EPFL

Thèse n° 10 568

# Developing structure-oriented models for enzymatic hydrolysis of lignocellulosic biomass

Présentée le 11 juin 2021

Faculté des sciences de base
Laboratoire des procédés durables et catalytiques
Programme doctoral en chimie et génie chimique

pour l'obtention du grade de Docteur ès Sciences

par

**Jessica Charlène ROHRBACH**

Acceptée sur proposition du jury

Prof. B. Fierz, président du jury
Prof. J. Luterbacher, directeur de thèse
Prof. T. Jeoh, rapporteuse
Prof. H. Mayes, rapporteuse
Prof. V. Hatzimanikatis, rapporteur

École polytechnique fédérale de Lausanne

2021

# Acknowledgements

This doctoral research has been quite a rollercoaster both research- and emotion-wise. Looking back, you realise that this journey would not have been the same without the people that surrounded you from the beginning or that you met along the way.

I would like first to express my deepest gratitude to Prof. Jeremy Luterbacher, who welcomed me in his research group and accompanied me throughout this journey. For always being supportive, finding the right words when the motivation was low and being indulgent with my chaotic organisation, I sincerely thank you.

I'm not sure I would have been able to go through this adventure without my two PhD acolytes from other labs, Adrian Pulgarin and Camille Wolf. Vinci lunch-debates, Arcadie coffee breaks, Sat late evening beers, chaotic Worms games, weekend hikes... these are memories that I'll always cherish. Adrian, thank you for wit, your empathy, your passionate spirit. Camille, thank you for your sassiness, your strength, your creativity. You both made me grow as a person and brought so much in my life, and I'll never be able to thank you enough for your friendship.

Working in a laboratory is way more than just research, and throughout the years you encounter vibrant minds and personalities. I'm thankful for my former coworkers that I'm lucky to count now among my friends; Jher Hau, for your support and assumed quirkiness, Ydna, for your fiery personality, Florent, for your ironic humour, Wu, for your playful mischievousness,

**Acknowledgements**

Stefania, for all the entertainment, support and amazingness you brought to this PhD journey that I wouldn't even know where to start, Lorenz, for your lovely fire-y smell, Louisa, for your kindness but most importantly your bold Choréoké talents, Jean, for your amazing impersonator and dancing skills, Chloé, for your enthusiasm tinted with cheekiness, Luka, for your on point music recommendations, Nathalie, for your help and piquant spirit, and for everyone else that made LPDC what it is.

I would like to also thanks all my friends who where here in the best and worst moments; the EPFL crew, Marine Bouduban, Sylvie Hauert and Solène Oberli, the Catherine gang, Freya Joessel, Amine El Amrani and Audrey Vuillemez, the Neuchâtel band, Vanessa Küffer, Jessica Oberson and Nathalie Veillard, and the Let's go on an Adventure! team, Ismaël Tall and Vincent Benoun. A PhD can be a rocky road, but no matter the distance, some friendships will carry you and push you to stay strong. Nathalie Veillard, thank you for your support, your wit and for the long empowering discussions that helped me grow. Vincent Benoun, thank you for having always been there despite all my antics and my egregious displays of *mauvaise foi*, I'll never be able to repay you for having been such a strong pillar during those years.

Last but not least, I would like to thanks from the bottom of my heart my family, Francine, Daniel and Jonathan Rohrbach, who made me the person I am today and supported me unconditionally even in my most cranky days. I dedicate this research to my parents, Francine and Daniel Rohrbach, whose absolute love and support carried me to this moment.

*Lausanne, January 15, 2021* *Jessica*

# Abstract

With the unsustainable use of fossil fuels increasing strains on human institutions and ecosystems, the development of a renewable energy alternative is of paramount importance. Second generation biorefineries, based on the production of fuels and chemicals from lignocellulosic biomass, appear as attractive alternative to their non-renewable counterparts. Within the multiplicity of existing valorisation routes, enzymatic hydrolysis of lignocellulosic biomass into its constituent sugars has generated considerable interest, notably in the context of biofuels production. However, major hurdles stemming from the intricate structures of the lignocellulosic feedstocks still impede large-scale deployment of this process.

The multiplicity of highly intertwined spatiotemporal factors impacting the enzymatic hydrolysis of lignocellulosic biomass represents a challenge to rationally design efficient hydrolysis process. Here, we propose a theory-based modelling framework relying on pore-diffusion and surface reaction to explore the effects of recognised bottlenecks on the enzymatic hydrolysis of lignocellulosic substrates. The model is based on a set of partial differential equations describing the evolution of the substrate morphology to investigate the interplay between experimental conditions and the physical characteristics of biomass particles as the reaction proceeds. The overall quantity of cellulase present in the hydrolysis mixture is carefully considered to investigate its interplay with the available accessible cellulose surface. Also,

non-uniformity in terms of cellulose accessibility and cellulose digestibility are introduced in the model to weight their influence on observed hydrolysis rates. Finally, deactivation mechanisms are considered through unproductive adsorption of cellulases on both cellulose and lignin fractions, with the existence of such phenomena alleged to be critical in the efficiency of the hydrolysis process.

Based on predictions of our model, we were able to confirm the critical role of cellulose accessibility, as defined by the combination of particle size, porosity and accessible cellulose surface, in dictating early reaction rates for a range of pretreated beech wood substrates. While high biomass loadings should be favoured to improve enzyme penetration in the substrates, high enzyme loadings going beyond the initial number of cellulose adsorption appeared beneficial in notably two cases: (i) to promote internal diffusion in large particles and (ii) counteract undesired enzyme adsorption on lignin. For the latter, a relatively low increase in enzyme loading was sufficient to offset the resulting slowdown. We also showed that the existence of structural heterogeneities, and in particular non-uniform pore volume distribution within the lignocellulosic samples, contribute to the rate slowdown observed at later stage of the hydrolysis, while not explaining it in its entirety. Unproductive adsorption to cellulose, coupled to decrease in the cellulase efficiency at the cellulose surface, appeared as major contributor to the rate slowdown. Overall, we show how the use of a theory-based model can help decouple and evaluate the effects of key factors in the enzymatic hydrolysis of lignocellulosic biomass. As such, our model can help pave the way towards efficient integrated rational design strategies for enzyme process engineering for biomass conversion.

# Résumé

Avec l'usage non durable d'énergies fossiles augmentant les pressions sur les institutions humaines et les écosystèmes, le développement d'une énergie alternative renouvelable est d'importance cruciale. Les bioraffineries de seconde génération, basées sur la production de carburants et produits chimiques à partir de biomasse lignocellulosique, apparaissent comme attrayantes pour remplacer leurs homologues non-renouvelables. Au sein de la multitude de chemins de valorisation existants, l'hydrolyse enzymatique de la biomasse lignocellulosique en ses sucres structuraux a suscité un intérêt considérable, notamment pour la production de bioéthanol. Cependant, des obstacles majeurs, notamment liés à la structure complexe des matières premières lignocellulosiques, bloquent le déploiement à large échelle de ce procédé. La multiplicité et l'interdépendance des facteurs spatiotemporels affectant l'hydrolyse enzymatique de la biomasse lignocellulosique représentent un défi dans la conception de procédés d'hydrolyse efficients. Ici, nous proposons un modèle théorique basé sur la diffusion et réaction en surface des enzymes au sein de la structure poreuse afin d'explorer les effets d'obstacles reconnus dans l'hydrolyse enzymatique de la biomasse lignocellulosique. Le modèle se base sur une série d'équations différentielles partielles décrivant l'évolution morphologique du substrat, afin d'étudier les liens entre conditions expérimentales et caractéristiques physiques du substrat au cours de la réaction. La quantité d'enzymes est soigneusement mise en perspective

avec la surface de cellulose disponible. De plus, les influences individuelles de non-uniformités en terme d'accessibilité et de digestibilité de la cellulose au sein du substrat sont analysées et comparées entre elles. Finalement, des mécanismes d'adsorption non-productifs d'enzymes à la surface de la cellulose et de la lignine sont introduit afin d'évaluer leur incidence sur la vitesse d'hydrolyse, ces deux effets ayant été avancés comme critiques sur l'efficience du procédé.

Basé sur les prédictions du modèle, nous avons confirmé le rôle central de l'accessibilité à la cellulose, facteur dépendant de la surface de cellulose disponible, de la porosité et taille de la particule, dans la phase initiale de l'hydrolyse. Bien qu'une grande concentration de particules en milieu réactionnel soit favorable pour une diffusion efficace des enzymes, un nombre d'enzymes excédent le nombre de sites d'adsorption sur la cellulose est avantageux afin de promouvoir une diffusion rapide dans les substrats larges et/ou compenser la perte d'enzymes irréversiblement adsorbées sur la lignine. Dans ce dernier cas, une relativement petite compensation est suffisante pour contrer cet effet indésirable. Nous montrons également que l'hétérogénéité du substrat, notamment en terme de porosité, contribue à la baisse de vitesse de réaction observée dans la phase ultérieure de l'hydrolyse, sans pour autant l'expliquer dans son entièreté. L'adsorption non-productive d'enzyme sur la cellulose, associée à une baisse d'efficience dans l'hydrolyse, apparaît comme un facteur majeur dans la perte de vitesse de réaction observée. Globalement, nous démontrons l'utilité de notre modèle pour découpler et évaluer les effets de facteurs clés dans l'hydrolyse enzymatique. En tant que tel, notre modèle contribue au développement rationnel de procédés basés sur l'hydrolyse enzymatique de la biomasse lignocellulosique.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Global energy situation: current and prospects

Faced with global warming, one of society's main challenges for the 21st century resides in the need to develop a sustainable energy alternative to current non-renewable resources. The radical modifications in agricultural and industrial practices at the heart of the industrial revolution in the mid 19th century have allowed rapid economic developments, but at a significant cost for the environment. With economies largely relying on fossil energy, together with increased energy demand and global consumption, concentrations of greenhouses gases (GHG) in the atmosphere have continuously increased to unprecedented levels, leading to a rapid warming of the Earth's climate[1] [Figure 1]. On the Earth's timescale, the swiftness of this anthropogenic change has not only pushed the limits of the natural ecosystems' resilience and adaptative capacity, but also increased strains on human systems. These strains have had and will continue to have detrimental impacts on political, economic and social structures[2–4]. Based on a business-as usual scenario, the average global temperature is predicted to increase to up 4°C by the end of the century, with geographical disparities leading to rise in temperature

far more significant in some regions[5]. In cases where there is a lack of immediate action, adaptative measures will likely prove insufficient to counter the adverse effects of this predicted global warming[2,3]. In this context, significant efforts must be directed to accelerate the decarbonisation of current economies to curb GHG emissions.

With mounting evidences of the disastrous impacts of the environmental crisis on societies, GHG mitigations, long associated with economic decline, now represent an economic opportunity for numerous countries[6–8]. While the deployment of clean and sustainable energy is technologically and economically feasible, significant challenges remain, which include the lack of global strong political commitments to tackle the issue[9]. Energy transition, i.e. the phasing out of a fossil-based energy system with a move towards the use of carbon-neutral resources, is most likely going to rely on a combination of technologies and energy sources rather than a single solution[10,11]. While share of renewables in the global energy supply have only marginally increased since 2010 (representing about 14% of the energy mix in 2017[12]), contributions of solar and wind power have grown rapidly in the past decade, led by the cost reduction of these technologies making them, if not cheaper, at least competitive to fossil-based electricity[13]. Coupled to extension of geothermal and dispatchable hydropower as well as the deployment of tidal/ocean power technologies, renewable electricity appears as a central element in the decarbonisation roadmap, thanks to its low cost and its potential usage across the energy sectors. With the redesign of power grids and innovations in storage technologies, a recent report has even predicted that up to 50% of the global primary energy mix could be supplied by renewable electricity by 2050[11].

As important as its role in a decarbonised economy could be, electrification of the energy system falls short in some key sectors. Energy intensive (e.g. steel and iron, cement) and petro-

chemicals industries, as well as the aviation and shipping sectors, require input feedstocks and high energy density that is difficult to achieve by electricity alone and thus will likely largely rely on carbon-based fuels for the foreseeable future[11]. Such cases call for the development of clean, economically competitive and widely available renewable energy- and carbon-sources in order to supplant their current dependence on unsustainable fossil resources. Currently, only a few candidates are being seriously studied as viable replacements for fossil fuels. These include green hydrogen, $CO_2$-to-chemicals through the use biomass, or generation of fuels and chemicals through chemical or electrochemical conversion of carbon dioxide[9,11].

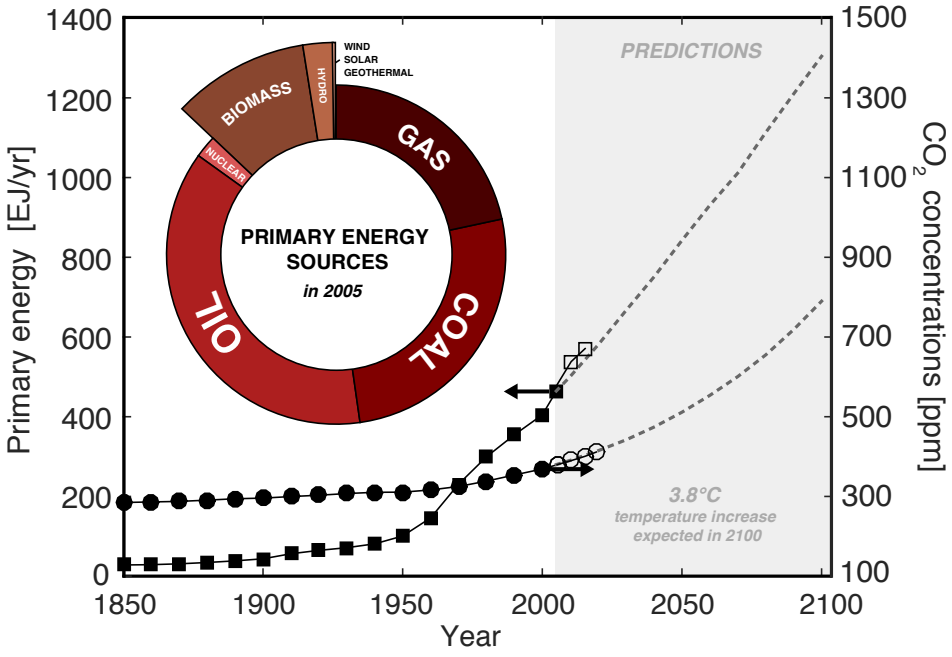



Figure 1.1 – *Evolution of global primary energy supply (■) set against atmospheric $CO_2$ levels (•) since the industrial revolution[5]. Projections to 2100 are taken from the shared socioeconomic pathways (SSP) database for the middle road model (SSP2 model with IAM marker), representative of a business-as-usual scenario from 2005[5]. For comparison, actual primary energy supply[12] (□) and atmospheric $CO_2$ concentrations[4] (○) are shown .*

### 1.1.1 Role of biomass in the energy transition

Thanks to its wide availability and versatility, biomass appears as a viable renewable carbon source to fossil feedstocks in sectors requiring carbon-based materials and/or high energy density dispatched on demand. Bioenergy, i.e. energy extracted from organic materials, has long been part of the primary energy mix to provide low-heat through direct combustion of wood and charcoal, mainly in domestic settings[14]. However, this traditional use of biomass, which still represents a significant energy source in developing countries, has been associated to numerous environmental and social issues. In addition to being an important source of indoor pollution, the unsustainable management of feedstocks have been a major cause of deforestation and decline in soil quality in the last decades, particularly in developing countries[9,15,16]. By contrast, the deployment of modern biomass usage, i.e. a use of biomass different from its direct combustion to produce low quality heat, can have many socio-economic positive impacts[14]. Utilisation of the chemical energy stored in biomass and more efficient generation of heat and electricity could allow biomass to supply energy for a wider range of sectors than just domestic heat use. This not only includes transport and shipping with the expansion of biofuels but also industrial heating processes with bio-coal[17–19]. As such, bioenergy can be a source of employment, secure current and long-term domestic energy supplies and reduce GHG emissions as long as feedstocks are managed in a sustainable fashion[9,17,18].

Selection and management of resources is a key pillar in the development of bioenergy and mitigation of GHG. Several conditions must be met for a bioresources to comply with sustainability and profitability. Assessing the viability of a specific feedstock is a complex process with outcomes highly dependent on the considered supply chain and the temporal management of the resource, as well as the political, social and economic context[9,18]. In particular, bioenergy

feedstocks should not threaten food security by competing with subsistence crops or increase the strain on local ecosystems by relying on an intensive use of resources such as water or soil nutrients. In addition, biomass represents a major carbon reservoir, with an estimate of 31% of anthropogenic carbon dioxide stored in forests alone every year, as well as an important carbon shrink[20]. This double role as both carbon source and carbon sink poses additional challenges in the assessment of the sustainable potential of biomass in the energy transition. As such, the question of appropriate forest management and the balance between afforestation, deforestation, conservation and restoration remains heavily debated in literature[21–24]. All these reasons have led to diverging estimates for the potential of bioenergy, with projected values ranging from less than 50 EJ/yr to up to more than 1000 EJ/yr in 2050[9,25,26]. As the exact extent to which bioenergy can sustainably cover future energy needs is still debated, prioritizing the use of low value biomass and unvalorized residues as biofeedstocks as well as applications optimising the energy extractable for a given material are likely important. In this perspective, five types of feedstocks are commonly put forward as promising bioenergy sources:

- Algae, both micro- and macro-algal feedstocks
- Agricultural residues including corn stover, wheat straw and rice husk among others
- Energy crops, such as switchgrass, silvergrass and bamboo
- Forestry residues resulting from timber felling and leftovers of wood-based products manufacturers
- Municipal and animal wastes, such as manure, wastewater sludges and domestic organic wastes

Each of these feedstocks presents their own geographic and application scope, associated to a different set of advantages and disadvantages[27–29]. In all cases, challenges remain for their large-scale implementation as energy sources, notably in terms of supply chain efficiency and technologies[17,18]. In particular, while technologies for heat and electricity generation from biofeedstocks are well established, the development of biorefineries, i.e. the production of biofuels and biochemicals, still require technologies that are currently under development to be cost-effective and scalable[30]. Moving away from the unsustainable use of readily available carbohydrates, such as those in corn or sugar cane, significant efforts are made towards the deployment of biorefineries based on lignocellulosic substrates, algae and other waste streams. Commonly, second-generation biorefineries refer to productions based on lignocellulosic substrates that are not used for food production and forming most of the energy crops or biomass residues, while third generation biorefineries rely on algae and microbial $CO_2$ fixation from waste streams gasification[31,32].

## 1.2 Lignocellulosic biomass for fuels and chemicals production

Second-generation biorefineries are based on the deconstruction of lignocellulosic biomass into their constituent building blocks and subsequent upgrading to a broad range of products, including fuels and added-value chemicals, as well as heat and power. Akin to oil refineries, lignocellulosic biorefineries have an important potential to produce a wide pool of commonly used platform chemicals, but also expand the actual chemical space, thus creating new markets[33]. However, the transition from oil to biomass creates a transformational challenge for chemical manufacture due to the inherent differences between both substrates; crude oil, a liquid composed mainly of hydrocarbons, offers readily available fuels but requires complex reaction schemes to obtain functionalised heteroatomic chemicals. By contrast, lignocellulosic

biomass, a heterogeneous solid, readily contains chemical complexity, but its oxygen-rich and aromatic molecules are embedded in an intricate structure making them hard to access and use [Figure 1.2]. Therefore, the implementation of second-generation biorefineries still faces significant technical challenges, notably due to the intricate structure of lignocellulosic biomass, which makes it recalcitrant to deconstruction[34–36]. This deconstruction's economic success largely lies in the successful valorisation of all three biopolymers constituting the lignocellulosic substrate into bulk and high-value chemicals, as well as materials[37,38].

### 1.2.1 Composition and structure

Lignocellulosic biomass refers to plant biomass composed mainly of three biopolymers: (i) cellulose, representing 20-50 wt. % of the dry plant matter, (ii) hemicellulose, ranging from 15 to 35 wt. %, and (iii) lignin, forming 10 to 30 wt. % of the structure [Figure 1.2]. These three major components account for up to 90 wt. % of the dry plant weight, with the remainder consisting of minor amounts of pectins, proteins, extractives and ashes[29,36,39]. Lignocellulosic biomass covers a wide range of plant families and represent thus a broad spectrum of constituent structures, including at the (i) micro- (cellulose, hemicellulose and lignin molecular structures), (ii) meso- (e.g. cell walls) and (iii) macro- (e.g. grass stalks, tree branches) scales[35]. Two main clades of lignocellulosic plants can be distinguished depending on their reproduction mode, with (i) gymnosperms ('naked' seed) comprising notably softwoods (e.g. conifers, ginkgo) and ferns and (ii) angiosperms ('enclosed' seed), comprising hardwood (e.g. oak, birch, beech, bamboo, palms) as well as grass (e.g. switchgrass, rice straw) and herbaceous plants (e.g. banana, pineapple)[36].

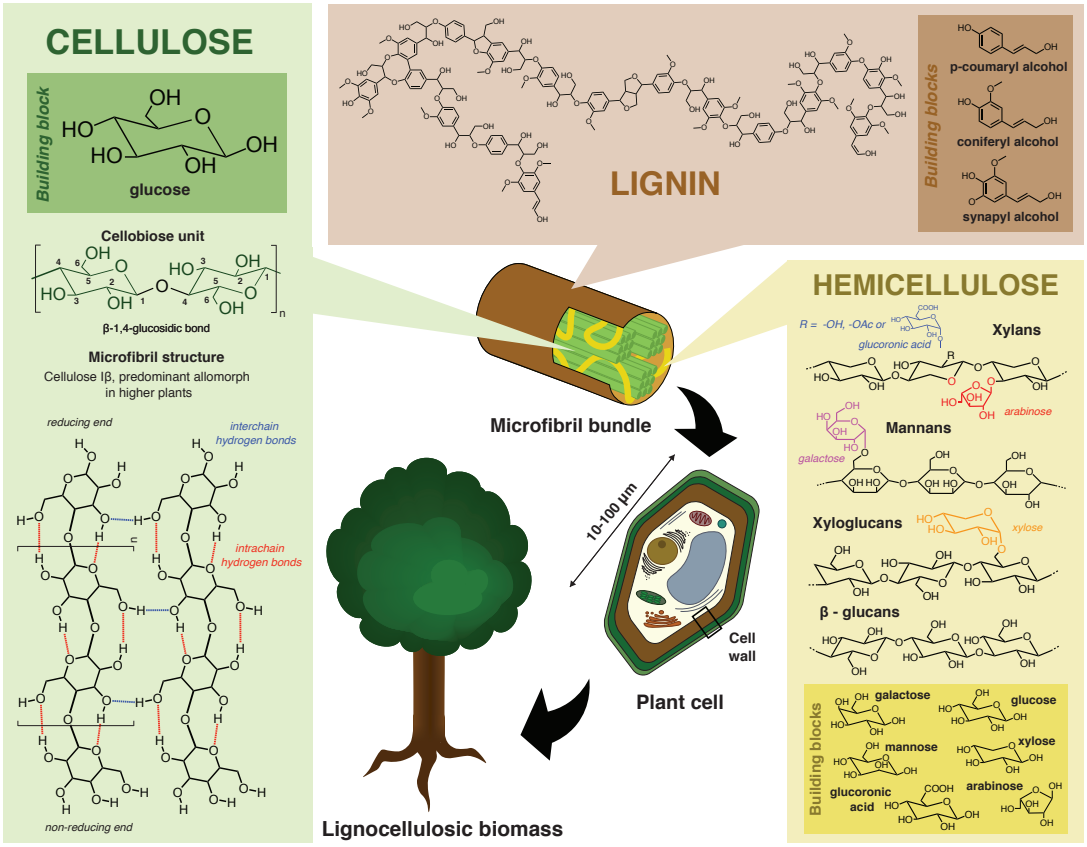



Figure 1.2 – *Schematic overview of the lignocellulosic biomass structure and its main structural components. Cellulose microfibrils, interwoven in a complex hemicellulose-lignin matrix, aggregate into large structures forming the basis of the plant cell wall. The latter offers physical resistance to the cell, allowing it to organise in highly specific tissues at the core of the different plant ultrastructures. Both cellulose and hemicellulose are polymers based on carbohydrates, contrasting to the polyphenolics chains constituting lignin.*

#### 1.2.1.1 Cellulose

Cellulose is a complex macromolecule composed of linear chains of glucose units covalently linked by $\beta(1,4)$-bonds. Glucan chains are polarised structures of variable length, presenting a reducing end formed by the unmodified $C_1$-hydroxyl group, and a $C_4$-hydroxyl group at

their non-reducing end [Figure 1.2]. Their degree of polymerisation (DP), i.e. the number of glucan monomers constituting the chain, is highly dependent on the source and the potential treatment applied to the cellulose to isolate it (making its native length difficult to measure), with typical values of DP lying between 100 to 10 000 glucan units[36]. Reported chain lengths for cotton and wood fibers range from 800 to 10 000 units, while extracted wood pulp exhibit smaller length between 300 and 1 700 monomers[40].

By providing a suitable conformation of the glucan units in the chain, the $\beta(1,4)$-glycosidic linkages allow cellulose strands to assemble into a crystalline structure, the microfibril, through a network of strong intra- and inter-molecular hydrogen bonds as well as weaker Van der Waals interactions [Figure 1.2]. While the exact size and structure of the microfibrils remain under discussion, recent studies tend to corroborate a 18 or 24 chains model of ~3.5 nm diameter for the cellulose microfibril, with possible fusion or aggregation leading to thicker complexes[41–43]. Within the microfibril, several types of chain arrangements can exist to form the crystalline complex, with variations in the relative position of adjacent chains (parallel or anti-parallel) and types of H-bonds forming the network leading to different cellulose density and degree of crystallinity (or index of crystallinity CrI)[44]. While the two crystal allomorphs $I_\alpha$ and $I_\beta$ are naturally occuring, cellulose II, $III_I$ ,$III_{II}$, $IV_I$, and $IV_{II}$ can be obtained by thermochemical treatments of the thermodynamically less stable native cellulose I[45,46]. In plants, both types of cellulose I, $\alpha$ and $\beta$, coexist with a predominance of the $I_\beta$ allomorph, along with amorphous and para-crystalline domains[47]. The proportion and distribution of these less ordered cellulose fractions depend on the plant species and affect the overall mechanical and physical properties of the macrostructure constituted by the cellulose fibers.

#### 1.2.1.2 Hemicellulose

Hemicelluloses comprise an ensemble of polysaccharides composed of pentoses (notably xylose and arabinose) and hexoses (such as glucose, galactose, mannose) ranging from linear homopolymers to highly branched heteropolysaccharides [Figure 1.2]. Depending on the composition of the polymer backbone, four main classes of hemicellulose can be distinguished[36,48]: (i) xylans, (ii) mannans, (iii) mixed-linkage $\beta$-glucans and (iv) xyloglucans .

Xylans are the predominant structures in hardwood as well as most grasses, and consist of chains of xylose linked by $\beta$(1,4)-xylosyl bond, that can either be linear (homoxylan) or substituted on the xylose with arabinose (arabinoxylan AX), glucoronic acid (glucoronoxylan GX) or both ((arabino)glucoronoxylan and (glucorono)arabinoxylan, AGX and GAX respectively). The xylose backbone can bear acetyl groups on both the O-2 and O-3 position of the xylose units in addition to possible methylation of the glucoronic acid decorations [Figure 1.2]. The composition of xylan, as well as the position of the modifications on the xylose chains, are strongly dependent on the biomass source. While homoxylan is rarely extracted in higher plants, glucoronoxylan is the dominant form of hemicellulose in woody dicots and xylan containing arabinose can be isolated from monocots and softwoods[48,49].

The main component of hemicellulose in softwood are backbones of mannans that consist of either a mixture of mannose and glucose (glucomannans GM) or mannose only, which are branched with galactose side-chains (galactoglucomannans GGM and galactomannan GaM respectively) [Figure 1.2]. In both cases, the galactose residues connect to mannose units in the main chain. As for cellulose, the last two groups of hemicellulose are based on a glucose backbone. While xyloglucans, a $\beta$(1,4)-glucosyl backbone branched with xylose residues, are a major component in vascular plants, the mixed-linkage-$\beta$-glucan hemicellulose are specific

to a type of monocots and consist of an unbranched backbone of glucose units connected either through $\beta(1,4)$- or $\beta(1,3)$-glucosyl linkage[38,49,50].

By contrast to the crystalline cellulose, hemicelluloses, due to the less ordered structure, presence of branching and lower degree of polymerisation, are amorphous polymers, which are less recalcitrant to hydrolysis[37]. However, the susceptibility of the hemicellulose fraction to dissolve will strongly depend on their intrinsic nature, determining not only their spatial organisation and interchain interactions, but also the strength of association with other biopolymers. As such, their composition, molecular weight as well as the presence of regular motifs in the succession of building blocks of main chain and relative spacing of the residues on the backbone will affect their recalcitrance to hydrolysis.

#### 1.2.1.3 Lignin

Lignin is an irregular polyphenolic polymer derived from three main building blocks, known as monolignols: (i) synapyl alcohol, (ii) coniferyl alcohol and (iii) p-coumaryl alcohol. Once linked, these three subunits are respectively referred to as the syringyl (S), guaiacyl (G) and p-hydroxyphenyl (H) subunits [Figure 1.2]. The relative quantity of these building blocks varies depending on the biomass source. While softwoods (e.g. ferns, conifers, ginkgo) mainly contain guaiacyl as subunits, hardwoods are composed of a mixture of syringil and guaiacyl with variable amounts of p-hydroxyphenyl[36].

By contrast to cellulose and hemicellulose, lignin presents a less regular structure with monolignols linked by various C-O and C-C linkages stemming from the free radical polymerisation mechanism that occurs during its biosynthesis. Due to its complexity and reactivity, the characterisation of lignin proves to be a challenge on its own, with lignin extraction methods altering its structure through intra- and inter-molecular condensation reactions, and a lack of

reliable analytical methods leading to uncertainties about its native structure[51]. Notably, while long assumed to be a highly branched polymer, recent advances in lignin isolation processes and analytical methods have pointed to a more linear model to represent the native lignin structure[52,53]. The reported molecular weight of lignin extracted with methods minimizing physicochemical modifications of the native structure, e.g. by enzymatic hydrolysis of the polysaccharide fractions, lead to values ranging from 2500 to 10 000 g/mol[54]. In addition to size and organisation, the relative ratio of monolignols play an important role in the physicochemical properties of the lignin, notably affecting the swelling capacity of the cell wall and the interaction of the lignin with its co-polymers.

#### 1.2.1.4 Plant ultrastructure

Driven by survival, plants have evolved into highly complex structures to protect their structural carbohydrates from both microbial and animal attacks, and thrive in diverse environmental conditions. These protective strategies have translated into the development of dedicated macrostructures, such as the plant vascular system to allow plant to grow tall to prevent being eaten by animals, that are in turn reflected into specialised plant cells[55]. At least 35 types of plant cells exist, with compositions and structures adapted to the tissue they compose[35]. Plant cell walls, the most abundant source of reduced renewable carbon on earth, are typically composed of three distinct layers protecting the organelles contained in the cell cyctoplasm: (i) the middle lamella, the outermost layer that helps cells to stick together, (ii) the primary cell wall (PCW), an extensible layer composed mainly of polysaccharides and (iii) the secondary cell wall (SCW), a rigid nanocomposite providing strength in many mature tissues[56,57]. The middle lamella is mainly composed of pectins, a galacturonic acid-rich branched heteropolysaccharides, forming a gel like structure between cells notably through

calcium-mediated cross-linkages[58]. Cell walls are dynamic systems, not only changing as a function of growth and tissue functions, but also in response to environmental variations. A comprehensive understanding of their organisation and composition still remains elusive despite major research efforts. PCWs provide a combination of strength and extensibility to allow cell morphogenesis while supporting cell expansion upon water uptake. These relatively thin protective layers are highly hydrated structures due to the presence of pectins and hemicellulose, mainly in the form of xyloglucans that create a pore network between the layers of cellulose microfibrils known as lamellae[56,57,59]. Pectins make up for up to 50% of the dry weight of PCWs, with fraction of cellulose and hemicellulose ranging between 15-40% and 20-30% respectively. While pectins dominates the composition of PCW, their contributions to the total dry weight remain low due to the predominance of SCW over PCW in plant tissues. While some cells retain only a primary cell wall once mature, most develop a thicker rigid layer deposited onto it, which is the secondary cell wall. To provide strength and protection from degradation, SCWs present a more complex layered structure, with compact cellulose lamellae embedded in an intricate biopolymer matrix composed of hemicellulose and lignin[56]. Experimental evidences suggest the existence of large cellulosic structures exceeding 10 and up to 60 nm diameter, likely resulting from the association of individuals microfibrils, in the SCWs of both soft- and hardwoods. While their formation mechanism remains somewhat unclear, the relative amount of lignin and hemicellulose in the cell wall affects the cellulose aggregate size[60]. In addition to its role in cellulose microfibril coalescence, hemicelluloses also act as a bridge between cellulose and lignin, interacting with both biopolymers through polar and non-polar forces and covalent interactions. The evenly substituted and regular conformation in hemicelluloses favour hydrophilic interactions with the ordered cellulose

through a network of hydrogen bonds, while denser, unevenly packed decorated hemicellulose chains are likely to associate with the hydrophobic surface of the cellulose surface. By contrast, hemicellulose is covalently cross-linked to lignin through different attachment points, leading to so-called lignin-carbohydrate complexes (LCCs), which create a tight structure that is hard to penetrate. These LCCs are not yet fully characterised and subject to debate in literature, with 5 types of bonds argued to exist based on ether (benzyl ethers, phenyl glycosides), ester (ferulate/coumarate esters and $\gamma$-esters) or acetal/hemiacetal- linkages[61,62].

#### 1.2.1.5 Biomass recalcitrance

Overall, difficulties in accessing the plants biopolymers arise from the various levels of complexity existing within the lignocellulosic biomass structure, from the biopolymers chemical compositions to the heterogeneities in plant tissue in terms of composition, density and cell organisation[63]. In addition to the inherent diversity existing across the plant kingdom and within the organisms themselves, biomass undergoes seasonal growth in most parts of the world with environmental conditions affecting the plant physiology. All these effects participate in forming so-called biomass recalcitrance, i.e. the overall hindrance impeding the bioconversion of lignocellulosic substrates, and the struggle to develop efficient, comprehensive processes for their deconstruction.

### 1.2.2 Biorefinery concept

Apart from direct thermochemical treatment of lignocellulosic biomass to produce high energy density carriers such as bio-coal, the most common approach for the production of chemicals and fuels from biomass revolves around four main steps that are pretreatment, fractionation, depolymerisation and upgrading[54,64]. Each of these steps can either be performed sequen-

tially or in a combined fashion [Figure 1.3]. The initial stage in the biorefinery consists of a pretreatment step altering the mechanical, physical and chemical properties of the lignocellulosic substrate at various degrees of severity to facilitate the separation of its structural biopolymers. Pretreatments cover a broad range of processes, from mechanical comminution of the biomass macrostructure to thermochemical treatments partially hydrolysing one or more of the biopolymers[65,66]. As such, pretreatment and fractionation are usually two steps that are closely intertwined. Broadly, fractionation approaches, based on either the solubilisation of the carbohydrates or the removal of lignin, lead to two main product streams that are (i) lignin-rich and (ii) carbohydrate-rich, the latter containing derivatives of the cellulose and hemicellulose fractions. Reflecting the complexity of the lignocellulosic substrate, no predominant fractionation strategy has yet emerged, and several methods have been developed based on either chemo- or bio-catalytic reactions to obtain the separated building blocks in their polymerized form or directly in their monomeric form. As the chemical and physical characteristics of the different fragments heavily depends on the process applied, the efficiency of the fractionation should be assessed in terms of the type of input feedstocks and target end-product that is desired. For example, delignification, made by thermochemical treatment, leaves the carbohydrates in the form of a solid pulp and lignin as a depolymerised oil or solid residue in form of a residue or precipitate. On the other end, soluble monosaccharides can be obtained by enzymatic or acid-catalysed hydrolysis of the cellulose and hemicellulose fractions, leaving lignin as a precipitate/residue with varying degree of alteration compared to its native structure[54].

