1; B: concentration of G_2 in solution, $x_{\mathrm{S}}(2,t)/2$, *vs.* time t; C: concentrations of G_6 in solution, $x_{\mathrm{S}}(6,t)/6$, *vs.* time t.

Fig. 5.2 shows the comparison of SS model results between the site number formalism with LP approximation and the chain number results for pure CBH1,2 enzymes. All model parameters are from Table 1, except that the total EG1 concentration is set to $u_1 = 0$. As seen in Figs. 5.2A-C, respectively, for initial DP-values $DP^{(o)} \geq 60$ the site number LP approximation is again remarkably accurate in reproducing chain number results for total insoluble substrate monomer, $x_V (= x_M)$, and for dissolved monomer concentrations $x_S(k)$ in soluble oligomers.

In the limit of a pure CBH1,2 enzyme system, $u_1 = 0$ and hence $\Gamma_{N,\sigma} = 0$, which in turn sets the first $x_{L,\sigma}$-term in *r.h.s.* of Eq. (4.10) to be 0. This quenches the major source of short-chain production ($\ell \leq \ell_D$), since in long-chain-limit, the majority of short chain concentration comes from *endo*-cuts, and their initial concentration is set to be at a "negligible" amount. Thus, except for cellobiose ($k = 2$), LP approximation gives zero short-chain concentrations $G_\sigma(\ell)$ for the entire duration of hydrolysis process. For cellobiose, it is easily understandable that as CBH1,2 can only cut off oligomers lengths $k_X = k_Y = 2$ from the non-reducing and reducing ends, respectively, cellobiose should comprise the majority of soluble monomers left in the solution. This result is verified in Fig. 5.2B and that (very small) fraction of $k = 5$ and $k = 6$ oligomers is simply approximated by zero (Fig. 5.2C). However, notice also, a small amount of $k = 5$ (cellopentose) and $k = 6$ (cellohexose) oligomers can be produced in the exact full chain calculations, as seen in Fig. 5.2C. However, in terms of the *overall* oligomer distribution, LP is actually a very good approximation to the exact full chain results for delta- or Gaussian-shaped initial distribution for realistic chain lengths $DP^{(0)} \geq 60$, since it reproduces the dominant $k = 2$ oligomer very accurately for realistic $DP^{(o)}$.

In addition, as a result of the LP approach in SS model, it will then give a zero rate $R_{L,\sigma}$ for the production of chains or chain ends from chain fragmentation processes. In the single-layer limit ($\eta_\sigma(\lambda) = 0$), the chain concentration $x_{L,\sigma}$ thus becomes t-independent in the LP approximation. As shown in Fig. 5.2D, this result agrees poorly with the chain number results in the case of short chains with $DP^{(o)} < 60$. However, under LCL conditions, *i.e.* for

larger $DP^{(o)}$-values, the full chain model with delta- or Gaussian initial distribution shape also predicts approximately t-independent x_L for most of the hydrolysis time: in comparing Figs. 5.2A and 5.2D for $DP^{(o)} \geq 60$, we note that the full chain x_L remains approximately constant until about $80-95\%$ of the substrate has been converted, which is then followed by a very quick downturn of x_L. Hence, the rather simple approximate LP result for x_L agrees, for most of the hydrolysis time with the full chain number result.

Both the site LP approximation and the exact full chain results for delta- or Gaussian-shaped initial distribution deviate noticeably in Fig. 5.2 from the exact full chain results for the GP chain length distribution. Compared to the other distributions, the GP distribution shows a slower loss of substrate monomers, as seen in Fig. 5.2A, and, at the same time, a faster loss in the total number of chains in Fig. 5.2D. This can be understood by noting that the GP distribution contains a larger fraction of its monomers in longer chains with $\ell > DP^{(o)}$, but larger fraction of its chains (and chain ends) in shorter chains with $\ell \ll DP^{(o)}$. Recall here that both the site LP approximation and the full chain delta- and Gaussian-shaped initial distributions assume that there are initially no short chains at all, whereas GP assumes that the shortest chains have the largest concentrations right from the start. In the mixed *endo-exo*-acting enzyme system shown in Fig. 5.1, this difference in the initial chain length distribution does not affect the hydrolysis rate significantly, since *endo*-cutting processes are very efficient (see Fig. 5.1E) in quickly producing a large population of short chains, even if short chains are initially absent. However, in the purely *exo*-acting enzyme system, the difference in the initial chain length distribution has a much more pronounced effect on the hydrolysis and chain loss rate, since it takes comparatively a much longer time for *exo*-cuts alone to produce short chains from long ones.

The primary pathology of the LP approximation which manifests itself in Fig. 5.2D is that, for purely *exo*-acting enzyme systems (and only for those!), the LP site number formalism fails to eventually remove the chain ends (x_L) from the substrate (*i.e.* x_L is constant), even after all substrate monomers (x_{M}) have been completely dissolved. The

persistence of these "phantom chains" in the LP approximation would be of no consequence if single-layer substrates are considered, as shown in Fig. 5.2D. However, in the case of a full multi-layer surface ablation model with hydrolytically evolving substrate morphology and a purely *exo*-acting enzyme system, this pathology currently still limits the applicability of the site number formalism. Thus in LP-based site number simulations for surface ablation models purely *exo*-acting enzyme systems, we must restrict ourselves to a short-time limit where only a few SAC surface layers are solubilized so that the accumulation of surface phantom chain ends remains a negligible artifact.

5.1.3 Chain Length Distributions

Fig. 5.1E shows the common logarithm of chain length concentration distribution profile from full chain model result, with mixed EG1-CBH1,2 enzymes, adopting Delta shape initial chain distribution with $DP^{(o)} = 100$, well within the LCL regime we introduced earlier. Time check points are selected across hydrolysis process from 10min all the way up to 5640min, which is equivalent to about 90% conversion. LP approximation predicts, under LCL condition, one should be able to extrapolate, near short chain neighborhood, longer chain length concentration with a linear logarithm relationship. As expected from the LP approximation, we observe almost perfect linear function of ℓ, at least for short chain lengths up to $\ell \leq 80 - 90$, not only at early times, but this relationship remains throughout the whole hydrolysis process. We have also tested such short-chain Poisson behavior under widely varying parameter conditions, including different rate and adsorption coefficients, different initial chain length distributions $Q(\ell)$ and initial DP (within LCL: $DP^{(o)} \gg \ell_S$), and different enzyme concentrations and mixing ratios, and results all justified the legitimacy of LP extrapolation. However, for the case of pure *exo*-acting CBH1,2 enzyme system, what is required in LCL condition, *i.e.* most short chains are quickly generated by *endo* cuts are no longer satisfied. For a pure *exo* system that started off with a non-Poissonian initial distribution shape $Q(\ell)$

will generally *not* evolve towards a Poisson shape, while a global Poissonian initial shape leads to exactly preserved Poisson shape thereafter, for any enzyme composition.

For these reasons, the LP approximation breaks down in purely exo-acting enzyme systems. Another indication of this breakdown is that the LP approximation incorrectly predicts the chain (end) concentration x_{L} to be constant in time for a pure exo-system even though the solid substrate monomer concentration x_{V} is decreasing. So, LP fails to account for the fact that the exo-cutting activity at the chain ends must eventually also lead to the disappearance of each chain if each chain contains only a finite number of monomers. This artifact also prevents us from using the LP approximation, and hence the chain formalism, to model surface ablation by pure exo-systems in substrate morphology models. If the chains are not properly removed from the surface then the phantom chain ends left behind would eventually cover up the entire surface and block further access to the remaining solid substrate material underneath.

In the presence of non-negligible amounts of *endo*-activity, a population of insoluble chains of *all* fragment chain lengths $\ell < \ell_{\mathrm{Max}}$, down to short chains with $\ell \sim \ell_{\mathrm{S}}$, gets produced immediately by the *endo*-cuts and this population very quickly evolves a Poisson distribution shape, regardless of the initial distribution shape $Q(\ell)$. Hence, in LCL the fragmentation kinetics becomes "universal", that is, independent of initial distribution shape, since a Poisson distribution shape is established, especially at short chain lengths, long before even a small fraction of the substrate has been hydrolyzed. This also explains why the mixed enzyme system is much less sensitive to both initial DP in the LCL regime and to initial chain length distribution shape than the pure *exo*-enzyme system.

5.2 Hydrolysis Controlled by Morphology

Now we are going to present hydrolysis results with substrate morphology included ($\eta_\sigma \neq 0$). We consider four morphology models, SS, MS, MM82-1, MM82-2 as described in Sec. 4.4. All four models are parameterized into the same degree of polymerization $DP^{\mathrm{o}} = 300$, and

the enzymatic surface accessibility fraction $\bar{F}_{\mathrm{a}}^{(\mathrm{o})} = 0.006$. In the case of SS model, parameter $\bar{F}_{\mathrm{a}}^{(\mathrm{o})}$ is incorporated as an ad hoc correction factor to counteract over exposed surface area, a method Zhang & Lynd (Zhang and Lynd, 2006) adopted to reduce EG1 enzyme-chain accessibility for the actual substrate. All other parameters are taken directly from Table 1.

Fig. 5.3 shows results of three surface ablation models (*i.e.* MS, MM82-1 and MM82-2) and the SS model with $\bar{F}_{\mathrm{a}}^{(\mathrm{o})} = 0.006$ applied, for the complete hydrolytic conversion process, using the low-concentration enzyme system E1. As illustrated in Figs. 3A by the total monomer concentration x_{V} in solid substrate, the overall hydrolytic conversion in the three surface ablation models is significantly slower than in the pure chain fragmentation single-layer model SS. Furthermore, there are significant differences in hydrolytic conversion times between the three surface ablation models: MS, representing a zero-width Gaussian, hydrolyzes faster than MM82-2 with a finite-width Gaussian initial-size distribution; and MM82-2 in turn, is faster than MM82-1 representing the much wider uniform initial-size distribution: the hydrolytic conversion time increases with the width of the initial SAC size ($\lambda_{\sigma}^{(\mathrm{o})}$) distribution. Fig. 5.4 plots essentially the same quantities, but in E200 system.