While the use of extracted cellulose pulp from wood has a long history tracing notably in papermaking, the upgrading of bio-based carbohydrates into platform chemicals and fuels

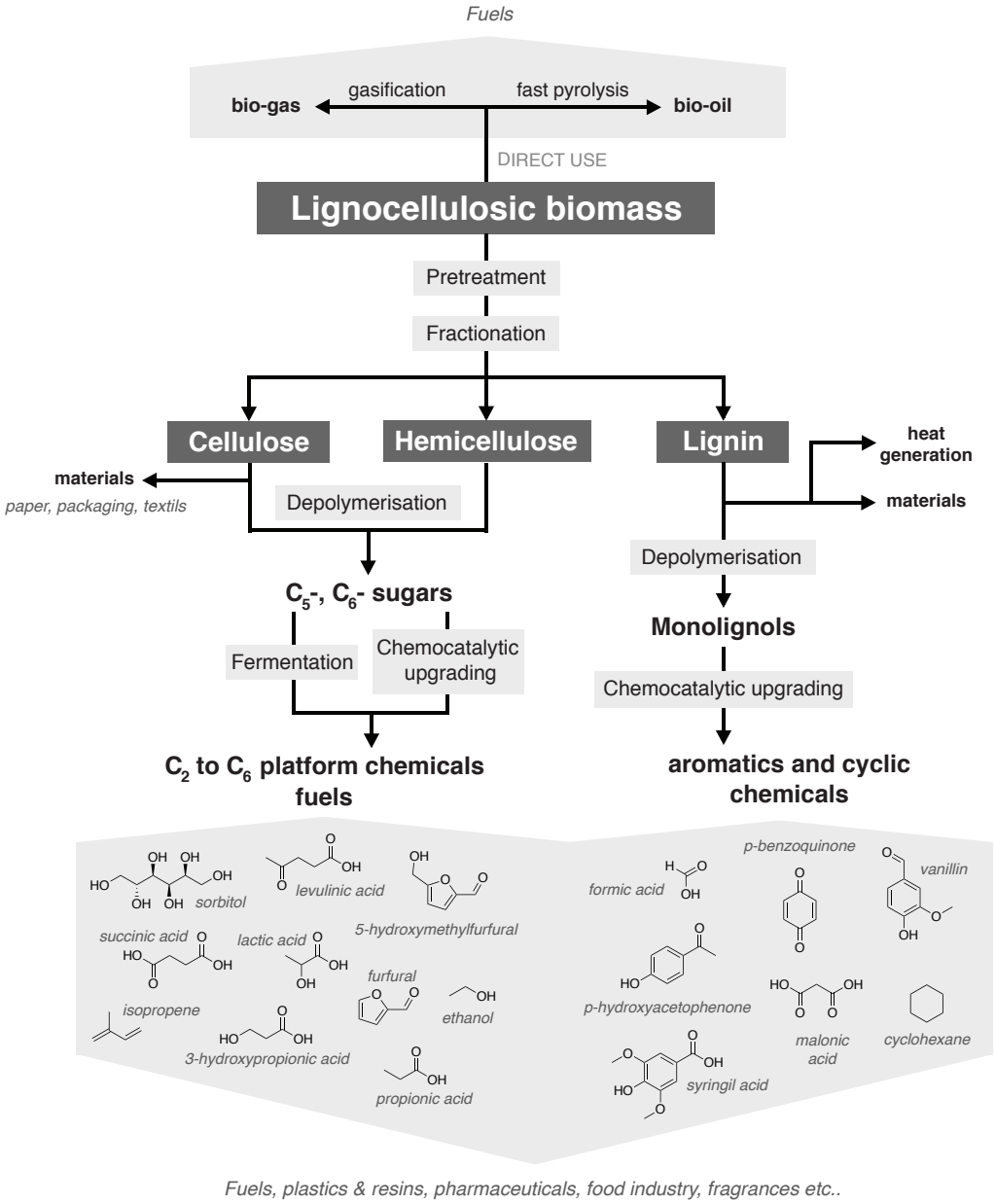



Figure 1.3 – *Broad overview of the biorefinery concept. Two mains routes can be distinguished in the use of lignocellulosic as energy source: (i) direct use through thermal treatment, leading mainly to fuels, and (ii) indirect use, based on a frationation-depolymerisation-upgrading sequence and leading to a vast pool of chemicals/materials.*

has a more recent history, with periods of strong interest in bio-based products related to shortage/crisis in the petroleum industry since the beginning of the 20th century[67]. Due to their reactivity and their presence in various metabolic pathways, numerous upgrading routes can be applied to carbohydrates[68,69]. Fermentative organisms can be used to notably produce various alcohols and organic acids from glucose through various metabolic pathways. In addition, a network of chemocatalytic reactions based on hydrolysis, dehydration, hydrodeoxygenation, hydrogenolysis and oxidation can be used to obtain additional building blocks from sugars[33]. These two approaches have led to a broad range of platform molecules. Some are used for fuels with upgrading strategies based on maximising the energy content of the end-product. Others target value-added chemicals, where the efficiency of the process relies on the atom- and functionality-conservation[69]. While both types of targets products are of interest, the choice of upgrading pathways to deploy within biorefinery is its own challenge and depends on multiple economic, technical and chemical factors. In 2010, the Department of Energy (DOE) highlighted 14 relevant $C_2$-$C_6$ targets from the broad spectrum of primary chemicals derived from the fermentation or catalytic upgrading of $C_5$- and $C_6$-carbohydrates, including not only fuel substitutes such as ethanol, but also potent platform chemicals such as various furans and organic acids[70].

In contrast, catalytic conversion of lignin monomers has long been overlooked, with lignin often considered as a low-value by-product of the carbohydrates stream. This lack of consideration has resulted from the lack of efficient lignin depolymerisation strategies, coupled to several delignification processes often strongly degrading the native lignin structure into something much more recalcitrant, limiting the use of residual lignin to a cheap source of energy for heat generation. However, with the valorisation of lignin increasingly recognised

as an essential element in the success of lignocellulosic biorefineries due to its significant presence within biomass, high energy density and its rich chemical composition, research on lignin has increased dramatically in recent years[71,72]. Developments of fractionation strategies that avoid the formation of a recalcitrant lignin have expanded the chemical space occupied by bio-derived compounds, notably offering access to high-valued aromatics[54,73–75].

## 1.3 Enzymatic hydrolysis of lignocellulosic biomass

For the carbohydrates stream, a major depolymerisation pathway is the enzyme-mediated hydrolysis of cellulose, leading to monomeric sugars available for chemical upgrading to added-value chemicals and/or fermentation into biofuels. Similar to first-generation of biofuel processes based on readily available glucan monomers, this route has been heavily investigated in the context of bioethanol production, where resulting sugars are converted into ethanol using fermentative organisms, notably *Saccharomyces cerevisiae*. Such conversion can be performed in a stepwise- or consolidated-fashion, with fermentative yeasts added conjointly with the hydrolysis enzymes (SSF, simultaneous sacharification and fermentation), mid-saccharification (HHF, hybrid hydrolysis and fermentation) or at the end of the hydrolytic process (SHF, separate hydrolysis and fermentation). Each method presents its own challenges, such as enzyme end-product inhibition in the case of SHF or finding operating conditions suitable for both enzymes and yeasts in SSF. A more extensive discussion on the fermentative stage can be found elsewhere[30,76], as the focus here will be on the enzymatic step.

While bacteria and fungi have co-evolved with plants to develop complex enzyme systems tailored for cell wall deconstruction, the development of commercially viable sugar production through the large scale enzymatic hydrolysis of lignocellulosic substrates still faces challenges. The conversion of native lignocellulosic materials into fermentable sugars via an enzymatic

route requires multiple steps. The choice of the feedstock can influence the pretreatment method used to decrease its recalcitrance to enzymes and hydrolysis conditions often need to be tailored to the resulting physico-chemical properties of the substrates[77].

The following sections detail the main pretreatment methods investigated in the context of the enzymatic hydrolysis of lignocellulose, the mechanisms behind it and the challenges arising from the interplay between enzymes and substrates properties as well as from operating conditions.

### 1.3.1 Pretreatment methods

One key bottleneck in the enzymatic hydrolysis of lignocellulosic biomass is the low accessibility of carbohydrates to the enzymes, requiring pretreatments that disrupt the intricate biopolymer matrix. Ideally, pretreatment should (i) lead to efficient downstream saccharification by increasing the enzyme penetration (ii) reduce heavy pre- and post-processing of the substrates, (iii) maximise the recovery of the different lignocellulosic fractions by avoiding the formation of undesired side-products, (iv) be environmentally-friendly and safe while minimising energy demand, (v) be applicable to a wide range of feedstocks and (vi) maintain low investment- and operational-costs. Various pretreatment methods have been applied to different types of substrates and are typically classified into five categories: (i) physical, (ii) chemical, (iii) physicochemical and (iv) biological [Figure 1.4] So far, none of them alone, or in combination with others, have fulfilled all the above conditions, with results often highly dependent on the properties of the incoming material[65,66,78].

#### 1.3.1.1 Physical pretreatments

Physical pretreatments are based on the use of mechanical forces to not only reduce the size of the substrate, but also alter its crystallinity and degree of polymerisation. For most types of lignocellulosic biomass, size reduction is an essential step leading to reasonably sized substrates (1-5 cm) that can be further processed, through pretreatment methods that may or may not be coupled with more extensive comminution.

Milling and extrusion are both established pretreatment methods using shear forces to disrupt the recalcitrant lignocellulosic structure. While material is ground with a system of rotating screws at a given temperature (<300°C) in the case of extrusion, milling is performed at ambient temperature on either dry or wet substrates by cutting (e.g. knife mill), crushing (e.g. mill ball), grinding (e.g. wood grinder) or vibrating mills (e.g. vibratory mill)[66]. Less commonly, ultrasonification[79] or microwave-assisted size reduction[80] can be applied to breakdown particles. In all cases, technical and operating conditions will strongly affect the efficiency of the process and the resulting physical alterations to the native material.

#### 1.3.1.2 Chemical pretreatments

Due to its chemical complexity, biomass can be subjected to several alternate chemical strategies to expose the embedded cellulose fibers to the cellulolytic attack. Acids, alkali and organic solvents can be used to break the linkages between the different biopolymers, while ionic liquids and deep eutectic solvents can disrupt the strong hydrogen bond networks existing in biomass.

**Dilute-acid pretreatment** (DAP) relies on the use of acids to hydrolyse the glycosidic bonds between the structural carbohydrates, in particular hemicellulose. To prevent significant

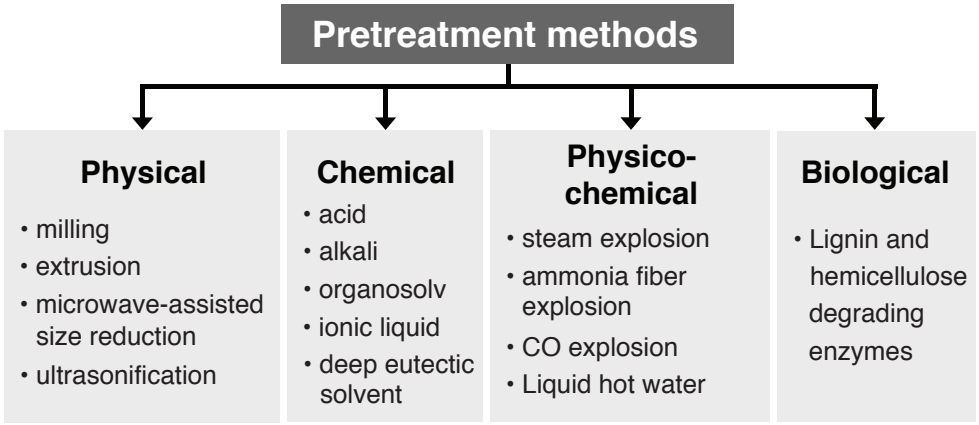



Figure 1.4 – *Possible classification of existing pretreatment applied to lignocellulosic substrates to increase their native low digestibility. Physical pretreatment is a prerequisite for most of the biomass, and is usually followed by a thermochemical and/or biological pretreatment step. To enhance the breakdown of the recalcitrant lignocellulosic substrate, different pretreatment steps can be applied sequentially before the enzymatic hydrolysis stage.*

operational issues related to the toxic and corrosive nature of acids, DAPs are often performed at low acid concentrations (less than 10%) and relatively high temperatures (100-250°C) for residence times ranging from seconds to hours. Inorganic acids (such as sulfuric, phosphoric or hydrochloric acids), and in particular sulfuric acid ($H_2SO_4$), have been largely favoured because of their lower cost, but their organic alternative (e.g. formic, oxalic acids) have also demonstrated their efficiencies in altering the lignocellulosic structure[81–83]. The severity of the pretreatment, i.e. the combination of time, acid concentration and temperature, can be adjusted to the feedstock type to maximise its digestibility by enzymes, while minimising negative outcomes. Notably, lignin reorganisation within the cell wall due to temperature exceeding its phase transition temperature has been shown to impact the subsequent hydrolysability of the pretreated material. Such repolymerisation of condensed lignin leads

to formation of lignin droplets deposited on the cellulose fibers and has been observed at pretreatment temperatures[84–86]. Increasing pretreatment severity also leads to higher amount of sugars degradation products (e.g. HMF, furans from the dehydration of carbohydrates). This degradation is accelerated by the dissolution of cellulose at higher severity conditions into monomeric carbohydrates, which are similarly subject to degradation. These undesired reactions lead not only to a loss of targeted carbohydrates, but also possibly to the formation of inhibitors in subsequent processing, which may require costly detoxification processes[87]. DAPs are efficient on both woody and herbaceous residues as well as softwoods. In the case of biomass presenting higher lignin content, such as softwoods, a two-stage pretreatment with variable severity can be applied to prevent extensive degradation of hemicellulosic sugars.

**Alkali pretreatment** , in contrast to DAP, selectively solubilises lignin, notably by targeting the ester bonds existing in the LCCs, and remove the acetyl and uronic acid side groups on the hemicellulose fraction. The process also causes the cellulose to swell, increasing its accessible surface area while decreasing its CrI and DP. Experimental conditions vary broadly depending on the feedstocks and the type of acid used, with reactions conducted at both ambient temperature and 160°C, for residence times ranging from days to minutes, respectively. While sodium hydroxide appears as the most effective base for lignin removal, other bases can be used to treat biomass, such as lime, for its low cost and safe handling, or ammonia, for its easy recovery[66]. Alkali pretreatments are best suited for low-lignin substrates (e.g. herbaceous and agricultural wastes), but their efficiency falls short when it comes to pretreatment of hardwoods.

**Organolsolv pretreatments** rely on various organic or organic-aqueous solvents mixtures to substantially reduce the lignin and hemicellulose content by hydrolysing internal bonds

and inter-linkages in both fractions, leaving highly accessible cellulose. Compared to DAPs, organosolv pretreatments solubilise lignin and hence prevent its recondensation on the cellulose surface. Organosolv pretreatments have been carried out in a wide range of organic solvents, with both those having low- and high-boiling points (e.g. ethanol, THF, acetone, acetic acid and ethylene glycol), with or without addition of a catalysts (acid, base or salts) depending on the solvent nature. Organosolv pretreatments are favourable for a selective fractionation of both soft- and hard-woods[88].

**Ionic fluids** (ILs) have been more recently applied to disrupt the lignocellulosic biomass structure by disrupting the H-bonds network existing between carbohydrates chains. While research on ionic liquids in the context of biomass pretreatment is still new compared to other methods, pretreatments notably based on imidazolium salts have proven their efficiency in improving the enzymatic hydrolysis of IL-pretreated lignocellulosic biomass[89]. The inherent nature of ILs presents numerous advantages (environmentally-friendly, thermostable, non-volatile, recyclable and reusable, versatile), but with major culprits of being expensive and toxic to enzymes. In this respect, the use of deep eutectic solvents, i.e. solvents resulting from a mixture of 2-3 salts presenting melting points lower than each of their individual components, have gained recent interest as alternative to ILs due to the lower cost of the chemicals composing them[90].

#### 1.3.1.3 Physicochemical pretreatments

Physicochemical pretreatments rely on both mechanical forces and chemical alterations to decrease the biomass recalcitrance to degradation. Specifically, in most cases, a sudden change in pressure can disrupt and swell the fibers, with in-situ generated acids or added catalysts promoting bond cleavage.

**Steam explosion** (SE) relies on the permeation followed by rapid depressurisation of the substrate with saturated steam at high temperature and pressure (160 – 240°C / 0.69-4.89 MPa / residence time: seconds to minutes). In a process called autohydrolysis, the hydrolysis of hemicellulose is promoted, in addition to the high pressure and temperature, by the release of acetic acid from the acetyl groups decorating the carbohydrates chains. This effect can be enhanced by the addition of an acid, such as $H_2SO_4$ or $CO_2$, in the reaction media.

**Ammonia fiber explosion** (AFEX) is a process similar to SE, where ammonia-soaked substrates (1:1 to 1:2 ratio biomass:ammonia / 60 – 100°C / residence time: 5-30 minutes) are subjected to sudden pressure release during ammonia fiber explosion pretreatment. And, as in alkali pretreatments, ammonia catalyses the dissolution of lignin and alters the cellulose structure, while forming minimal quantities of inhibitors and degradation products.

**$CO_2$-explosion** relies on the physical properties of supercritical $CO_2$ to penetrate and disrupt the lignocellulosic fibers upon depressurisation, but also to form carbonic acid in the presence of water to hydrolyse bonds.

**Liquid hot water pretreatment** relies on similar mechanisms than SE with the major distinction of not relying on a pressure release to disrupt the lignocellulosic structure. Here, the high pressure is used to prevent water evaporation from the reaction media at high temperature (160 – 230°C), and promote the dissolution of hemicellulose and lignin. By maintaining a neutral or slightly acidic pH (pH 4-7), liquid hot water pretreatment avoid the formation of toxic compounds, but at the cost of high water required.

#### 1.3.1.4 Biological pretreatments

Numerous fungi and bacteria rely on the deconstruction of plant cell wall for their survival with the use of highly selective enzymatic machineries, i.e. cellulases, hemicellulases and lignin-degrading enzymes, and metabolic pathways based on carbohydrates or lignin monomers or both. In the context of pretreatment, the use of hemicellulases in conjunction with ligninolytic enzymes, such as laccases and various types of peroxidases, can promote the access to the protected cellulose fraction. Such pretreatments can be carried out using an enzyme mixture or directly the whole lignocellulolytic organism. While highly selective, enzymatic pretreatments are slow process with residence time spanning days to obtain digestible substrates[66].

### 1.3.2 Carbohydrate-active enzyme mediated saccharification

A wide consortium of organisms, mainly within the fungi and bacteria kingdoms, have developed enzymatic strategies to deconstruct the structural components of plants, and particularly carbohydrates. Biomass degrading enzyme systems are composed of an ensemble of extracellular proteins working synergistically to deconstruct the plant cell walls, either as free enzymes blends or as cellulosomes, i.e. complex scaffolding structure containing up to 100 enzymes attached to the organism cell wall. Free enzymes secreted by filamentous fungi have drawn the most attention within the vast pool of enzymatic machineries relevant in the context biomass conversion[91–93]. This attention has been notably due to their high activity on cellulose and their innate secretion from their producing microorganism, which can be exploited by enzyme producers. In particular, enzymes excreted by *Trichoderma reesei* (commonly abbreviated as *T. reesei*) have been the focus of much industrial interest, leading them to be the main components in most current commercial enzyme cocktails, such as DuPont Accellerase 1500

or Novozymes Cellic CTec2. Significant enhancement in cellulase production by the fungi over the years, through notably strain improvement, protein engineering and/or mix with other organisms secretions, have led to potent commercial enzyme cocktails working synergistically at temperature around 50°C in slightly acidic conditions (pH 5) to hydrolyse cellulosic substrates[94]. Typically, enzyme blends are a combination of cellulases, hemicellulases and other accessory enzymes which help overcome the recalcitrant nature of the lignocellulosic substrates[95] [Figure 1.5].

#### 1.3.2.1 Cellulases

Three types of enzymes constitute the core of *T. reesei* cellulolytic secretion and thus many enzyme blends: (i) cellobiohydrolases (CBHs), which sequentially depolymerise cellulose strands from their free ends, typically into cellobiose units, (ii) endoglucanases (EGs), which hydrolyse internal glucosidic linkages, and (iii) $\beta$-glucosidases, which cleave the released cellobiose in solution into glucose. CBHs and EGs are both part of the glycoside hydrolyses (GH) class, regrouping families of catalytic protein hydrolysing glucosidic bonds between carbohydrates. These two types of enzymes work synergistically to deconstruct the insoluble cellulose, with EGs creating new substrates for CBHs. This synergism enhances the global activity of the blend to higher levels than the simple combination of each separate enzyme activity. GHs often present a modular architecture, with a catalytic domain (CD), tailored to the targeted substrate, connected via a flexible linker to a non-hydrolytic carbohydrate binding module (CBM), protein showing high affinity to cellulose[91].

Two primary CBHs, which are predominant constituent of the excreted blend by *T. reesei* (typically > 70 wt. %), can be distinguished depending on their directionality: (i) Cel7A, which hydrolyses from the reducing end to non-reducing end of the cellulose strand in a mechanism

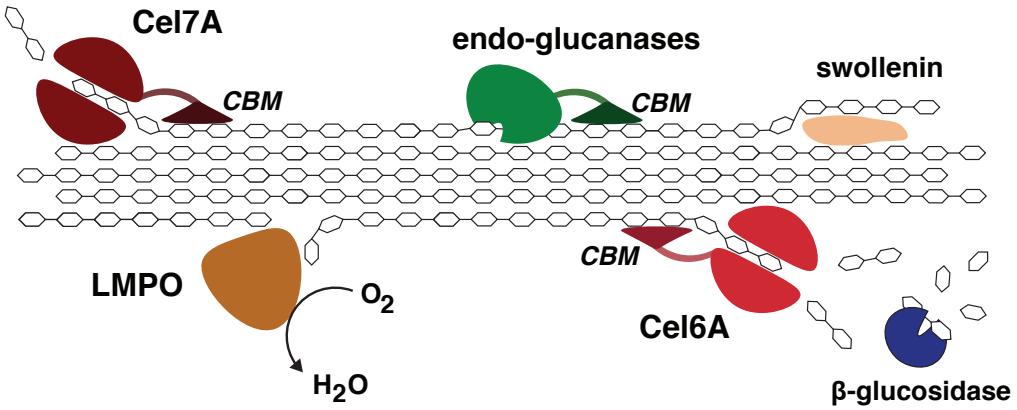



Figure 1.5 – *Schematic representation of the synergistic cellulolytic action of enzymes on cellulose chains. Exoglucanases, such as Cel7A and Cel6A, anchor onto the cellulose with the help of a carbohydrate-binding module (CMB) to processively hydrolyse cellulose from their chain ends into cellobiose, which in turn is cleaved into glucose by β-glucosidases in solution. Two other enzymes help the cellulose degradation by creating new chain ends in the carbohydrate polymers via a hydrolytic (endoglucanases) or oxidative (lytic polysaccharide monooxygenases, LMPOs) process. Lastly, swollenins enhance the cellulose digestibilty by disrupting the strong H-bonds networks bestowing cellulose its crystalline properties. Recent studies attributes swollenin endo-acting catalytic properties, with release of internal glucose units, which are not shown here*[96]*. Hemicellulolytic activity is similar, with addition of specific enzymes targeting the variety of side-chains decorating the carbohydrate backbone.*

that retains the stereochemistry of the sugars, and (ii) Cel6A, which starts depolymerisation from the non-reducing end through a mechanism that inverts the stereochemistry. Despite many unknowns, extensive studies on Cel7A, which has been to shown to play a pivotal role in the observed cellulose depolymerisation rate[97], have led to some important mechanistic insights on the action of these enzymes. Specifically, the productive cycle appears to be a

sequence of six distinct steps[36]: (i) binding to cellulose through the CBM, (ii) surface diffusion until a suitable chain end is located and recognized by the CD; (iii) threading of the cellulose fiber (i.e. decrystallisation), where a single cellulose chain is partially detached from the crystal into a tunnel containing the catalytic suite in the CD; (iv) complexation to form the active enzyme-substrate complex leading to (v) hydrolysis and release of cellobiose units as the cellulase slides along the chain in a processive movement before (vi) detachment from the surface. Two important factors have been highlighted in both experimental and computational as limiting the efficiency of the enzyme: the processivity and more particularly the catalytic step[97]. The processivity, representing the number of catalytic cycles that the enzyme performs without releasing the substrate, typically averages around 50 and is dependent on the substrate properties.

Such a stepwise mechanism, resulting from the modular nature of the enzyme that is common to many cellulases (~ 40%), helps overcome the challenging insoluble nature of cellulose. While the CBM plays a crucial part as driver for enzyme adsorption on the cellulose surface, the extent of its role on the hydrolysis rate is not yet fully clear, in particular its influence in the unproductive adsorption of cellulase as well as on the decrystallisation process[36]. In contrast to Cel7A and Cel6A containing CBM tailored to bind to ordered cellulose structure (Type A), CBMs of EGs secreted by *T. reesei* can present either affinity to crystalline region (Type A) or to more disordered/amorphous regions (Type B). At least five main EGs (Cel7B, Cel5A, Cel12A, Cel61A, and Cel45A) are produced by the fungi (representing ~ 15 wt. %), which present CDs tailored to the cleavage of internal bonds with an active site located in a deep groove rather than a tunnel[98,99]. The activity of $\beta$-glucosidases is weak in the native *T. reesei*, and is usually enhanced in enzyme cocktails to prevent GH end-product inhibition by high

cellobiose concentration in the media. In particular, fungal extracts from *Aspergillus Niger* are rich in $\beta$-glucosidases, which can for instance be directly incorporated into the commercial preparation or combined as separate commercial blends.

#### 1.3.2.2 Hemicellulases

Due to its richer chemical and structural composition, the enzymatic hydrolysis of hemicellulose requires a broader range of hydrolytic enzymes to both cleave backbone side-groups and depolymerise the chains. Similar to cellulose, hemicellulose depolymerisation, and in particular its main component xylan, relies on the synergistic action of enzymes, with endoxylanases releasing small xyloooligomers which are, in turn, subject to further depolymerisation in solution by $\beta$-xylosidases. In this case, endo-acting hemicellulases are at the core of the hydrolytic process, with exoxylanases that usually present weak activities. Simultaneously, a myriad of accessory hemicellulases act on specific backbone decorations to un-branch the main chains, including families of enzymes such as $\alpha$-glucoronidases, acetyl esterases and $\alpha$-L-arabinofuranosidases[100].

#### 1.3.2.3 Accessory enzymes and proteins

More recently, several non-hydrolytic proteins have been recognised as important contributors to enzyme synergism and positively impact the overall hydrolysis rate. In particular, lytic polysaccharide monooxygenases (LPMOs) use an oxidative mechanism based on copper-activated molecular oxygen to randomly cleave glycosidic bonds in both cellulose and hemicellulose[101]. The copper active site can then be restored by lignin[97]. These enzymes have been added to commercial enzyme cocktails to enhance the synergism, as in the case of Novozymes Cellic CTec2 and CTec3. Based on a different mechanisms, swollenins, which

are proteins found in few fungi families and notably in *T. reesei*, help loosen the tight cell wall structure by disrupting the H-bonds network, with more recent findings even attributing them a catalytic activity. In such cases, swollenins appeared to processively hydrolyse cellulose chain from mid-chain positions[96].

### 1.3.3 Factors influencing the enzymatic hydrolysis

As a hetereogeneous process, numerous factors can impact the efficiency of the enzymatic hydrolysis of lignocellulosic biomass[35,102,103]. While inherently interlinked, these parameters are commonly distinguished into two classes: substrate-related and enzyme-related [Figure 1.6]. Experimental conditions can also affect the hydrolysis rate, adding an extra layer of complexity to understanding how to obtain high rates and glucose titers [Figure 1.6].

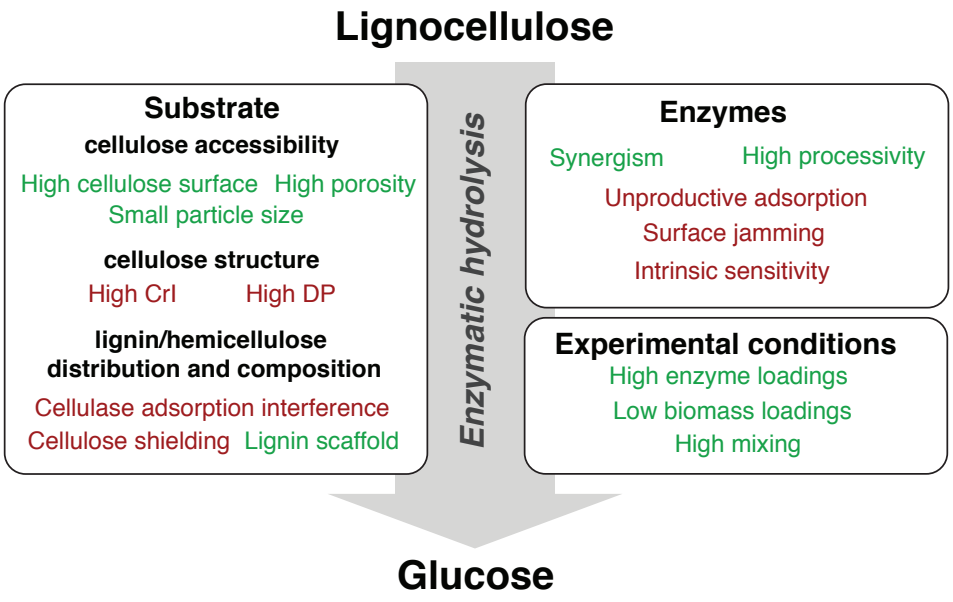



Figure 1.6 – *Factors influencing the enzymatic hydrolysis of lignocellulosic biomass, classified into three categories: substrate- and enzyme-related, as well as arising from processing conditions. Parameters enhancing the efficiency of the hydrolytic process are shown in green, while factors presenting a negative impact on the reaction are associated to the red.*

#### 1.3.3.1 Structure-related factors

From a physical point of view, hydrolysis rates are limited by three main factors: (i) the cellulose accessibility, (ii) the cellulose structure and (iii) the distribution and composition of co-biopolymers.

**Cellulose accessibility** is a key factor that has often been put forth to explain hydrolysis rates and pretreatment effects, at least for early reaction times[104–107]. This global parameter, resulting from the combination of particle size and porosity, as well as available cellulose surface, reflects the capacity of cellulases to reach their substrate over the course of the hydrolysis. Size reduction is pivotal for efficient enzymatic hydrolysis to take place, not only by decreasing the enzyme diffusional path, but also improving the rheological properties of the reaction media[108–110]. In terms of cellulases accessibility, the beneficial impact of size reduction is related to the feedstock characteristics, with evidences pointing towards a limited effect on the hydrolysis once a certain size threshold is reached[111]. Disruption and removal of the protective hemicellulose-lignin matrix from cellulose not only increase its available surface, but also enhance the substrate digestibility by creating a network of pores in the lignocellulosic substrate more or less accessible to cellulases[112–114]. Both network geometry, reflected by the tortuosity, and pore size distribution play a role on the penetration of enzymes in the substrates and the hydrolytic product release in the bulk solution. With typical cellulase size assumed to be ~ 5.1 nm, two pore size ranges can be distinghuised in the substrate, with accessible pores presenting diameters > 5.1 nm and the remaining being inaccessible porosity. Depending on the substrate characteristics and experimental conditions, both type of porosity demonstrated either strong or no correlation to glucose yields, while smaller pore ranges in some cases associated to unfavourable enzyme pore entrapment[107,115,116]. High tortuosity,

affecting the diffusion efficiency, can negatively impact the hydrolysis rates, notably through increased product inhibition and impediment to enzymes accessing their substrate.

**Cellulose structure** itself affects its deconstruction with its degree of crystallinity and degree of polymerisation impacting the enzyme action[117–122]. Even if they are both known to have an impact, their exact role and importance in observed hydrolysis rates are still not fully understood, notably due to the difficulties to accurately measure and study them experimentally[123,124]. A major culprit lies in the fact that changes in the DP and CrI during pretreatment and over the course of hydrolysis are often accompanied by other structural modifications, making it difficult to probe their inherent impact on the hydrolysis rate. On one hand, enzyme synergism and adsorption capacity are expected to change depending on the cellulose characteristics. On the other hand, lower DP is assumed to facilitate cellulose degradation by forming weaker H-bonds networks, and lower CrI, associated with higher proportion amorphous regions, is usually related to higher hydrolysis rate. Overall, experimental studies have either strongly correlated these two parameters with increases or decreases in the hydrolysis rates, or shown minimal impact on the glucose release, which makes understanding their effects very challenging.

**Distribution and composition of lignin and hemicellulose** In addition to acting as physical barriers limiting the access to cellulose, hemicellulose and lignin can directly interact with cellulases. Irreversible adsorption of cellulases and enzyme inhibition by free phenolic hydroxyl groups on lignin have both been associated with the negative impact of lignin on the hydrolysis rate[125–127]. While overall, delignification is favourable, some studies have correlated extensive lignin removal to a decrease in hydrolysis rate (over-delignification), with an optimal amount of lignin being seemingly important to structurally maintaining the pore network[114].

The composition of lignin, and particularly its S/G ratio, impacts both the polymer strength and interactions with enzymes, but a clear picture of its effect is yet to be established. Similarly, the presence of hemicellulose is usually negatively correlated with increasing hydrolysis rates, with evidence of acetyl-decorations interfering with enzymes, in addition to hemicellulose just shielding cellulose[128]. Whether the presence of hemicellulose or lignin has the most impact is not clear, as their removal is usually closely interconnected.

#### 1.3.3.2 Enzyme-related factors

Because enzymes are quite sensitive to experimental conditions, several factors can contribute to the commonly observed decrease in the hydrolysis rate as the reaction proceeds. In addition to their intrinsic thermal sensitivity, enzymes have been prone to inhibition by both hydrolysis products and residual compounds produced during the pretreatment step. More recently, enzyme deactivation at the liquid-air interface of the reaction setup was highlighted as impacting the glucose release[129]. On cellulose, cellulases can also encounter multiple hurdles limiting their efficiency to catalyse the hydrolysis of glucan linkages: unproductive adsorption[126,127], surface jamming[130], which occurs at high enzyme surface concentration, and decrease in processivity due to presence of surface obstacles[131,132]. All these effects have been proposed as being potentially important reaction impediments.

The composition of the enzyme cocktail also plays a pivotal role for the process efficiency, and should be seen in conjunction with the substrate characteristics to achieve deconstruction in a highly synergistic manner[133,134].

#### 1.3.3.3 Experimental conditions

As a heterogeneous catalytic reaction, the efficiency of the enzymatic hydrolysis of a lignocellulosic substrate strongly relies on an effective mass transfer. Cellulase diffusion toward their cellulosic substrate is partially affected by the three parameters that are mixing, solids loading and enzyme loading. Mixing has been shown to impact digestion not only by improving mass transfer, but also by enhancing particle breakdown through shear forces[135]. While high biomass loadings are favoured for the economic viability of the process, the resulting increase in viscosity of the mixture coupled with the lack of free water and increased concentration of enzyme inhibitors typically significantly reduce the observed hydrolysis rates[87]. This negative effect, generally appearing as the substrate loading exceeds > 15 wt. %, has been widely observed and is commonly referred to as the high-solid effect[136–138]. In parallel, low enzyme dosages that are targeted industrially to reduce costs have similarly led to significantly reduced rates and substrate conversion[105,139].