In Fig. 5.3, the crucial point to emphasize here is that, in *all* four models, we have assumed the *same* chain fragmentation mechanism, with the *same* kinetic rate coefficient and enzyme parameters, and the *same* macroscopic substrate parameters, that is, the same initial molar amount of substrate $x_{\mathrm{V}}(t^{(\mathrm{o})})$, the initial degree of polymerization $DP^{(\mathrm{o})}$ and the same initial enzyme surface accessibility fraction, $\bar{F}_{\mathrm{a}}^{(\mathrm{o})}$, respectively. Clearly, the hydrolysis kinetics is very substantially dependent upon "other" factors, beyond the rate coefficient, enzyme adsorption or macroscopic substrate ($DP^{(\mathrm{o})}$ and $\bar{F}_{\mathrm{a}}^{(\mathrm{o})}$) parameters used in single-layer chain fragmentation models (Okazaki and Moo-Young, 1978; Zhang and Lynd, 2006). The substrate morphology is one such critically important factor determining the overall hydrolytic conversion time.

It is evident from Fig. 5.3 that the simulation results within the morphologic surface ablation models (MS and MM) are similar to each other, but quite different from the non-

morphologic SS model. From the inset of Fig. 5.3B and C, we can see for very short hydrolysis times (up to $\sim$ 180min) the behavior of the three morphologic surface ablation models is almost identical As explained later, this characteristic time scale of $\sim$ 180min corresponds to the hydrolytic fragmentation of the initially accessible fraction of substrate material, residing in the outermost SAC layers. While the three morphologic models diverge from each other thereafter, this divergence is much less pronounced than their profound differences from the non-morphologic SS model at longer time scales. In particular, the non-morphologic model predicts a much higher solubilization rate, which can be understood as a consequence of the fundamental neglect of the obstruction of enzyme access to the chain ends.

The corresponding results for the E200 enzyme system shown in Fig. 5.4 are qualitatively very similar to results from the E1 system: the three morphologic surface ablation models have much longer hydrolytic conversion time than the SS model. There also significant differences in hydrolytic conversion times between the three surface ablation models. Due to the 200-fold increase in enzyme concentration, the reaction rates are scaled up, and the overall times scales are scaled down, by a factor of order 100. From the inset figure of Fig. 5.4B and C, it can be seen that, even on very short hydrolysis time scales, the non-morphologic SS model is not a good approximation to the morphologic surface ablation models for the E200 enzyme system.

Hence, the hydrolytic conversion of cellulose substrate is crucially impacted by the substrate morphology. The above results also demonstrate very clearly that non-morphologic models can only be relied upon for the low-enzyme limit regime and only for very short time scales, up to the hydrolysis of the initial accessible fraction of substrate material. On time scales required to achieve substantial or near-complete hydrolytic conversion, or at higher enzyme loading, non-morphologic models are likely to fail.

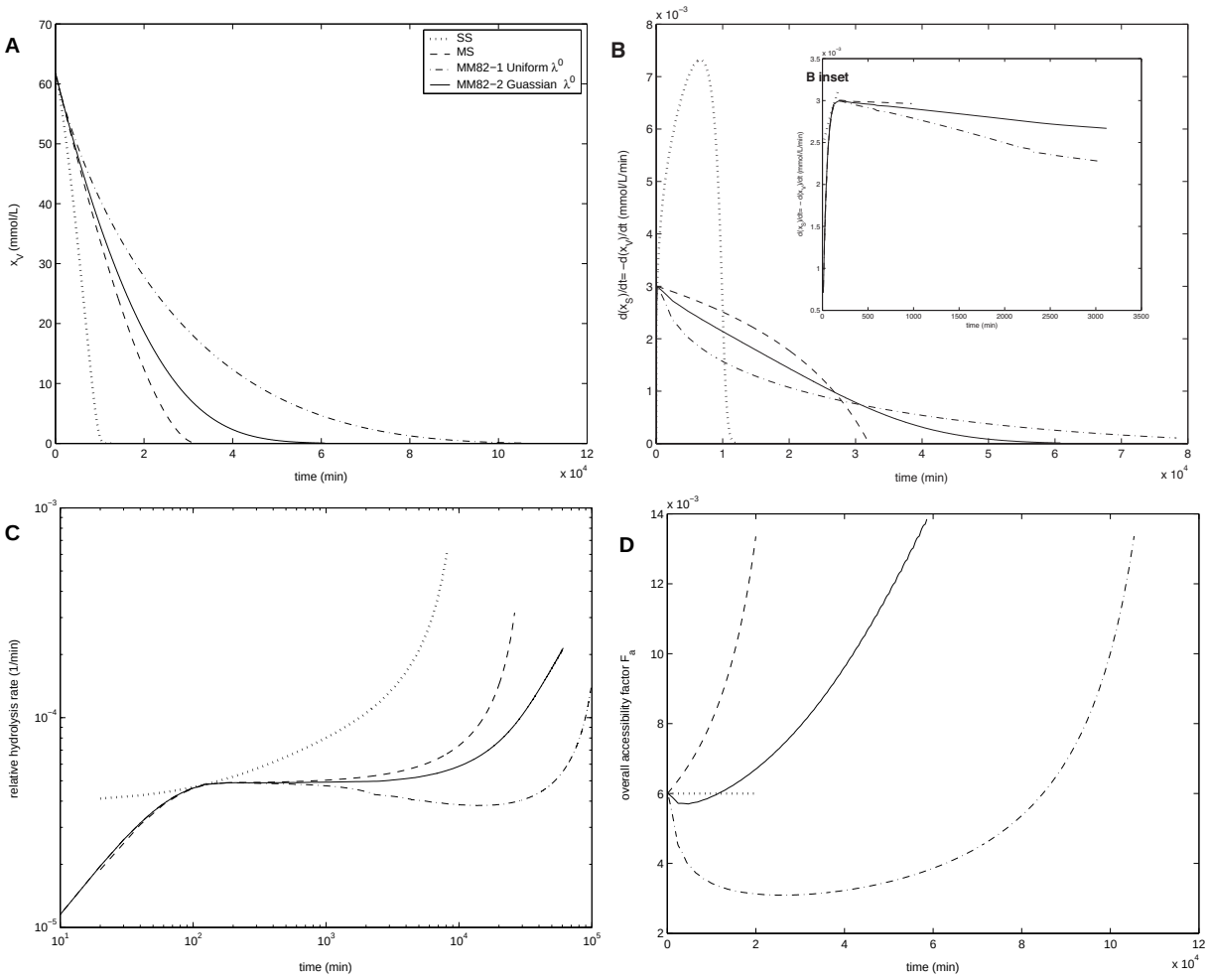

Figure 5.3: Hydrolysis Controlled by Morphology - E1 system. Simulation results of the MM82-1, MM82-2, MS and SS models, with $DP^{(0)} = 300$ and $\bar{F}_{\mathrm{a}}^{(\mathrm{o})} = 0.006$, for the E1 enzyme system. Plotted as functions of time t are A: total monomer concentration in solid, x_{V}; B: hydrolysis rate $|dx_{\mathrm{V}}/dt|$; C: relative hydrolysis rate Γ_{rel}; D: overall accessibility fraction $\bar{F}_{\mathrm{a}}$;

5.3 Two-Time Scale Behavior

Fig. 5.3B shows the conversion rate $|dx_{\mathrm{V}}/dt|$ as a function of hydrolysis time for the E1 enzyme system. All four models show a very rapid rise in their initial conversion rate at very early times. However, in the SS model, this rise continues unabated until about $t \sim 6600$min where a maximum rate is reached, followed by a decline on a similar time scale, through

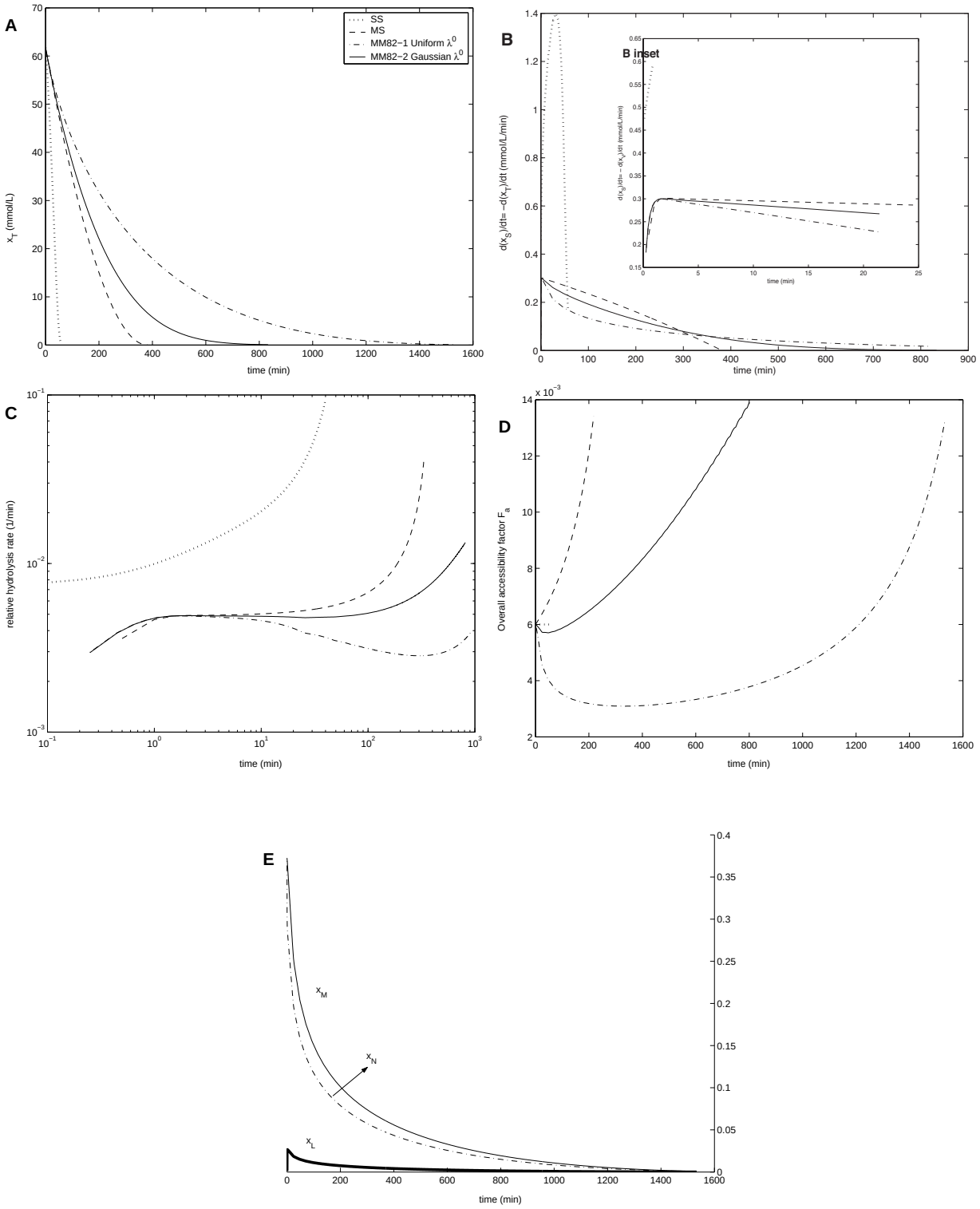

Figure 5.4: Hydrolysis Controlled by Morphology - E200 system. Simulation results of the MM82-1, MM82-2, MS and SS models, with $DP^{(0)} = 300$ and $\bar{F}_{\rm a}^{(\rm o)} = 0.006$, for the E200 enzyme system. Plotted as functions of time t are A-D: see descriptions of Fig. 3; E: concentrations of total surface exposed N sites, $x_{\rm N}(t) \equiv \sum_\sigma x_{{\rm N},\sigma}(t)$, total surface exposed non-reducing ends, $x_{\rm L}(t) \equiv \sum_\sigma x_{{\rm L},\sigma}(t)$, and total surface exposed monomers, $x_{\rm M}(t)$.

completion of hydrolysis until about 11000min. By contrast, in the three surface ablation models, the early rapid rise is abruptly arrested and a much lower maximum rate is reached already at a much earlier time, $t \sim 180$min, followed by a very slow drop-off for about $25000 - 60000$ minutes, consistent with the overall much longer conversion times in the surface layer ablation models. These results strongly suggest that the hydrolysis kinetics in the surface layer ablation models exhibits two quite distinct characteristic time scales: the very short, early-arrest time scale, and the much longer hydrolysis completion time scale, indicated, *e.g.*, by the 90%-conversion times.

This two-time-scale behavior is also clearly seen in Fig. 5.4B, for the E200 enzyme system, in all three morphologic surface ablation models, indicating that this is a common feature of the morphologic models, regardless of enzyme concentrations. The result in Fig. 5.4B is qualitatively very similar to that of Fig. 5.3B, with approximately 100-fold reduction in time, as mentioned before. The early arrest of the reaction rate in the surface ablation models occurs at ~ 1.7min.

In the surface layer ablation models, the steric obstruction of enzyme accessibility is not "mimicked" by the *ad hoc* correction factor as was done in previous modeling studies (Okzaki and Moo-Young, 1978; Zhang and Lynd, 2006). Rather, reduced accessibility results naturally from the actual substrate morphology, *i.e.*, from the fact that only surface-exposed sites are available for enzyme adsorption. Inspection of the early arrest and downturn of the ablation rate, near ~ 180min for the E1 enzyme system and ~ 1.7min for the E200 enzyme system, in simulation results Fig. 5.3B and 5.4B of the three surface ablation models reveals that this time corresponds to a 0.6% conversion of total substrate, for both enzyme systems; and this 0.6%-fraction is exactly equal to the initial fraction $\bar{F}_{\mathrm{a}} = 0.006$ of substrate material exposed in the outermost SACs layers at the start of hydrolysis. Hence, the early-arrest time scale, for both enzyme systems is clearly associated with the hydrolytic chain fragmentation and ablation of the outermost SAC layer.

Up to this "outermost layer ablation time", enzymatic cuts of *endo*-acting EG1 generates a large number of new chain ends, compared to initially existing native chain ends, which stimulates the activities of *exo*-acting CBH1 and CBH2. This cooperative work between *endo*-*exo*- enzymes causes the rapid increase in the production rate of soluble oligomers seen at early times in the insets of Fig. 5.3B and Fig. 5.4B. This is also clearly shown in Fig. 5.4E, where x_M and x_N decrease monotonically, and x_L shows a rapid rise during the early hydrolytic stage, with a time scale equal to that of the early-arrest time scale seen in $|dx_{\mathrm{V}}/dt|$. After that early-arrest time, the steric obstruction by only partially ablated overlaying material affects and persists for all subsequent layers being ablated and hence controls the ablation rate for the entire remaining hydrolytic conversion time. Consequently, the rate of new surface exposure, that is the $-\bar{R}_\sigma \eta_\sigma(\lambda_\sigma) g_{N,\sigma}(\lambda_\sigma)$-term in Eq. (4.7), *not* the enzymatic chain fragmentation, is the rate limiting factor for most of the remaining hydrolytic conversion time. This result clearly has technological implications: to substantially improve the performance of hydrolytic conversion, one may have to consider not only a re-engineering of the available enzyme systems, but also a re-engineering of the substrate morphology.

For ablation of the outermost SAC layer, only the total surface area, or surface site concentration, and the total ablation rate from all SACs of all geometry classes combined are relevant. Consequently, the MM and MS models of identical initial $\bar{F}_{\mathrm{a}}$-, x_{V}- and DP-values should exhibit the same early-arrest short-time behavior arising from the outermost layer ablation. However, on the much longer overall hydrolysis time scales the three surface ablation models are evidently very different from each other, since the replenishment rate of digested substrate material at the SAC surfaces in these models is quite different because of the effects of the different morphology distributions and their evolution under hydrolysis. As shown in Figs. 5.3A and B, the overall hydrolysis time scales show several fold differences between the three morphologic surface ablation models. This can be easily understood since the overall conversion time is controlled by hydrolysis of the large-size SACs. Hence, even though the initial accessible surface x_{M} and $\bar{F}_{\mathrm{a}}$, is the same in all three models, the uniform

MM82-1 model contains a larger fraction of its substrate in large SACs than the Gaussian MM82-2 and the (Delta-function!) MS model.

5.4 Hydrolysis Slow-Down and Morphology Evolution

Rapid decline in cellulose hydrolysis rate, and in the corresponding production rate of soluble glucose equivalent, as shown in Figs. 5.3B and 5.4B, is a feature that has been frequently observed in real hydrolysis experiments, and is referred to as hydrolysis slowdown in general.(Desai and Converse, 1997; Hong *et al.*, 2007; Lynd *et al.*, 2002; Yang *et al.*, 2006; Zhang and Lynd, 2004). Some part of this effect found in real substrates has been attributed to a loss of enzyme activity, either due to enzyme degradation/inactivation or due to enzyme inhibition by the hydrolysis-generated soluble monomer and oligomer products. However, the experiments, in which neither enzyme degradation/inactivation nor product inhibition appears operative, suggest that a significant part of the effect is in fact due to hydrolysis-induced changes in the substrate itself that can not be explained by loss of enzymatic activity or product inhibition (Valjamae et al., 1998; Zhang et al., 1999). Zhang et al. (1999) explained that by declining substrate reactivity caused by substrate heterogeneity, where more easily degradable substrate was depleted. Valjamae et al. (1998) tried to explain the rate decline in terms of steric hindrance due to nonproductive cellulase adsorption, as well as surface erosion after extended hydrolysis. Some studies (Desai and Converse, 1997; Yang *et al.*, 2006), however, showed that the substrate is as reactive as its initial state, implying that substrate reactivity is not the cause of the slowdown in hydrolysis. Therefore, the cause of hydrolysis slowdown is still uncertain, and whether or not there is a change in substrate reactivity is also in debate.

Zhang and Lynd (2004) stated that "It is widely observed that the heterogeneous structure of cellulose gives rise to a rapid decrease in rate as hydrolysis proceeds, ...", and "... it would seem logical to expect that the declining reactivity of residual cellulose during enzymatic hydrolysis is a result of factors such as less surface area and fewer accessible chain

ends ...". Indeed, our present simulation results of morphology-evolving surface ablation models confirms these expectations. As shown in Fig. 5.4E, both x_M (which is a measure of exposed accessible surface area) and x_L (after the short outermost surface ablation time) decrease as hydrolysis proceeds. Please note that in our simulation, neither enzyme degradation/inactivation nor product inhibition are present. Thus the heterogeneous solid structure of cellulose, *i.e.*, the steric obstruction of access to the inner, below-surface chains, does contribute to the phenomenon of hydrolysis rate decline in our model. Specifically, from the Eqs. (4.5) and (4.4), the rate of solubilization $|dx_{\mathrm{V},\sigma}(\lambda_\sigma)/dt|$ for SACs of size λ_σ is roughly proportional to their surface area, $x_{\mathrm{M},\sigma}(\lambda_\sigma)$. Since the exposed surface of every individual SAC will shrink as the hydrolysis proceeds, the total exposed surface area, thus the total solubilization rate $|dx_{\mathrm{V}}/dt|$ will decrease.

However, solid substrate structure heterogeneity alone can not explain the whole picture. As we can see from Figs. 5.3B and 5.4B, the extent of the loss of substrate reactivity is quite different among the three morphology models. Among them, the MM82-1 model exhibits the deepest depression from loss of substrate reactivity, while the MS model exhibits the least amount of reactivity loss. As discussed before, the MS and MM82-1 models represent two limiting cases of the Gaussian MM model, namely the "zero-width" limit and the "infinite-width" limit" respectively. A real morphology size distribution would likely fall in between these two extremes. In the MS model, all SAC units of the substrate have the same initial size and the same size at any time along the hydrolysis, as enzymatic surface ablation shrinks every SAC size and so does its accessible surface area. As a result, there must exist some other mechanism that also contributes to the relatively steeper decline in cellulose hydrolysis rate in the MM82-1 model.