### 1.3.4 Challenges

To be economically viable, enzymatic hydrolysis should ideally be performed at high solid loadings, resulting in high glucose titers which minimise processing cost, coupled with low enzyme dosages to limit the expenses associated with their production. However, variabilities within and across lignocellulosic substrates coupled to the specificities of numerous enzymatic machineries have led to struggles in achieving high sugar conversion in an economically viable manner. In particular, experimental hydrolysis rates typically show a slow down after an initial burst in the glucose release, blocking the path towards full substrate conversion on relevant timescales and leading to high enzyme use and low solid loadings to maintain reasonable

yields[140–142]. Within the pool of factors influencing enzymatic deconstruction, no single one can fully correlate to hydrolysis trends. Further layers of complexity are notably brought by (i) the difficulty of experimentally investigating each of these parameters independently, as targeted modifications of one parameter usually impact the others, and (ii) experimental observations being highly dependent on the substrate native structure and treatment history, as well as the composition of enzyme cocktail used. All this has led to conclusions between different studies appearing sometimes as incoherent. For example, CrI has been reported to strongly correlate to observed initial hydrolysis rate for various pretreated lignocellulosic substrates, while others attributed to CrI a less critical role[143].

While several strategies have been used to counterbalance recognised negative effects and improve the overall glucose release – such as the addition of surfactants to stabilise enzymes and counteract their unspecific adsorption to lignin or the engineering of enzymes presenting higher thermostability –, the lack of a comprehensive understanding of the enzymatic hydrolysis of biomass still impedes the emergence of efficient integrated rational design strategies for enzyme process engineering for biomass conversion.

## 1.4 Modelling of the enzymatic hydrolysis

To develop such improved process engineering of enzymatic processes, improved understanding of the complex enzymatic hydrolysis process would be required. Notably, our lack of understanding of the sharp reaction rate decrease observed over hydrolysis time contributes to the difficulty of maintaining high hydrolysis productivity, which increases overall processing cost. Numerous modelling strategies have tried to address this lack of understanding by helping to unravel the numerous mechanisms that control hydrolysis rates, but the multiplicity and temporality of the factors impacting the heteregeneous catalytic deconstruction of

lignocellulosic biomass have impeded the emergence of a comprehensive mechanistic model. As such, the predictive accuracy of current models often relies on a trade-off between computational cost and the level of detail used for the reaction mechanism. Additional complexity is introduced by the difficulty of accurately measuring system parameters experimentally, such as the different kinetic parameters of enzymes (e.g. rate constants, processivity values) or the exact physico-chemical properties of the substrate, which can lead to over-fitting of models when estimating many unknown parameters.

Within this context, extensive research has approached this system in a bottom-up manner, focusing on the interactions between cellulases and cellulose only, while others have integrated the entire, complex lignocellulosic substrate to test several assumptions. In both cases, models cover a broad range of mathematical approaches, from fully empirical to their inclusion of detailed kinetics mechanisms and/or substrates physical characteristics.[144,145].

### 1.4.1 Pure cellulosic substrates

Enzymatic hydrolysis of cellulose has been modeled at different levels of detail. On one end of the spectrum, quantum mechanics/molecular mechanics simulations (QM/MM) have helped unravel the energetics of the catalytic hydrolysis of glucosidic linkages by cellulolytic enzymes[146]. On the other end, empirical methods, despite providing limited mechanistic insights, can help predict hydrolysis rates over ranges of experimental conditions[144]. While both of these methods are important in terms of either fundamental understanding or industrial application, respectively, they cannot be broadly used, notably due to the extensive computational cost associated to QM/MM simulations and the high dependence of empirical methods on the set of data they were based/trained on.

In this context, mechanistic models have been pivotally important in helping detangle the

influence of the numerous parameters playing a role in the enzymatic hydrolysis of cellulose, and highlighting the specific factors limiting experimentally observed rates[144,145]. Core assumptions and the level of kinetics that are detailed vary widely between the numerous models proposed to explain the hydrolysis trend observed over the course the reaction. Some modelling strategies focuse mainly on the cellulase-cellulose interactions with a detailed description of the enzymatic kinetic mechanism. Others are based on the description of the physical evolution of the substrate, which is used to explain/predict the hydrolysis rate. Overall, while initial hydrolysis rates have been correlated to the rate of the cellulase complexation/decomplexation of cellulases with cellulose and available adsorption sites, more recent studies have reported that evolutions in cellulose's heterogeneous behaviour over time were major factors behind the commonly observed rate decline[147,148].

While modelling of the enzymatic hydrolysis of pure cellulosic substrates is critical in the understanding of the process, findings obtained might not be directly transposable to process based on the more complex lignocellulosic material. In such case, determining-factors rationalising hydrolysis rates in the case of pure cellulosic substrates might be outweight by factors related to (i) the presence of other biopolymers and their interactions with both cellulose and enzymes, (ii) distribution of cellulose within the substrate and (ii) cellulose ultrastructure within the hemicellulose-lignin matrix that might not be well reflected by the extracted pure cellulosic substrates.

### 1.4.2 Lignocellulosic substrates

While kinetic modelling of the enzymatic hydrolysis of cellulosic substrates have gathered a lot of attention due to their relative simplicity, fewer studies have examined the more complex deconstruction of cellulose within the complex lignin-hemicellulose scaffold. In most cases,

the influence of the complete lignocellulosic structure has been approached from a more empirical perspective or as a perturbation in the cellulose hydrolysis kinetics[144,149]. For example, Vani et al.[150] used artificial neural networks trained on a dataset for alkali-pretreated rice straw to predict optimal hydrolysis conditions in terms of particle size, biomass- and enzyme-loadings, while Wojtusik et al.[151] used fractal kinetics to compare the digestibility of three substrates (corn stover, wheat straw and cardoon stems) subjected to either DAP or water-ethanol extraction. Based on a phenomenological model, Lischeske et al.[152] used a two-population particle model with distinct digestibility to predict sugar release from hydrolysis of DAP-pretreated corn stover in both batch- and continuous-settings. While such models are of value for the implementation of more efficient lignocellulosic biomass to sugars pathway, they give limited mechanistic insights, notably on the rate slowdown observed at higher conversion and at high solid loadings.

## 1.5 Objectives

Alluding to the numerous unknowns that limit opportunities to rationally design enzymatic hydrolysis processes for biomass conversion, this research proposes the use of a theory-based model to evaluate the relative importance of known bottlenecks in the enzymatic hydrolysis of lignocellulosic biomass. As emphasized before, heterogeneities in the lignocellulosic substrate, coupled to the complexity of the enzymatic machineries, represent a challenge to discriminate critical from non-critical factors limiting the hydrolysis rate. Reflecting the intricacy of the reactional system, no single parameters alone has been able so far to rationalise depolymerisation trends, such that various enzyme-substrate interactions have to be considered at different scale and time to explain the process. From an experimental point of view, while numerous methods have been developed to assay the physico-chemical properties

of both substrates and enzymes, assessing the extent of the role of assayed factors in dictating reaction rates represents a challenge. Due to the close intertwining of parameters and unknowns in the relationships existing between them, changes in one parameter are shown to affect others, as well as the overall dynamics of interactions. Coupled with the wide range of lignocellulosic substrates considered as suitable feedstocks for biorefineries, no consensus has yet emerged on a set of critical factors to target for designing efficient enzymatic hydrolysis process. However, such knowledge is critical for tailoring pretreatments to the lignocellulosic substrate types, as well as carefully adjust enzyme cocktails and experimental conditions to the properties of the pretreated substrates. As such, modelling appear as a pivotal tool to explore factors dictating the hydrolysis rate and decoupling their individual roles. While the detailed interactions of cellulases and cellulose have focused a lot of in silico efforts, the ensuing key findings are not directly transposable to the more complex lignocellulosic substrate. To the best of the author's knowledge, the impact of the entire lignocellulosic structure in the context of the enzymatic hydrolysis of lignocellulose has been included in models only in an implicit way, with the presence of the hemicellulose-lignin scaffold treated as a perturbation in the cellulose hydrolysis kinetics or based on an empirical/phenomenological approach. As such, they provide rather coarse relationships between substrate's specificities and hydrolysis efficiency. In this context, we propose a theory-based model describing the enzymatic hydrolysis of lignocellulosic biomass from the substrate's perspective to link the substrate's physical characteristics to observed hydrolysis trend. As such, the model will allow to test hypothesis related to the individual impact of recognised bottlenecks and the extent to which they limit the depolymerisation rates.

The description and extent of this work can be translated into 3 main objectives:

- **Objective 1** Based on the work of Luterbacher et al.[153,154], a general model based on pore-diffusion and surface reaction will be developed to describe the enzymatic hydrolysis of lignocellulosic substrate in a range of experimental conditions. This model will describe the process from the substrate perspective, assuming at first a simplified, aggregated cellulolytic mechanism. The interplay between enzyme loading, solid loading and cellulose accessibility will be investigated. In particular, the total amount of enzymes in the system will be carefully accounted for, allowing to model situations where enzymatic hydrolysis is carried out in enzyme- or substrate-limiting conditions (i.e. where the number of enzyme or number of adsorption sites respectively limits the hydrolysis rate).
- **Objective 2** Adopting a bottom-up approach, the developed model will be made more complex, notably to investigate the effect of structural heterogeneities in the lignocellulosic substrate on the cellulose degradation rate, with emphasis on their implication on the rate slowdown observed at later stages.
- **Objective 3** Lastly, enzyme-related deactivation mechanisms will be included in the model framework, to allow a more comprehensive view of the heterogeneous catalytic process. In particular, we will explore the role of unproductive cellulase on both the cellulose and the lignin fractions of the lignocellulosic substrate.

The reason behind adopting a bottom-up approach, where the model is progressively complexified, is mostly to prevent overfitting while evaluating biomass structure related factors. By aggregating at first all cellulase-cellulose interactions into a single parameter and ensuring that it adequately represent the complex process, we are hence able to reduce the total of fitted parameter in our model.

# Chapter 2

# Impact of internal mass transfer on enzymatic hydrolysis rate

## 2.1 Introduction

In this chapter, we use a theoretical framework to investigate the effect of internal mass transfer, which arise due to experimental conditions and the substrate's physical characteristics, during the early rates of the enzymatic hydrolysis of lignocellulosic biomass. Several mechanistic studies had investigated the relationship between initial rates, and enzyme- or biomass-loadings in the case of cellulosic substrates to extract information on the cellulose hydrolysis kinetics[155–158]. In these works, experiments were conducted on rather small substrates (typically Avicel, a microcrystalline cellulose with particle size lying around 50 μm), which consequently presented low diffusional resistance. In particular, bursts in hydrolysis rates have been predicted at early hydrolysis stages of cellulosic substrates in enzyme-limited conditions[155]. By contrast, lignocellulosic substrates will inherently lead to more significant mass transfer limitations, due to the presence of other biopolymers increasing the substrate

volume for a given mass of cellulose. This adds complexity in the diffusion path and increases diffusion length for the cellulase to reach the cellulose binding sites, which will increase the time scale at which surface saturation will occur.

Here, we present a diffusion-reaction model describing the enzymatic hydrolysis of lignocellulosic biomass particles with a focus on the substrate's physical evolution. Our goal is to evaluate the interplay between cellulose accessibility to cellulase and glucose release [Figure 2.1]. Our analysis focuses at first on early hydrolysis stages, where modifications in the substrate's reactivity due to structural heterogeneities or enzyme deactivation mechanisms are minimal. Both effects are further developed in subsequent chapters. While such diffusion-reaction systems involving porous media have been subject of extensive research efforts in many domains - from chemical engineering for the evaluation of catalyst deactivation[159] to the study of minerals deposition in hydrology/geology[160] to only cite a few -, we use here a similar formalism to describe the enzymatic hydrolysis of lignocellulosic biomass. The philosophy behind the model's development and the estimation of its parameters is to include enough information to capture the complexity of the lignocellulosic substrates, while minimising the number of model parameters to both keep an effective model and avoid overfitting. As the emphasis is on lignocellulosic biomass structure effects, the hydrolysis mechanism is reduced to its simplest formulation, i.e. an adsorption-desorption cycle triggering a loss of cellulose mass, and collect all complexity into one lumped, fitted parameter. As such, our approach can be seen as complementary to detailed studies on pure cellulosic substrates. Specifically, we seek to expand the study of the importance of recalcitrance arising from the structure of the lignocellulosic substrates, including cellulose accessibility to cellulases within these substrates and understand the differences that may arise in comparison with pure cellulosic substrates.

## 2.2 Modelling framework

Before presenting a detailed derivation of the working equations, the key hypothesis of the model are summarised. All symbols are defined and summarized in Table 2.1.

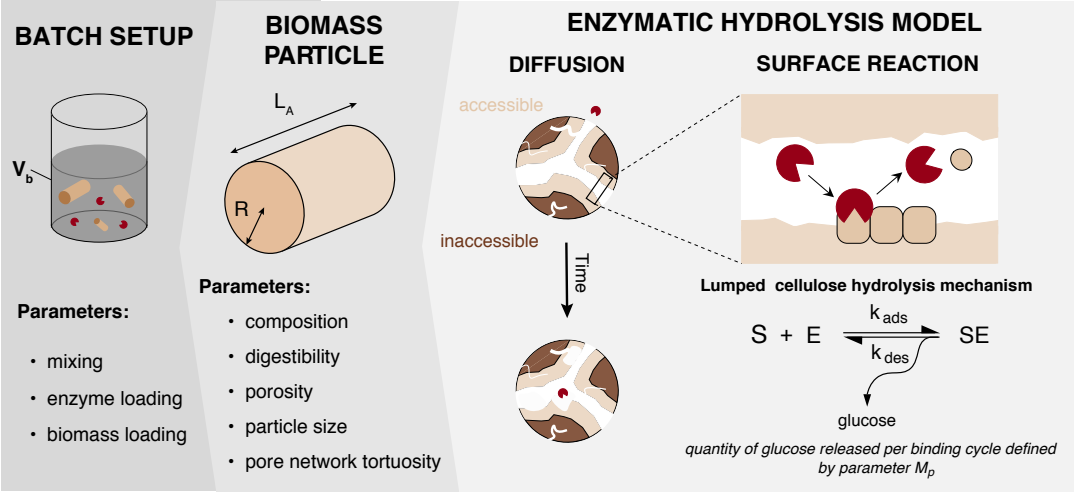



Figure 2.1 – *Schematic representation of the assumed biomass geometry and enzymatic hydrolysis mechanism.*

In continuity to the work of Luterbacher et al.[153,154], biomass fragments are modeled as non-shrinking, porous cylinder of radius $R$ evenly composed of lignin, hemicellulose and cellulose. The model assumes that only a certain weight fraction $x_{A,M}$ of biomass, based on the final glucose yield $Y_g$, is susceptible to be hydrolysed over the course of the reaction. The rest is considered to be recalcitrant to the enzymatic action. Hence, the cylindrical particle can be divided in an accessible volume $V_{cyl,A}$, containing by definition all pores presenting a diameter $d_p$ large enough to accommodate enzymes, and an inaccessible segment of volume $V_{cyl,IV}$ [Figure 2.1]. The pores too narrow to allow the diffusion of the cellulases are homogeneously distributed throughout the entire biomass fragment. Pore diffusion refers to internal mass transfer, that, based on initial calculations, is much more significant than external mass

transfer, which is ignored in this work (see Appendix A.1).

In addition to assumptions made for the particle morphology, a stepwise model for the enzymatic degradation is used: the enzyme that has reached the enzyme surface through pore diffusion, can adsorb, react and desorb [Figure 2.1]. No distinction is made between cellulose adsorption sites, with the catalytic process being identical for all sites. Cellulose hydrolysis is hence implicitly expressed through a fitted, lumped parameter $M_p$ representing the glucose release per enzyme binding cycle. This parameter accounts indistinctly for effects related to hydrolysis including: the distinct type of enzymes in the cocktail, synergism, individual mechanisms of these cellulases and local specificities in the cellulose structure (enzyme processivity, differentiated adsorption sites, cellulose structural heterogeneities, enzyme synergism etc..) and, as such, represents an overall measure of the cellulose's susceptibility to be digested by a given enzyme cocktail. Mathematically, we can demonstrate that this parameter is equivalent to the intrinsic processivity of a lumped enzyme representing the collective action of the enzyme cocktail on cellulose (see Appendix A.2). Hence, from the substrate's perspective, the enzymatic process translates as a change in biomass porosity, with notably consequences of products released in solution, such as inhibition or crowding within pores, not accounted for. A more detailed comparison of the prior model[153,154] with one developed here can be found in Appendix A.3.

| Symbol | Description | Value | Units |
|---|---|---|---|
| $b_l$ | Biomass loading | Dep. var. | [g cm$^{-3}$] |
| $C_E(r,t)$ | Total enzyme concentration (in fluid and adsorbed at the surface) | Dep. var. | [mol cm$^{-3}$] |
| $C_E^F(r,t)$ | Enzyme concentration in fluid (pores or bulk) | Dep. var. | [mol cm$^{-3}$] |
| $C_E^S(r,t)$ | Enzyme concentration adsorbed at the cellulose surface (per total cylinder volume) | Dep. var. | [mol cm$^{-3}$] |
| $C_{E,max}^S(r,t)$ | Maximal enzyme concentration adsorbed at the cellulose surface (per total cylinder volume) | Dep. var. | [mol cm$^{-3}$] |

| | | | |
|---|---|---|---|
| $C_{E,0}$ | Initial bulk concentration, defined by both enzyme- and biomass-loading | Dep. var. | [mol cm$^{-3}$] |
| $D_E^{bulk}$ | Cellulase diffusivity in bulk solution taken from Luterbacher et al.[153] | $7.35 \cdot 10^{-5}$ | [cm$^2$ min$^{-1}$] |
| $D_E^{eff}(r,t)$ | Effective cellulase diffusivity in pores, function of the pore network characteristics as defined in Eq. (2.5) | Dep. var. | [cm$^2$ min$^{-1}$] |
| $D_E^{pore,j}$ | Cellulase diffusivity in pores of diameter $d_{p,j}$, as defined in Eq. (2.6) | Dep. var. | [cm$^2$ min$^{-1}$] |
| $\overline{D}_E^{pore}$ | Average cellulase diffusivity in pores, function of substrate pore size distribution (Eq. (2.6)) | Dep. var. | [cm$^2$ min$^{-1}$] |
| $D_E$ | Time independent part of the effective cellulase diffusivity in pores, as defined in Eq. (2.5) | Dep. var. | [cm$^2$ min$^{-1}$] |
| $d_\mathrm{p}$ | Pore diameter, measured experimentally | Dep. var. | [Å] |
| $d_E$ | Enzyme diameter in solution[107] | 51 | [Å] |
| $e_l$ | Enzyme loading | Dep. var. | [-] |
| $H_{glu}$ | Hydrolysis factor for glucose | 0.9 | [-] |
| $k_{ads}$ | Cellulase surface adsorption rate[161] | $3 \cdot 10^{10}$ | [cm$^3$ mol$^{-1}$ min$^{-1}$] |
| $k_{des}$ | Cellulase surface desorption rate[154] | 0.068 | [min$^{-1}$] |
| $L_A$ | Accessible cylinder length | Dep. var. | [cm] |
| $M_p$ | Number of glucan monomer liberated during one enzyme binding-reaction cycle | Fitted var. | [-] |
| $MM_{glu}$ | Molar mass of glucose | 180 | [g mol$^{-1}$] |
| $n$ | Grid size | Dep. Var. | [-] |
| $N$ | Sampling number in Monte-Carlo analysis | Dep. Var. | [-] |
| $n_{E,0}$ | Total number of moles in the system, function of the enzyme loading | Dep. var. | [mol] |
| $n_{E,max}^S(r,t)$ | Maximum number of adsorption sites at time $t$ and radius $r$ | Dep. var. | [mol] |
| $r$ | Radial distance | Indep. var. | [cm] |
| $R$ | Particle radius, measured experimentally via sieving or from explicit particle size distribution | Dep. var. | [cm] |
| $S_p$ | Pore wall surface as defined in Eq. (2.12) | Dep. var. | [cm$^2$] |
| $S_c(r,t)$ | Accessible cellulose surface, dependent on cellulose fraction and pore wall surface (Eq. (2.13)) | Dep. var. | [cm$^2$] |
| $S_{cyl}$ | Outer cylinder suface | Dep. var. | [cm$^2$] |
| $t$ | Time | Indep. var. | [min] |
| $V_{cyl,X}$ | Accessible ($X = A$) or inaccessible ($X = IV$) biomass volume, function of the experimental final glucose yield | Dep. var. | [cm$^3$] |
| $V_B$ | Volume of the bulk solution, expressed via the biomass loading (Eq. (2.19)) | Dep. var. | [cm$^3$] |
| $V_p$ | Pore volume | Dep. var. | [cm$^3$] |
| $x_{A,M}$ | Mass fraction of accessible biomass, function of the final glucose yield | Dep. var. | [-] |
| $x_{C,V}$ | Initial volume fraction of cellulose, function of the substrate composition | Dep. var. | [-] |

| | | | |
|---|---|---|---|
| $z$ | Longitudinal coordinate | Indep. var. | [-] |
| $Y_g$ | Experimental final glucose yield | Dep. var. | [-] |
| $\chi_E^X$ | Fluid ($X = F$) or surface ($X = S$) concentrations relative to the initial bulk concentration $C_{E,0}$ | Dep. var. | [-] |
| $\varepsilon(r,t)$ | Porosity, measured experimentally | Dep. var. | [-] |
| $\varepsilon_X$ | Initial ($X = 0$) and final ($X = \infty$) porosity | Dep. var. | [-] |
| $\eta(t)$ | $L^2$ normalised error | Dep. var. | [-] |
| $\varphi$ | Angular coordinate | Indep. var. | [-] |
| $\Omega_{cyl}$ | Cylindrical integration domain | Indep. var. | [-] |
| $\rho_C^{IV}$ | Density of cellulose (including void), function of the porosity | Dep. var. | [$cm^2$] |
| $\sigma$ | Maximum cellulase surface concentration on cellulose | $2.10 \cdot 10^{-12}$ | [$mol\,cm^{-2}$] |
| $\tau$ | Tortuosity | Fitted var. | [-] |

Table 2.1 – *List of symbols and their sources. For variables depending on experimental measurements, reference to relevant equations are indicated.*

### 2.2.1 Derivation

Within this framework, the set of equations describing the simultaneous diffusion and reaction of enzymes in the substrate over time is obtained by performing a mole balance over a cylindrical particle defined by $\Omega_{cyl} = \{(r,\varphi,z) \in \mathbb{R}^3 : 0 \leqslant r \leqslant R, 0 \leqslant \varphi \leqslant 2\pi, 0 \leqslant z \leqslant L_A\}$:

$$\begin{aligned} \frac{d}{dt}\iiint_{\Omega_{cyl}} C_E(r,t)dV &= \frac{d}{dt}\iiint_{\Omega_{cyl}} \left[\varepsilon(r,t)C_E^F(r,t) + C_E^S(r,t)\right] dV, \\ &= \frac{d}{dt}\iiint_{\Omega_{cyl}} \left[\varepsilon(r,t)C_E^F(r,t)\right] dV + \frac{d}{dt}\iiint_{\Omega_{cyl}} C_E^S(r,t)dV. \end{aligned} \tag{2.1}$$

Here, $\varepsilon(r,t)$ is the particle porosity and $C_E(r,t)$, $C_E^F(r,t)$ and $C_E^s(r,t)$ represent the total-, fluid- and surface-enzyme concentration inside the particle, respectively. Any edge effects are disregarded, i.e. diffusion is considered in a cylinder of infinite length. The total change in enzyme quantity inside the particle, represented by left-hand side of Eq. (2.1), corresponds, at any given time, to the net flux of enzyme passing through the particle surface $S_{cyl} = \partial\Omega_{cyl}$

at that same time. Using the divergence theorem, we can relate the net flux $\boldsymbol{F}_i$ of a specie $i$ through a closed surface $S = \partial\Omega$ to the volume integral of the divergence $\nabla \cdot \boldsymbol{F}_i$ over an enclosed domain $\Omega$:

$$\oiint_{\partial\Omega} \boldsymbol{F}_i \cdot \hat{n}\, dS = \iiint_{\Omega} \nabla \cdot \boldsymbol{F}_i\, dV. \tag{2.2}$$

At the same time, Fick's first law can be used to relate a species concentration $C_i$ to its diffusive flux:

$$\boldsymbol{F}_i = -D_i \nabla C_i, \tag{2.3}$$

where $D_i$ is the diffusivity of specie $i$. With these two equations (Eq. 2.2-2.3), the total concentration of enzyme inside the particle becomes:

$$\frac{d}{dt} \iiint_{\Omega_{cyl}} C_E(r,t)\, dV = \oiint_{\partial\Omega_{cyl}} \boldsymbol{F}_E \cdot \hat{n}\, dS = \iiint_{\Omega_{cyl}} \nabla \left[\varepsilon(r,t) D_E \nabla C_E(r,t)\right] dV, \tag{2.4}$$

where $D_E^{\text{eff}}$ is the effective enzyme diffusivity in the particle pores and $\hat{n}$ represents the normal to the surface $S$. In this equation, $D_E$ is the time independent part of the effective diffusivity $D_E^{\text{eff}}(r,t)$ parameter that is dependent on the average pore diffusivity $\overline{D}_E^{\text{pore}}$, the tortuosity $\tau$ of the pore network and the evolving porosity $\varepsilon(r,t)$[162]:

$$D_E^{\text{eff}}(r,t) = \frac{\varepsilon(r,t)\overline{D}_E^{\text{pore}}}{\tau} = \varepsilon(r,t) D_E. \tag{2.5}$$

Here, $\overline{D}_E^{\text{pore}}$ is estimated empirically from the pore size distribution specific to the biomass material and the diameter $d_E$ of the solute, in this case cellulases, in the following way[162]:

$$\overline{D}_E^{\text{pore}} = \frac{\sum_{j \geqslant d_E}^{\infty} D_E^{\text{pore},j}\, V_{p,j}(t=0)}{\sum_{j \geqslant d_E}^{\infty} V_{p,j}(t=0)} = D_E^{bulk}\, \frac{\sum_{j \geqslant d_E}^{\infty} (1-\frac{d_E}{d_{p,j}})^4\, V_{p,j}(t=0)}{\sum_{j \geqslant d_E}^{\infty} V_{p,j}(t=0)}. \tag{2.6}$$

While the effect of pore widening on the effective diffusivity inside the particle is accounted for in the model (Eq. 2.5), the average pore diffusivity as defined in Eq. (2.6) is assumed constant over the course of the reaction. Upon insertion of Eq. (2.4) and, using the Leibniz integral rule, Eq. (2.1) becomes:

$$\nabla\left[D_E^{\text{eff}}(r,t)\nabla C_E^F(r,t)\right] = \frac{\partial}{\partial t}\left[\varepsilon(r,t)C_E^F(r,t)\right] + \frac{\partial C_E^S(r,t)}{\partial t} \tag{2.7}$$

After explicit development of the right-hand side of Eq (2.7) and rearrangement, the time evolution of the enzyme's fluid concentration inside the particle can be expressed as:

$$\begin{aligned}\frac{\partial C_E^F(r,t)}{\partial t} &= D_E\left[\frac{\partial^2 C_E^F(r,t)}{\partial^2 r} + \frac{1}{r}\frac{\partial C_E^F(r,t)}{\partial r} + \frac{1}{\varepsilon(r,t)}\frac{\partial C_E^F(r,t)}{\partial r}\frac{\partial \varepsilon(r,t)}{\partial r}\right] \\ &\quad - \frac{1}{\varepsilon(r,t)}\frac{\partial C_E^S(r,t)}{\partial t} - \frac{C_E^F(r,t)}{\varepsilon(r,t)}\frac{\partial \varepsilon(r,t)}{\partial t}.\end{aligned} \tag{2.8}$$

Here, only radial diffusion is assumed to occur ($L_A >> R$), reducing thus the cylindrical del formulae to their radial part only:

$$\text{gradient}: \quad \nabla f = \frac{\partial f}{\partial r}\hat{r} + \frac{1}{r}\frac{\partial f}{\partial \varphi}\hat{\varphi} + \frac{\partial f}{\partial z}\hat{z} \xrightarrow[\text{diffusion}]{\text{only radial}} \nabla f = \frac{\partial f}{\partial r}\hat{r} \tag{2.9}$$

$$\text{Laplacian}: \quad \nabla^2 f = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial f}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 f}{\partial \varphi^2} + \frac{\partial^2 f}{\partial z^2} \xrightarrow[\text{diffusion}]{\text{only radial}} \nabla^2 f = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial f}{\partial r}\right) \tag{2.10}$$

where $(\hat{r}, \hat{\varphi}, \hat{z})$ is the standard unit vector and $f(r, \varphi, z)$ represents an arbitrary scalar function. The three effects, which are mathematically described in right-hand side of Eq. (2.8), and control the time evolution of the enzyme's fluid concentration, are: (i) the diffusion into the particle, (ii) the adsorption-desorption process from the cellulose surface, and (iii) the dilution of enzymes within pores due to increases in void volume being created during the hydrolysis. The model was then further developed to describe the time evolution of porosity and bound enzymes remain with time. The latter is described by a time dependent Langmuir isotherm, with the concentration of enzymes at the cellulose surface given by:

$$\frac{\partial C_E^S(r,t)}{\partial t} = k_{ads} C_E^F(r,t) \left[ C_{E,max}^S(r,t) - C_E^S(r,t) \right] - k_{des} C_E^S(r,t), \tag{2.11}$$

where $k_i$ designates the adsorption ($i = ads$) or desorption ($i = des$) rate constant and $C_{E,max}^S(r,t)$ is the maximum possible concentration of enzymes bound to the surface at a given time $t$, derived from the initial number of accessible binding sites. The latter is computed from the pore size distribution measured for a specific substrate using purely geometrical assumptions. Based on the hypothesis that pores have a double-slit geometry[163], the pore surface $S_{p,j}$ for a pore presenting a diameter/width $d_{p,j}$ and volume $V_{p,j}$ can be computed as follows:

$$S_{p,j} = \begin{cases} \dfrac{2V_{p,j}}{d_{p,j}} & \text{for } d_{p,j} \leq 121\,\text{Å} \\ \dfrac{V_{p,j}}{d_{p,j}} & \text{for } d_{p,j} > 121\,\text{Å} \end{cases} \tag{2.12}$$

In this study, the distinction is made between pores that can accommodate 1 or 2 cellulases within their diameter by considering 1 or 2 pore surfaces within a slit geometry (see Figure A.2 in Appendix A.4). The total pore surface $S_{p,tot}$, given by the sum of individual pore surface $S_{p,j}$,

is then converted into a number of cellulase adsorption sites by assuming that the footprint of a cellulase on the cellulose surface $\sigma$ is equivalent to $5X5$ nm[164]. Considering a homogeneous distribution of cellulose within the particle, we obtain:

$$n^{S}_{E,max}(r,t) = \sigma x_{c,V} S_{p,tot}(r,t) = \sigma S_c(r,t), \tag{2.13}$$

where $x_{c,V}$ is the cellulose fraction and $S_c(r,t)$ is the available cellulose surface. As the reaction proceeds and the accessible cellulose is hydrolysed, the available surface for enzyme binding was assumed to gradually decrease along with the cellulose content. As the decrease of cellulose content is directly proportional to the increase in porosity (i.e. the dissolution of cellulose into soluble sugars during the enzymatic hydrolysis increases the internal volume), the available number of binding sites on the cellulose surface decreases linearly with the increase in porosity as cellulose is consumed:

$$S_C(r,t) = S_C(r,0)\frac{\varepsilon_\infty - \varepsilon(r,t)}{\varepsilon_\infty - \varepsilon_0}, \tag{2.14}$$

where $\varepsilon_\infty$ corresponds to the final porosity, i.e. when the whole accessible cellulose is digested. Since, for simplicity, cellulose hydrolysis was assumed to occur (and is accounted for) as enzymes detach from the surface, the time evolution of the porosity, which tracks the progression of hydrolysis, can be equated to the rate of enzyme desorption:

$$\frac{\partial \varepsilon(r,t)}{\partial t} = \frac{k_{des} C^{S}_{E}(r,t) M_p M M_{glu} H_{glu}}{\rho^{IV}_{C}}, \tag{2.15}$$

where $M_p$ represents the number of glucan monomers released in solution per binding cycle. Here, the term is expressed as mass of cellulose degraded per binding cycle per mole of enzyme through the hydrolysis factor $H_{glu}$ and molar mass of glucose $MM_{glu}$, and $\rho_C^{IV}$ is the total cellulose density (accounting for inaccessible pore volume, i.e. with diameter smaller than one of a cellulase). In summary, the following system of coupled partial differential equations (PDEs) composed of Eqs. (2.8),(2.11) and (2.15) describes the diffusion of enzymes inside a porous biomass particle and its subsequent hydrolysis:

$$\begin{cases} \dfrac{\partial C_E^F(r,t)}{\partial t} = D_E \left[ \dfrac{\partial^2 C_E^F(r,t)}{\partial^2 r} + \dfrac{1}{r}\dfrac{\partial C_E^F(r,t)}{\partial r} + \dfrac{1}{\varepsilon(r,t)}\dfrac{\partial C_E^F(r,t)}{\partial r}\dfrac{\partial \varepsilon(r,t)}{\partial r} \right] \\ \qquad\qquad - \dfrac{1}{\varepsilon(r,t)}\dfrac{\partial C_E^S(r,t)}{\partial t} - \dfrac{C_E^F(r,t)}{\varepsilon(r,t)}\dfrac{\partial \varepsilon(r,t)}{\partial t}, \\ \dfrac{\partial C_E^S(r,t)}{\partial t} = k_{ads} C_E^F(r,t) \left[ C_{E,max}^S(r,t) - C_E^S(r,t) \right] - k_{des} C_E^S(r,t), \\ \dfrac{\partial \varepsilon(r,t)}{\partial t} = \dfrac{k_{des} C_E^S(r,t) M_p MM_{glu} H_{glu}}{\rho_C^{IV}}. \end{cases} \tag{2.16}$$

This system can then be solved if defined boundary (BC) and initial (IC) conditions are provided. Assuming that all enzymes are initially contained in the bulk solution $V_B$, we have:

$$IC \begin{cases} C_E^F(r,0) = 0 \quad \forall\, r \neq R \qquad C_E^F(R,0) = C_{E,0} \\ C_E^S(r,0) = 0 \quad \forall\, r \\ \varepsilon(r,0) = \varepsilon_0 \quad \forall\, r \neq R \qquad \varepsilon(R,0) = 0. \end{cases} \tag{2.17}$$

The depletion of enzymes in the bulk solution is here accounted for by integrating the flux of enzymes passing through the biomass particle external surface $S_{cyl}$, which ensures continuity at the particle's solid-liquid interface. Coupled to a no-flux boundary condition at the center

of the particle, as longitudinal diffusion is neglected, the boundary conditions are expressed as:

$$BC \begin{cases} \left.\dfrac{\partial C_E^F(r,t)}{\partial r}\right|_{r=0} = 0 \\ \left.\dfrac{\partial C_E^F(r,t)}{\partial t}\right|_{r=R} = -\dfrac{S_{cyl}}{V_B}\varepsilon(R,t)D_E\left.\dfrac{\partial C_E^F(r,t)}{\partial r}\right|_{r=R} \\ \left.\dfrac{\partial \varepsilon(r,t)}{\partial r}\right|_{r=0} = 0. \end{cases} \tag{2.18}$$

Here, any external mass transfer phenomena are neglected (see Appendix A.1). Also, because the external surface usually represents only a small fraction of the accessible cellulose surface area, and because it is difficult to measure, enzyme adsorption on the external surface of the biomass particle is neglected (see Appendix A.5). The calculation results are used to chart the time-course of enzymatic hydrolysis.

### 2.2.2 Initial parameters determination

Parameters extracted from literature include adsorption- and desorption-rate constants and cellulase diffusivity in bulk solution. Parameters fitted to the data include tortuosity and average mass of cellulose liberated per binding cycle. The remaining variables are calculated based on available experimental data, including the measured initial accessible pore volume and cellulose fraction, used to compute initial cellulose surface area, and final glucose yield, used to predict the amount of cellulose that can be hydrolysed.

The enzyme $e_l$- and biomass $b_l$-loading are two key experimental parameters, and are used to define the initial enzyme bulk concentration,

$$C_{E,0} = \frac{n_{E,0}}{V_B} = \frac{e_l \sigma \int_0^R S_c(r,0)\, dr}{b_l}, \tag{2.19}$$

with $n_{E,0}$ being the total number of moles of enzymes in the system. Here, the enzyme loading is expressed in terms of the initial moles of adsorption sites per mass of substrate, i.e. an enzyme loading of 2 corresponds to twice the amount of enzymes required to cover all initially accessible binding sites on the cellulose surface.

### 2.2.3 Numerical implementation

Numerical solutions for the system of PDEs are obtained using the Method of Lines[165], in which all but one variable are discretised, leading to a system of decoupled ordinary differential equations (ODEs) for which efficient solvers exist. Here, the spatial coordinate representing the particle radius is partitioned into $n$ regions, with layer $n+1$ representing the surrounding bulk solution. The resulting $3(n+1)$ coupled ODEs, mirroring the initial PDEs system, are then solved in a dimensionless form by scaling the surface and fluid concentrations with the initial bulk concentration:

$$C_E^F(r,t) \longrightarrow \chi_E^F(r,t) = \frac{C_E^F(r,t)}{C_{E,0}} \quad \text{and} \quad C_E^S(r,t) \longrightarrow \chi_E^S(r,t) = \frac{C_E^S(r,t)}{C_{E,0}}, \tag{2.20}$$

leading to an equivalent patially non-dimensional system of coupled equations (upon insertion in Eq. (2.16),(2.17) and (2.18)):

$$\begin{cases} \dfrac{\partial \chi_E^F(r,t)}{\partial t} = D_E \left[ \dfrac{\partial^2 \chi_E^F(r,t)}{\partial^2 r} + \dfrac{1}{r}\dfrac{\partial \chi_E^F(r,t)}{\partial r} + \dfrac{1}{\varepsilon(r,t)}\dfrac{\partial \chi_E^F(r,t)}{\partial r}\dfrac{\partial \varepsilon(r,t)}{\partial r} \right] \\ \qquad\qquad - \dfrac{1}{\varepsilon(r,t)}\dfrac{\partial \chi_E^S(r,t)}{\partial t} - \dfrac{\chi_E^F(r,t)}{\varepsilon(r,t)}\dfrac{\partial \varepsilon(r,t)}{\partial t}, \\ \dfrac{\partial \chi_E^S(r,t)}{\partial t} = k_{ads}\chi_E^F(r,t)C_{E,0}\left[\chi_{E,max}^S(r,t) - \chi_E^S(r,t)\right] - k_{des}\chi_E^S(r,t), \\ \dfrac{\partial \varepsilon(r,t)}{\partial t} = \dfrac{k_{des}\chi_E^S(r,t)C_{E,0}M_p MM_{glu}H_{glu}}{\rho_C^{IV}}, \end{cases} \tag{2.21}$$

with associated IC and BC:

$$IC \begin{cases} \chi_E^F(r,0) = 0 \quad \forall\, r \neq R \qquad \chi_E^F(R,0) = \chi_{E,0} \\ \chi_E^S(r,0) = 0 \quad \forall\, r \\ \varepsilon(r,0) = \varepsilon_0 \quad \forall\, r \neq R \qquad \varepsilon(R,0) = 0. \end{cases} \tag{2.22}$$

$$BC \begin{cases} \left.\dfrac{\partial \chi_E^F(r,t)}{\partial r}\right|_{r=0} = 0 \\ \left.\dfrac{\partial \chi_E^F(r,t)}{\partial t}\right|_{r=R} = -\dfrac{S_{cyl}}{V_B}\varepsilon(R,t) D_E \left.\dfrac{\partial \chi_E^F(r,t)}{\partial r}\right|_{r=R} \\ \left.\dfrac{\partial \varepsilon(r,t)}{\partial r}\right|_{r=0} = 0. \end{cases} \tag{2.23}$$

This new system of working equations is then solved using the ODE solver ode15s[166] in Matlab[167]. The discretisation of the radial dimension into $n$ sub-regions $\omega_i = \{(r,\varphi,z) \in \mathbb{R}^3, i \in [1,n] : (i-1)R/n \leq r \leq iR/n, 0 \leq \varphi \leq 2\pi, 0 \leq z \leq L_A\}$ as proposed by the method of lines is then based on the central finite difference, where the first $f'(x)$ and second $f''(x)$ derivatives are approximated as:

$$f'(x) \approx \frac{f(x+h) - f(x-h)}{2h} \quad \text{and} \quad f''(x) \approx \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} \tag{2.24}$$

where $h$ is the size of the discretized step and $f(x)$ is the function of interest.

The convergence of the system against the discretisation parameter $n$ is assessed to ensure reliable estimations of glucose release over time for relevant ranges of model parameters. As analytical solutions are not available for the system of equations considered here, the correctness of the implemented algorithm is evaluated by verifying proper conservation of

enzymes over the course of the simulation through the following criterion $\eta$:

$$\eta(t) = \frac{\left|n_{E,tot}(t) - n_{E,0}\right|}{n_{E,0}} \tag{2.25}$$

For each simulations, discretisation number is chosen to ensure that deviations in the total number of moles of enzymes in the system remain within $< 2\%$ of the initial loading, while minimising computational cost (see Appendix A.6).

### 2.2.4 Error propagation

To account for uncertainties associated with experimental measurements, and in particular those associated with the experimental determination of the pore volume distribution, a Monte-Carlo uncertainty analysis was performed in silico. To this end, the initial standard error associated with the accessible pore volume measured experimentally was sampled assuming a normal distribution using the function normrnd in Matlab with sampling number $N$. Simulations were then run for each of the $N$ generated data sets, which allowed for computing uncertainties on the predicted glucose yields.

## 2.3 Results and Discussion

### 2.3.1 The relationship between enzyme loading and cellulose accessibility

In previous work[153,154], not considering enzyme depletion in the bulk provided rather good predictions of initial yields and quantitatively confirmed the importance of surface accessibility as a key parameter on the hydrolysis. However, with the more complicated model developed here, we demonstrate that enzyme depletion can strongly impact the rate of cellulose hydrolysis especially for specific biomass structures [Figure 2.2]. We compare predictions

from our previous work with those generated using the current improved model on the same data sets[107,153,154] (see Appendix A.7).

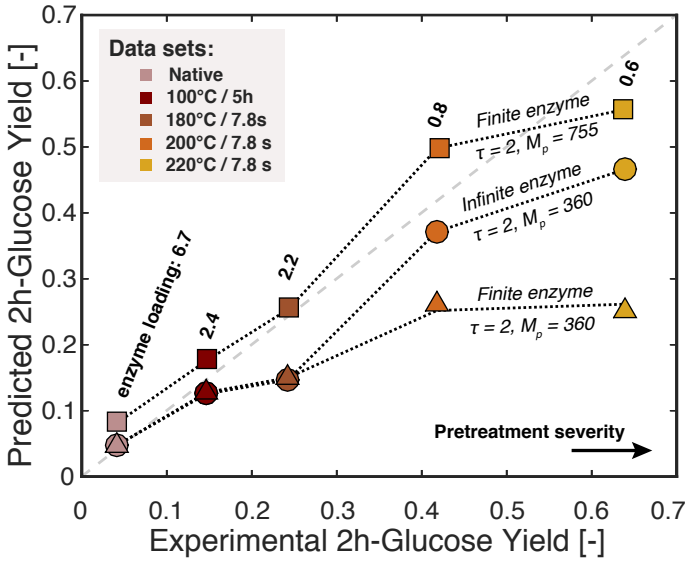



Figure 2.2 – *Comparison of predicted initial glucose yields to experimental data taken from literature for native and acid-pretreated (1% acid) mixed hardwood at increasing degree of severity*[107]*. Predictions from both models – accounting for finite or infinite enzyme loadings – are reported keeping fitted parameters unchanged from the infinite enzyme fit ($M_p$ = 360, $\tau$ = 2). Also shown are the early glucose yield predictions considering finite enzyme with refitted parameter ($M_p$ = 755, $\tau$ = 2) in terms of least-square fittings over the whole range of substrates considered. Enzyme loadings normalized by the number of adsorption sites are shown for each substrates and correspond to an experimental enzyme loading of 92.5 mg cellulase/g.*

Using the fitted parameters ($M_p$ = 360 and $\tau$ = 2) from the model assuming infinite enzyme loadings (no change in the bulk concentration of enzymes), accounting for enzyme depletion within the new framework worsens early yield predictions for the more severely pretreated hardwood, while generating similar results for mildly pretreated and native substrates [Fig-

ure 2.2]. For the latter, the initial excess cellulases in terms of adsorption sites ($e_l$>2) makes the constant enzyme bulk concentration hypothesis reasonable. For more digestible substrates, the internal mass transfer is slowed by enzyme depletion in the bulk and to incomplete initial surface coverage, which significantly slows the initial glucose release compared to a case where no enzyme depletion is accounted for. As we will discuss below (see subsection 2.3.2), this fractional coverage of available cellulase binding sites can play a significant role in controlling the rate. Due to the difference in pretreatment severity and thus porosity, a significant difference exists across various substrates in number of adsorption sites. As a result, the ratio of enzyme loadings to the number of available adsorption sites at the start of the reaction can vary significantly for the different substrates even if the enzyme loading per mass of substrate is kept constant (0.925mg/ml)[107] (see Appendix A.7) [Figure 2.2]. These differences, in turn, lead to significant variations in the maximum possible coverage of the enzymes at the start of the reaction, and thus can reduce the initial rate for cases where enzyme depletion leads to less than full coverage at the beginning of the reaction.

Improved predictions were obtained by refitting the parameters to the new model. The combined least square fitting of both parameters on the whole set of data led to new values of $M_p$ = 755 glucan monomers liberated per binding cycle of one enzyme and $\tau$ = 2 (see Appendix A.8). These parameters allowed for an accurate prediction of early glucose yields in good agreement with experimental observations [Figure 2.2]. The significant increase of $M_p$ can be attributed to the previous overestimation of the number of enzyme present in the system. Assuming an infinite number of enzyme in silico when in reality a limited quantity was present lead to an underestimation of the hydrolytic capacity of the enzyme. Interestingly, the pore network complexity had an important impact on the mildly pretreated substrates,

while having a more minor influence on both native and more severely pretreated substrates (see Appendix A.8). In the case of the low accessibility extreme, the reaction rate was governed by a poor cellulose accessibility due to the low porosity for the native substrate. For the more severely pretreated wood samples, which was the highly accessible extreme, the rate was almost purely governed by the surface reaction rate and the pore network complexity played a fairly limited role.

Even though pore connectivity has been shown to increase (i.e. decreasing tortuosity) with the severity of acid-pretreatment of populus substrates (0.1 M sulfuric acid (SA) / 160 °C / 5-60 min)[168] and similar treatments increase cellulose digestibility[169–171], no clear trends were predicted by the fitting of $\tau$ and $M_p$ for individual substrates (see Appendix A.8). In addition, a single set of fitted parameter allowed us to accurately predict initial rates for a range of acid-diluted pretreated substrates, suggesting that similar modifications in the pore network structure and cellulose susceptibility to digestion occur independently to pretreatment severity. However, this explanation should be treated cautiously, as no information on the particle size distribution (PSD) after pretreatment was available and predictions were based on the size of the native substrate (with diameter of 25 µm). Even though this important structural factor can play a significant role on early glucose yield (see section 2.3.3), it seems unlikely to drastically shift predictions in this case, as the already small size of the native particle is expected to exhibit low diffusional resistance.

### 2.3.2 High enzyme- and biomass-loadings as drivers for enzyme penetration

To evaluate the model's predicted effect of experimental parameters on the course of the reaction, cellulose depolymerisation was run in silico for relevant ranges of enzyme- and biomass-loadings, using as model substrate an acid-pretreated mixed hardwood (dilute-acid

pretreatment at 220°C for 7.8s). This substrate had a small particle size (assumed to be 25 µm, due to sieving) with a high digestibility (>80% glucose yield after 24 hr).

For such small particles, both internal and external mass transfers are expected to be minimally limiting the rate (see Appendix A.1). Nevertheless, varying enzyme loadings for a fixed solid loading – in this case 2% – strongly affects the course of hydrolysis. When the amount of cellulase is low compared to the number of adsorption sites ($e_l$<1), a specific pattern in the glucose release emerges; up until the surface is completely covered with enzymes, the rate of cellulose degradation is limited by the number of adsorbed cellulases, which is controlled by the quantity of free enzymes in the pore lumen [Figure 2.3 a/b]. Once the number of enzymes matches the available amount of adsorption sites in the system, the rate rapidly increases to match the maximum surface reaction rate. This rate only decreases once the quantity of remaining accessible cellulose decreases significantly, towards the end of hydrolysis. This initial rate transition becomes less noticeable with increasing enzyme loadings, as the surface becomes saturated more rapidly.

For systems where the initial amount of enzymes matches or exceeds the number of adsorption sites on the cellulose surface ($e_l \geq 1$), further increasing the cellulase loading only slightly improves the rate of glucose release. This slight improvement in rate is due to the increase in enzyme concentration gradient in the particle within the 30 minutes of this multi-hour reaction. Glucose generation is almost solely reaction-limited for this particle size, as the entire available cellulose surface is rapidly covered leading to this small effect. Without deactivation mechanisms, the model shows that, at a certain point, working with large enzyme excesses does not improve performance. Importantly, the model allows to clearly extract diffusion effects from these other enzymatic phenomena, which we are discussed in the following

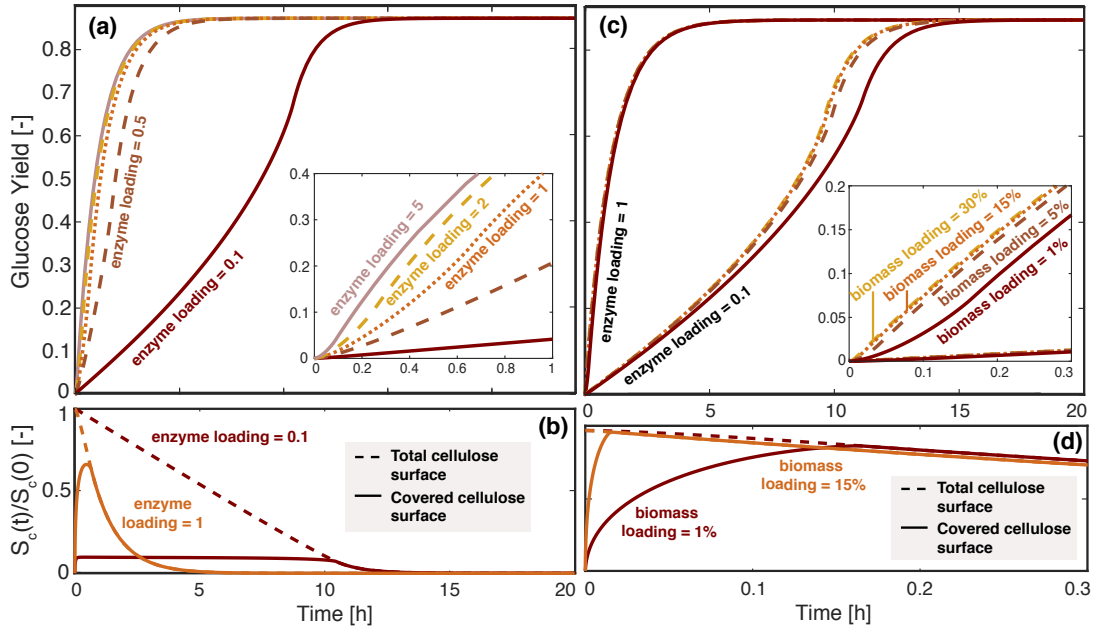



Figure 2.3 – *Predicted glucose yield and surface coverage with time. Glucose yield as a function of time for (a) different enzyme loadings at a constant solid loading of 2%, and (c) different biomass loadings. (b)/(d) Illustrations of the progressive coverage of the cellulose surface for the chosen simulations: for (d), the case of an enzyme loading of 1 was used.*

Chapters (see Chapter 3 and 4).

To illustrate these effects, a set of experimental data was generated on dilute-acid pretreated beech wood (1% sulfuric acid / 160°C / 30min) by varying the enzyme loading (from $e_l$ = 0.3 to 2 in terms of surface initial coverage, corresponding to 13-85 FPU/g) for a set biomass loading of 7% dry matter (DM) [Figure 2.4] (see Appendix A.7 and A.9-A.12 for methods and substrate's characterisation). Wet sieving after pretreatment ensured a narrow particle size distribution around 400 µm and cellulose accessibility was determined by solute exclusion to be about 24 $m^2\,g^{-1}$ (see Appendix A.12). Using the set of parameters Mp = 755 and $\tau$ = 2 fitted on data set from the literature[107,153,154](see subsection 2.3.1), predicted early glucose yields showed

good agreement with experimental data, even when working in "enzyme-limited" conditions [Figure 2.4]. Our ability to accurately capture this kinetic data demonstrated the dependence of the rate on enzyme coverage and the dependence of this coverage on enzyme loading.

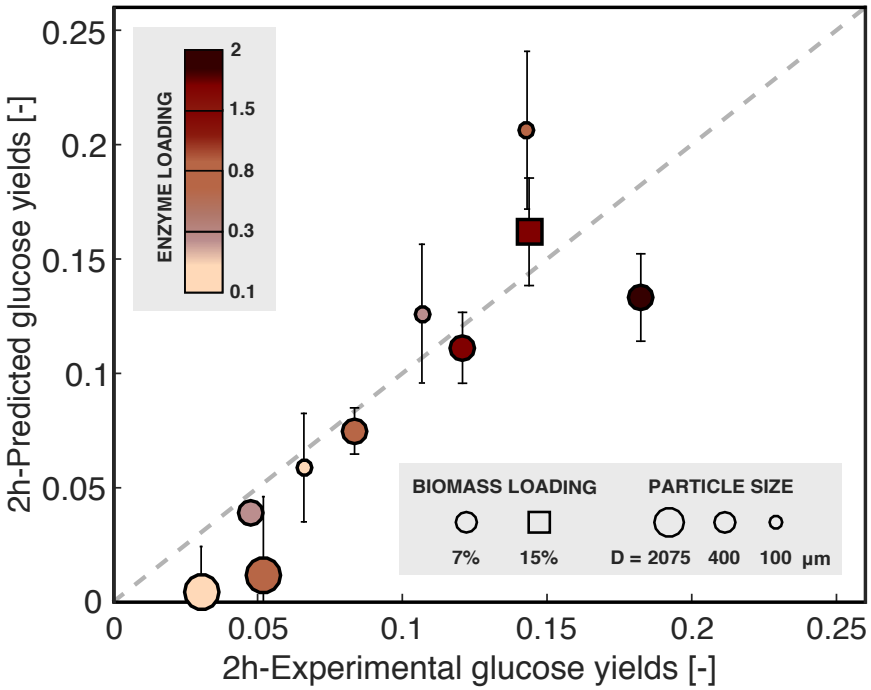



Figure 2.4 – *Comparison of predicted initial glucose yields and experimental data for a range of dilute-acid pretreated beech wood substrates (1% SA / 160°C / 30min) (see Appendix A.7). Experiments and corresponding simulations were performed over a range of enzyme- and biomass-loadings. Standard errors were obtained by propagating the uncertainties associated with pore volume measurements.*

However, a larger deviation was observed for high enzyme loading ($e_l$=2). This deviation could reflect the limitations of the assumptions underlying the quantification of both available cellulases and cellulose adsorption sites (i.e. double-slit pore geometry, enzyme surface footprint, homogeneous distribution of component throughout the particle, no distinction between different cellulases constituting the enzyme cocktail) which affected our estimate of minimum

enzyme loading required to get the maximal initial rate. In this case, the decrease in predicted yields suggests an underestimation of the number of accessible adsorption sites. Interestingly, the initial number of adsorption sites for all substrates was estimated between 0.12 to 1.5 $\mu mol\,g^{-1}$ of cellulose, which is on the lower end of productive binding measured on cellulosic substrates by Cel7A adsorption, (0.1 to 10 $\mu mol\,g^{-1}$)[158]. Recent adsorption measurements on lignocellulosic substrates, using two different types of recombinant CMB-proteins on steam exploded pretreated birch/beech wood mixtures, have led to protein coverage of cellulose up to 20 $\mu mol\,g^{-1}$ of cellulose[172], with individual coverage by specific CBM-recombinant varying between 5.1-13.5 $\mu mol\,g^{-1}$ of cellulose. While these lower estimates could rationalise the mismatch observed at higher enzyme loading, other structure- or enzyme-related simplifications could contribute to both over and underestimating the true enzyme loading.

By contrast to variations in enzyme loadings, changes in biomass loadings, where the enzyme-to-biomass ratio is kept constant, showed more limited effects on the course of hydrolysis [Figure 2.3c]. When working at relatively high enzyme loading ($e_l$=1), the reaction was mainly controlled by the rate of enzyme adsorption and desorption with diffusion from bulk only playing a limited role. A higher biomass loading increased the mass transfer rate by accelerating the saturation of the cellulose surface because of the higher enzyme gradient within the particle. However, this phenomenon only marginally improved the rate of glucose release in the first few minutes of hydrolysis [Figure 2.3d]. In cases where the enzyme loading was low ($e_l$=0.1), increasing the concentration gradient by increasing the biomass loading was beneficial at hydrolysis times beyond 5 h [Figure 2.3c], which is when the system becomes more diffusion limited after initially being almost entirely reaction limited due to the low surface coverage. To test the ability of the model to predict early glucose yields upon changes

to biomass loadings, enzymatic hydrolysis was performed at a high loading of 15%DM with an enzyme loading of 1.5 for the DAP pretreated beech substrate [Figure 2.4]. The model was able to capture the significant increase in glucose titers observed experimentally, confirming the beneficial effect of working at relatively high concentrations of both biomass and enzyme to drive internal mass transfer. In this case, unaccounted adverse effects related to change in rheological properties[136,173] and increased enzyme deactivation[136,174–176] that might have occurred when decreasing amount of free water in the system, did not appear to limit the glucose release. However, such effects are expected to become more pronounced as the biomass loadings increases even more, and not taking them into account might further limit a model's predictive ability.

### 2.3.3 Role of particle size reduction on internal mass transfer

In this last part, we integrate the role of mass transfer effects as the substrate's size increases. To illustrate this, we can assume the same model substrate (dilute-acid pretreatment at 220°C for 7.8s), but vary the particle radius in silico to assess the effects on glucose release. Similar to contradictory results from the literature[136,150,175], the important role played by the substrate's size on the hydrolysis rate depends on the experimental conditions. While increasing enzyme- and biomass-loading favours the enzyme penetration into the substrate by reinforcing the concentration gradient throughout the particle, their effect varies with particle size [Figure 2.5]. Compared to particles that had an intermediate radius in the range of sizes considered here, changing biomass loading had a limited effects for both small (R<10 µm) and large (R≥0.1 cm) particles [Figure 2.5a]. Increasing solid loading leads to increasing enzyme concentration in the bulk when the enzyme-to-substrate ratio is kept constant. However, for the larger particles, this extra driving force is not sufficient to compensate for increasing internal mass transfer

limitations. In such cases, working at enzyme loadings beyond full coverage of adsorption sites helps compensate for limited enzyme penetration into the substrates by maintaining high enzyme concentration in solution and thus limiting any enzyme depletion in the bulk throughout the reaction [Figure 2.5b]. These important internal mass transfer limitations could contribute to the lag sometimes observed experimentally in the glucose release as the dry matter loads are increased[136]. In such cases, the progressive liquefaction of the substrate could be reducing large particles to smaller sizes which would transition the overall process from a system that is severely limited by internal mass transfer to one that is less so. This would translate to an accelerating hydrolysis rate. When compared to experimental results for dilute-acid pretreated beech wood (1% sulfuric acid / 160°C / 30min) ranging from 100 to 2075 μm in radius (see Appendix A.7), the model provided reasonable predictions for early glucose yields [Figure 2.4]. As previously observed, predictions considering higher enzyme loadings exhibit larger deviations from experimental values. However, by contrast to the 400 μm particle discussed above, predictions for the smaller substrate considered (100 μm) at high enzyme loadings overestimated observed yields. In this case, other phenomena, such as deposition of lignin on the cellulose surface or enzyme jamming, may counterbalance the initial underestimation of binding sites.

Overall, even though mismatches occur in the more extreme cases of experimental conditions tested, the strong correlations obtained when predicting early glucose yields over a range of data sets using a single set of fitted parameters validates core modelling assumptions. In this regard, the modelling results can aid in the design of efficient hydrolysis process. For example, working at increasing enzyme loading to allow a rapid degradation of cellulose in relatively small sized substrates only works to a point. In these cases, loading more cellulases

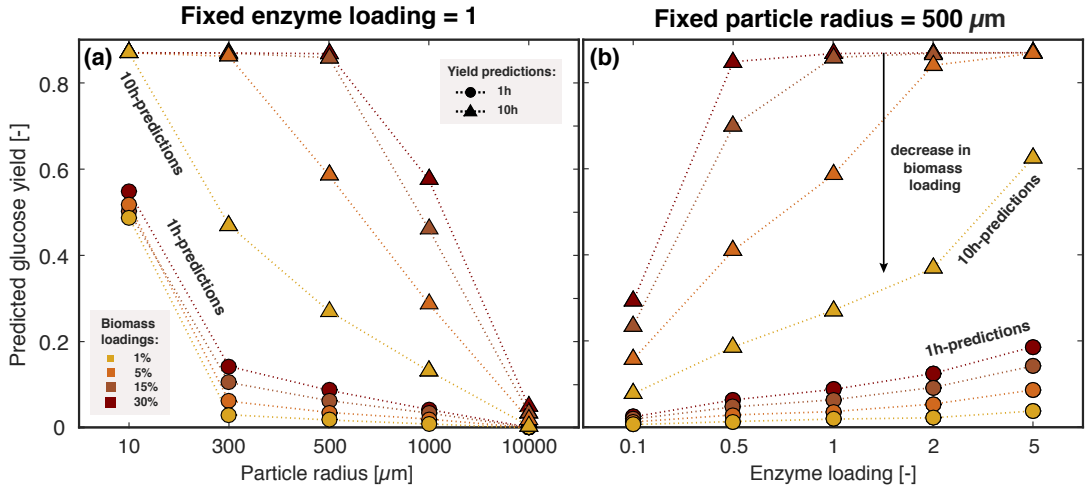



Figure 2.5 – *Effect of particle radius and experimental conditions on the predicted yields. Glucose yields after 1 hr and 10 hr of hydrolysis, for (a) variable particle radius as well as biomass loading at fixed enzyme loading, and (b) fixed particle radius with variable enzyme- and biomass-loading.*

than there are available adsorption sites brings only little benefit. In contrast, for larger size substrates exhibiting important diffusional resistance, high enzyme loadings that are beyond full coverage appear beneficial to increase early glucose release. Tailoring of enzyme quantity to cellulose accessibility is thus not only important for improving process, but also to compare the pretreatment's ability to promote cellulose degradation. For the latter, since glucose release observed for a given substrate will strongly depend on the enzyme's ability to completely cover the initial cellulose surface, one should consider comparing systems based on enzyme loadings reported per available surface area instead of per mass of cellulose.

### 2.3.4 Model variables

#### 2.3.4.1 Rate constants

In this work, we assumed a general desorption rate constant, with no distinction made between desorption and decomplexation rate. This global desorption rate constant used was $k_{des}$=0.011 $s^{-1}$, which lies within the range values discussed in literature. In particular, Nill et al. reported values for TrCel7A lying between 0.125-0.067 $s^{-1}$ for the desorption step and 0.0007-0.14 $s^{-1}$ for decomplexation[155]. In this particular research, several measured rate constants from multiple studies were reviewed and compared[155].

Comparing values for the binding rate constant is less straightforward, as its value depends on the definition of adsorption sites. The binding rate constant (which is defined here as an average with no distinction between adsorption and complexation) is expressed in terms of moles of enzymes per volume of substrate per minute, which is the same basis used to describe the concentration of adsorption sites on the cellulose in the model. This definition makes direct comparisons with literature values challenging. However, by assuming a particle with density including of 0.9 in solution and a loading between 2-10%, the desorption rate constant used in this study can be estimated at around 0.01-0.05 $\mu M^{-1} s^{-1}$, which is within the same order of magnitude reported for adsorption rates of 0.097-0.33 $\mu M^{-1} s^{-1}$ and complexation rates of 0.018-0.029 $\mu M^{-1} s^{-1}$ by Nill et al[155]. The slightly lower adsorption rates obtained can likely be explained by the fact that this parameter includes all enzymes, including some enzymes that bind and others that don't or may bind irreversibly to other parts of lignocellulosic biomass.

#### 2.3.4.2 Catalytic efficiency and $M_p$

Due to the simplified mechanism assumed in this study, direct comparison of the catalytic efficiency agglomerated in the lumped parameter $M_p$ with values from literature is challenging. Here, $M_p$ represents the susceptibility of cellulose to be hydrolysed by a specific enzyme cocktail, and implicitly englobes all phenomena that can impact the cellulose hydrolysis by cellulases, stemming from both the cellulose ultrastructure (cristallinity, type adsorption sites celluloses available, presence of lignin droplet etc.), the enzymes constituting the cellulase cocktail (e.g. processivity on the cellulose surface, catalytic rate constant, mechanism of action...) and their mutual interactions (synergism, jamming...). As such, $M_p$ should not be directly compared to either catalytic rate constants or reported apparent processivities of cellulases on cellulose. However, it is still of interest to examinate where our fitted parameter lies compared to values of processivities in particular, if only for evaluating its physical relevance. While values reported here for Mp (~500 glucan monomer release per binding cycle, representing ~250 cellobiose units) are ~5-10-fold greater than usual values of apparent processivity for cellulases estimated in literature[117,155]. This discrepancy can be likely attributed to the fact that, by representing the whole enzyme cocktail and not a specific enzyme, $M_p$ is accounting for synergistic effects between the cellulases constituting the enzyme blend and also agglomerating both first (cellulose to cellobiose)- and second (cellobiose to glucose)-hydrolysis steps. By contrast, our $M_p$ values fall well below reported intrinsic processivity for TrCel7A on crystalline cellulose ( 4000), which is in line with expected decrease in hydrolysis efficiency for cellulose surface presenting heterogeneities[117]. Even though our $M_p$ parameter cannot be directly correlated to processivities values, we see that its values are physically not out of range.

## 2.4 Conclusions

Exploring the effect of various experimental and structural parameters in silico highlighted the significant role of internal mass transfer as the substrate size increases and/or the enzyme loading decreases. In such cases, diffusion of cellulases to the available cellulose surface limits the rate of glucose release, with an increase in rate observed once the cellulase quantity matches that of available adsorption sites. These modelling results highlight the importance of considering multiple process parameters simultaneously and tailoring experimental conditions to the substrate specificities when designing an enzymatic reaction to maximise rates and discussing pretreatment efficiencies.

Implicitly expressing the enzymatic action through the use of a lumped, fitted parameter appeared to be sufficient to reveal internal mass transfer phenomena, which impact early hydrolysis rates, while keeping simulations computationally tractable. The evolution of surface accessibility is not able to solely capture late trend hydrolysis. However, validation of the model at early stages of the hydrolysis, in which the substrate structure can be assumed, with fairly high confidence, to remain unchanged and in which enzyme deactivation is assumed to be minor, is necessary to verify that the model is accurately capturing native structural features. Here, we demonstrated that the simple diffusion-reaction model that was developed was robust enough to predict initial glucose release rates for a range of substrates and experimental conditions based on a single set of fitted parameters $M_p$ and $\tau$. As such, this model constitutes a basis to further investigate the role of particle morphology evolution throughout hydrolysis and later stage kinetics by introducing additional mechanisms.

# Chapter 3

# Modelling structural heterogeneities in the lignocellulosic substrate

## 3.1 Introduction

In this chapter, we add complexity to the model developed in Chapter 2 to evaluate the role of lignocellulosic structures on later stage hydrolysis rates. In addition to the inherent non-uniform structure of the plant cell structure and tissues, initial size reduction and thermochemical pretreatment can enhance the native heterogeneity across the substrate. In particular, (i) cellulose accessibility[86,107,114,177,178], which is controlled by a combination of particle size, porosity and maximum fraction of cellulose susceptible to hydrolysis, and (ii) cellulose ultrastructure[120,121,147,155], which is notably related to its crystallinity, degree of polymerisation and overall surface morphology, are two important global parameters frequently use to explain hydrolysis trends. As it is usually difficult to measure these parameters at the microscale, only average, bulk measurement are available to describe the pretreated lignocellulosic substrate. In addition, deconvoluting the concurrent effects of these struc-

tural parameters is similarly challenging experimentally. In this context, we test the effect of structural heterogeneities across substrate and their effect on hydrolysis rates using the kinetic and mass transfer model developed in Chapter 2. In particular, this chapter focuses on parameters related to the initial cellulose accessibility, i.e. fraction of hydrolysable cellulose and the particle- and pore-size distribution, and discuss the extent of both their combined and individual contribution to the observed rate.

## 3.2 Modelling framework

Based on the model developed in Chapter 2, we investigated the potential effect of the substrate physical heterogeneities on the hydrolysis rate. In particular, the model is modified to include (i) the effect of the particle fragmentation upon pretreatment, (ii) variable pore volume distributions between particles and (iii) variable fraction of hydrolysable cellulose across the particle fragments constituting the wood sample [Figure 3.1]. These effects are considered both individually or in a combined manner. Conceptually, increasing the pore volume for a given particle initially increases the accessible cellulose for the given particle, by not only facilitating enzyme diffusion, but also increasing available accessible cellulose surface. By contrast, decreasing the fraction of hydrolysable cellulose does not impact initial cellulose accessibility, but represents a case where hindrances to the complete hydrolysis arise over the course of reaction.