To further analyze the differences between the three multi-layer surface ablation models, we have plotted in Figs. 5.3C and 5.4C the relative substrate hydrolysis rate Γ_{rel}, for E1 and E200 enzyme system respectively,

$$\Gamma_{\mathrm{rel}} \equiv -\frac{1}{x_{\mathrm{V}}}\frac{dx_{\mathrm{V}}}{dt} \ . \tag{5.1}$$

In Figs. 5.3C and 5.4C, the early rapid rise is again arrested in all three substrate ablation models at the outermost layer ablation time. Beyond that point, $\Gamma_{\rm rel}$ in the MM82-1 model drops noticeably below its early arrest value; $\Gamma_{\rm rel}$ in the MS models continues to rise, albeit with a markedly slower growth rate; and $\Gamma_{\rm rel}$ in the MM82-2 model falls between MS and MM82-1.

The decline of $\Gamma_{\rm rel}$ seems to indicate a decrease in the effective substrate reactivity for the uniform initial-λ distribution model MM82-1: the hydrolysis rate $|x_{\rm V}/dt|$ declines faster than the remaining substrate concentration $x_{\rm V}$ itself. This is the model with the widest, distribution of initial SAC sizes $\lambda_\sigma^{(\rm o)}$, extending with uniform weight from $\lambda_1^{(\rm o)} = 20$ to $\lambda_{82}^{(\rm o)} = 1640$. By contrast, in the MS and MM82-2 model, where all SAC units of the substrate have either the same initial size or a narrower, Gaussian size distribution, there is no, or only a very weak $\Gamma_{\rm rel}$ depression.

In Figs. 5.3D and 5.4D, we show the hydrolytic evolution of the overall accessibility fraction $\bar{F}_a$, for E1 and E200 enzyme system respectively. For the zero-width distribution MS model, $\bar{F}_a$ increases monotonically; for the finite-width Gaussian distribution MM82-2 model, $\bar{F}_a$ at first declines very slightly for a short time and then increases; and for the widest uniform distribution MM82-1 model, $\bar{F}_a$ declines most strongly and it has the longest duration of decline. Hence, the depression of $\bar{F}_{\rm a}$ increases with increasing width of the morphology distribution. The proportionality of solubilization rate and surface area also implies, by Eq. (5.1), that the relative hydrolysis rate $\Gamma_{\rm rel}$ is proportional to the accessibility fraction $\bar{F}_{\rm a}$. Hence, a decline of $\bar{F}_{\rm a}$ during early hydrolysis implies a corresponding decline in $\Gamma_{\rm rel}$.

To understand why MM82-1 and MM82-2 models exhibit a decline in the accessibility fraction $\bar{F}_{\rm a}$ and thus a depression in the relative hydrolysis rate $\Gamma_{\rm rel}$, Wen (2009) attributed the differences to "morphological heterogeneity". He suggested that the heterogeneity in substrate sizes leads to a preference of hydrolyzing smaller SAC units, thus leads to the loss of overall accessibility fraction. Model MM82-1 has essentially larger amount of smaller SAC

units than its counterpart MM82-2, therefore more surface-to-volume ratio are lost, which induces a steeper decline in $\bar{F}_{\mathrm{a}}$.

It has been proposed in previous studies that cellulose material contains two types of cellulose fractions that differ distinctly in their susceptibility to cellulase enzymatic attack. The basic idea here is that some types of, *e.g.*, amorphous, cellulose are easier to hydrolyze and other types, of, say, highly crystalline cellulose, are harder to hydrolyze (Gonzalez et al., 1989; Nidetzky and Steiner, 1993; Scheiding et al., 1984). Thus, if a (hypothetical) material contains both a substantial "fast-hydrolyzing" *and* a substantial "slow-hydrolyzing" substrate fraction, the fast early hydrolysis of the "fast" substrate results in the decline in hydrolysis rate at later times when only the "slow" substrate fraction remains. This "two-substrate" hypothesis attributes the difference of substrate reactivity specifically to the differences in the crystallinity of the two hypothesized fractions (Gonzalez et al., 1989; Nidetzky and Steiner, 1993; Scheiding et al., 1984). However, this picture has not yet been experimentally supported (Lynd et al., 2002). Although the existence of two (or multiple) substrate fractions of different reactivity within real pre-hydrolysis materials is presently uncertain, our simulation results imply that there do exist different hydrolysis rates among different substrate fractions, which are differentiated simply by volume, surface size and surface-to-volume ratios of their respective accessible substrate compartments. The fact that most of the substrate surface in real cellulosic materials is indeed comprised by internal surfaces (Zhang and Lynd, 2004), strongly suggests that this proposed "fractionation of substrate reactivity by geometry" may in fact be a ubiquitous feature of these materials.

5.5 Enzyme Concentration Scale-Up

The E1 enzyme set used above corresponds to a low-enzyme limit. It is speculated that the amount of cellulase required to achieve reasonable hydrolysis rate for real applications can be substantial (Lynd *et al.*, 2002). Specifically, Mandels (1985) estimated that for *T. Reesei* cellulase system, 3% by mass of the initial amount of cellulose is required. Here, we have

examined the cellulose enzymatic hydrolysis process in a mimic industrial environment by using the E200 enzyme set, with a 200-fold increase in concentrations in relative to the E1 set.

The corresponding results for the E200 enzyme system shown in Fig. 5.4 are qualitatively very similar to those shown in Fig. 5.4, except for the overall, approximately 100-fold reduction in time scales already discussed. Note that the reaction speeds should scale exactly linearly with enzyme concentrations as long as the enzyme-substrate system remains in the low-enzyme limit. However, the E200 system is already in the intermediate-to-high enzyme regime where the enzymes compete for available substrate sites, rather than substrate sites competing for enzymes. Consequently, the scale up in the reaction speed in going from E1 to E200 is neither exactly linear nor is it the same in all the four model on all time scales.

An analysis of the initial enzyme adsorption equilibrium shows that the free N and L site concentrations decrease from 99% in E1 to around 50% in E200, *i.e.*, almost 2 fold, relative to the total N and L site concentrations. On the other hand, the adsorbed enzyme fractions do not change much for the *exo*-acting enzymes, CBH1 and CBH2, and decrease from 53% in E1 to 37% in E200 for the *endo*-acting EG1. As a result, the initial concentration of ES complex, and hence the initial enzymatic cutting rates, show a 200-fold scale-up for CBH1,2 and a 139-fold scale-up for EG1 in E200 in relative to in E1, provided that the reaction rate coefficients remain unchanged.

The depression of Γ_{rel} in the MM82-1 model is somewhat more pronounced in the E200 system. A weak and brief Γ_{rel}-depression is now also seen in the Gaussian-distributed MM82-2 model in Fig. 5.4C. Both of these results suggests that higher enzyme concentrations tend to favor hydrolysis slow-down behavior. This is probably due to greater cooperativity between *endo*- enzyme and *exo*- enzymes under high enzyme concentrations, in consistent with experimental studies (Nidetzky *et al.*, 1994; Woodward *et al.*, 1988). As we can see, both the hydrolysis rate and the relative hydrolysis rate are much higher in the E200 system after the initial rapid rise, compared to those in the E1 system, as shown in Figs. 5.4B and

C and 5.3B and C. Thus, more and faster surface ablation on relatively smaller size SACs can be expected in the E200 system. This leads to more difficulty of substrate sites supply after the initial highest rate, and consequently the deeper decline in the hydrolysis rate, in relative to the lower E1 enzyme system.

For industrial applications, there is of course always a trade-off between the cost of enzyme concentrations added and the benefit from better performance of the operation using more enzymes. To explore possible performance optimization applications of our modeling approach, we have also simulated, in addition to the E1 and E200 systems, an enzyme system, labelled E200/50, where only EG1 is increased 200-fold, but CBH1,2 are increased *only* 50-fold. The 90% conversion time in this E200/50 system differs by less than 1% from that in the E200 system. Thus from an economic point of view, one should never use the E200 enzyme system in an industrial application, since lower enzyme usage in E200/50 gives the same performance.

While the E200/50 system's enzyme composition deviates from the naturally occurring composition found in living microbial cells, this does not necessarily mean that the natural composition is not at optimum under *in vivo* conditions. It is possible that the enzymatic activities exhibited *in vivo* are different from technologically relevant *in vitro* environments, and that they may be subject to regulation by the cells. Thus, it is quite possible that substantial improvements of hydrolysis cost/performance under technologically relevant *vitro* conditions can be achieved by our modeling approach, even for enzyme systems that have already been optimized, by nature, for *in vivo* performance. If process operation and economic parameters are available our modeling framework can provide a useful tool for a more detailed process optimization and design, by allowing us to perform systematic computational searches of parameter space for optimal processes and enzyme utilization. This will be the focus of future work.

5.6 Quasi-steady state Analysis

In previous section, we have demonstrated two-time scale behavior exhibited in the conversion rate $|dx_{\mathrm{V}}/dt|$ as a function of hydrolysis time: after a steep rise of hydrolysis rate, it is abruptly arrested at some maximum rate and then gradually drop to zero that spans over the rest of hydrolysis process. This two-time-scale behavior can also be interpreted as the very rapid establishment of a quasi-steady state (between surface chain fragmentation and new surface chain exposure rates) , followed by a very slow gradual decay of the substrate particle size λ_σ through a series of such quasi-equilibrated surface states. Thus, we propose here an approximate quasi-steady state treatment to model this quasi-equilibrated decay over long time scales. In this approach, the quasi-steady state is assumed to be already established from the very beginning and the very short, rapid initial build-up phase of the quasi-steady state is neglected.