Briefly, for a given substrate, the average accessible pore volume measured experimentally is distributed throughout differently sized particles of substrate as a function of the particle size in a way ensuring that, globally, the combined accessible pore volumes quantitatively match the average bulk value (see Appendix B.1). Similarly, the fraction of hydrolysable cellulose, given by the final glucose yield, is varied as a function of the particle size, but in a stepwise

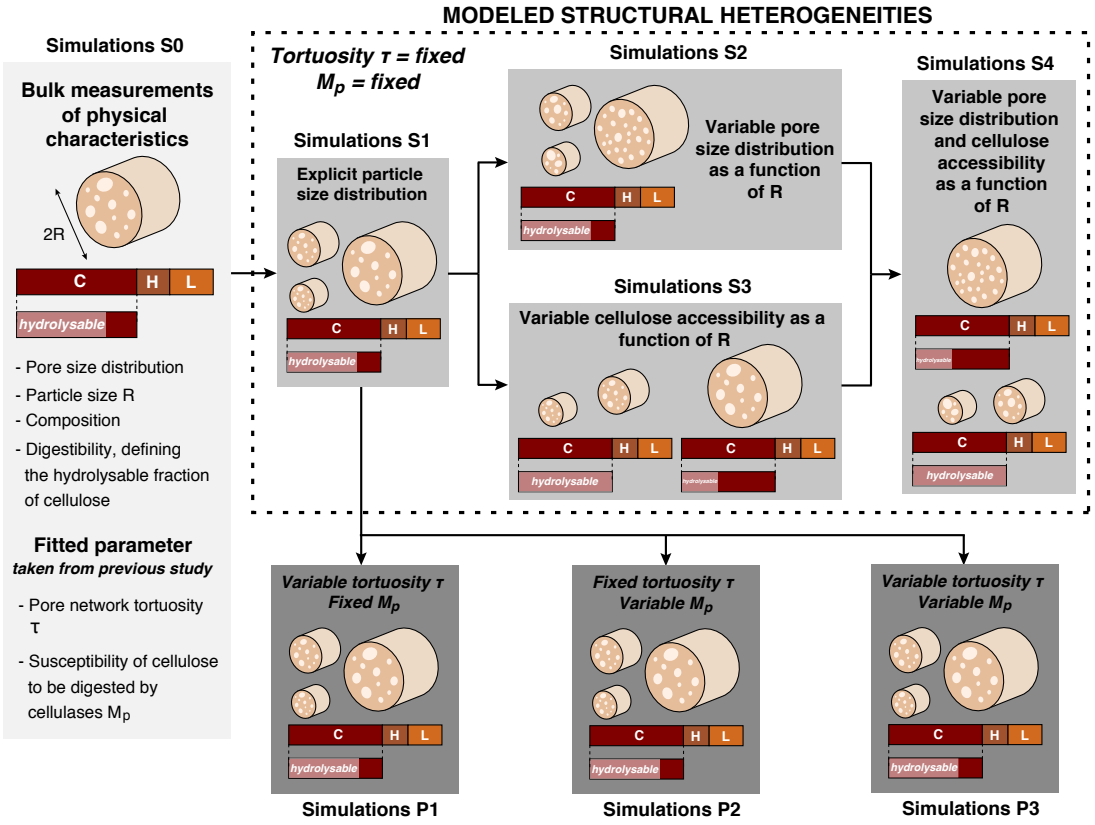



Figure 3.1 – *Schematic representation of the modelled substrate in the various simulations performed. First, the model used average physical characteristics to model the substrate (S0). Following this, structural heterogeneities are progressively included (S1-S4) in the model while keeping the two fitted parameters, representing the pore network tortuosity* $\tau$ *and cellulose susceptibility to be digested* $M_p$*, fixed: simulations S1 include the particle size distribution (PSD), simulations S2 include PSD and non-uniform pore volume distribution, simulations S3 include PSD and non-uniform cellulose accessibility and simulations S4 include all of these effects. Finally, the two parameters* $M_p$ *and* $\tau$ *are also allowed to vary (P1-P3). For the latters, simulations include the PSD.*

manner; the cellulose contained in particles presenting a radius below a defined threshold, $R_{thres}$, is considered fully accessible, while the remaining particles contained reduced fractions of hydrolysable cellulose (see Appendix B.2). Here, we explore non-uniformity in the fraction of hydrolysable cellulose by assuming an extreme case where some particles contain 100% digestible cellulose, while others have a fraction of recalcitrant cellulose at the end of the reaction. Along with this rule, the total amount of cellulose susceptible to digestion is distributed across the particles constituting the sample to match the defined, average cellulose hydrolysability. Particle size distribution are included based on a weight average of simulations for specific particle radii (see Appendix B.3 and B.4). In this study, we refer to the maximum amount of cellulose that can be digested by cellulases for a given substrate as the fraction of hydrolysable cellulose. As such, it is mathematically defined as the final glucose yield. This parameter is not to be confused with the cellulose accessibility that represents the ease with which cellulases reach their substrate and arises from a combination of particle size, porosity and cellulose surface area.

The modelling strategy applied here is as follows: in a first stage, the aforementioned structural heterogeneities are investigated while maintaining the values of the two fitted parameters $M_p$ and $\tau$ constant [Figure 3.1, simulations S0-S4]. The values of these parameters are based on those obtained in Chapter 2, $M_p = 755$ and $\tau = 2$. This is to avoid a situation where a prior re-fitting of these two parameters will compensate for structural effects, and thus prevent the study of such effects. As the particle size distribution is based on experimental observations, its contribution is always included in all simulations, except ones meant to compare the effect of the inclusion particle fragmentation, which instead use the dry, native radius of the substrate [Figure 3.1, simulations S0]. In a second stage, the values of the two fitted parameters

are adapted to the simulations to get insights on (i) the model's sensitivity to the pore network tortuosity (via $\tau$) and (ii) the susceptibility of cellulose to be digested by the given enzyme cocktails (through $M_p$) [Figure 3.1, simulations P1-P3]. Contributions from other mechanisms impacting the hydrolysis rate – and not included in the model framework – might potentially impact the values of fitted parameters, which could obscure the extent to which they affect the process in reality. While such interferences cannot be excluded, important variations across substrates and/or over the course of the reaction will still provide useful insights.

Note that, for sake of clarity throughout the following sections, simulation references as defined in Figure 3.1 are used as a guide for the reader. All simulations were performed with a grid size of n=50, which was sufficient to ensure the conservation of the total number of enzymes in the system over time. If not stated otherwise, values for the two model variables that are the tortuosity $\tau$ and $M_p$, the variable representing the number of unit glucose released per enzyme binding cycle, were taken as those fitted in using the finite enzyme scenario in Chapter 2, i.e. $\tau$ = 2 and $M_p$ = 755. Uncertainties in the experimental pore volume distribution were propagated as described previously (see Appendix B.5).

## 3.3 Results and Discussion

### 3.3.1 Substrates characteristics

To evaluate the effect of substrate morphology, and more specifically the initial cellulose accessibility to cellulases on the rate of glucose depolymerisation, a set of pretreatment with various degree of severity were performed on the native wood (see Appendix B.6). Briefly, the native beech wood was subjected to DAP with varying residence time (1%SA / 160 °C / 15, 30 or 60 min) and an organosolv pretreatment (1.6% HCl in dioxane / 80°C / 5h). For sake

of readabillity, the different DAP pretreated substrates are referred to by their pretreatment time, e.g. DA-30min corresponds to native wood treated at 160°C for 30 min with 1% SA. As previously discussed (see Chapter 2), enzyme loading plays an important role in mass transfer limitations. Here, hydrolysis are based on the same experimental enzyme loading for all substrates, leading to maximal loading of enzyme compared to the number of initially avalaible sites ranging from 0.5 for the organosolv pretreated sample, to 9.5 for the native substrate. Importantly, mild (DA-15min) and more severely DAP pretreated substrates (DA-30min,DA-60min and PH-DA-30min) lead to enzyme loadings of 2.5 and 1.4 respectively.

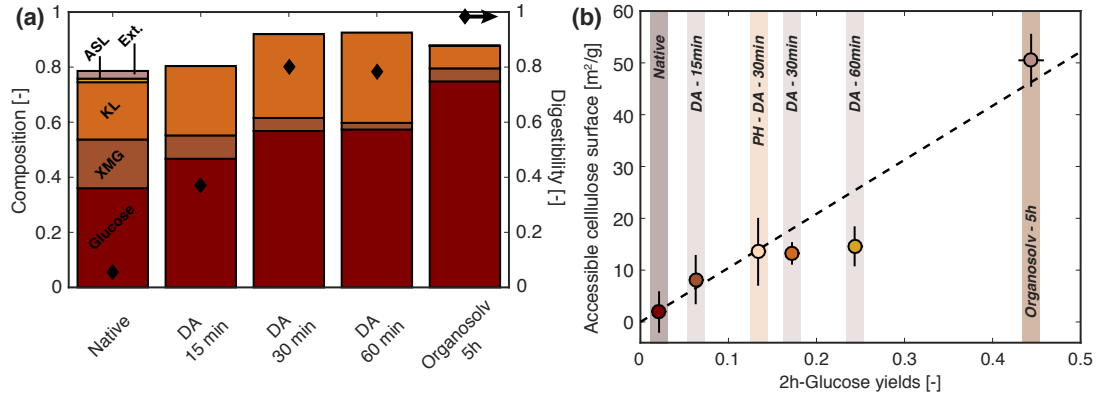



Figure 3.2 – *(a) Composition and final glucose yields for the different substrates used in this study. (b) Correlation between measured accessible cellulose surface and initial glucose yields, with associated experimental standard errors in both cases.*

Both composition and particle morphology were affected by the pretreatment, which resulted in increased cellulose content as most of the hemicellulose and some lignin were removed [Figure 3.2a]. As expected, the organosolv process removed a significant portion of the lignin from the substrate, while the mild DAP affected the composition less significantly [Figure 3.2a]. In these aqueous processes, wood recalcitrance is lowered by the rapid removal of hemicellulose, leading to the disruption of the protective hemicellulose-lignin matrix and lignin

redistribution throughout the particle. These changes caused significant structural changes, as evidenced by the increased porosity [Figure B.5 in Appendix B.8 and Figure 3.2b] and modifications in the particle size distribution [Figure 3.3 and Figure B.6 in Appendix B.8]. The native substrate had a relatively homogeneous volume distribution between 265 and 565 µm [Figure 3.3]. But, after mild DAP, particle fragmentation created many smaller particles. Nevertheless, the main contribution to the total mass was still due to larger particles (particles with diameter ≥ 365 µm represented about 70% and 60% of the total volume for the DA-15min and DA-30min, respectively). However, the distribution progressively tightened around smaller mean particle diameters (shifting to 190 µm for the DA-60min compared from around 365 µm for other DAP pretreated substrates) as the conditions became harsher. Wood subjected to the organosolv process lead to a fine particle suspension with particles having a mean diameter of 15 µm [Figure B.7 in Appendix B.8].

Accessible cellulose surfaces, derived from both pore volume distribution analysis and substrate composition, correlated well with initial glucose yields [Figure 3.2b], confirming previous studies[5]. Interestingly, while extending the DAP reaction time from 30 to 60 minutes only marginally affected the initial accessible cellulose surface (from 13.2 to 14.6 $m^2 g^{-1}$ respectively), the initial glucose yield increased by 40% [Figure 3.2b] emphasizing the importance of other factors explaining trends in early rates. Structural alterations of cellulose, such as change in crystallinity and degree of polymerization, and/or lignin redistribution and formation of pseudo-lignin, have all been reported to be pretreatment-driven changes influencing biomass recalcitrance[169].

Changes in substrate morphology related to pore network complexity, and contributions from both small and large porosities as well as particle size could also affect early digestibility.

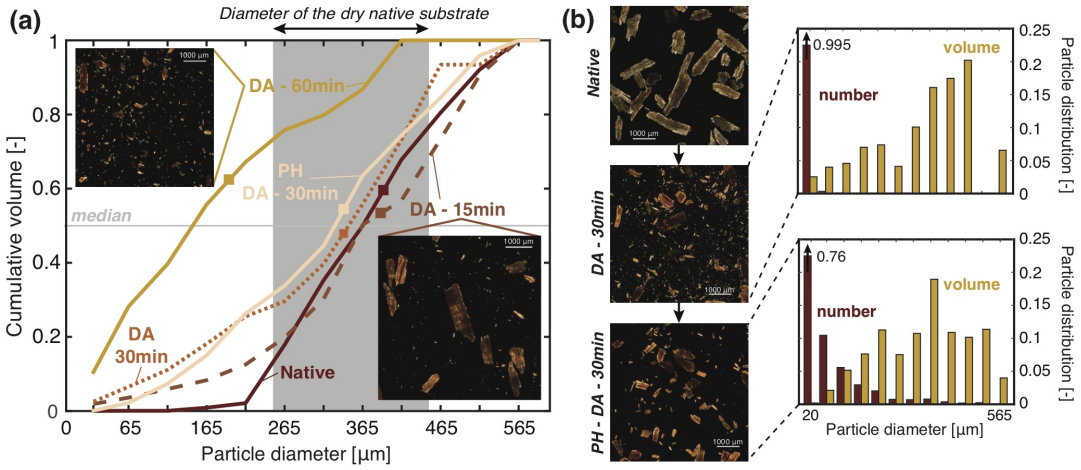



Figure 3.3 – *Substrate's particle size distribution. (a) Cumulative particle volume for the dilute-acid pretreated and native substrates as a function of particle diameters, with indicators of both median (horizontal line) and mean values (shown as ■) of the distribution. (b) Change in particle size distribution after 2 hr of enzymatic hydrolysis for the sample pretreated with dilute-acid for 30min.*

In this particular case, a significant shift towards smaller ranges in the PSD [Figure 3.3 and Figure B.6 in Appendix B.8] coupled to an increase in the accessible pore volume by 16.7% [Figure B.5 in Appendix B.8] was observed between the DA-30min and DA-60min substrates. This could increase cellulose accessibility in DA-60min by facilitating enzyme diffusion inside the particle, as further discussed in the following section.

To gain some insight into the evolution of the particle morphology during the course of enzymatic hydrolysis, digestion was interrupted after 2h and the pore- and particle-size distribution of the recovered solids were characterized [Figures 3.2b and 3.3]. The solids were then subjected once again to hydrolysis conditions to complete the cellulose depolymerization. This experiment (PH-DA-30min) was carried out with the DA-30min substrate, as it initially presented an intermediate cellulose accessibility and particle distribution among the

pretreated wood samples considered in this study. Compared to its native counterpart (i.e. DA-30min), the PH-DA-30min exhibited a significant increase in porosity – a 51% increase in accessible volume [Figure B.5 in Appendix B.8] – and a more homogeneous PSD, with fewer of the largest and smallest diameter particles compared to the sample before enzymatic treatment [Figure 3.3b]. In particular, the smallest particles no longer represented the overwhelming majority of particles that were observed, suggesting that they underwent rapid digestion along with a progressive fragmentation of the larger ones, leading to an overall similar mean particle diameter. Further enzymatic hydrolysis of the PH-DA-30min substrate lead to decreased early glucose yields by about 20% compared to its un-hydrolyzed equivalent (i.e. DA-30min), even though no significant changes in cellulose accessibility were observed. The model is notably used to explore this difference in digestibility for substrates, which exhibited an a priori similar initial cellulose accessibility to cellulases but different porosity and PSD.

### 3.3.2 Importance of particle breakdown upon pretreatment to enhance hydrolysis rate

As discussed in the Chapter 2, initial depolymerisation rates were closely related to the accessibility of cellulose, and in particular to the input particle size. For the substrates considered here, explicitly including the PSD in the simulations confirmed the significant role played by particle breakdown upon pretreatment. In mild pretreatment conditions (considered here to be DA-15min and DA-30min), differences in early digestibility between substrates appear to be largely controlled by changes in porosity, with relatively minor changes in particle size distribution having a minor impact on the predictions [Figure 3.4a]. In these cases, the average size of the dry native substrate is sufficient to represent the PSD reasonably well, hence leading to accurate yield predictions when $R_{native}$ is used as model input. By contrast, for the two

more severely pretreated substrates, i.e. DA-60 min and organosolv samples, the significant shift in the PSD towards smaller particle diameter significantly improved early hydrolysis rates [Figure 3.4a, Figure 3.3]. The important role of particle size breakdown is notably illustrated in the difference in yield gap observed in early stage of hydrolysis between DA-30min and DA-60min. Even though both DA-30min and DA-60min samples presented initially similar accessible cellulose surface areas, the increase in the pretreatment time led to a significant shift in the PSD, with the creation of many smaller particles [Figure 3.3]. When simulations were based on the average diameter of the native particle as model input (simulations S0, see Figure 3.1), the slight increase in porosity between the two substrates [Figure B.5 in Appendix B.8] led to very similar yield prediction, underestimating the difference seen experimentally [Figure 3.4a]. Using the measured PSD as the model input (simulations S1, see Figure 3.1) decreased mass transfer rates at early times due to the increase in the number of small particles for DA-60min and led to higher yields at 2 hr of hydrolysis, which more closely matched experimental results.

Interestingly, the percentage of increase in rate observed due to particle breakdown shows a rather weak dependence to the initial porosity, with decrease in particle size leading to the similar increase in rate across all substrates [Figure B.8 in Appendix B.8]. However, this translates to more significant improvement in glucose yields for more porous substrates. In other words, for substrates where the cellulose accessibility is limited by a low porosity, decreasing the particle size will only marginally improve hydrolysis yields. In contrast, in a regime where internal diffusion is not limited by pore volume, particle breakdown has the potential to greatly enhance fast glucose release. In all cases, even for the finer powder that mostly contained cellulose resulting from the organosolv pretreatment, the high internal

porosity and/or particle size still justify neglecting the external particle surface (representing only ~ 1% of the total available surface in the case of the organosolv pretreated substrate as computed in Appendix A.5).

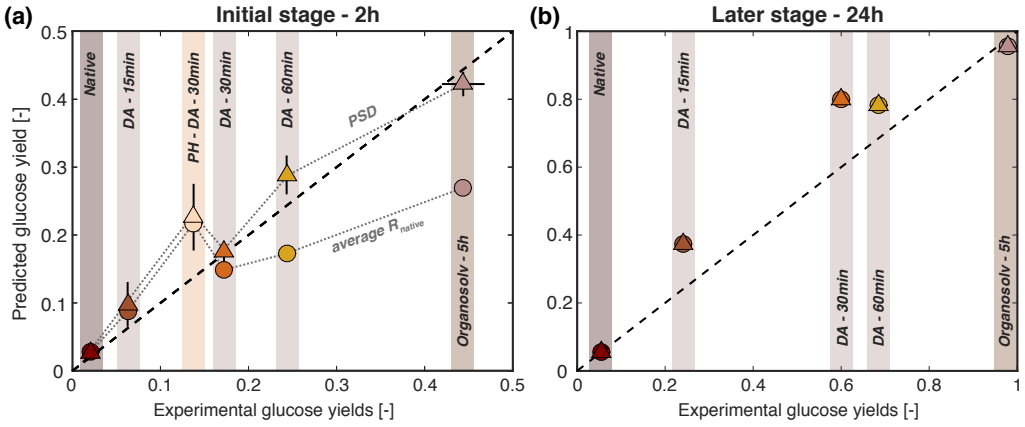



Figure 3.4 – *Impact of particle size on predicted yields. Comparison between predicted and observed glucose yields after (a) 2 hr and (b) 24 hr of enzymatic hydrolysis for native and pretreated substrates, as well as the partially hydrolyzed sample for different model assumptions. Model assumptions for the substrate particle size were either: assuming the same radius as the dry native wood (simulations S0, see Figure 3.1), or using the experimentally measured PSD (simulations S1, see Figure 3.1). Standard errors were obtained by propagating the uncertainties associated with pore volume measurements (see Appendix B.5).*

Compared to yields at 2 hr, predictions for 24 hr hydrolysis times were overestimated for most cases [Figure 3.4b], with the exception of the organosolv pretreated substrate. In this particular case, where particles presented high digestibility, the final glucose yield was already reached at 24 hr, which agreed with the model's prediction. In other cases, the model's only mechanism for slowing the hydrolysis – the gradual reduction of the available cellulose surface as the reaction progressed – appeared to be insufficient alone to explain the observed slow-

down. This discrepancy could notably point to an uneven recalcitrance of the cellulose (i.e. the remaining cellulose at the end of the reaction was more recalcitrant than the initially accessible cellulose due to change in its ultrastructure upon hydrolysis) or to an uneven distribution of cellulose accessibility across particles, both of which will be explored with our model.

### 3.3.3 Structural heterogeneities as obstacles in late stage hydrolysis

With its role in increasing the substrate's accessibility to cellulases, pretreatment can potentially amplify the heterogeneity of the lignocellulosic substrate, adding to the inherent non-uniformity of the plant structure. In addition to the observed particle fragmentation, we explored the potential effects of structural heterogeneities in the various substrate particles constituting the wood sample on the reaction rate. In particular, non-uniform pore volume distributions or/and cellulose fractions that were distinct in their ability to be hydrolysed were both incorporated in silico while still matching their average bulk measurements. From the rapid disappearance of small particles observed in the partially hydrolysed substrate [Figure 3.3], the modelled heterogeneities were generated to assign higher cellulose accessibility to smaller particles in the wood sample, i.e. higher porosity or/and higher hydrolysable cellulose fractions were assigned to smaller particles, with overall cellulose accessibility decreasing with increasing particle radius. Non-uniform pore volume distributions were randomly generated under the constraint that the particle porosity has to decrease with increasing particle size (see Appendix B.1). Concerning the hydrolysability of cellulose, we assume at first that particle with diameter smaller than 200 µm contain cellulose that are 100% digestible. This threshold represent particle containing only few layers of cells (the size of a plant cell typically ranging between 10 and 100 µm), for which we assume complete cellulose hydrolysis.

Initial rates were only marginally impacted by assuming heterogeneous porosities and cellu-

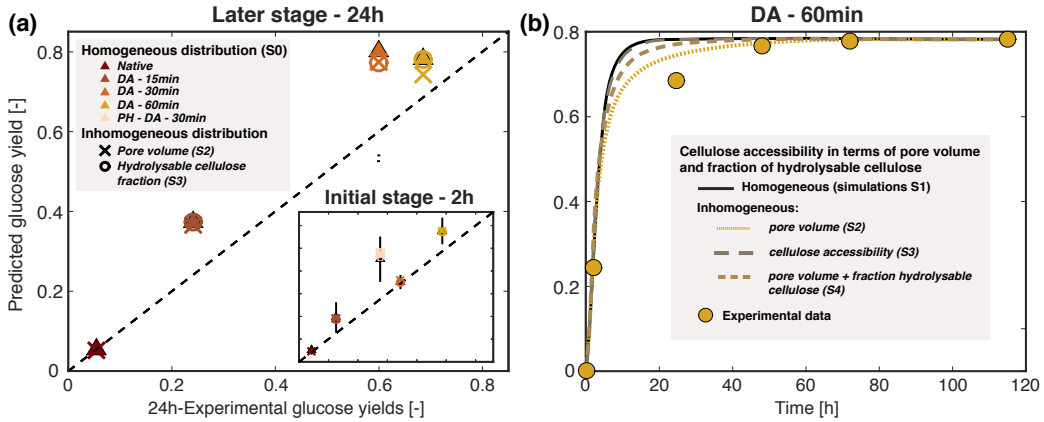



Figure 3.5 – *Impact of pore volume and cellulose content heterogeneities on predicted yields. (a) Experimental glucose yields for various substrates correlated with model predictions based on scenarios accounting for pore volume and cellulose accessibility heterogeneities within the sample: while numerous randomly generated distributions were considered for modelling heterogeneities in pore volumes, only the best fit to the experimental data in terms of least-square is shown. (b) Effect of substrates morphological heterogeneities on the glucose yields over time for the DA-60min substrate.*

lose digestible fractions [Figure 3.5a]. By contrast, creating heterogeneities in the cellulose accessibilities across particles slightly improved predictions at 24 hr. Attributing higher accessibility in smaller particles led to more mass transfer limited particles being dominant in the later stages of hydrolysis, which slowed down the process as time advanced leading to better predictions of experimental yields at later time for all pretreated substrates [Figure 3.5b]. Unsurprisingly, this effect is more pronounced for more severely pretreated substrates, which had wider particles size distributions, as illustrated by the case of DA-60min [Figure 3.5b and Figure B.9 in Appendix B.8]. Between these two modelled phenomena, discrepancies in accessible pore volume contributed the most to the rate slowdown, while variations in the

fraction of hydrolysable cellulose played only a minor role if any, even when they were varied around their assumed values. While several combinations of pore volume distribution as function of particle size were generated for each substrate, those showing the largest discrepancies between particles of different sizes led to the best predictions [Figure B.10 in Appendix B.8]. Interestingly, in the case of DAP-60min, coupling both effects and shifting $R_{thres}$ with a lower value (from 200 µm to 100 µm) lead to similar rate predictions than in the case of variations of pore accessibility only. However, in this case, the difference between specific pore volume distributions within the sample is less significant [Figure B.11 in Appendix B.8]. While differences in cellulose accessibility, in terms of accessible pore volume and fraction of hydrolysable cellulose, could partially explain the slowdown in hydrolysis, it fails at capturing the whole extent of the phenomenon by itself.

### 3.3.4 Time-evolution of cellulose-cellulases interplay as apparent key contributors to the rate slowdown, with peripheral role of the pore network complexity

Now that the influence of potential structural heterogeneities on the hydrolysis rate has been explored, the role of the two fitted parameters, that are the tortuosity of the pore network $\tau$ and the cellulose's susceptibility to be digested by a given enzyme cocktail $M_p$, is evaluated. We investigate both influences on hydrolysis yields based on model including only the particle size distribution [Figure 3.1, simulations P1-P3].

The impact of the pore network complexity, having the potentiality to affect diffusion and hence initial rates, is first assessed by varying the tortuosity parameter while keeping $M_p$ = 755 fixed [Figure 3.6a]. By simulating the random orientation of pores, values of tortuosity greater than unity, but reasonably bound to between 1 and 7[179], contribute to slowing diffusion inside

wood fragments. As expected, increasing tortuosity similarly lowered predicted glucose yields for all substrates, leading to shifted 2 hr conversions without altering the observed trend, with exception of the native and organolsolv samples [Figure 3.6a]. In these two extreme cases, where the reaction rate is governed by a poor cellulose accessibility for the native substrate and, for the organosolv substrate, purely by the surface reaction rate as cellulose is highly accessible, the pore network complexity played a fairly limited role. Even though pore connectivity has been shown to increase with the severity of acid-pretreatment of populus substrates (0.1M SA / 160°C / 5-60min) with decreases in tortuosity of up to 70%[168], the prediction trend observed here between the different DAP does not suggest a clear relationship between tortuosity and pretreatment severity for a fixed value of $M_p$ [Figure B.12 in Appendix B.8].

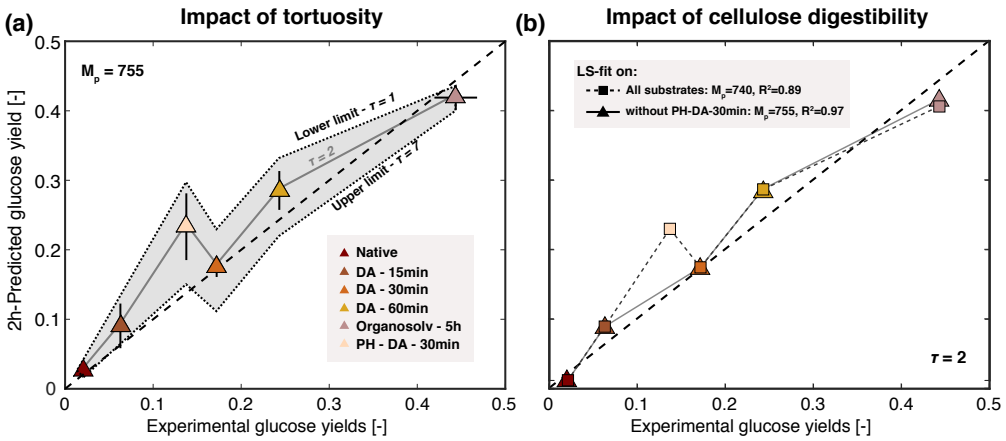



Figure 3.6 – *Impact of tortuosity and cellulose digestibility on predicted yields. (a) Impact of the fitted variable τ on initial glucose yield predictions. (b) Least-square fitting of parameter $M_p$ based on two different scenarios: considering all substrates, without PH-DA-30min substrate and without both PH-DA-30min. For each fit, both the optimal value of $M_p$ and goodness-of-fit $R^2$ are shown.*

By contrast, modifying the susceptibility of cellulose to be hydrolysed by fitting the parameter

$M_p$, while keeping the tortuosity value equal to 2, had a significant effect on the course of cellulose hydrolysis predicted by the model. The influence of $M_p$ was modelled in three different ways to attempt to explain experimental observations: (i) using uniform value of $M_p$, (ii) using non-uniform value of $M_p$ that varied as a function of the particle size, with two distinct initial populations of particles and (iii) setting a conversion-dependent value of $M_p$, that introduced an evolving cellulose hydrolysability as the reaction progressed. A non-uniformity of $M_p$ varying with particle size would indicate distinct pretreatment effects on the cellulose structure depending on the substrate size. In contrast, the role of the evolution of the cellulose ultrastructure upon digestion, which impacts the efficiency of cellulases, would be represented by in a change of $M_p$ over time.

First, allowing $M_p$ to vary as a function of the particle diameter improved long-term predictions, but at the expense of those for early yields. As was the case when varying the hydrolysable cellulose fraction [Figure 3.1 simulation S3], particles presenting a radius below a given threshold were all associated with a unique, high value of $M_p$, while a lower value of this parameter was attributed to larger particles (see Appendix B.7). While all possible combinations of radius thresholds coupled to high and low $M_p$ were then simulated, none of them led to significantly improved predictions over time for all pretreatments combined. In all cases, decreasing the value of $M_p$ to improve the fit of the hydrolysis kinetics at longer times worsened the predictions at early times. This suggests that no significant difference in cellulose's susceptibility to be digested exist between the particles constituting the sample, with pretreatment affecting in a uniform manner the cellulose structure within the substrate.

Interestingly, while a simple refitting of a uniform value $M_p$ did not significantly improve late stage predictions, the performance fit's for 2h-yields decreased when including the partially

hydrolysed sample PH-DA-30min [Figure 3.6b]. Performing a separate fit on all substrates except the partially hydrolysed one led to a net improvement in the model's performance [Figure 3.6b]. Moreover, individual fittings assigned a lower $M_p$ value to the PH sample compared to other substrates ($M_p$=410 for PH-DA-30min vs. 560-785 for all other samples), which indicated that it might contain somewhat more recalcitrant cellulose to the given enzyme cocktail. This was further corroborated by simultaneous fitting of both $\tau$ and $M_p$, with the inclusion of the PH substrate lowered the fit's performance [Figure 3.7]. One hypothesis that could corroborate this observation is that more recalcitrant cellulose is left after a rapid hydrolysis of highly digestible cellulose within the first hours of reaction. Such change in cellulose susceptibility to be hydrolysed could notably arise from a depletion of accessible binding sites by non-productively bound enzymes and/or removal of a fraction of cellulose easier to hydrolyse, as suggested by studies on pure cellulosic substrates[147,180].

To quickly evaluate the likelihood of such a phenomenon without necessarily constructing a physically realistic simulation, we tested letting $M_p$ change with the time-evolving porosity, such that the susceptibility of cellulose to be digested by the enzyme cocktail gradually decreases with the cellulose digestion (see Appendix B.7). A simple exponential decay of $M_p$ over the course of the hydrolysis greatly improved the late stage predictions for both DA-30min and DA-60min. In the fitting range leading to reasonable predictions for both DA-30min and DA-60min substrates, the exponential decay fit was not able to properly capture the slowdown for the DA-15min. While this might be related to the higher enzyme loading used in terms of initial adsorption sites, a more realistic dependence of digestibility on the remaining fraction of cellulose should be defined to allow a detailed discussion on the implication of (i) cellulose heterogeneity, (ii) enzyme loading and (iii) their interplay on the hydrolysis rate. In any case,

this quick test demonstrates that changing the susceptibility of cellulose to be digested by enzymes over time can help, at least partially, explain the decrease in hydrolysis rate observed at later hydrolysis stages, at least partially [Figure 3.7b].

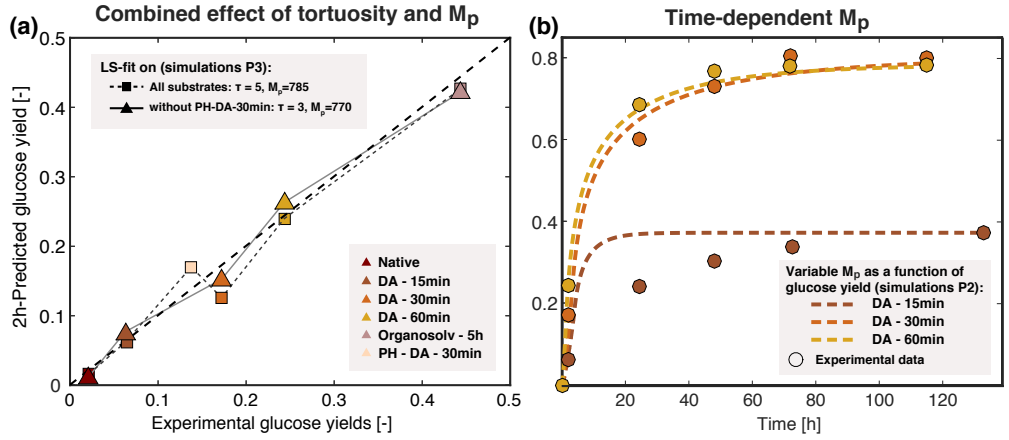



Figure 3.7 – *Model parameter fitting. (a) Combined least-square fit of the variables* $\tau$ *and* $M_p$ *based on initial glucose yield predictions. (b) Predictions based on a time-dependent* $M_p$*, represented by a simple exponential decay function dependent on porosity, same for all substrates (initial* $M_p$*=755 with decay constant of 3.5).*

## 3.4 Conclusion

Including potential structural heterogeneities in our model of lignocellulosic substrates improved predictions of glucose release upon enzyme action over time. In addition to confirming the well-known dependence of initial rate on cellulose accessibility, we show that, without impacting early rates, potential non-uniformity in cellulose accessibility within the substrate can contribute to rate slow-down observed at later stage. Interestingly, accurate yield predictions could be obtained for both organosolv and DAP pretreated substrates, which highlights the flexibility of the model to predict early glucose yields for different pretreatment methods.

Importantly, these successful predictions highlight the importance of structural features on the substrate digestibility, which occur regardless of the pretreatment method. While including heterogeneities could not entirely capture the whole extent of the slow down, first results suggests that a change in cellulose susceptibility to be hydrolysed with conversion could further capture the increased biomass recalcitrance over time, which could not be explained solely by the gradual decrease in cellulose surface as the depolymerisation proceeds.

# Chapter 4

# Investigating cellulose recalcitrance within the lignocellulosic substrate

In this chapter, we investigate the potential role played by changes in cellulose susceptibility to be digested by cellulases in the observed rate slow down over the course of the reaction. Cellulose depolymerisation has been widely modelled in literature in an attempt to explain observed hydrolysis trends, with modelling strategies presenting varying degrees of complexity in terms of enzyme-substrate interactions and cellulose morphology[155,181]. In those studies, initial heterogeneities in the cellulose ultrastructure as well the evolution of the cellulose surface properties have been proposed as key elements contributing to the rate decline over time. Recently, Ahamed et al.[147] differentiated cellulose into populations with hydrolysis rates dependent on the crystallinity and DP of the cellulose fractions to predict Avicel depolymerisation rates by Cel7A alone or complemented with Cel5A, Cel7A and/or BG. By contrast, Nill et al.[148] correlated the hydrolysis rates of various cellulosic substrates to change in productive binding capacities of TrCel7A, with two distinct populations of productive binding sites that

followed an exponential depletion over time, which explained the rate decline. In addition to dynamic changes in the cellulose surface properties, product inhibition has also been put forward as a contributing factor to the rate deactivation. However, this effect alone was not sufficient to capture the whole extent of this slowdown[182]. In this context, we use our model to explore the potential of these additional effects affect the effectiveness of cellulases to depolymerise cellulose embedded in the lignin-hemicellulose matrix and explain experimentally observed rate slowdowns. To this end, we add various new mechanisms to the model, including unproductive adsorption to the lignocellulosic substrate, coupled to decline in the efficiency of enzymatic action over the reaction extent and explore their effect on soluble sugar yields over time.