Before we present quasi-steady state formalisms and related algorithm, we first make an analogous example to help illustrate the purpose behind this approach. Suppose we have a tank with a stream of water pouring in from the top at a constant rate or very slowly varying rate of $\mathbf{V}_{\mathrm{in}}(t)$; and at the same time, the water is leaking from the bottom at a rate that is proportional to the height of liquid contained in the tank $\mathbf{h}$, and can be expressed as $\mathbf{V}_{\mathrm{out}} = \alpha_{\mathrm{out}}\mathbf{h}$. (Illustrated in Fig. 5.5) Both rates are measured in units of volume / time. Assume we start with an empty tank, the tank will be slowly filled with water, until to some equi-

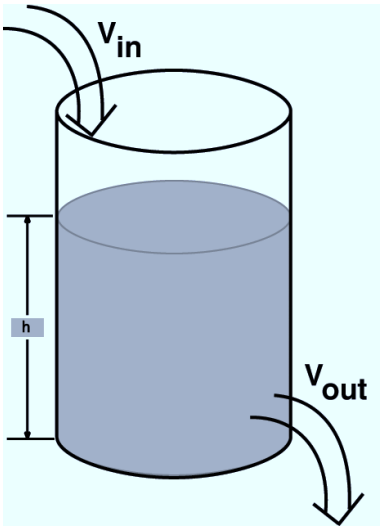

Figure 5.5: Quasi-steady state in water tank example. *Quasi-steady state state in water tank where the rate of water pouring in equals the rate of water leaking out at the bottom.*

librium state where the rate of water pouring
in approximately equals the rate of water
leaking out. Let us consider a very short time interval dt, during which the change of $\mathbf{h}$ is given as:

$$\mathbf{S} \times \mathrm{d}\mathbf{h} \;=\; (\mathbf{V}_{\mathrm{in}}(t) - \mathbf{V}_{\mathrm{out}})\mathrm{dt} = (\mathbf{V}_{\mathrm{in}}(t) - \alpha_{\mathrm{out}}\mathbf{h})\mathrm{dt} \tag{5.2}$$

where $\mathbf{S}$ is denoted as the cross section area of the tank. Straightforwardly, for the purpose of solving for $\mathbf{h}$ at equilibrium, we set $\mathrm{d}\mathbf{h}/\mathrm{dt}$ to be zero. Via Eq.(5.2), after rearrangement, we have

$$\frac{\mathrm{d}\mathbf{h}}{\mathrm{dt}} \;=\; \frac{\mathbf{V}_{\mathrm{in}}(t) - \alpha_{\mathrm{out}}\mathbf{h}}{\mathbf{S}} \cong 0 \tag{5.3}$$

$$\mathbf{h} \;\cong\; \frac{\mathbf{V}_{\mathrm{in}}(t)}{\alpha_{\mathrm{out}}} \tag{5.4}$$

The quasi-steady state approximation is valid here if the relative rate of change of the in-flow is very slow compared to the equilibration rate constant, *i.e.*, if $(1/V_{in})dV_{in}(t)/dt \ll \alpha_{out}/S$. The quasi-steady state approximation reduces differential rate equations to algebraic equations. In this simple example, we are able to obtain $h(t)$ without even having to solve any differential rate equation. Following the same reasoning, we would like to see if we can simulate the hydrolysis as a whole without its initial setup phase.

We first discuss the setup in chain number formalism, which is consisted of a close set of equations from Eqs. (4.27) and (4.9). Following aforementioned conclusion that hydrolytic evolution is inert to pre-hydrolysis substrate chain number distribution, we conveniently adopt here the delta function initialization where all substrate chain number are of the same length, $\ell^{(\mathrm{o})}$. In light of quasi-equilibrated water tank analogue given above, we set $\dot{G}_\sigma(\ell)$ in Eq. (4.27) to be 0:

$$\dot{G}_\sigma(\ell) = \dot{G}_\sigma^{(\mathrm{fra})}(\ell) - \bar{R}_\sigma \eta_\sigma(\lambda_\sigma) Q_\sigma(\ell, \lambda_\sigma) / \sum_{j=\ell_{\mathrm{S}}}^{\infty} j Q_\sigma(j, \lambda_\sigma) = 0 \tag{5.5}$$

It is important to notice that with normalization condition inherited from $P_\sigma(\ell)$ construction, it is obvious that the following equality holds true:

$$\sum_{\ell=\ell_{\rm S}}^{\ell_{\rm mid}} P_\sigma(\ell) = 1 - \sum_{\ell=\ell_{\rm mid}+1}^{\ell_{\rm max}} P_\sigma(\ell), \qquad \ell_{\rm S} \le \ell_{\rm mid} \le \ell_{\rm max} \tag{5.6}$$

Rewriting Eqs. (4.10) using Eq. (5.6), via Eq. (4.27), we can thus solve for $G_\sigma(\ell)$ recursively from $\ell_{\rm max}$ to $\ell_{\rm S}$ given the value of λ_σ. With a complete profile of $G_\sigma(\ell)$ in hand, through Eqs. (4.16)-(4.18), one can calculate $\bar{R}$, which in turn enters into Eq. (4.9) for λ_σ rate equations. All $G_\sigma(\ell)$ are thus eliminated as dynamical variables from the rate equation system. They are now algebraically dependent on the λ_σ-variables. The resulting closed set of λ_σ rate equations generates the full time evolution of all chain number variables $G_\sigma(\ell)$.

In site number formalism, the number of independent variables, *i.e.* λ_σ, $x_{L,\sigma}$, $G_\sigma(\ell_{\rm S})$ and $G_\sigma(\ell_{\rm S}+1)$, is much smaller than in the chain number formalism with LP approximation. Their aforementioned coupled rate equations are now set to equal 0 in our quasi-steady state analysis:

$$\dot{x}_{L,\sigma} = R_{L,\sigma} - \bar{R}_\sigma\ \eta_\sigma(\lambda_\sigma)\ g_{L,\sigma}(\lambda_\sigma) = 0 \tag{5.7}$$

$$\begin{aligned} \dot{G}_\sigma(\ell) &= \Gamma_{{\rm N},\sigma}\, x_{L,\sigma} \left[2 - 2\sum_{k=\ell_{\rm S}}^{\ell} P_\sigma(k) - (\ell-1)\ P_\sigma(\ell)\right] \\ &\quad + \Gamma_{{\rm X},\sigma}\, x_{L,\sigma} \left[P_\sigma(\ell + k_{\rm X}) - P_\sigma(\ell)\right] \\ &\quad + \Gamma_{{\rm Y},\sigma}\, x_{L,\sigma} \left[P_\sigma(\ell + k_{\rm Y}) - P_\sigma(\ell)\right] \\ &= 0 \qquad\qquad\qquad\qquad\qquad \text{for } \ell = \ell_{\rm S}, \ell_{\rm S}+1 \end{aligned} \tag{5.8}$$

Again, under LP approximation, there is no surface exposure term entering rate equations for $G_\sigma(\ell_{\rm S})$ and $G_\sigma(\ell_{\rm S}+1)$. Eqs. (5.7)-(5.8) then form a closed set of quasi-steady state equations under site number formalism.

Fig. 5.6 shows four complete hydrolysis profiles for MS model with DP=100, 300, 1000 and 2000, respectively. Enzyme concentrations, substrate initial morphology data ($\lambda^{\rm o}$, $x_{\rm V}$), among other variables are directly calculated from Table 1. (We are dropping the subscript $_\sigma$,