## 4.1 Modelling framework

Based on the model developed in Chapter 2, we modified the surface reaction scheme to include the possibility for enzyme to adsorb onto the cellulose surface without initiating cellulose depolymerisation [Figure 4.1]. The description of the added symbols and their value can be found in Table 4.1.

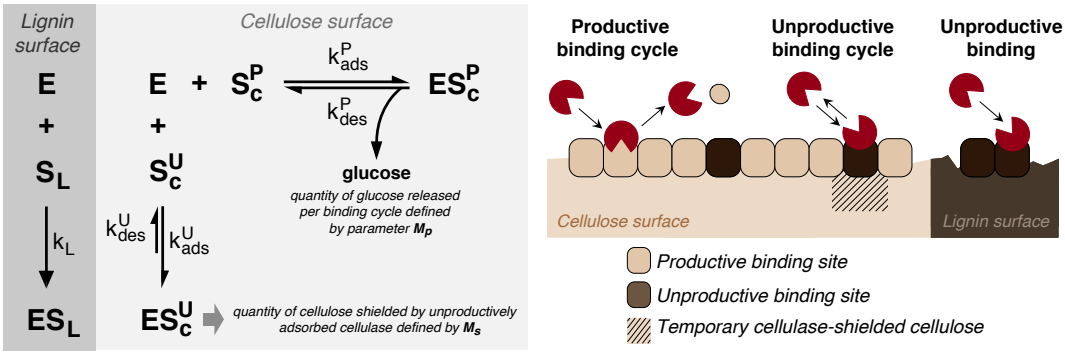



Figure 4.1 – *Overview of the enzymatic hydrolysis mechanism assumed in this chapter, accompanied by a representative scheme of the different types of binding being considered.*

Within this construction, the core hypothesis defining the enzymatic action hold, i.e. the synergistic conversion of cellulose to glucose by the enzyme cocktail is lumped into the action of a single cellulase, whose value is fitted to represent the complex enzyme system's effect on cellulose. The mechanism behind the added unproductive binding path is equivalent to the one assumed for the productive binding, with no distinction made between adsorption and complexation, nor desorption and decomplexation. While non-productive binding can be considered with both cellulase only adsorbed onto the surface or complexed with a cellulose chain[155], no distinction is made here between these two possible states, with a single unproductive adsorption rate constant $k^U_{ads}$ assumed to be equal to its productive counterpart $k^P_{ads}$:

$$k^U_{ads} = k^P_{ads}. \tag{4.1}$$

As such, in this simplified reaction scheme, we assume that interaction of cellulase with cellulose is not dependent on whether the enzyme initiate depolymerisation. In addition, enzyme stalling, i.e. productively bound enzyme ceasing its hydrolytic activity after a number of catalytic cycles, is not considered here.

We explore here the impact of cellulases present on the cellulose surface without participating to the hydrolytic effort by considering four possible parameters [Figure 4.1]: (i) the extent of time the cellulase stays on the surface before desorbing by varying the value of $k^U_{des}$, (ii) the amount of cellulose $M_s$ shielded by the unproductively bound cellulase's presence on the surface, (iii) the impact on the efficiency of productively bound cellulase by varying $M_p$ and (iv) the ratio between productive and unproductive binding sites on the cellulose surface. All these parameters are fitted to experimental data (see Appendix C.1). Conceptually, an unproductively bound cellulase in our model is assumed to reduce the quantity of accessible

cellulose surface to other enzymes due to its footprint. In addition, it is assumed to reduce the overall amount of cellulose that is susceptible to degradation by physically hindering the access of not only cellulose chains it covers but also surrounding ones to other cellulases. The latter effect is represented by the parameter $M_s$. By constituting obstacles on the surface, unproductively bound cellulases may hinder the processive nature of productively bound cellulases. While processivity is not explicitly included in our model, its effect, among others, is reflected in the lumped parameter $M_p$, which we adjusted to specifically include the effect on processivity. As the probability for productive cellulases to encounter an unproductively bound cellulases increase with their surface coverage, we now define $M_p$ so as to linearly decrease with the increase in surface concentration of unproductively bound cellulases $C^S_{E,U}(t)$ from its initial value $M_{p,0}$:

$$M_p(t) = M_{p,0}\frac{C^S_{E,U}(t)}{C^S_{E,max}(t)}. \tag{4.2}$$

In addition, to explore variations in the interactions between the substrate and cellulases, we investigated the potential effect of unspecific adsorption of cellulases onto lignin [Figure 4.1]. This effect that has been widely reported in literature as impeding the efficiency of the enzymatic hydrolysis of lignocellulosic substrates[125]. Cellulase is believed to bind to lignin through electrostatic-, H-bonds- and/or hydrophobic interactions. The reversible nature of the binding to lignin appear as being highly dependent on notably the nature of the lignin, temperature as well as the type of cellulase and the organism it was extracted from[126,127,183,184]. Here, similarly to the modeling of the adsorption on cellulose, unproductive or un-specific adsorption of cellulases on lignin is treated without considering any specific mechanism. In addition, we assume that cellulase can adsorb irreversibly on lignin. More details on the parameter optimisation can be found in Appendix C.1.

| Symbol | Description | Value | Units |
|---|---|---|---|
| $E$ | Free enzyme in solution | - | - |
| $C^S_{E,U}$ | Concentration of unproductively bound on the cellulose surface (per total cylinder volume) | Dep. var. | [mol cm$^{-3}$] |
| $C^S_{E,max}$ | Maximal enzyme concentration adsorbed at the cellulose surface (per total cylinder volume) | Dep. var. | [mol cm$^{-3}$] |
| $k^P_{ads}$ | Cellulase surface productive adsorption rate from cellulose[161] | $3 \cdot 10^{10}$ | [cm$^3$ mol$^{-1}$ min$^{-1}$] |
| $k^U_{ads}$ | Cellulase surface unproductive adsorption rate from cellulose, assumed as identical to its productive counterpart $k^P_{ads}$ | $3 \cdot 10^{10}$ | [cm$^3$ mol$^{-1}$ min$^{-1}$] |
| $k^P_{des}$ | Cellulase surface productive desorption rate[154] | 0.068 | [min$^{-1}$] |
| $k^U_{des}$ | Cellulase surface unproductive desorption rate | Fitted var. | [min$^{-1}$] |
| $k_L$ | Cellulase surface unproductive desorption rate from lignin | Fitted var. | [min$^{-1}$] |
| $S^P_c$ | Number of productive adsorption site on the cellulose surface | Fitted var. | [-] |
| $S^U_c$ | Number of non-productive adsorption site on the cellulose surface | Fitted var. | [-] |
| $S_L$ | Number of non-productive adsorption site on the lignin surface | Dep. var. | [-] |

Table 4.1 – *List of complementary symbols and their sources. Signification of symbols discussed in previous chapter can be found in Table 2.1 in Chapter 2.*

### 4.1.1 Results and Discussion

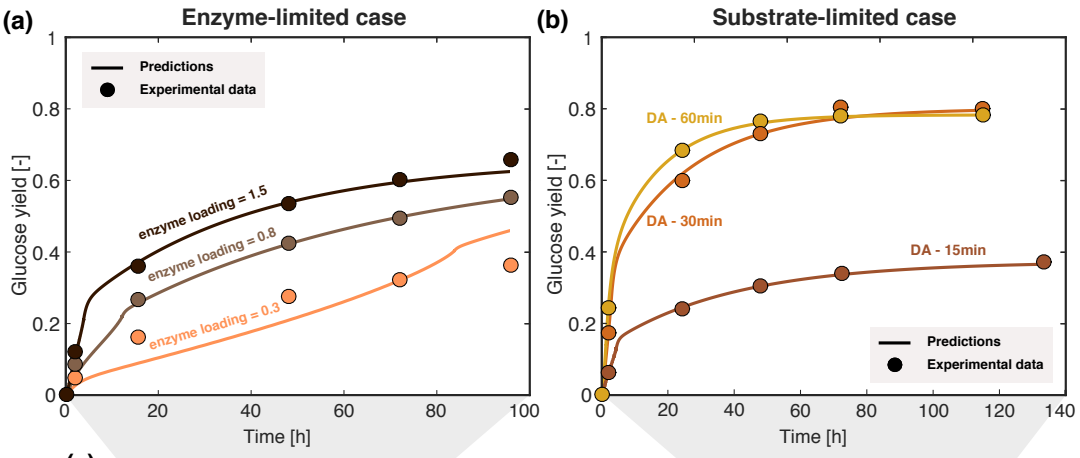
To investigate the role of cellulase-cellulose interactions and their evolution over the course of the hydrolysis, we considered datasets generated previously that we generated for both enzyme-limiting and substrate-limiting conditions. For the latter, we specifically used experiments with DAP pretreated beech wood with an increasing pretreatment severity (see Appendix B6) and hydrolysed in the presence of an excess of cellulases in terms of initial surface coverage (i.e. $e_l > 1$). For the experimental datasets covering enzyme-limiting conditions, we use results obtained from enzymatic hydrolysis carried out at three enzyme loadings on the post-pretreatment wet sieved DAP-30min beech wood substrate (see Appendix A7).

Simulations were based on measured bulk physical characteristics of the substrates with explicit inclusion of the particle size distribution (simulations type S1, see Figure 3.1).

### 4.1.2 Effects of unproductive bound cellulases on rate slowdown

Including an unproductive cellulase adsorption mechanism onto the cellulose surface helped capture the rate slowdown observed over the course of hydrolysis [Figure 4.2 a/b]. Importantly, as rather minor variations were observed between optimal fitted values obtained for each of the substrates that were considered and, to reduce parametrisation, we further assumed here that the enzyme desorption process is independent on the substrates (see Appendix C.2). As such, the unproductive desorption rate constant was set to a mean value obtained from initial individual optimisations across substrates [Figure 4.2c]. By contrast, parameters related to the cellulose characteristics – $M_p$, $M_s$ and the fraction of productive binding sites – were individually fitted. Interestingly, the best fits over the whole range of datasets were obtained when the unproductive sites represented a relatively small fraction of the available cellulose surface (~4-18%) coupled to low unproductive desorption rate (~90-95% slower than the productive desorption) [Figure 4.2c]. Due to the reversibility of the unproductive binding mechanism, the model's predictions show low sensitivity to values of $M_s$ representing the amount of cellulose shielded from hydrolysis from cellulase sitting on the surface [Figure 4.2c]. Interestingly, not including a change in hydrolysis efficiency as a function of the fraction of surface covered by unproductively bound enzyme led to similarly good predictions, but with optimal parameters for each substrates and conditions showing significant discrepancies between them and no rationale trend (see Appendix C.3).

Forcing unproductively adsorbed cellulases to desorb at similar rates to their productively bound counterparts or faster could not explain the hydrolysis trends, even when the fraction




(c)

| DA-30min, wet sieved ~ 400 um | | | SUBSTRATE | DA-15min | DA-30min | DA-60min |
|---|---|---|---|---|---|---|
| 0.3 | 0.8 | 1.5 | ENZYME LOADING | 2.5 | 1.5 | 1.4 |
| *fixed to 0.003* | | | $k^{U}_{des}$ [$s^{-1}$] | *fixed to 0.003* | | |
| 1360 I 240 | 1565 I 240 | 1590 I 310 | $M_{p,0}$ I $M_{p,\infty}$ | 1590 I 260 | 1600 I 410 | 785 I 330 |
| - | | | $M_S$ [g/mol] | - | | |
| 0.84 | 0.82 | 0.87 | Fraction of productive sites | 0.84 | 0.91 | 0.96 |

Figure 4.2 – *Predictions considering an unproductive cellulase adsorption pathway at the cellulose surface. Best individual parameter fitting for (a) a given substrates at different enzyme loadings and (b) different substrates in excess enzyme conditions. Table (c) gives an overview of the parameters for the best resulting fits. By assumption, $k^{U}_{des}$ was fixed to a predefined value and values of $M_s$ are not indicated as changes in their value shows no impact on the predicted yields.*

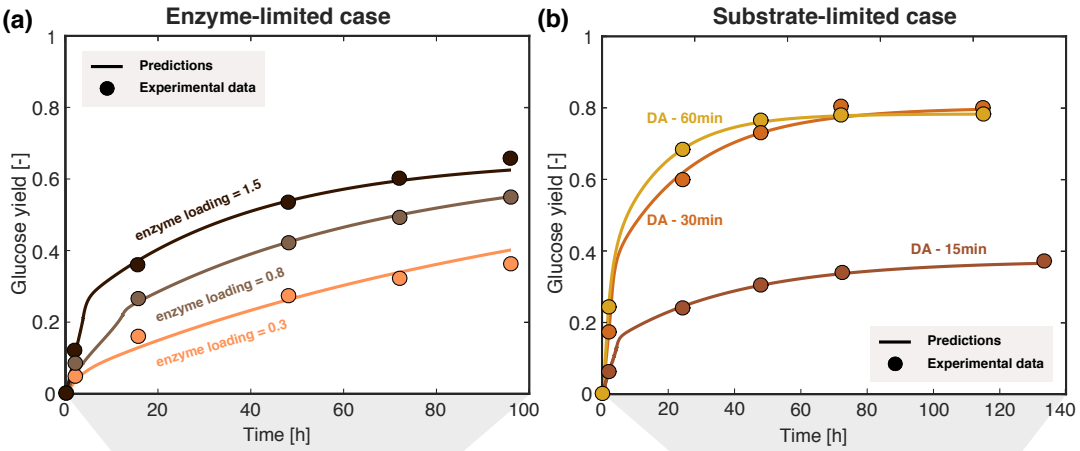
of unproductive adsorption sites at the cellulose surface was increased (see Appendix C.4). These results suggest that, even though the existence of a "fast" unproductive binding cycle of cellulase on the cellulose cannot be excluded, a prolonged stay of cellulases on the cellulose surface would better explain the rate observed, by decreasing both available surface for adsorption and the effectiveness of productively bound enzymes. While the latter is mathematically related to the presence of unproductively bound cellulase at the cellulose surface, the decline in cellulose hydrolysis effectiveness as represented by an evolving value $M_p$ over

the course of the reaction may contain other deactivation effects due to the lumped nature of this parameter. In particular, as the reaction proceeds, an evolving value of $M_p$ is expected to capture changes in synergism or observed rate related to change in the cellulose physical characteristics, defined notably by the DP and CrI. Another observed effect decreasing the observed hydrolysis rate, that is expected to be captured by our decreasing value of $M_p$ is the enzyme surface jamming, as cellulose coverage increases and with it the probability of the action of a productively bound cellulase to be disturbed by the presence of others[130]. Within our current model formulation, it is not possible to discriminate between such mechanisms. Detailed description of the cellulose physical characteristics in terms of DP and CrI notably would be necessary to explore this, but are both challenging to obtain experimentally in an accurate way. As such, we show here that heterogeneities on the cellulose surface that evolve with the extent of reaction can at least capture the rate slowdown, but the detailed physical origin of such changes is yet to be determined. Nonetheless, such methodology allows to pointing out to specific set of parameters representing an interest to be further investigated.

Compared to the previously assumed mechanism (see Chapter 3), fitted initial values of $M_p$ increased about two-fold for all substrates, except for the more severely pretreated substrate DA-60min [Figure 4.2]. In this case, a lower proportion of unproductive adsorption sites coupled to somewhat more recalcitrant cellulose compared to the other substrates appear as the best fit to explain the observed rate slowdown. This observation suggests that higher degree of severity, while it improves mass transfer by enhancing particle breakdown without showing significant change in porosity [Figure 3.2 and 3.3], negatively alters the cellulose susceptibility to be digested. As such, improvements made in surface accessibility might, at a certain point, be counterbalanced by undesired alteration of the cellulosic fraction.

### 4.1.3 The role of cellulase unproductive binding to lignin

Interestingly, while individual fits lead to rather good agreement with experimental data for all the substrates in terms of least squares, predicted hydrolysis trend for the more strongly enzyme-limited conditions had trouble capturing the experimental trend [Figure 4.2a]. In particular, final glucose yields obtained experimentally showed strong correlation with the enzyme loading, suggesting a deactivation mechanism not explainable by any reversible mechanism on the timescale considered here. As such, we explore the potential impact of a well-known effect that occurs when the enzymatic hydrolysis is carried out on the whole lignocellulosic structure, the unproductive/non-specific adsorption of cellulase on the lignin polymer. Interestingly, optimal values for $k_L$, required to adequately capture the rate slow-down in the case of the two enzyme-deficient conditions with enzyme loadings < 1, were $\sim 10^{-4} - 10^{-6}$ times the productive adsorption rate constant on cellulose. When added in the substrate-limited cases, this mechanism had virtually no impact on the predicted hydrolysis rates. As such, this unproductive mechanism becomes minor as the enzyme loading is increased above initial full coverage (i.e. $e_l > 1$). This effect decreases also as the relative fraction of lignin surface decreases compared to cellulose, which typically occurs as the pretreatment severity increases [Figure 4.3]. The quite significant discrepancy ($10^2$) between the two fitted values of $k_L$ can be symptomatic of other deactivation mechanisms being important for the substrate hydrolysed with lower amount of cellulase ($e_l = 0.3$) as, in such enzyme-limited case, loss of active enzymes will strongly impact observed yields.




(c)

| DA-30min, wet sieved ~ 400 um | | | SUBSTRATE | DA-15min | DA-30min | DA-60min |
|---|---|---|---|---|---|---|
| 0.3 | 0.8 | 1.5 | ENZYME LOADING | 2.5 | 1.5 | 1.4 |
| *fixed to 0.003* | | | $k^U_{des}$ [$s^{-1}$] | *fixed to 0.003* | | |
| 1290 I 540 | 1500 I 240 | 1590 I 310 | $M_{p,0}$ I $M_{p,\infty}$ | 1590 I 260 | 1600 I 410 | 785 I 330 |
| - | | | $M_S$ [g/mol] | - | | |
| 0.91 | 0.82 | 0.87 | Fraction of productive sites | 0.84 | 0.91 | 0.96 |
| 2e6 | 1e4 | *set to 2e6* | $k_L$ [cm / (mol s )] | *set to 2e6* | | |

Figure 4.3 – *Predictions considering an unproductive cellulase adsorption pathway at the cellulose surface coupled to irreversible adsorption on lignin. (a) Predictions for a given substrates at different enzyme loadings. Best individual fitting of $k_L$ for cases where $e_l < 1$. For $e_l > 1$, $k_L$ was set to one fitted in the case where $e_l$=0.3, where the binding to lignin is the largest (b) Predictions at different substrates in excess enzyme conditions. Again, $k_L$ is set to its largest fitted value. Table (c) an overview of the parameters used in the simulations. By assumption, $k^U_{des}$ was fixed to a predefined value and values of $M_s$ are not indicated as change in their value shows no impact on the predicted yields.*

### 4.1.4 Conclusion

While potential structural heterogeneities could only partially rationalise the rate slowdown, modeling the evolution of the interplay between cellulases and cellulose confirms that this could at least be a principle factor to explain transient hydrolysis trends. While the simplified

mechanistic framework in its current form does not provide careful details of the cellulose physical characteristics, it allows us to determine the potential extent to which a given factor/mechanism can slow down the rate. In particular, we showed that unproductive enzymes that undergo a prolonged stay on the cellulose surface seem more likely to impede the hydrolytic rates, while "fast"-desorbing unproductive enzymes couldn't explain hydrolysis trend. In addition, modeling shows that unproductive adsorption of cellulase on lignin should at least be considered a candidate for explaining the observed rate and yield decline when working in enzyme-limiting conditions, highlighting the importance of developing strategies reducing the negative impact of lignin present in the system.

# Chapter 5

# Conclusion

In summary, we demonstrate the use of a theory-based framework to decouple and evaluate the impact of key factors in the enzymatic hydrolysis of lignocellulosic biomass. This approach allowed us to explore how the presence of the lignin-hemicellulose scaffold affects the hydrolysis of lignocellulose, without introducing an unreasonable number of model and fitted parameters. By correlating the physical characteristics of the substrates to observed yields, this model can be used as a diagnostic tool to assess the plausibility that various mechanisms truly can limit the reaction efficiency in specific experiments. As such, our model represent a useful tool not only to investigate the role played by specific factors in the hydrolysis efficiency, but also to test hypotheses related to the substrate's physical characteristics that are hard to determine or study individually in an experimental setting.

## 5.1 Key takeaways

By notably comparing our predictions to a series of experiments conducted on pretreated beech wood substrates and covering both enzyme-limiting and substrate-limiting conditions,

we were able to highlight to the following key findings:

- From in silico investigation on the importance of internal mass transfer in Chapter 2, we highlighted the importance of tailoring experimental conditions to substrate's characteristics to efficiently hydrolyse the lignocellulose. High enzyme- and biomass-loadings are both important driving forces to enhance enzyme penetration and increase hydrolysis rate, but their effects greatly depends on the particle size. High biomass loadings should be favoured for both small- (R<300um) and middle-ranged (300<R<1000um) substrates to enhance enzyme diffusion while minimising the use of enzymes. In such cases, working at enzyme loadings exceeding the full coverage of the cellulose surface (i.e. $e_l$>1) does not bring a significant benefit. For larger particles (R>1000um), important increase in biomass loading is unable to offset the significant internal mass transfer, but high enzyme loadings improve enzyme penetration by maintaining a high concentration gradient within the particle.

- From comparing the ability of DAP to promote the enzymatic hydrolysis in Chapter 3, we put into perspective the key role of particle fragmentation with porosity. Particle breakdown upon pretreatment improves hydrolysis rates, but its effect can be strongly limited by a low porosity: decreasing particle size to enhance hydrolysis in low porosity materials only marginally improves glucose yields. As such, efficient pretreatment should concurrently decrease particle size and increase porosity.

- By exploring further the substrate characteristics in Chapter 3, we highlighted the potential contribution of structural heterogeneities within the biomass sample, in terms of porosity and fraction of recalcitrant cellulose, to the rate slowdown observed at extended hydrolysis stage. While such non-uniformities are challenging to assay experimentally

as separation methods affect the particle characteristics, their existence cannot be excluded. Lignocellulosic biomass is inherently inhomogeneous and pretreatments can potentially enhance this non-uniformity by fragmenting and more severely affecting smaller fragments.

- Finally, we explored in Chapter 4 cellulose-related recalcitrance to rationalise experimental observations from enzyme- and substrate-limiting conditions. Unproductively adsorbed cellulases on the cellulose, coupled to irreversible adsorption on lignin, appeared to be a plausible cause for the rate slowdown observed at late stage hydrolysis. From the obtained fittings, it appears that working at enzyme loadings around $e_l \sim 1.4$ is already enough to offset this undesirable effect.

## 5.2 Outlook

To improve process engineering of enzymatic hydrolysis of lignocellulosic biomass, a comprehensive understanding of the reaction would be required. For such intricate heterogeneous reaction, modelling appears as a key tool to explore the multiplicity of parameters impacting the reaction efficiency. In its current formulation, our model could be useful to notably help design efficient pretreatment methods, by emphasizing which key parameters or combined set of parameters critically limit hydrolysis rate. In addition, this modeling approach forms a basic framework that can be further complexified to evaluate how the detailed interactions between cellulose-cellulase, which dictate hydrolysis rates for pure cellulosic substrates, play a role in the more complex lignocellulosic substrate, or explore external mass transfer in industrially relevant cases of very high solid loadings where viscosity will begin to play an important role of biomass loadings.

# Appendix A

# Appendix for Chapter 2

## A.1 Effect of external mass transfer

The potential impact of external mass transfer on the overall reaction rate of cellulose depolymerisation was evaluated using the modified Weisz-Prater criterion for external mass transfer[179]:

$$\frac{r_{obs} R_p}{\overline{k}_c C_{E,bulk}} < \frac{0.15}{n}, \tag{A.1}$$

where $r_{obs}$ is the observed reaction rate, $R_p$ the particle radius, $C_{E,bulk}$ the enzyme bulk concentration, $n$ is the reaction order and $\overline{k}_c$ the mass transfer coefficient. Satisfying this relationship indicates that internal mass transfer or reaction controls the reaction and that external mass transfer can be neglected.

The mass transfer coefficient $\overline{k}_c$ can be estimated from a correlation dependent on both the Schmidt $Sc$ and Reynolds $Re$ numbers:

$$Sc = \frac{\overline{\mu}}{\rho D_E} \quad \text{and} \quad Re = \frac{u \rho L}{\overline{\mu}}, \tag{A.2}$$

through the Sherwood number $Sh$:

$$Sh = \frac{\overline{k}_c L}{D_E} = 2 + 0.6 Re^{1/2} Sc^{1/3}, \tag{A.3}$$

leading to:

$$\overline{k}_c = \frac{2D_E}{L} + \frac{0.6 D_E^{2/3} u^{1/2} \rho^{1/6}}{L^{1/2} \overline{\mu}^{1/6}}. \tag{A.4}$$

While the viscosity $\overline{\mu}$ and density $\rho$ of the sodium citrate buffer containing the enzymes is assumed to be same as those for water, the fluid velocity is calculated based on the flask internal radius ($R_{flask} = 1.25 \cdot 10^{-2}$ m) and incubator rotation speed ($\omega_{incubator}$=120 rpm):

$$u = \frac{2R_{flask}\omega_{incubator}}{60}, \tag{A.5}$$

and the particle characteristic length L is assumed to be $2R_p$. As a first approximation, we assumed that the observed rate corresponded was linked to the average internal rate of reaction, i.e. the rate of cellulose depolymerisation $r_{dep,cellulose}$, which is ultimately what the model predicts:

$$r_{obs} \approx -\frac{1}{M_p} r_{dep,cellulose} = \frac{1}{M_p} \frac{dC_{glucose,released}}{dt}. \tag{A.6}$$

Since we are comparing enzyme mass transfer to its participation in the surface reaction, the quantity of glucose generated per amount of time is divided by the factor $M_p$ to calculate the quantity of enzyme participating in the reaction per time. This factor accounts for the fact that several glucose molecules are generated for one enzyme going through a binding cycle.

The Weisz-Prater criterion was evaluated for the initial predicted reaction rate (the fastest rate,

and thus the most likely to be controlled by external transfer) for two extreme particle sizes:

$$\begin{cases} R_p = 10\mu m: & \dfrac{r_{obs}R_p}{\overline{k}_c C_{E,bulk}} = 9.4\cdot 10^{-7} < 0.15, \\ R_p = 1cm: & \dfrac{r_{obs}R_p}{\overline{k}_c C_{E,bulk}} = 2.4\cdot 10^{-4} < 0.15. \end{cases} \tag{A.7}$$

In both cases, predicted initial rates with respect to enzyme concentration, closely followed a first order reaction rate, thus n=1 was used as a first approximation. Altogether, the values were far below the criteria requirement. Therefore, even though the estimation of $r_{obs}$ and $\overline{k}_c$ were based on several approximations, we can safely assume that external mass transfer plays a negligible role on enzymatic hydrolysis and can safely be ignored.

## A.2 Comparison of parameter $M_p$ and processivity

In our model formulation, the release of glucose in solution resulting from the hydrolytic action of cellulases is expressed as a change in particle porosity $\varepsilon(r,t)$. Within the model framework, such glucose release is accounted for as the enzyme desorbs from the surface. Mathematically, change in the particle porosity can thus be expressed as:

$$\frac{\partial \varepsilon(r,t)}{\partial t} = \frac{k_{des}C_E^S(r,t)M_p MM_{glu}H_{glu}}{\rho_C^{IV}}, \tag{A.8}$$

where $k_{des}C_E^S(r,t)$ is the enzyme desorption rate from the cellulose surface, $M_p$ is an adimensional parameter representing the number of glucan unit released per enzyme desorbing from the surface (i.e. mole of glucose released per mole of enzymes desorbing), $MM_{glu}$ and $H_{glu}$ are the molar mass and hydrolysis factor of glucose, respectively, and $\rho_C^{IV}$ is the density of cellulose including the pore volume, which varies with porosity. These two constants and

variable are essentially responsible for translating the mass of glucose released into a pore volume that is created as a result of cellulose hydrolysis and this resulting glucose release. By dividing both side of Eq. (A.8) by these three terms, we can equivalently reformulate this equation describing a change in porosity into an equation describing a change of glucose concentration:

$$\frac{\rho_C^{IV}}{MM_{glu}H_{glu}}\frac{\partial\varepsilon(r,t)}{\partial t} = \frac{\partial}{\partial t}\left[\frac{\varepsilon(r,t)\rho_C^{IV}}{MM_{glu}H_{glu}}\right] = \frac{\partial C_{glucose}(r,t)}{\partial t} = k_{des}C_E^S(r,t)M_p. \tag{A.9}$$

This translation allows us to compare our model more directly to a general and common mechanistic model of cellulose hydrolysis[155], where the glucose release is dictated by the processive action of productively bound cellulases hydrolysing cellulose into glucose at a rate given by the catalytic rate constant $k_{cat}$ [Figure A.1]:

$$\frac{\partial C_{glucose}(r,t)}{\partial t} = k_{cat}C_E^S(r,t). \tag{A.10}$$

Comparing the common mechanistic model (Eq. (A.10)) and our transformed equation model (Eq. (A.9)), we obtain the following equivalence:

$$k_{cat} = k_{des}M_p. \tag{A.11}$$

On an ideal cellulose polymer (i.e. where the catalytic action of an adsorbed cellulase is independent of its location on the cellulose surface), the number of catalytic events $n_{int}$ that a cellulase can perform before desorbing from the surface (i.e. the so-called intrinsic

processivity), is related to both catalytic and desorption constant[117]:

$$n_{int} = \frac{k_{cat} + k_{des}}{k_{des}}. \tag{A.12}$$

Combining Eqs. (A.11) and (A.12), we obtain:

$$k_{cat} = k_{des}(n_{int} - 1) = k_{des} M_p, \tag{A.13}$$

leading to the following equivalence:

$$M_p = n_{int} - 1. \tag{A.14}$$

Considering that values of intrinsic processivity for crystalline cellulose fall in the range of ~4000, $M_p$ and $n_{int}$ are essentially equal[117]. Therefore, $M_p$ should be understood as an estimate of the average intrinsic processivity per enzyme for a mixed enzyme cocktail. Because the definitions differ slightly and because intrinsic processivity is defined in the literature for a given enzyme as opposed to a mixture, we used a different designation in this work. Nevertheless, mathematically, $n_{int}$ and $M_p$ are essentially equivalent in both model formulations.

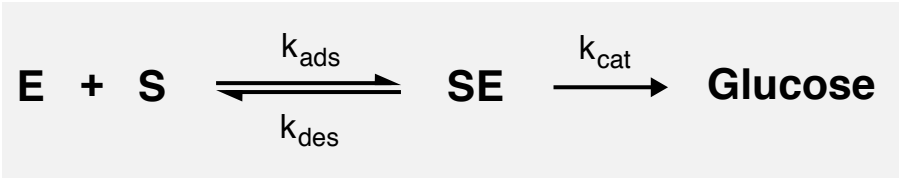



Figure A.1 – *General mechanistic model for cellulose hydrolysis*[155].

## A.3 Model: previous model formulation

The new formulation of the model described in this thesis is based on the same core hypothesis than ones assumed in the previous work of Luterbacher et al.[153,154], with three notable exceptions; first, in the prior work, enzyme loadings were defined through their initial bulk enzyme concentration, which was assumed to stay constant as the reaction proceeds (i.e. "infinite" enzyme model), leading to potential overestimations of glucose yields in "enzyme-limited" reaction conditions and/or as the diffusional path within the substrates increases. Second, the effect of pore widening on the cellulase concentration within the substrate was disregarded in the prior model. While this assumption was reasonable for the "infinite" enzyme model, it may lead to overestimation of rates when reaction rates are strongly limited by the low number of enzymes in solution. In such cases, decreasing the concentration gradient throughout the particle may impact the efficiency of mass transfer. Finally, this current model accounts for the possibility of two accessible pore surfaces within the infinite-slit pore model for pores presenting a diameter larger than twice the size of a cellulase. For the sake of clarity, the evolution of surface coverage $\theta(r,t)$ and particle density $\rho(r,t)$ described by Luterbacher et al.[153,154] are expressed in terms of surface concentration $C_E^S(r,t)$ and porosity $\varepsilon(r,t)$ in this thesis. This terminology change does not affect any underlying assumption or model results.

## A.4 Estimation of the number of accessible binding sites from porosity

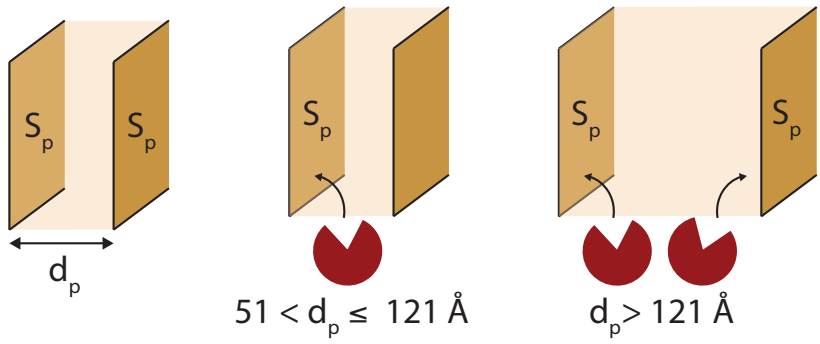



Figure A.2 – *Schematic representation of the double-slit pore geometry, and how the pore's surfaces $S_p$ are accounted for depending on the pore diameter $d_p$ and the size of a cellulase*

## A.5 External vs. internal accessible cellulose surface

Assuming a cylindrical geometry for the biomass, the relationship between internal and external surface can be derived as follows. The external surface of a cylindrical particle $S_{cyl,ext}$ of radius $R_{cyl}$, not considering the edge surface, can be related to its volume $V_{cyl}$, and consequently its mass $M_{cyl}$,

$$S_{cyl,ext} = \frac{2V_{cyl}}{R_{cyl}} = \frac{2M_{cyl}}{\rho_{cyl}R_{cyl}}, \tag{A.15}$$

where $\rho_{cyl}$ is the particle density including the void created by the pores. This lead to the ratio between external and internal surface being:

$$\frac{S_{cyl,int}}{S_{cyl,ext}} = \frac{S_{cyl,int}\rho_{cyl}R_{cyl}}{2M_{cyl}}. \tag{A.16}$$

As external surface is likely to be more important for particles exhibiting a low porosity and a small diameter, we consider here the extreme cases of a 10 μm particle with a porosity of 6 $m^2\,g^{-1}$ of biomass (i.e. the typical accessibility of native mixed hardwood), leading to a ratio of 29.5. In this extreme case, in which cellulose was also assumed to be evenly distributed in the particle, the error on the calculated accessible surface would be around 2%.

## A.6 Model implementation – verification and convergence

While increasing the discretisation number improves accuracy, a trade-off value ensuring less than 2% deviation in the enzyme mole balance was chosen in order to alleviate computational cost and keep simulation time reasonable. For most systems considered here, this condition was already fulfilled with n=50. Notable exceptions occurred for simulations that used low enzyme loadings ($e_l < 1$). In these cases, because the pore enzyme concentration tended to be particularly low due to significant surface adsorption, a finer grid was required to maintain numerical stability. Figure A.3 shows convergence of the system as a function of the model parameters for a small particle, with different enzyme loadings. For lower enzyme loadings and/or higher biomass loadings (not pictured) the criterion was met only for a discretisation number n above 2000. However, n=50 was actually sufficient for most conditions. Notably, as a consequence of the finite time step and the fact that results are essentially exact up to machine precision, $\eta$ does not tend to zero, but to a finite value.

## A.7 Model – experimental inputs

Table A.1 shows an overview of the available data found in literature for the various substrates used as inputs for the model in this work[107]. The overall digestibility, defining the accessible cellulose fraction, is taken as the reported 24h-glucose yields. Particle radii are assumed to

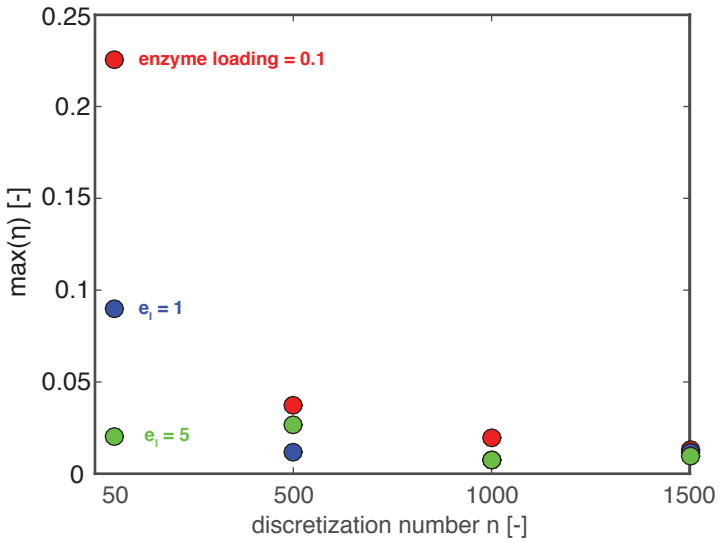



Figure A.3 – *Convergence as a function of discretisation number for particle exhibiting a radius of 10* µm *and biomass loading of 1%.*

correspond to the average size between sieves used to screen the native biomass particles (i.e. R=25 µm). Enzymatic hydrolysis was carried out at a low solid loadings of 2% in citrate buffer pH=4.8 at 50°C for 24h, using cellulase powder from the *Trichoderma Reesei* Rutger's C strain (92.5mg/100ml) complemented by a source of $\beta$-glucosidase (0.1ml).

| **Substrate** | **2h-glucose yield [-]** | **24h-glucose yield [-]** | **Cellulose fraction [-]** | **Accessible cellulose surface** $[m^2\,g^{-1}]$ |
|---|---|---|---|---|
| Native (90% Birch/10% Maple) | 0.0410 | 0.1530 | 0.42 | 6.3 |
| DA-1%sulfuric acid/100°C/5hr | 0.1470 | 0.2180 | 0.55 | 21.6 |
| DA-1% sulfuric acid/180°C/7.8s | 0.2430 | 0.4260 | 0.60 | 23 |
| DA-1% sulfuric acid/200°C/7.8s | 0.4180 | 0.8540 | 0.67 | 59 |
| DA-1% sulfuric acid/220°C/7.8s | 0.64 | 0.87 | 0.62 | 80.9 |

Table A.1 – *Characteristics of the various substrates used in this study*[107,185]

In addition to these data, a new data set was generated from beech wood, the characteristics of which are summarized in Table A.2. The final glucose yield was assumed to be that reached after 120h when the enzymatic hydrolysis was performed at high enzyme loadings to avoid incomplete cellulose degradation stemming from enzyme-related limitations. In any case, sensitivity analysis performed to evaluate this influence of change in the final yields on the predictions of the 2h-glucose yields showed only a marginal effect [Figure A.4]. The recorded concentration of cellobiose throughout hydrolysis was marginal in comparison to the glucose concentrations, as illustrated in Figure A.5.

| **Substrate diameter [*lower limit,upper limit*] [μm]** | **Final glucose yield - 120h [-]** | **Cellulose fraction [-]** | **Accessible surface [$m^2\,g^{-1}$]** |
|---|---|---|---|
| [50,150] | 0.8380 | 0.61 | 35.8 ± 5.1 |
| [300,500] | 0.6330 | 0.62 | 24.7 ± 3.6 |
| [1000,3150] | 0.6773 | 0.61 | 33.9 ± 3.1 |

Table A.2 – *Characteristics of the pretreated beech wood substrates generated in this study*

Experimental enzyme loadings were expressed as normalized to the number of initially accessible binding sites on the cellulose surface assuming an average molecular mass of 51.8 kDa for the cellulases making up the enzyme cocktail used experimentally[186].

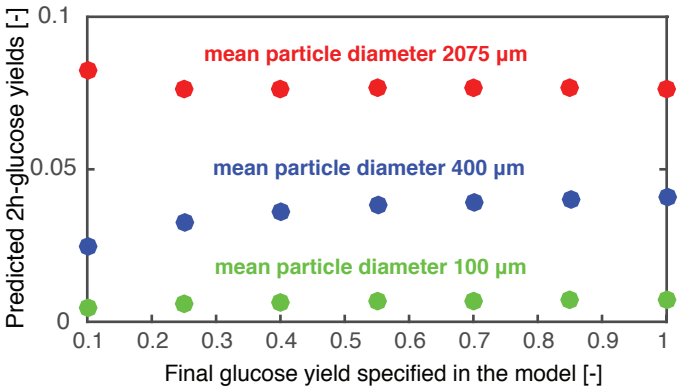



Figure A.4 – *Predicted 2h-glucose yields as a function of the specified final glucose yield in the model for pretreated beech wood (1%SA / 160°C / 30min) with different particle sizes. Simulations were performed for a low enzyme loading of 0.3.*

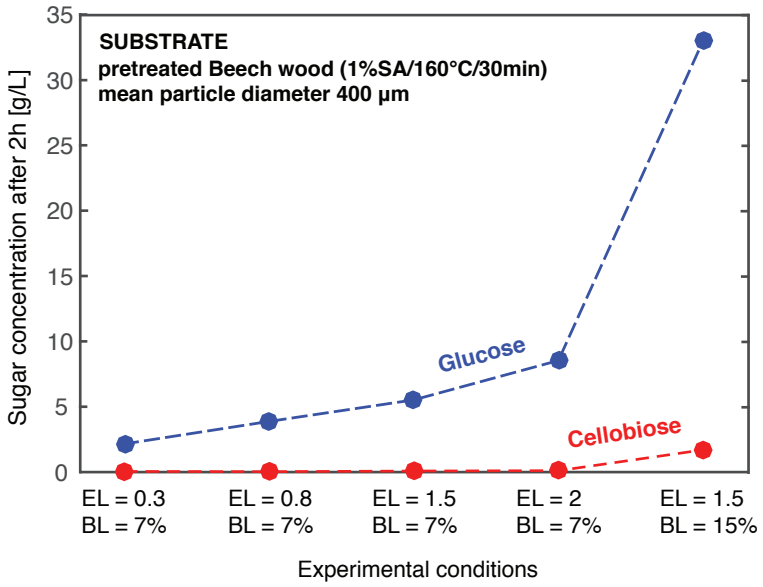



Figure A.5 – *Experimental glucose and cellobiose concentrations measured after 2h of hydrolysis for pretreated beech wood (1%SA / 160°C / 30min) for several enzyme and biomass-loadings.*

## A.8 Individual and combined optimization of $M_p$ and $\tau$

Tortuosity values were rationally bound between 1 and 7, which represent situations from the case where all pores are aligned with the diffusive flux to extreme cases of tortuosity[179]. When fitting $M_p$ over the whole set of data, only marginal differences in fit were obtained for values of tortuosity up to 4-5 [Figure A.6], with slight decreases in goodness-of-fit when increasing the diffusion resistance. The choice of the optimal parameter to use for simulations was then purely based on the slightly improved $R^2$ value (leading to the optimal value of $M_p$ = 755 and $\tau$ = 2). When instead performing individual fits for each substrate [Figure A.7], we did not observe any clear trend based on substrate characteristics (e.g. pretreatment severity), indicating again that the dependence on accessibility dominated the kinetics and that small changes in $M_p$ or $\tau$ are likely due to uncertainty in parameter estimation rather than the result of physical phenomena.

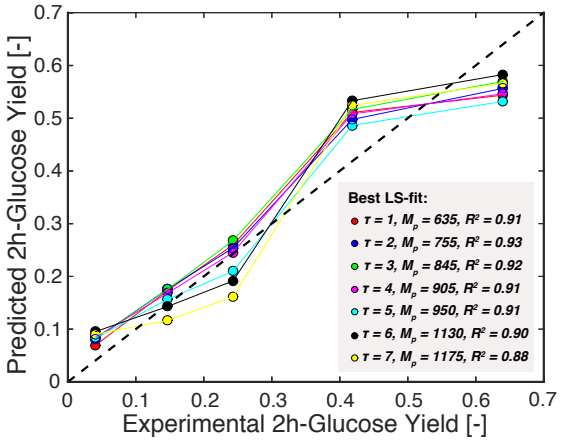



Figure A.6 – *Optimal value of $M_p$ in terms of least-square fitting for each $\tau$ considered to predict early glucose yields, with indication of the goodness-of-fit parameter $R^2$. Fits were performed on the whole data set. Predictions were performed for the mixed hardwood substrates considered in the study*[107].

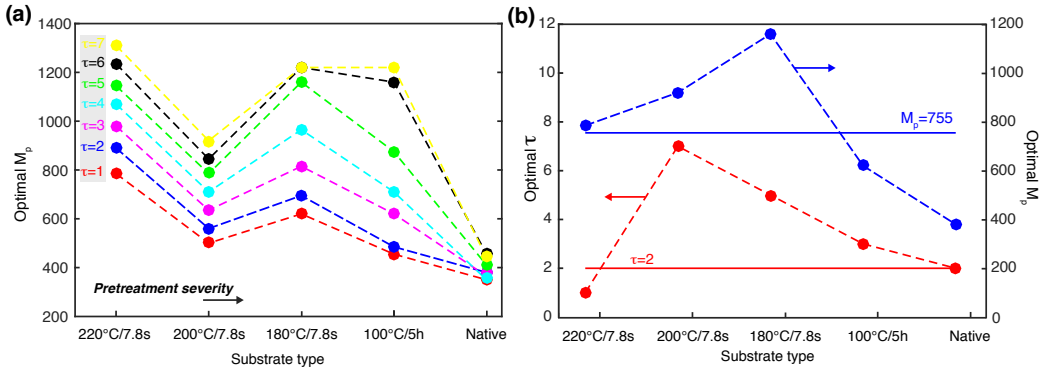



Figure A.7 – *(a) Optimal values of $M_p$ for a range of $\tau$ after least-square fitting to predict early glucose yields for different pretreatment conditions. (b) Combined optimal values of $M_p$ and $\tau$ after least-square fitting to predict early glucose yields for different pretreatment conditions. For comparison, the values of these parameters ($M_p$ = 755 and $\tau$ = 2) on which calculations are based in this study are shown with the horizontal lines. In both cases, fits were performed on each substrate individually. Predictions were performed for mixed hardwood substrates considered in the study*[107]

## A.9 Pretreated beech wood

Air-dried beech wood (Fagus sylvatica) chips collected from Zollikofen (Switzerland) were first milled to pass through a 2-mm screen. These particles were further sieved and those between 250 and 450 µm in diameter were retained as the so-called native substrate. This substrate was further processed using dilute-acid (DAP) –1wt% sulfuric acid (SA/Merck, 100732) at 160°C– in 60ml glass reactors at a loading of 2g of dry substrate per 20 ml acid solution for 30 minutes, followed by Büchner filtration and extensive washing with purified water (Milli-Q grade). To allow fiber swelling, wood particles were pre-soaked overnight at 4°C in the pretreatment solution. Wet pretreated wood sample was wet sieved (300-500 µm diameter) under purified

water (Milli-Q grade) and then kept for a maximum of two weeks in sealed plastic bags at 4°C prior to further utilization, to avoid drying and degradation. Composition analysis of the pretreated wood sample was performed according to the LAP procedure published by the National Renewable Energy Laboratory[187] (see section A.10)

## A.10 Compositional Analysis

Compositional analysis followed the LAP procedure published by NREL[187]. Briefly, dried substrates were subjected to a two-step acid hydrolysis, starting with the incubation of samples ( 0.3 g) in 7.5ml of acid solution (72% SA) for 2h at 30°C and 120 rpm, followed by dilution down to 3% sulfuric acid concentration before proceeding with the reaction at 121°C for 1h in an autoclave. Lignin was then separated from the soluble sugars by filtration and dried overnight at 105°C for quantification, while cellulose and hemicellulose contents were estimated through determination of sugar concentration in the filtrate by HPLC analysis (Bio-Rad Aminex HPX-87H column, 5mM $H_2SO_4$ mobile phase). The moisture content of all samples was initially determined by drying the wood particles overnight at 105°C and weighing the dried mass at room temperature after the sample was left for 2h in a desiccator.

## A.11 Enzymatic hydrolysis

Enzymatic hydrolysis was carried out in a citrate buffer (pH=4.8/Sigma, C1909, 71402) as described previously[188] using a commercial enzyme blend (150 FPU/g, Cellic CTec2, Novozyme, Denmark/Sigma, SAE0020) at various enzyme loadings. In addition, tetracycline (Sigma, 87128) and cycloheximide (Sigma, C7698) were added to the reaction medium to avoid undesired bacterial and fungal growth, respectively. Protein content was assayed to 55.1mg protein/ml according to the Bradford method[189] using the commercially available Pierce

Coomassie protein assay kit (ThermoFisher, 23200).

## A.12 Pore size distribution

To avoid any change in porosity that might occur due to hornification during drying, pore size distribution was determined in wet conditions using a modified batch solute exclusion technique[163,190]. A series of suitably sized PEGs as well as glucose (Sigma, G8270) were used as molecular probes [Table A.3]. Wet wood samples ($m_{wood,wet}$=0.4 g) were incubated with the probe solution in ultrapure water ($V_{probe}$=0.35 ml, $C_{probe,init}$=50 g/L) for 3 hours with occasional mixing, followed by which, the supernatant was recovered by centrifugation (2500 rpm, 15min) through 3 µm centrifugal filters. The resulting solution was further diluted with Milli-Q water and the final probe concentration $C_{probe,final}$ was measured using a refractive index detector (Viscotek VE 3580) connected to a syringe pump – with an injection volume of 3ml at 0.5ml/min at a detection temperature of 35°C. The remaining solids were then washed and dried to record the dry weight and moisture content $x_{wat}$. For each set of recorded data, the refractive index of a blank solution obtained by incubation of the wood samples with Milli-Q water was used to correct the signal for any soluble material that could interfere with quantification before computing pore volumes. Prior to any measurements on the native substrate, the particles were soaked for at least 48h to allow fiber swelling, with daily water changes to avoid bacterial/fungal contamination.

From the difference of concentration between the initial stock solution $C_{probe,init}$ and the solution left in contact with the wood sample $C_{probe,final}$, inaccessible pore volume for a given probe $i$ was given by:

$$V_{inacc,i} = V_{probe} - V_{wood,wet} x_{wat} - \frac{C_{probe,init} V_{probe}}{C_{probe,final}} \tag{A.17}$$

Measurement of the fiber saturation point (FSP) was performed using a probe that was assumed to be unable to penetrate into the pores. In this case, we used a probe with a size of 560Å, and used the following formula to calculate the pore volume accessible to probe $i$:

$$V_{acc,i} = V_{inacc,560\text{Å}} - V_{inacc,i} \tag{A.18}$$

| **Molecular probe** | **Molecular weight**[a] [$\mathrm{g\,mol^{-1}}$] | **Diameter**[b] [Å] |
|---|---|---|
| Glucose | 180 | 8 |
| P6000 | 6000 | 51 |
| P35000 | 35000 | 121 |
| P600000 | 600000 | 560 |

Table A.3 – *Characteristics of the relevant probes employed in this study for the pore size distribution determination.* [a] *From manufacturer (Merck).* [b] *Extrapolated from Neuman et al.*[191]*, assuming an ellipsoidal shape for the polymer in solution*

# Appendix B

# Appendix for Chapter 3

## B.1 Model – inclusion of pore volume heterogeneities

For each substrate, $j = 1, 2, \cdots, M$ accessible pore volume distributions $\{V_{Aj,R1}, V_{Aj,R2}, \cdots V_{Aj,RN}\}$ were sampled from the solution space defined by the following system of linear equations:

$$\begin{pmatrix} w_{R1} \\ \vdots \\ w_M \end{pmatrix} \cdot \left(V_{Aj,R1} \cdots V_{Aj,RN}\right) = \begin{pmatrix} V_{A,experimental} \\ \vdots \\ V_{A,experimental} \end{pmatrix} \tag{B.1}$$

and associated set of constraints $\{V_{Aj,R1} \geq V_{Aj,R2} \geq \cdots \geq V_{Aj,RN}\}$ for $\{R_1 < R_2 < \cdots < R_N\}$. Here, $V_{Aj,Ri}$ is the accessible volume to a cellulase in a particle of radius $R_i$ and contributing to the overall volume of the sample with a fraction $w_{Ri}$ [Figure B.1].

On average, we impose that, over the whole sample, the accessible volume must match the one measured experimentally ($V_{A,experimental}$). Also, the minimal porosity generated is constraint to be larger or equal to one measured on the native substrate. Solutions with $V_{Aj,Ri}$<0 were then removed from this initial sampling. The large number of distributions remaining were

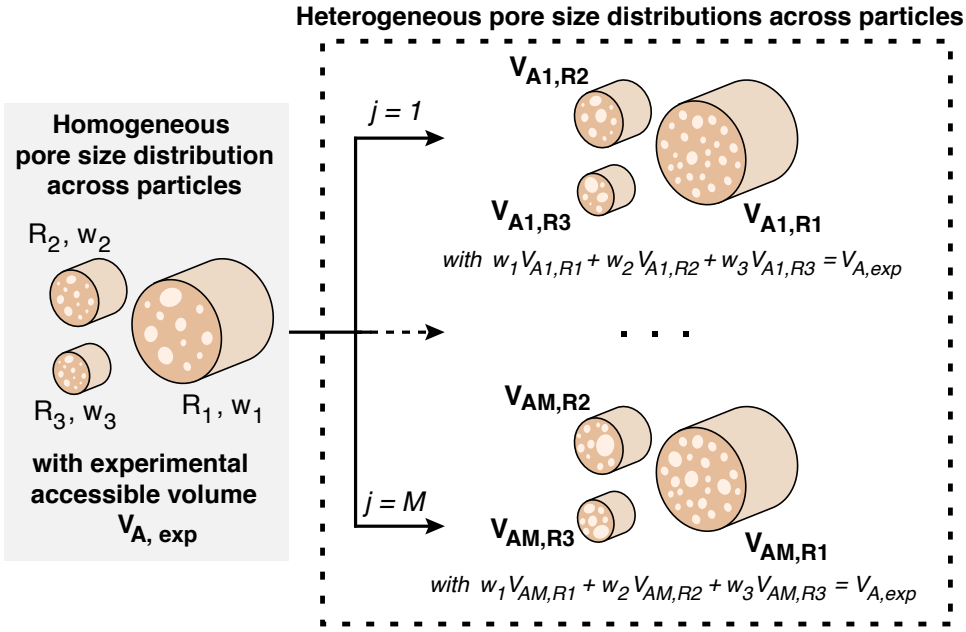



Figure B.1 – *Schematic representation of the generation of variable pore size distributions within the lignocellulosic sample. As an illustration, 3 different particles are here distinghuised as a function of their radius $R_i$, each representing $w_i$ weight fraction of the sample. Experimental measurement of pore volume distribution lead to an average bulk value of $V_{A,exp}$. For each partition $j = \{1,2,\cdots,M\}$ generated, the combination of the specific accessible pore volume generated for each particle radius must match the experimentally measured value $V_{A,exp}$.*

then clusterised based on distance using the built-in function cluster in Matlab. A cutoff value of 3 was chosen to sort distributions as it led to a good trade-off between the number of cluster and the homogeneity of distribution inside each individual cluster. The cutoff value set the maximum distance allowed between distributions within a given cluster. Only one simulation was then run for each cluster to alleviate computational cost, where we chose the most representative distribution for this specific group (i.e. the centroid of the cluster in terms of distance) [Figure B.2]. The ratio between the total pore volume that can accommodate one

(51 Å < R > 102Å) or two enzymes (102 Å < R) was matched to the one measured experimentally. Number of simulations per substrate ranged from 20 for the native sample to 169 for the organosolv sample.

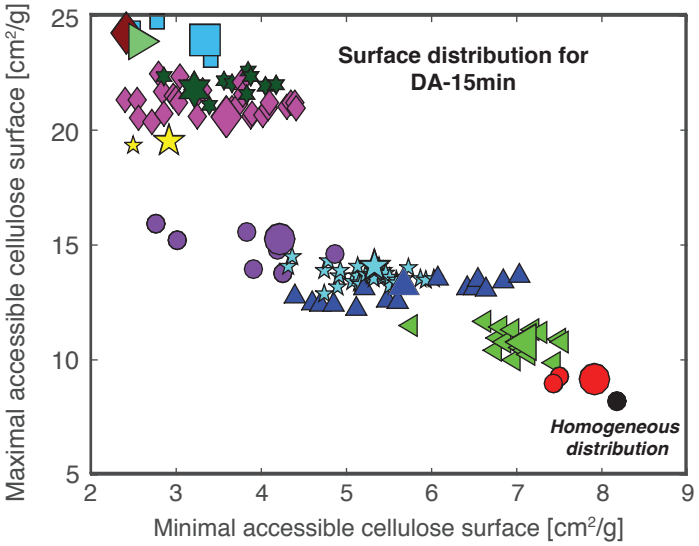



Figure B.2 – *Spatial visualisation of the different pore volume distributions generated for the case DA-15min substrate. Each point represents a single distribution with its coordinates corresponding to the accessible cellulose in the largest particle (minimum surface) and the smallest particle (maximum surface). For comparison, the original homogeneous distribution is shown. Each cluster is shown with a unique colour, with the representative (centroid) distribution of each one highlighted as the largest point within the subgroup.*

## B.2 Model – inclusion of variable cellulose hydrolysablity

The overall cellulose accessibility $x_a$ for a given substrate was defined as equal to the final yield of cellulose conversion $y_{g,final}$:

$$x_a = y_{g,final}, \tag{B.2}$$

However, this variable was allowed to vary between two values as a function of the particle size. To conduct a first sensitivity analysis, only two values were used to test the effect of letting this accessibility vary with particle size, rather than representing the more complicated physical reality of the samples. Each particle presenting a radius smaller or equal to the chosen threshold of 100 µm is assumed to be completely accessible, i.e. $x_{a,R\leq 100\mu m}$=1. For particles larger than this limit, a reduced cellulose accessibility ($x_{a,R>100\mu m} < y_{g,final}$) was calculated. This accessibility ($x_{a,R>100\mu m}$) was computed from the particle size distribution, where the total volume of accessible cellulose contained in the completely accessible particle (R≤100µm) $V_{cellulose,acc,R\leq 100\mu m}$ was first subtracted from the total volume of accessible cellulose $V_{cellulose,acc,total}$, to obtain the total volume of cellulose contained in particles presenting restricted cellulose accessibility $V_{cellulose,acc,R\leq 100\mu m}$:

$$\begin{aligned} V_{cellulose,acc,R\leq 100\mu m} &= V_{cellulose,acc,total} - V_{cellulose,acc,R\leq 100\mu m} \\ &= y_{g,final} \cdot V_{cellulose,total} - x_{a,R\leq 100\mu m} \cdot V_{cellulose,R\leq 100\mu m} \\ &= y_{g,final} \cdot V_{cellulose,total} - V_{cellulose,R\leq 100\mu m} \end{aligned} \tag{B.3}$$

Assuming that accessible cellulose is evenly distributed in large particles (R>100µm), $x_{a,R>100\mu m}$ is then given by:

$$x_{a,R>100\mu m} = \frac{V_{cellulose,acc,R\leq 100\mu m}}{y_{g,final} \cdot V_{cellulose,R\leq 100\mu m}}, \tag{B.4}$$

ensuring that, overall, the total accessible cellulose corresponds to the final glucose yield.

## B.3 Model – inclusion of particle size distribution

The model developed was extended to include the effects of the particle size distribution and possible heterogeneities in terms of accessible pore volume and cellulose accessibility as a function of particle diameter. Simulations for different mean particle diameters measured experimentally were run in parallel and then merged according to their fraction of total volume in the distribution:

$$y_{g,predicted}(t) = \sum_{i=1}^{N} y_{g,i}(t)\, w_{Ri}, \tag{B.5}$$

where $y_{g,i}$(t) is the glucose yields predicted for a wood particle presenting a diameter $d_i$ which contributed a fraction $w_{Ri}$ to the total sample volume. The final predicted yield $y_{g,predicted}$(t) is sum over all mean diameter represented in the distribution.

## B.4 Particle size distribution

Images of wet wood samples were acquired in dual polarization using a digital camera (Canon EOS 5D/Camera Adapter 1.25x) mounted on a macroscope (Leica Wild M420 Macroscope/Apo-zoom 1:6 Objective) with 20x magnification. Samples were sealed onto microscopic slides to avoid evaporation. Wood sample chosen for analysis underwent the same procedure as the one used for enzymatic hydrolysis, i.e. substrates were filtered and extensively washed after the pretreatment process. The resulting pictures were then processed with ImageJ using the built-in Particle Analyzer plugin, indexing and classifying particles according to their projected surface with a detection threshold of 100 μm$^2$ [Figure B.3]. Particle diameters were then carefully measured individually to guarantee that the diameter corresponded to the width perpendicular to wood fibers as defined by the model and to insure that it was representative of

the particle size as particle never had perfectly cylindrical shapes. Overlapping particles were disregarded, and in cases of highly asymmetrical particles where large differences in diameter were observed across the particle, a mean value was assumed according to the particle volume. For each sample, three slides were prepared and a similar quantity of particles (in terms of total volume and not particle number) were analyzed in each case to ensure a representative particle distribution. Particles analyzed were binned within 50 µm ranges. Finally, by assuming a cylindrical shape, particle distributions measured by particle number were converted into volume-based particle distributions. In the particular case of the organosolv pretreatment, which lead to a finer particle distribution, the particle size distribution was estimated based on images acquired with an optical microscope (Zeiss Axiolab Pol Series, 400x magnification).

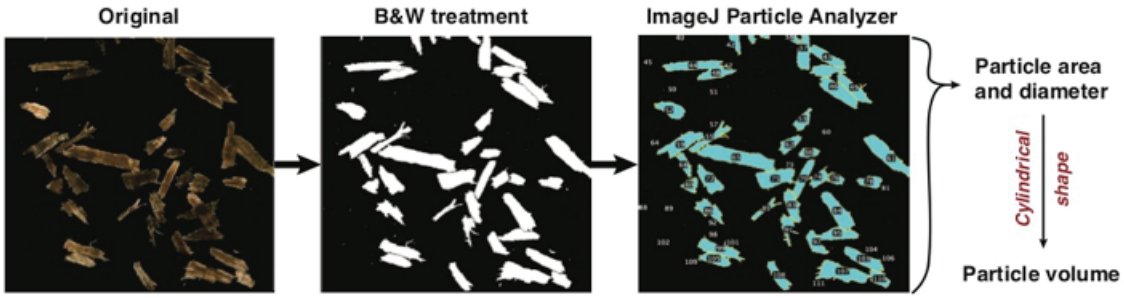



Figure B.3 – *Overview of the different steps for data analysis of wood particle images to determine the particle size distribution.*

## B.5 Model – Error propagation

A Monte-Carlo uncertainty analysis with sampling number of 50 was found to ensure that the average pore volume calculated from the generated sampling distribution led to the correct average pore volume and associated standard error, while limiting computational cost (Figure B.4). Note that, for sake of simplicity, Figure B.4 shows the accessible surface calculated

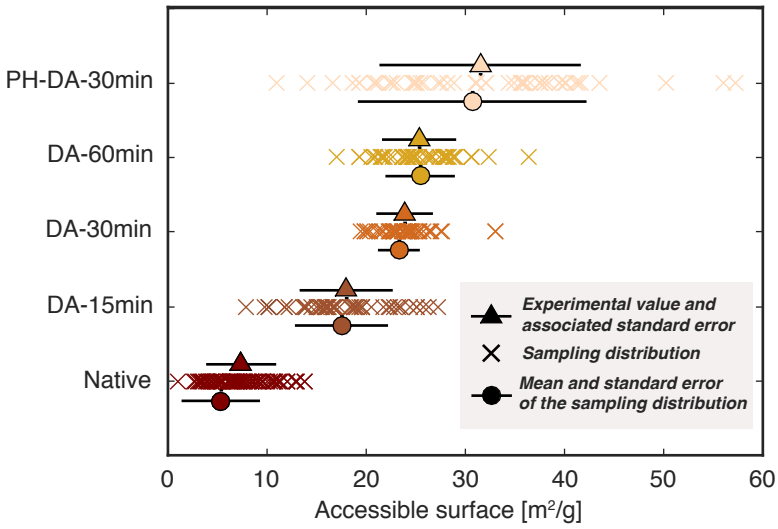



Figure B.4 – *Sampling of the standard error of the accessible surface for error propagation.*

each sampled accessible pore volume. Simulations were then run for each of the generated sets, allowing thus to computing uncertainties on the predicted glucose yields.

## B.6 Native and pretreated substrates

Air-dried beech wood (Fagus sylvatica) chips collected from Zollikofen (Switzerland) were first milled to pass through a 2-mm screen. These particles were further sieved and those between 250 and 450 µm in diameter were retained as the so-called native substrate. This substrate was further processed using dilute-acid (DAP) and organosolv pretreatments to obtain a wide range of materials with different degrees of accessibility to cellulases. Dilute-acid pretreatments with 1wt% sulfuric acid (SA/Merck, 100732) at 160°C were performed in 60 ml glass reactors at a loading of 2 g of dry substrate per 20 ml acid solution for several residence times – 15, 30 and 60 minutes –, followed by Büchner filtration and extensive washing with purified water (Milli-Q grade). To allow fiber swelling, wood particles were pre-soaked overnight at 4°C in the pretreatment solution.

For the organosolv process, 15 g of dry native biomass was hydrolysed using hydrochloric acid (HCl 37%wt, 6.3ml/Merck, 100317) in dioxane (135 ml/Sigma, 33147) under reflux for 5h at 80°C in the presence of water (9 ml). To avoid extensive lignin condensation onto the substrate, the resulting solids were filtered and washed with dioxane mid-reaction before restarting the pretreatment in similar conditions. After reaction, wood particles were filtered and washed with acetone to remove dissolved lignin prior to extensive washing with Milli-Q water. Wet pretreated wood samples were kept for a maximum of two weeks in sealed plastic bags at 4°C prior to further utilisation, to avoid drying and degradation.

Partially hydrolysed (PH) substrate was obtained by stopping the enzymatic action after 2 h following a procedure developed elsewhere[140]. Briefly, once the particles were removed and washed, cellulase proteolysis was performed by adding a Pronase E mixture (Sigma, P5147) mixing the particles overnight at 37°Cin a phosphate buffer (pH=7.4/Sigma, P3619). The proteases were then inhibited by a proteinase inhibitor cocktail (2 h, 37°C/Sigma, P2714). The resulting digested wood substrate was recovered by filtration and successively washed by Milli-Q water and a 1.0 M NaCl solution.

For all substrates, enzymatic hydrolysis was carried out at a biomass loading of 2% and enzyme loading of 60 FPU/glucan of Cellic CTec2, following the procedure previously described (see Appendix A.11).

Composition analyses of wood samples were performed according to the LAP procedure published by the National Renewable Energy Laboratory[187] (see Appendix A.10). For the analysis of the native substrate, extractives were removed as described in the same NREL procedure prior to compositional analysis to avoid interferences of said extractives in the mass balances of the various fractions[187].

## B.7 Model – fitting of $M_p$

**$M_p$ as a function of particle radius.** Optimal values of the parameter $M_p$ were obtained using combinatorial optimisation. Specifically, a set of simulations with relevant values of $M_p$ (from 350 to 800 with 15 units increments) for the different particle size constituting the sample was generated, and then combined in a way that allowed this parameter to vary with the particle diameter, as in the case of the cellulose accessibility (see section B.1). In this case, we imposed that a greater value of $M_p$ (higher cellulose digestibility) was attributed to particle presenting a diameter smaller than a given threshold while a lower $M_p$ (lower cellulose digestibility) was attributed for particles larger than the threshold. This threshold was varied from 20 µm to 565 µm with increment of 50 µm, following the different size categories defined in the particle size distribution. For each of these thresholds, a set of 2 optimal $M_p$ values, corresponding to the value of $M_p$ for particle presenting a diameter smaller than the threshold and one for larger particle, were obtained through least-square fitting to the experimental data.

**$M_p$ as a function cellulose conversion.** As a first approximation, the gradual decrease of $M_p$ as the cellulose is hydrolysed is expressed through a simple exponential decay,

$$M_p(r,t) = f[\varepsilon(r,t))] = M_{p,0} exp(-\alpha\varepsilon(r,t))) \tag{B.6}$$

Optimal values of $M_{p,0}$ and $\alpha$ were obtained for each substrate using the lsqnonlin solver in Matlab. To obtain reasonable initial guess for the optimisation, crude pre-fitting of an exponential decay as a function of the porosity was performed on values of $M_p$ fitted on early yields for the DA-30min and PH-DA-30min. For both DA-30min and DA-60min, the optimisation lead to several combinations of fitted parameter $[M_{p,0}, \alpha]$ giving reasonable

predictions, with a couple of them shared between the two substrates. However, no common set of parameter lead to reasonable predictions when considering DA-15min. In this case, increasing predictions accuracy for the DA-15min by adjusting [$M_{p,0}$, $\alpha$ ] was detrimental to the predictions for the DA-30min and DA-60min.