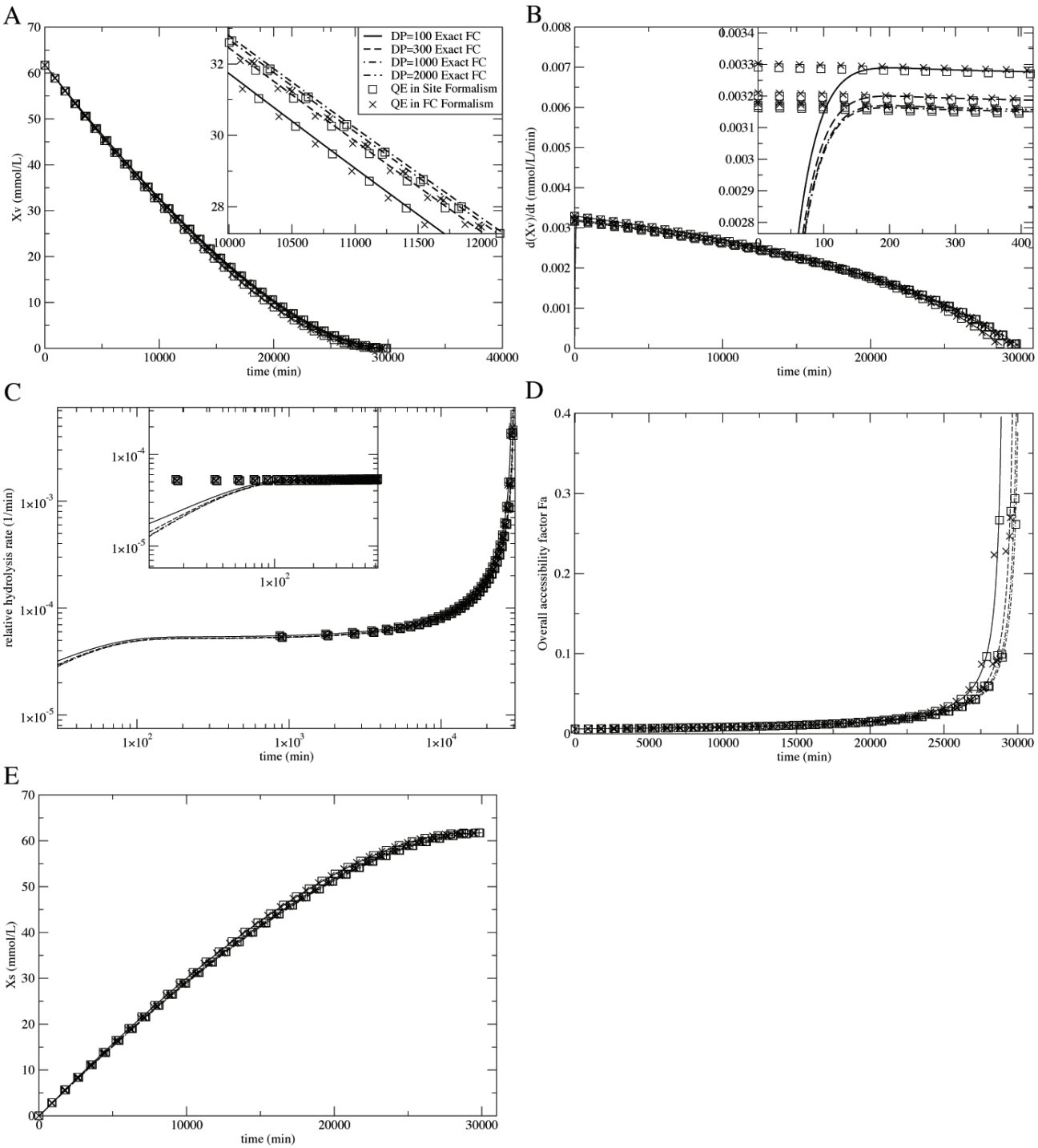

Figure 5.6: Quasi-steady state analysis. Comparison of the MS model results from exact chain number formalism against quasi-steady state models with chain-number and site-number formalisms, respectively, under different initial delta chain length distributions, for the mixed EG1-CBH1,2 enzyme systems. All enzyme concentrations are adopted from Table 1 directly (E1 system). In panels A-E, although almost indistinguishable, full lines, dashed lines, dot-dash-dot lines and dash-dot-dash lines are for exact chain number results with DP = 100, 300, 1000 and 2000; square represents the simulated data from quasi-steady state site number formalism and cross represents that from quasi-steady state full chain formalism. A: total monomer concentration $x_{\mathrm{V}}(\mathrm{t})$ in solid versus time t; (B) hydrolysis rate $|dx_{\mathrm{V}}/dt|$; (C) relative hydrolysis rate Γ_{rel}; (D) overall accessibility fraction F_{a}; (E) concentration of soluble monomers $x_{\mathrm{S}}(\mathrm{t})$ in solid versus time t.

since we are dealing with MS model with single geometry here) Fig. 5.6A plots the relationship between total monomer concentration (x_{V}) versus time. As foregoing analysis shows, for hydrolytic processes among different DP but otherwise identical morphologic surface ablation models (identical F_{a}, x_{V}), complete hydrolysis profile should look very similar to each other; in other words, these four situations only differ from each other by how chains are segmented within their respective substrates, and only differ on the initial concentrations of x_{L}. Therefore, the general rates at which *endo*-acting enzymes cut chains are roughly the same, as well as the replenishment rate of digested substrate material at the SAC surfaces. Consequently, the hydrolysis profile almost overlap with each other, with minor differences for complete hydrolysis time, which can be easily explained by small disparities in *exo*-acting enzyme activities.

Overall, it is very clear that both quasi-steady state in site number formalism and full chain formalism reproduce the exact full chain solution quite well. The inlet in Fig. 5.6A magnifies within a certain time frame, further indicating the good agreements between exact solution and our approximated quasi-steady state solutions, with the differences being at most 1% across all different DP values tested. However, according to the inlet in Fig. 5.6A, the approximations unanimously tend to underestimate the exact solution at corresponding time t. Such agreements are also observed on other quantities including the hydrolysis rate in Fig. 5.6B; relative hydrolysis rate in Fig. 5.6C; overall accessibility factor in Fig. 5.6D and soluble monomer concentrations in Fig. 5.6E.

It is critical to notice, in Fig. 5.6B we show the quasi-steady state approximation for conversion rate $|dx_{\mathrm{V}}/d\mathrm{t}|$ for the same four situations described above. Similar to what we have observed before, the exact solutions exhibited a rapid rise followed by a sudden arrest, which leads to a very slow decaying plateau phase. The abrupt arrest in hydrolysis rate signifies the turning point where all surface exposed monomers have been fragmented, and then the surface exposure term begin to dominate the hydrolysis process. However, the equilibrium establishment is completely circumvented under quasi-steady state approximation. For quasi-

steady state data points, we observe an almost linear extension from prehydrolysis t^o onto where would be the turning point for exact solutions. This is as if even for the outer most layer on each SAC surface, starting from the beginning of hydrolysis, the rate of monomer/site concentration loss from fragmentation is almost the same as the rate of newly exposed monomer/site from surface ablation, hence a series of quasi-steady states, throughout the whole hydrolysis process.

5.7 Degree of Synergy

We previously studied a system with a mixture of EG1 and CBH1,2 enzyme system, as well as systems under the effect of pure CBH1,2 enzyme to test the LP approximation. Theoretically speaking, any substrate is susceptible to hydrolytic effect from cellulase systems, with almost any combination possible. Simply a matter of time, either pure *endo*− or pure *exo*− or a mix of both will attack their respective adsorption sites on cellulose, and eventually break down all substrate into glucose or cellobiose. However, as evolution progresses, mother nature created a pot of different enzymes, with their functions closely related or even overlapping sometimes, that generates an optimal result. This additive effects among many enzymes are called synergism.

We are interested in finding out how prominent synergism acts between three enzymes from *Trichoderma reesei*. EG1, an *endo*-acting enzyme adsorbs to X-, Y- and O-sites in our five site model while CBH1,2, as *exo*-acting enzyme, only adsorbs to X-(for CBH2) and Y-(for CBH1) sites respectively. While the majority of chains are of length on a higher order of magnitude than ℓ_{S}, most sites on cellulose substrate are N-sites and thus only accept adsorption from EG1. This creates less of a competition between three enzymes than a situation where most chains are now on the same order as ℓ_{S}, where the ratio between N-sites and X-/Y-sites are considerably smaller and it is more likely for *endo*- acting enzyme EG1 to attack X-,Y-sites. These two different scenarios are likely to happen at the very beginning of hydrolysis and towards the completion of hydrolysis, respectively.

Our approach focuses on different concentration ratio between the above three enzymes. In order to be consistent with Table 1, we define u_1 to be concentration of EG1, in units of mmol/L; u_2 to be concentration of CBH2, and u_3 to be that of CBH1. Then we construct three constants u_{total}, α_1 and α_2 from:

$$u_{\text{total}} = u_1^{\text{o}} + u_2^{\text{o}} + u_3^{\text{o}} \tag{5.9}$$

$$\alpha_1 = \frac{u_2^{\text{o}}}{u_2^{\text{o}} + u_3^{\text{o}}} \tag{5.10}$$

$$\alpha_2 = \frac{u_3^{\text{o}}}{u_2^{\text{o}} + u_3^{\text{o}}} \tag{5.11}$$

where the superscript $^{\text{o}}$ denotes the values calculated from Table 1. For the purpose of identifying enzyme synergism between *endo*- and *exo*-acting enzymes, we are keeping the ratio α_1 and α_2 defined above constant while twitching the enzyme ratios between EG1 and the sum of CBH1,2. First, let us introduce variable $\phi \in [0, 1]$, with 0 stands for pure *exo* enzymes system and 1 entirely consisted of pure *endo* enzyme.

$$\begin{aligned} u_1 &= \phi u_{\text{total}} \\ u_2 &= \alpha_1(1-\phi)u_{\text{total}} \\ u_3 &= \alpha_2(1-\phi)u_{\text{total}} \end{aligned} \tag{5.12}$$

with $\phi = 0.0$ to 1.0, with an increment of 0.1.

Degree Synergy (DS) is a quantity defined to indicate the degree of cooperativity between enzymes and can be formulated as the ratio between hydrolysis rate from enzyme mixture and the sum of hydrolysis rates from each individual enzymes. In our calculations, we differentiate enzymes only to the extent of *endo*- vs. *exo*-activities. Depending on the common factor between all hydrolysis rates, we can calculate DS based on hydrolyzed monomer concentration or hydrolysis time respectively, formally shown as:

$$\text{DS}_{\text{t}} \equiv \frac{\text{t}_{\text{mixture}}^{-1}}{\text{t}_{\text{endo}}^{-1} + \text{t}_{\text{exo}}^{-1}} \tag{5.13}$$

$$\text{DS}_{x_{\text{V}}} \equiv \frac{\Delta x_{\text{V},mixture}}{\Delta x_{\text{V},endo} + \Delta x_{\text{V},exo}} \tag{5.14}$$

where t_{mixture}, t_{exo} and t_{endo} denote the amount of time it takes to reach the same substrate conversion rate within each enzyme configuration; while $\Delta x_{\text{V},mixture}$, $\Delta x_{\text{V},endo}$ and $\Delta x_{\text{V},exo}$ stand for total hydrolyzed monomer concentrations for the same time checkpoint in each respective enzyme configuration.

Fig. 5.7 shows results of degree synergy calculated in full chain quasi-steady state model, using the low-concentration enzyme system E1. Fig. 5.7A illustrates DS_{t} evaluated at 10% and 90% hydrolysis conversion time checkpoints; Fig. 5.7B shows $\text{DS}_{x_{\text{V}}}$ according to hydrolyzed monomer concentrations at 10% and 90% hydrolysis conversion time checkpoints for mixture enzyme system. Both figures show a degree synergy larger than 1 across all enzyme concentration profiles, ranging from pure endo to pure exo. This suggests that at least to some degree, there is synergism observed between endo and exo activities, regardless of their respective concentration ratio. It is quite noticeable that, also in both graphs, there is a peak situated immediately after pure-endo ($\phi = 0$) enzyme configuration, and DS value then sharply depressed into a neighborhood near value of 1. It is equivalent to say that, under low enzyme concentration, maximum synergy is achieved when the majority of enzyme concentrations ($> 90\%$) are contributed by *exo*-acting enzymes. This result agrees quite well with what Converse (Converse, 2004) have found in a sense that under low enzyme concentrations where the majority of sites are unabsorbed with enzymes, *exo*-activity from CBH1,2 does not depend on chain ends produced from *endo* enzymes. Thus hydrolytic productivity should be proportional to *exo*-acting enzymes that chops off chain ends into soluble oligomers. Furthermore, we have also studied various other proportionally increased/decreased enzyme systems (relative to E1). The results are exactly replications of what we have found in E1, as reaction speed scale accordingly with enzyme concentrations as long as we restrain ourselves strictly in the low-enzyme limit.

For degree synergy comparison between different time checkpoints, we have noticed, in both figures, earlier time checkpoint exhibits a higher degree of synergy than later time checkpoint. Such phenomena is only noticeable in Fig. 5.7A for high DS values (*i.e.*, $0 \leq$

$\phi \leq 0.2$), but much more pronounced in Fig. 5.7B. This can be straightforwardly explained, again, as a loss of substrate sites to enzyme ratio during the hydrolysis. In an extreme case, where very high enzyme concentrations almost occupy all free sites on substrate SAC surfaces, we expect to see competitions between *endo*- and *exo*-enzymes to an extent where DS value calculated would start to dampen. It is then our future work to involve solving the full enzyme-substrate equilibrium equations, and study the degree of synergy under medium to high enzyme concentrations.

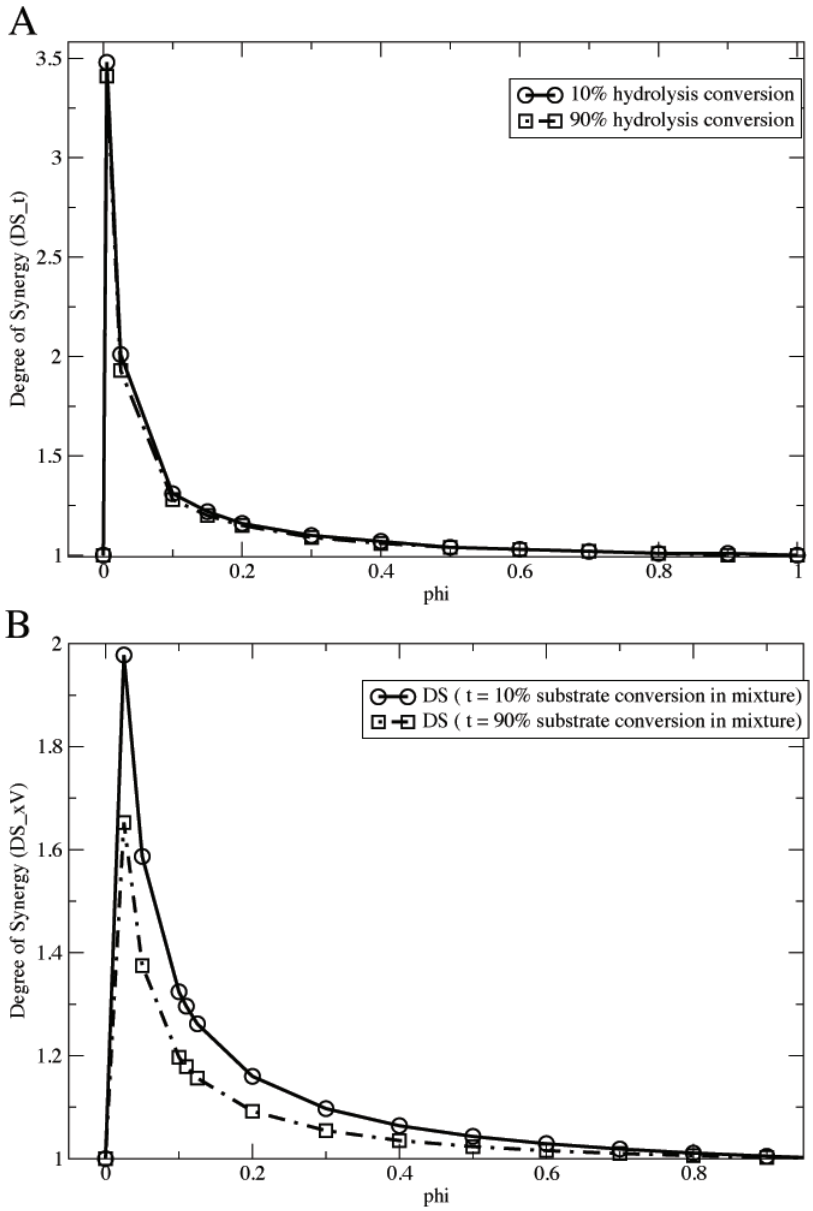

Figure 5.7: Degree of Synergy. Simulation results of Degree of Synergy, defined in Eq. (5.13)-(5.14), in MS model with full chain quasi-steady state, for the E1 enzyme system. (A)$\mathrm{DS_t}$ vs. ϕ; (B) $\mathrm{DS}_{x_\mathrm{V}}$ vs. ϕ.

Chapter 6

Conclusion

We first developed a general theoretical framework modeling enzymatic hydrolysis on solid substrates. Unlike our predecessors, for the first time we coupled enzymatic fragmentation with time-dependent substrate morphology. This formalism explicitly included how enzymatic degradation can affect the size of substrates, and how the morphology evolution in turn impact on hydrolysis rate. Other than that, an essential feature of this formalism is its ability to capture the effects of random spatial substrate heterogeneity, present in all pre-hydrolysis/pretreated substrates. Among these, we have acknowledged in our model: random distribution of enzyme-accessible internal surface area associated with randomly sized SACs in the substrate morphology; random spatial distribution of non-cellulosic contaminants within SACs; and/or random spatial distributions of the glucan chain degree of polymerization; random spatial distributions of the degree of chain ordering; and random distributions of hydrolysis time scales result from all the foregoing random spatial inhomogeneities of the substrate. To account for all above mentioned randomly distributed substrate geometries, we introduced SAC geometry σ and then represent macroscopic substrate as a population of discreet SAC geometry classes. We adopted geometry-specific layer variable λ_σ to describe SAC sizes and its rate equation governs time-dependent SAC geometry evolution.

Based on previous work (Zhang Lynd, 2006), we enhanced full chain number formalisms with our surface ablation model, in which the ad hoc factor has been replaced with surface exposure term, that is essentially governed by λ_σ time evolution. To further illustrate chain fragmentation kinetics, we invented the site number formalism where we keep track of site concentrations, which is on the order of ℓ_C, whereas in full chain formalism, variables relating

to chain length concentrations could easily go beyond hundreds, if not thousands. This site number mechanism thus benefits us with a significantly reduced-size ODE system, compared to earlier conventional chain number representation, at a small price of losing detailed chain length distribution profiles during the hydrolysis. Both formalisms open up a new gateway to simulate hydrolysis beyond initial conversion stage, extended to a simulation of full near-complete substrate conversion process.

We first tested our site number formalism against exact full chain formalism, and established fundamental equivalency between two formalisms. Also, local Poisson approximation was well validated under various combinations of model parameters.

Then we simulated on surface layer ablation models and observed in MS, MM82-1 and MM82-2 models two distinctive hydrolysis time scale: the characteristic short single outermost layer ablation time and the much longer overall hydrolysis time. The short single-layer ablation time scale corresponds to an early rapid rise in hydrolysis rate $|dx_{\mathrm{V}}/d\mathrm{t}|$, which is suddenly arrested at a maximum hydrolysis rate and followed by a much prolonged complete hydrolysis process. This phenomena is inevitably an consequence of hydrolyzing solid substrates where inner substrates are not accessible to enzymatic attack until outer layer are ablated. We refer to this as *substrate solid structural heterogeneity*.

Moreover, we have proved hydrolysis process largely depend on random substrate morphology. We, again, adopted three surface ablation models, where we kept the same specific accessible internal surface area, the same degree of polymerization, the same hydrolytic enzyme system with the same rate coefficients, but only varying the substrate morphology distribution, with MS being an extreme delta distribution, MS82-1 being another extreme with truncated uniform width distribution and MS82-2 the Gaussian shaped morphology distribution. Among three models, we witnessed vast differences in overall hydrolysis time, maximum hydrolytic rate, overall accessible surface evolution, etc. It is only natural to suggest that, in addition to substrate solid structural heterogeneity, it is equally important to recognize the effect of *substrate morphology heterogeneity*. In reality, these two effects could

be operational along with product-inhibition, substrate-induced enzyme inhibition factors, in bringing about the observed hydrolysis slow-down phenomena.

In regards to short-time scale phenomena, we proposed a quasi-steady state analysis that approximated the hydrolysis as a whole with a series of quasi-steady states. This approach circumvents modeling the rapid equilibration processes on very early short time scales, but gives accurate results on long time scales to near-complete substrate solubilization. The quasi-steady state results are in excellent agreement with the exact solution.Lastly, we investigated the synergism between EG1 and CBH1,2, and observed clear evidence for cooperativity under the low enzyme concentration regime.

Overall, our modeling approach opens up a new perspective into cellulosic hydrolysis process that not only we extended reliable hydrolysis simulation to a near-completion state, but also accredited "substrate solid structural heterogeneity" and "substrate morphology heterogeneity" for hydrolysis slow down, which is universally experienced in industrial production. However, our work could be substantially validated/enhanced if future experimental research shall lend insight into hydrolytic evolution of critical substrate and hydrolysis parameters. (*e.g.* overall accessibility fraction, surfaces-exposed degree of polymerization. Given the availability of these data, we will then be able to directly compare our model simulations to real experimental data. Furthermore, as we stated earlier, the importance of substrate morphology distribution renders further investigations into substrate's macroscopic/microscopic structure of essential importance. Combined with above all would we have a more clear picture and detailed understanding of the hydrolytic conversion process.

Nomenclature

$a_{\nu,\sigma}$, $A_{\nu,\mu,\sigma}$	decomposition parameters used in Eqs. (3.31) & (3.34)
$b_{\nu,\sigma}$, $B_{\nu,\mu,\sigma}$	decomposition parameters used in Eqs. (3.31) & (3.34)
$B_{\mathrm{V},\sigma}$	molar volume prefactor, $\equiv C_\sigma c_{\mathrm{V},\sigma}$
C_σ	class-σ SAC concentration, mM
	(moles of SACs in class-σ per reactor volume)
$c_{\mathrm{V},\sigma}$	volume prefactor to calculate $n_{\mathrm{V},\sigma}$
$d_{\mathrm{A},\sigma}$	ablation dimension for class-σ SACs
$d_{\nu,\sigma}$, $D_{\nu,\mu,\sigma}$	decomposition parameters used in Eqs. (3.31) & (3.34)
Dx_{V}	λ_σ-density distribution function of $x_{\mathrm{V},\sigma}$
$F_{\mathrm{a},\sigma}$	fraction of accessible G_1 for class-σ SACs, $\equiv n_{\mathrm{M},\sigma}/n_{\mathrm{V},\sigma}$
$\bar{F}_{\mathrm{a}}$	overall accessibility fraction of accessible G_1, $\equiv n_{\mathrm{M}}/n_{\mathrm{V}}$
$f_{\nu,\sigma}$	type-ν site fraction on class-σ SAC surfaces, $\equiv n_{\nu,\sigma}/n_{\mathrm{M}}$
$g_{\nu,\sigma}$	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
$G_\sigma(\ell)$	concentration of G_ℓ exposed on class-σ SAC surfaces, mM, $\equiv C_\sigma H_\sigma(\ell)$
$H_\sigma(\ell)$	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

$\ell_{\rm S}$	minimum insoluble chain length
$\langle \ell \rangle_\sigma$	average chain length for chains exposed on class-σ SAC surfaces
$\ell_{\rm L}$, ($\ell_{\rm R}$)	length of terminal segment that L-end (R-end) can affect
$\ell_{\rm LR}$	$\ell_{\rm LR} \equiv \ell_{\rm L} + \ell_{\rm R} - 1$
$\ell_{\rm T}$	$\ell_{\rm T} \equiv \max(\ell_{\rm S}, \ell_{\rm L} + \ell_{\rm R} - 1)$
$\ell_{\rm C}$	$\ell_{\rm C} \equiv \max(\ell_{\rm S}, \ell_{\rm L} + \ell_{\rm R} - 1) + \max(\ell_{\rm L}, \ell_{\rm R}) - 2$
$\ell_{\rm E}$	$\ell_{\rm E} \equiv \max(\ell_{\rm L}, \ell_{\rm R}) - 1$
$\ell_{\rm D}$	$\ell_{\rm D} \sim \ell_{\rm C}$, but $\ell_{\rm D} \geq \ell_{\rm C}$
$L_{\kappa,\mu}$	adsorption coefficient for (κ, μ)ES complex, 1/mM
$m_{\kappa,\mu,\sigma}$	number of (κ, μ) ES complexes per class-σ SAC
$M_{\rm MD}$	population size of SAC geometries
$n_{{\rm M},\sigma}$	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
$n_{{\rm S},\sigma}$	number of dissolved G_1 produced per class-σ SAC
$n_{{\rm V},\sigma}$	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
$\Delta \bar{N}_{\nu,\mu,\sigma}$	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
$p^{({\rm I})}_{\mu,\sigma}$, $p^{({\rm L})}_{\mu,\sigma}$, $p^{({\rm R})}_{\mu,\sigma}$, $p^{({\rm S})}_{\mu,\sigma}$	contributions to probability for finding type-μ, on class-σ SAC surfaces from the interior, L-terminal and R-terminal segments from short chains with $\ell \leq \ell_{\rm LR}$, 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
$R_{\nu,\sigma}$	production rate of type-ν site, mM/min, $\equiv C_\sigma V_{\nu,\sigma}$
$\bar{R}_\sigma$	negative rate of monomer loss ($\bar{R}_\sigma < 0$) into solution, mM/min
u_κ	total type-κ enzyme concentration, mM
v_κ	free type-κ enzyme concentration, mM
$V_{\mathrm{H},\sigma}(\ell)$	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)
$V_{\mathrm{S},\sigma}(k)$	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
$\bar{V}_\sigma$	negative rate of monomer loss into solution per class-σ SAC
w_ν	number of monomers represented by one type-ν site
$x_{\mu,\sigma}$	concentration of type-ν site exposed on class-σ SAC surfaces,

	mM, $\equiv C_\sigma\, n_{\nu,\sigma}$
$x_{\rm M}$	total concentration of G_1 exposed surfaces, mM
$x_{{\rm M},\sigma}$	concentration of G_1 exposed on class-σ SAC surfaces, mM
$x_{\rm V}$	total concentration of G_1 in solid substrate, mM
$x_{{\rm V},\sigma}$	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)
$\Gamma_{\mu,\sigma}$	enzyme cutting rate factors defined by Eqs. (4.12)-(4.14)
$\Gamma_{\rm rel}$	relative hydrolysis rate, $\equiv -dx_{\rm V}/dt/x_{\rm V}$
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, $\equiv x_{{\rm V},\sigma}/x_{\rm V}$
ζ	bond 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
MM	"**M**ultiple-layer, **M**ultiple-geometry" model
MM82-1	the MM model with uniform distribution of monomer concentration per geometry class,($i.e.\ \xi_\sigma^{(o)} = 1/82\ \ \forall\, \sigma$)
MM82-2	the MM model with Gaussian distribution of monomer concentration per geometry class through Eq. (4.36)
MS	"**M**ultiple-layer, **S**ingle-geometry" model
ODE	ordinary differential equation
R-end	reducing end of a cellulose chain, also called "right" end
SAC	smallest accessible compartment
SAV	smallest accessible void
SS	"**S**ingle-layer, **S**ingle-geometry" model

Bibliography

[1] 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* CBH1 on Avicel and bacterial microcrystalline cellulose. *Biores Technol* **60**:169-178.

[2] Chang MM, Chou TYC, Tsao GT. 1981. Structure, pretreatment and hydrolysis of cellulose. *Adv Biochem Eng* **20**:15-42.

[3] 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.

[4] 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.

[5] Substrate factors limiting enzymatic hydrolysis. In: Saddler JN, editor. Bioconversion of forest and agricultural plant residues. Wallingford: CAB International . pp. 93-106

[6] Converse AO, Optekar JD. 2004. A synergistic kinetics model for enzymatic cellulose hydrolysis compared to degree-of-synergism experimental results. *Biotechnol Bioeng* **42**:145-148.

[7] Desai SG, Converse O. 1997 Substrate reactivity as a function of the extent of reaction in the enzymatic hydrolysis of lignocellulose. *Biotechnol Bioeng* **56**:650-655.

[8] 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.

[9] Grethlein HE. 1985. The effect of pore size distribution on the rate of enzymatic hydrolysis of cellulose substrates. *Bio Technol* **3**:155-160.

[10] Gonzalez G, Caminal G, de Mas C, Lopez-Santin J. 1989. A kinetic model pretreated wheat straw saccharification by cellulose. *J Chem Technol Biotechnol* **44**:275-288.

[11] 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**:325-324.

[12] 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.

[13] Hong J, Ye X, Zhang YHP. 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.

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

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

[16] Mandels M. 1985. Applications of cellulase. *Biochem Soc Trans* **13**:414-416.

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

[18] 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.

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

[20] Nidetzky B, Steiner W. 1993. A new approach for modeling cellulase cellulose adsorption and the kinetics of the enzymatic hydrolysis of microcrystalline cellulose. *Biotechnol Bioeng* **42**:469-479.

[21] Nidetzky B, Sterner W, Hayn M, Claeyssens M. 1994. Cellulose hydrolysis by the cellulase from *Trichoderma reesei*: A new model for synergistic interaction. *Biochem J* **298**:705-710.

[22] 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.

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

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

[25] 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.

[26] 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.

[27] Puls J, Wood Tm. 1991. The degradation pattern of cellulose by extracellular cellulases of aerobic and anaerobic microorganisms. *Biores Technol* **36**:15-39.

[28] Scheiding W, Thoma M, Ross A, Schugerl K. 1984. Modeling of the enzymatic hydrolysis of cellobiose and cellulose by a complex enzyme mixture of *Trichoderma reesei* QM9414. *Appl Microbiol Biotechnol* **20**:176-182.

[29] Valjamae P, Sild V, Pettersson G, Johansson G. 1998. The initial kinetics of hydrolysis by cellobiohydrolases I and II is consistent with a cellulose surface–Erosion model. *Eur J Biochem* **253**:469-475.

[30] 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.

[31] Wood TM. 1975. Properties and Mode of action of cellulases. *Biotechnol Bioeng Symp* **5**:111-137.

[32] Woodward J, Lima M, Lee NE. 1988. The rold of cellulase concentration in determining the degree of synergism in the hydrolysis of microcrystalline cellulose. *Biochem J* **255**:895-899.

[33] Yang B, Willies DM, Wyman CE. 2006. Changes in the enzymatic hydrolysis rate of Avicel cellulose with conversion. *Biotechnol Bioeng* **94**:1122-1128.

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

[35] Zhang Y-HP, Lynd LR. 2006. A functionally based model for hydrolysis of cellulose by fungal cellulase. *Biotechnol Bioeng* **6**:1510-1515.

[36] Zhang S, Wolfgang DE, Wilson DB. 1999. Substrate heterogeneity causes the nonlinear kinetics of insoluble cellulose hydrolysis. *Biotechnol Bioeng* **66**:35-41.

[37] Zhou W, Schüttler HB, Hao Z, Xu Y. 2009. Cellulose hydrolysis in evolving substrate morphologies I: A general modeling formalism. *Biotechnol Bioeng* **104**:261-274.

[38] Zhou W, Hao Z, Xu Y, Schüttler HB. 2009. Cellulose hydrolysis in evolving substrate morphologies II. Numerical results and analysis. *Biotechnol Bioeng* **104**:275-289.