## B.8 Additional Figures

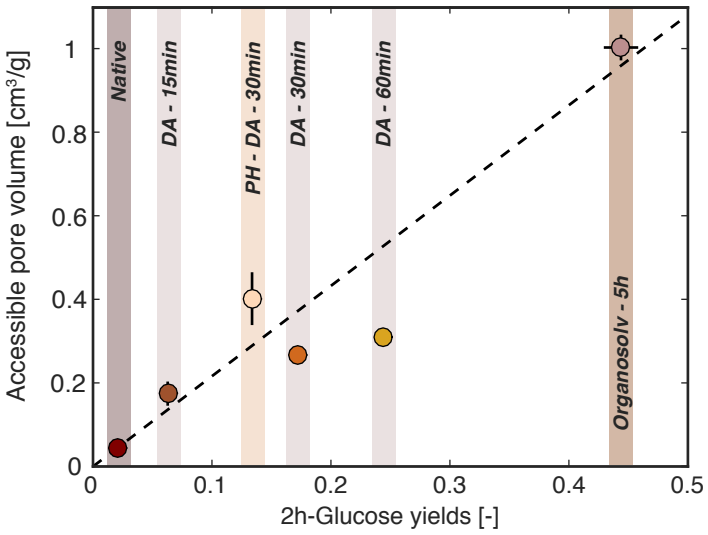



Figure B.5 – *Accessible pore volume to a cellulase as a function of initial glucose yields.*

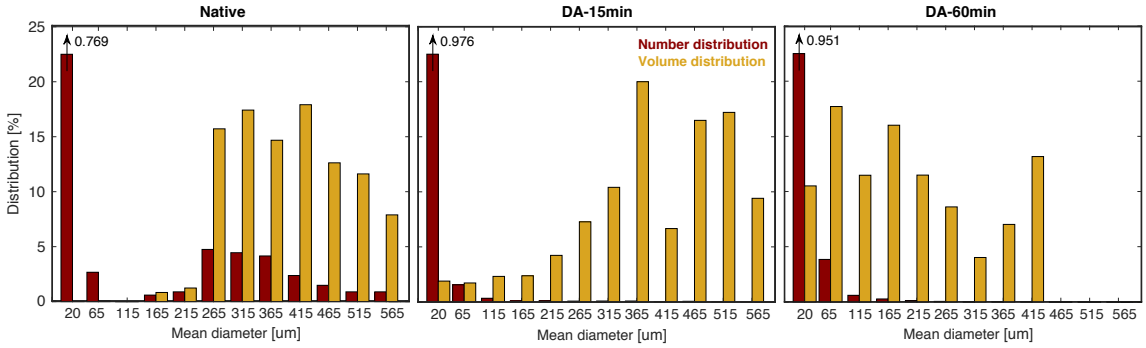



Figure B.6 – *Number- and volume-based particle size distribution as a function of particle diameter for native, DA-15min and DA-60min samples.*

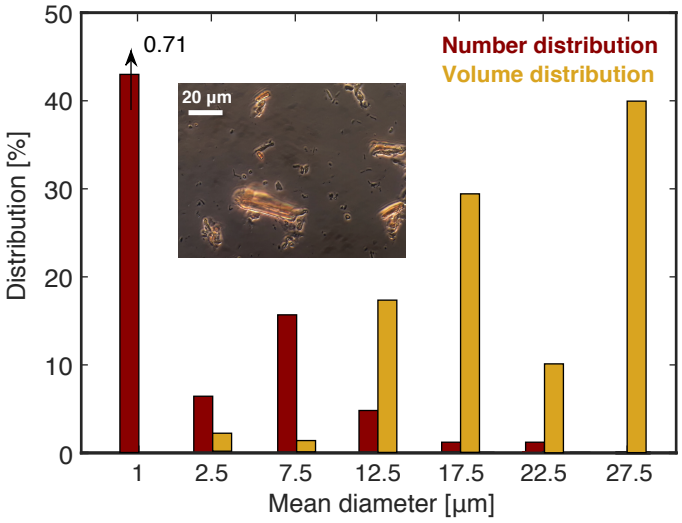



Figure B.7 – *Number- and volume-based particle size distribution as a function of particle diameter for the organosolv pretreatment (Dioxane-5hr).*

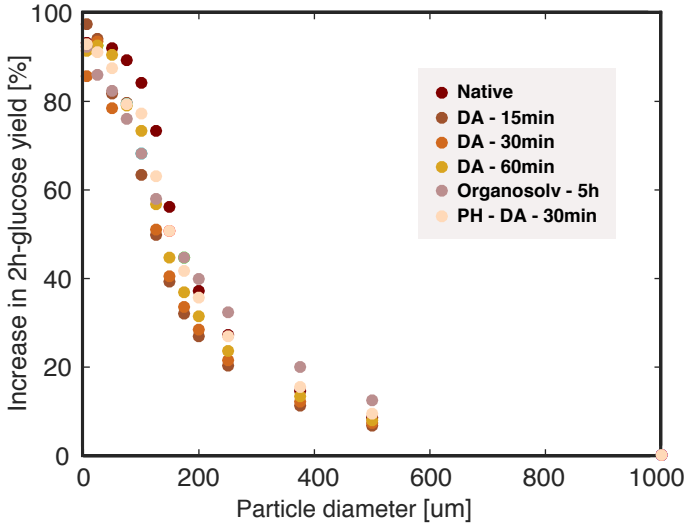



Figure B.8 – *Predicted percentage of increase in early glucose yield upon change in particle radius for all the substrates considered in this study. The porosity of each substrate is kept constant upon change in particle size in the simulations (simulation S0).*

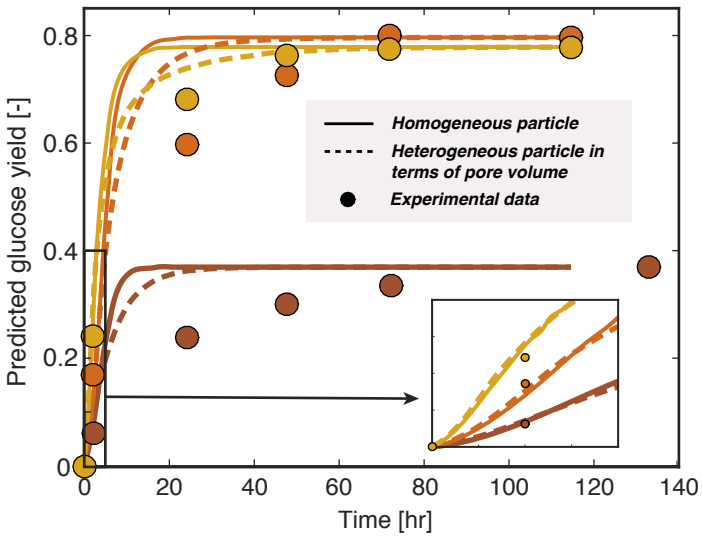



Figure B.9 – *Glucose yield as a function of time for the DAP substrates. Simulation parameters* $\tau$*=3,* $M_p$*=770.*

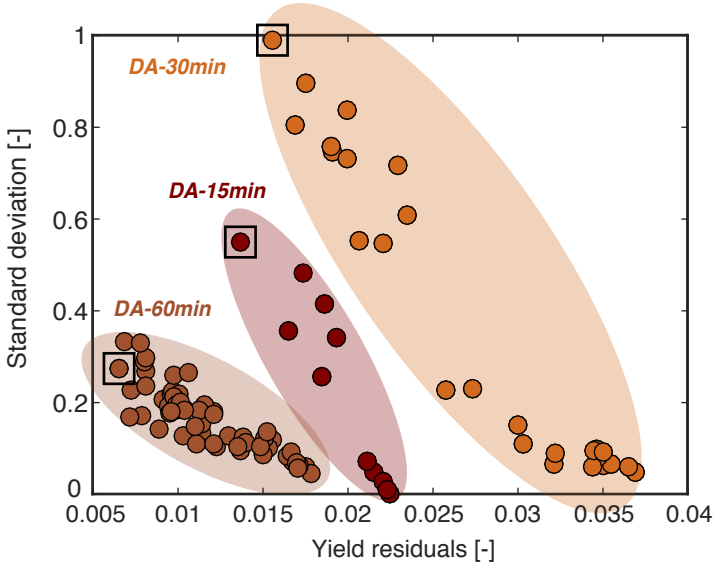



Figure B.10 – *Correlation between the model's prediction error and the standard deviation of the pore volume distribution. For each cluster, the most inhomogeneous distribution (i.e. with the highest standard deviation) leads to the lowest prediction errors.*

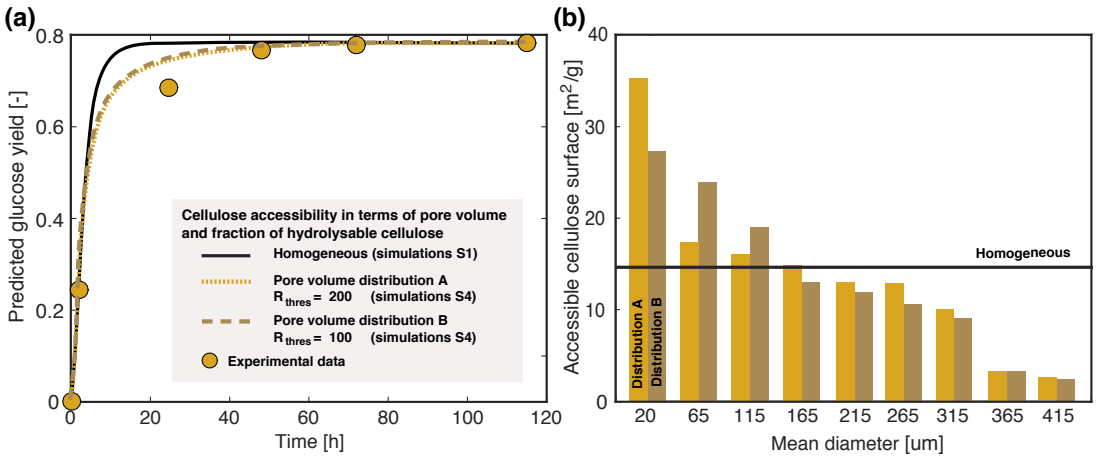



Figure B.11 – *(a) Predicted glucose yields for DA-60min under the two main assumptions: (i) uniform and (ii) non-uniform porosity and cellulose hydrolysability within the substrate. For the latter, two fits are presented leading to similar glucose hydrolysis rate but with varying degree of non-uniformity in terms of porosity and distribution of hydrolysable cellulose. (b) Corresponding pore volume distributions (A and B) as a function of the particle radius leading to the best LS-fit. For sake of readibility, pore size distributions within a particle are here expressed in terms of accessible cellulose surface. For comparison, value corresponding to the homogeneous distribution is also shown.*

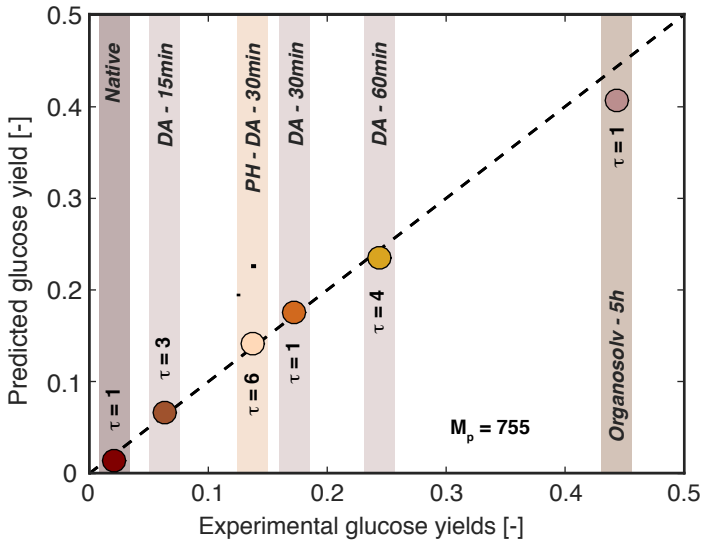



Figure B.12 – *Optimal value of tortuosity τ for all substrates for a given value of $M_p$ of 755*

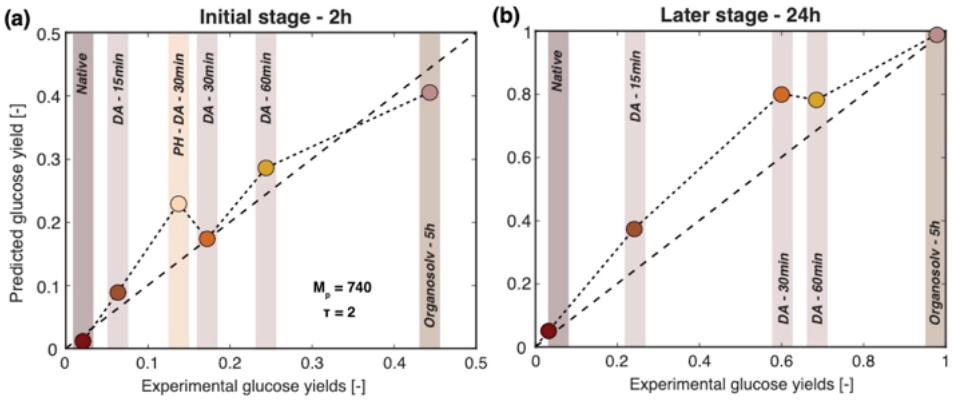



Figure B.13 – *Least-square fitting considering all substrates over time using a single value of $M_p$, leading to an optimal value of $M_p$ = 710*

# Appendix C

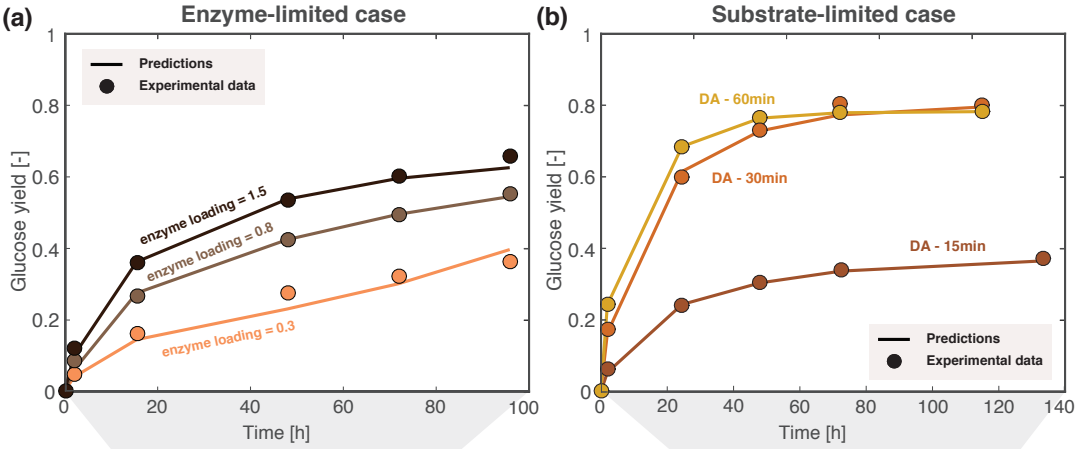
# Appendix for Chapter 4

## C.1 Modelling - Parameter optimisation

In a first stage, optimal values of $k_{des}^{U}$, $M_p$, $M_s$, $S_c^P$,$S_c^U$ were simultaneously fitted for each substrate using the lsqnonlin solver in Matlab. In a second stage, the unproductive adsorption constant on lignin $k_L$ is also included to the fitting scheme. To prevent the optimisation being stuck to a local minimum and ensure a good coverage of the parameter space, optimisation calculations were performed from various combinations of initial guess within realistic bounds. $k_{des}^{U}$ was allowed to vary from 10x faster to 100x slower than its productive counterpart to evaluate the effect of short- and long-stay of cellulases onto the cellulose surface, while the fraction of both productive- and unproductive binding sites on the surface was set to cover situation were all sites are either productive or non-productive. $k_L$ is allowed to assume lower or equal value to the adsorption rate constant on cellulose. From previously fitted values, values ranging from 700 to 3000 were allowed for $M_p$. Finally, within ranges of other parameters and assumed mechanism, values of $M_s$ showed no notable influence on the predicted yields.

## C.2 Best individual fittings - Unproductive binding and variable $M_p$

Results of individual fittings of the parameters for each substrates [Figure C.1].



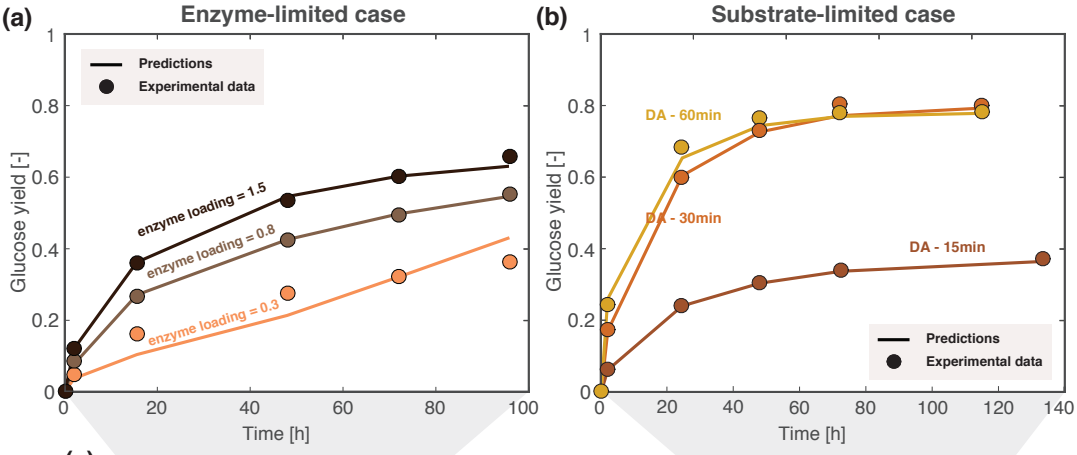

(c)

| DA-30min, wet sieved ~ 400 um | | | SUBSTRATE | DA-15min | DA-30min | DA-60min |
|---|---|---|---|---|---|---|
| 0.3 | 0.8 | 1.5 | ENZYME LOADING | 2.5 | 1.5 | 1.4 |
| 0.001 | 0.002 | 0.003 | $k^U_{des}$ [$s^{-1}$] | 0.003 | 0.003 | 0.006 |
| 1185 | 1580 | 1500 | $M_{p,0}$ | 1345 | 1600 | 1135 |
| | - | | $M_S$ [g/mol] | | - | |
| 0.93 | 0.84 | 0.86 | Fraction of productive sites | 0.85 | 0.90 | 0.90 |

Figure C.1 – *Predictions considering an unproductive cellulase adsorption pathway at the cellulose surface. Best individual parameter fitting for (a) a given substrates at different enzyme loadings and (b) different substrates in excess enzyme conditions. Table (c) gives an overview of the parameters for the best resulting fits. $M_s$ are not indicated as changes in their value shows no impact on the predicted yields.*

## C.3 Best individual fittings - Unproductive binding only

Results of individual fittings of the parameters for each substrates, considering only unproductive binding to cellulose [Figure C.2].



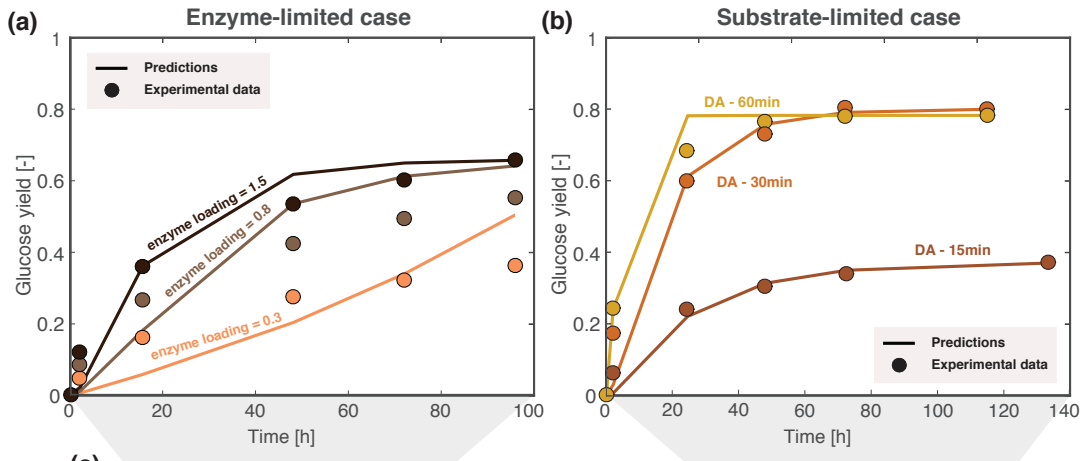

(c)

| DA-30min, wet sieved ~ 400 um | | | SUBSTRATE | DA-15min | DA-30min | DA-60min |
|---|---|---|---|---|---|---|
| 0.3 | 0.8 | 1.5 | ENZYME LOADING | 2.5 | 1.5 | 1.4 |
| 0.01 | 0.02 | 0.03 | $k^{U}_{des}$ [$s^{-1}$] | 0.02 | 0.03 | 0.03 |
| 1860 | 4400 | 2135 | $M_{p,0}$ | 3070 | 2270 | 960 |
| | - | | $M_S$ [g/mol] | | - | |
| 0.73 | 0.51 | 0.72 | Fraction of productive sites | 0.54 | 0.73 | 0.90 |

Figure C.2 – *Predictions considering an unproductive cellulase adsorption pathway at the cellulose surface. Best individual parameter fitting for (a) a given substrates at different enzyme loadings and (b) different substrates in excess enzyme conditions. Table (c) gives an overview of the parameters for the best resulting fits. $M_s$ are not indicated as changes in their value shows no impact on the predicted yields.*

## C.4 Best individual fittings - Low fraction of productive binding on cellulose

In this particular test, we force the unproductive rate constant $k_{des}^{U}$ to be equal or larger to its productive counterpart $k_{des}^{P}$, with a dominant number of unproductive site at the cellulose surface (i.e. fraction of unproductive site is bounded between 0.5-1) in the parameter optimisation [Figure C.3]. We explore here the possibility of having only a minor fraction of binding leading to hydrolysis.




(c)

| DA-30min, wet sieved ~ 400 um | | | SUBSTRATE | DA-15min | DA-30min | DA-60min |
|---|---|---|---|---|---|---|
| 0.3 | 0.8 | 1.5 | ENZYME LOADING | 2.5 | 1.5 | 1.4 |
| 0.22 | 0.26 | 0.32 | $k_{des}^{U}$ [$s^{-1}$] | 0.22 | 0.42 | 0.67 |
| 990 | 990 | 985 | $M_{p,0}$ | 990 | 1000 | 1020 |
| | - | | $M_S$ [g/mol] | | - | |
| 0.10 | 0.10 | 0.10 | Fraction of productive sites | 0.10 | 0.10 | 0.10 |

Figure C.3 – *Predictions considering an unproductive cellulase adsorption pathway at the cellulose surface. Best individual parameter fitting for (a) a given substrates at different enzyme loadings and (b) different substrates in excess enzyme conditions. Table (c) gives an overview of the parameters for the best resulting fits. $M_s$ are not indicated as changes in their value shows no impact on the predicted yields.*

# List of Acronyms and Abbreviations

**AFEX** Ammonia Fiber Explosion

**AGX** (Arabino)GlucoronoXylan

**AX** Arabinoxylan

**BC** Boundary Conditions

**CBM** Carbohydrate-Binding Module

**CBH** CelloBioHydrolase

**CD** Catalytic Domain

**CrI** Crystallinty Index

**DAP** Dilute-Acid Pretreatment

**DOE** Department Of Energy

**DM** Dry Matter

**DP** Degree Of Polymerisation

**EG** EndoGlucanase

**FSP** Fiber Saturation Point

**GH** Glycoside Hydrolase

**GHG** Greenhouses Gases

**GaM** GalactoMannan

**GAX** (Glucorono)ArabinoXylan

**GGM** GalactoGlucoMannan

**GM** GlucoMannan

**GX** GlucoronoXylan

**HHF** Hybrid Hydrolysis and Fermentation

**HMF** HydroxyMethylFurfural

**IC** Initial Conditions

**IL** Ionic Liquid

**LCC** Lignin-Carbohydrate Complex

**LPMO** Lytic Polysaccharide Monooxygenase

**ODE** Ordinary Differential Equation

**PCW** Primary Cell Wall

**PDE** Partial Differential Equation

**PH** Partially Hydrolysed

**PSD** Particle Size Distribution

**SA** Sulfuric Acid

**SCW** Secondary Cell Wall

| | |
|---|---|
| **SE** | Steam Explosion |
| **SHF** | Separated Hydrolysis and Fermentation |
| **SSF** | Simultaneous Saccharification and Fermentation |
| **SSP** | Shared Socioeconomic Pathways |
| **THF** | TetraHydroFuran |
| ***T. Reesei* or *Tr*** | *Trichoderma Reesei* |
| **QM/MM** | Quantum Mechanics/Molecular Mechanics simulations |

# jessicarohrbach

Nationality Swiss

Born 06.09.1990

Marital Status Single

## contact

Rue Cité-Devant 10
1005 Lausanne, VD
Switzerland

+41 79 746 98 02

jessica.rohrbach@epfl.ch

## languages

French mother tongue

English good level written and oral (C1/B2)

German intermediate level (B2)

Serious and hard-working student with a Master of Science from the EPFL (Switzerland) having a keen interest in Computational Chemistry and Biochemistry, as well as Physical Chemistry. Strong desire to increase her skills and knowledge in these fields in a working environment.

## education

2015-2020 (*expected*) PhD, Institute of Chemical Sciences and Engineering
Ecole Polytechnique Fédérale de Lausanne (EPFL), Basic Sciences Faculty

2012-2014 Master of Science in Molecular and Biological Chemistry
Ecole Polytechnique Fédérale de Lausanne (EPFL), Basic Sciences Faculty

Options: Computational and Physical Chemistry

Master project under the supervision of Prof. Jiri Vaníček in the Laboratory of Theoretical Physical Chemistry (LCPT): "*Accelerating quantum dynamics calculations of time-resolved electronic spectra by optimising Gaussian bases*"

Semester project under the supervision of Prof. Jiri Vaníček in the Laboratory of Theoretical Physical Chemistry (LCPT): "*Evaluation of absorption spectra using semiclassical methods beyond the Condon approximation*"

Average grade: 5.71/6

2009-2012 Bachelor of Science in Chemistry and Chemical Engineering
Ecole Polytechnique Fédérale de Lausanne (EPFL), Basic Sciences Faculty

Options: Computational Chemistry, Biological & Biophysical Chemistry and Synthetic & Analytical Chemistry

Semester project under the supervision of Prof. Ursula Roethlisberger in the Laboratory of Computational Chemistry and Biochemistry (LCBC): "Characterisation of promising Ir(III) complexes in the perspective of their use in OLED devices"

Average grade: 5.54/6

2006-2009 High School Degree
Lycée Denis-de-Rougemont in Neuchâtel, Switzerland

Specific Option: Biology & Chemistry

Complementary Option: Economics and Law

Maturity thesis: "Mendel's work: from theory to practice"

## experience

2013 Teaching Assistant
Bachelor course "Equilibres et reactivités Chimiques"

2011-2012 Staff member of the association Baramine

## awards

2012 — Excellence Mention

Awarded to the top 10 students during Bachelor cycle in Chemistry and Chemical Engineering

## technical skills

IT literacy

General: Microsoft Office (Word, Excel, PowerPoint), LaTeX

Programming: Fortran 90, basics in C, Bash and Python

Software: Mathematica, Matlab, Chemdraw

## interests

Arts, cooking, swimming, music

(1) IPCC, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P., Eds.; Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA, 2013.

(2) IPCC, *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Field, C. et al., Eds.; Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA, 2014, 1132pp.

(3) IPCC, *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Barros, V., Field, C., Dokken, D., Mastrandrea, M., Mach, K., Bilir, T., Chatterjee M.and Ebi, K., Estrada, Y., Genova, R., Girma, B., Kissel, E., Levy, A., MacCracken, S., Mastrandrea, P., White, L., Eds.; Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA, 2014, 688pp.

(4) NOAA *State of the Climate: Global Climate Report for Annual 2019*; tech. rep.; 2019.

(5) Riahi, K. et al. *Global Environmental Change* **2017**, *42*, 153–168.

(6) Gielen, D.; Boshell, F.; Saygin, D.; Bazilian, M. D.; Wagner, N.; Gorini, R. *Energy Strategy Reviews* **2019**, *24*, 38–50.

(7) Obama, B. *Science* **2017**, *355*, 126–129.

(8) Burke, M.; Hsiang, S. M.; Miguel, E. *Nature* **2015**, *527*, 235–239.

(9) IPCC, *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Edenhofer, O. et al., Eds.; Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA, 2014.

(10) Kittner, N.; Lill, F.; Kammen, D. M. *Nature Energy* **2017**, *2*, 1–6.

(11) IRENA *Global Renewables Outlook: Energy transformation 2050 (Edition: 2020)*; tech. rep.; Abu Dhabi: International Renewable Energy Agency, 2020.

(12) OECD Primary energy supply (indicator)., https://data.oecd.org/energy/primary-energy-supply.html, Accessed: 2020-12-06, 2020.

(13) IRENA *Renewable Power Generation Costs in 2019*; tech. rep.; Abu Dhabi: International Renewable Energy Agency, 2019.

(14) Goldemberg, J.; Teixeira Coelho, S. *Energy Policy* **2004**, *32*, 711–714.

(15) Wang, Q. *Renewable and Sustainable Energy Reviews* **2009**, *13*, 2562–2570.

(16) Bailis, R.; Drigo, R.; Ghilardi, A.; Masera, O. *Nature Climate Change* **2015**, *5*, 266–272.

(17) IRENA *Reaching zero with renewables: Eliminating CO2 emissions from industry and transport in line with the 1.5C climate goal*; tech. rep.; Abu Dhabi: International Renewable Energy Agency, 2020, p 216.

(18) IRENA *Advanced biofuels. What holds them back?*; tech. rep.; Abu Dhabi: International Renewable Energy Agency, 2019.

(19) Cheng, B. H.; Huang, B. C.; Zhang, R.; Chen, Y. L.; Jiang, S. F.; Lu, Y.; Zhang, X. S.; Jiang, H.; Yu, H. Q. *Science Advances* **2020**, *6*, 1–8.

(20) Griscom, B. W. et al. *Proceedings of the National Academy of Sciences* **2017**, *114*, 11645–11650.

(21) Favero, A.; Daigneault, A.; Sohngen, B. *Science Advances* **2020**, *6*, DOI: 10.1126/sciadv.aay6792.

(22) Cornwall, W. *Science* **2019**, *15*, 792–795.

(23) Bastin, J.-F.; Finegold, Y.; Garcia, C.; Mollicone, D.; Rezende, M.; Routh Devin Zohner, C. M.; Crowther, T. W. *Science* **2019**, *365*, 76–79.

(24) Schlesinger, W. H. *Science* **2018**, *359*, 1328–1329.

(25) Long, H.; Li, X.; Wang, H.; Jia, J. *Renewable and Sustainable Energy Reviews* **2013**, *26*, 344–352.

(26) Searle, S.; Malins, C. *GCB Bioenergy* **2015**, *7*, 328–336.

(27) Carpenter, D.; Westover, T. L.; Czernik, S.; Jablonski, W. *Green Chemistry* **2014**, *16*, 384–406.

(28) Jung, K. A.; Lim, S. R.; Kim, Y.; Park, J. M. *Bioresource Technology* **2013**, *135*, 182–190.

(29) Somerville, C.; Youngs, H.; Taylor, C.; Davis, S. C.; Long, S. P. *Science* **2010**, *790*, 790–793.

(30) Raud, M.; Kikas, T.; Sippula, O.; Shurpali, N. J. *Renewable and Sustainable Energy Reviews* **2019**, *111*, 44–56.

(31) Bhatia, L.; Bachheti, R. K.; Garlapati, V. K.; Chandel, A. K. **2020**.

(32) Liu, Z.; Wang, K.; Chen, Y.; Tan, T.; Nielsen, J. *Nature Catalysis* **2020**, *3*, DOI: 10.1038/s41929-019-0421-5.

(33) Brun, N.; Hesemann, P.; Esposito, D. **2017**, 4724–4738.

(34) Zoghlami, A.; Paës, G. **2019**, *7*, DOI: 10.3389/fchem.2019.00874.

(35) Himmel, M. E.; Ding, S.-y.; Johnson, D. K.; Adney, W. S. *Science* **2007**, *315*, 804–808.

(36) Chundawat, S. P.; Beckham, G. T.; Himmel, M. E.; Dale, B. E. *Annual Review of Chemical and Biomolecular Engineering* **2011**, *2*, 121–145.

(37) Schutyser, W.; Renders, T.; Van Den Bosch, S.; Koelewijn, S. F.; Beckham, G. T.; Sels, B. F. *Chemical Society Reviews* **2018**, *47*, 852–908.

(38) Singhvi, M. S.; Gokhale, D. V. **2019**, 9305–9320.

(39) Hames, B. R., *Biomass compositional analysis for energy applications. In Biofuels: methods and protocols; Methods in Molecular Biology vol. 581*; Humana Press: New York, NY, 2009, pp 145–167.

(40) Klemm, D.; Heublein, B.; Fink, H.-P.; Bohn, A. Cellulose: Fascinating Biopolymer and Sustainable Raw Material., 2005.

(41) Jarvis, M. C. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2018**, *376*, DOI: 10.1098/rsta.2017.0045.

(42) Schneider, R.; Hanak, T.; Persson, S.; Voigt, C. A. *Current Opinion in Plant Biology* **2016**, *34*, 9–16.

(43) Nixon, B. T.; Mansouri, K.; Singh, A.; Du, J.; Davis, J. K.; Lee, J. G.; Slabaugh, E.; Vandavasi, V. G.; O'Neill, H.; Roberts, E. M.; Roberts, A. W.; Yingling, Y. G.; Haigler, C. H. *Scientific Reports* **2016**, *6*, 1–14.

(44) Festucci-Buselli, R. A.; Otoni, W. C.; Joshi, C. P. *Brazilian Journal of Plant Physiology* **2007**, *19*, 1–13.

(45) Hayashi, J.; Sufoka, A.; Ohkita, J.; Watanabe, S. *Journal of Polymer Science: Polymer Letters Edition* **1975**, *13*, 23–27.

(46) Gardiner, E. S.; Sarko, A. *Canadian Journal of Chemistry* **1985**, *63*, 173–180.

(47) Yang, B.; Dai, Z.; Ding, S.-y.; Wyman, C. E. *Biofuels* **2011**, *2*, 421–450.

(48) Ebringerova, A. *Macromolecular Symposia* **2005**, *232*, 1–12.

(49) Wierzbicki, M. P.; Maloney, V.; Mizrachi, E.; Myburg, A. A. **2019**, *10*, 1–29.

(50) Hu, F.; Ragauskas, A. *Bioenergy Research* **2012**, *5*, 1043–1066.

(51) Bertella, S.; Luterbacher, J. S. *Trends in Chemistry* **2020**, *2*, 440–453.

(52) Balakshin, M.; Capanema, E. A.; Zhu, X.; Sulaeva, I.; Potthast, A.; Rosenau, T.; Rojas, O. J. **2020**, 3985–4001.

(53) Ralph, J.; Lapierre, C.; Boerjan, W. *Current Opinion in Biotechnology* **2019**, *56*, 240–249.

(54) Schutyser, W.; Renders, T.; Van Den Bosch, S.; Koelewijn, S. F.; Beckham, G. T.; Sels, B. F. *Chemical Society Reviews* **2018**, *47*, 852–908.

(55) Sarkar, P.; Bosneaga, E.; Auer, M. *Journal of Experimental Botany* **2009**, *60*, 3615–3635.

(56) Alberts, B and Johnson, A and Lewis, J and al., *The Plant Cell Wall*; New York: Garland Science: 2002.

(57) Cosgrove, D. J.; Jarvis, M. C. **2012**, *3*, 1–6.

(58) Daher, F. B.; Braybrook, S. A. *Frontiers in Plant Science* **2015**, *6*, 1–8.

(59) Meents, M. J.; Watanabe, Y.; Samuels, A. L. *Annals of Botany* **2018**, *121*, 1107–1125.

(60) Wightman, R.; Dupree, P.; Lyczakowski, J. J. **2019**, *10*, 1–14.

(61) Giummarella, N.; Pu, Y.; Ragauskas, A. J.; Lawoko, M. *Green Chemistry* **2019**, *21*, 1573–1595.

(62) Zhao, Y.; Shakeel, U.; Saif Ur Rehman, M.; Li, H.; Xu, X.; Xu, J. Lignin-carbohydrate complexes (LCCs) and its role in biorefinery., 2020.

(63) Himmel, M. E.; Xu, Q.; Luo, Y.; Ding, S.-Y.; Lamed, R.; Bayer, E. a. *Biofuels* **2010**, *1*, 323–341.

(64) Chundawat, S. P. S.; Lipton, M. S.; Purvine, S. O.; Uppugundla, N.; Gao, D.; Balan, V.; Dale, B. E. *Journal of Proteome Research* **2011**, *10*, 4365–4372.

(65) Agbor, V. B.; Cicek, N.; Sparling, R.; Berlin, A.; Levin, D. B. *Biotechnology Advances* **2011**, *29*, 675–685.

(66) Baruah, J.; Nath, B. K.; Sharma, R.; Kumar, S.; Deka, R. C.; Baruah, D. C.; Kalita, E. *Frontiers in Energy Research* **2018**, *6*, 141.

(67) Songstadt, D. D.; Lakkshmanan, P.; CHen, J.; Gibbons, W.; Hughes, S.; Nelson, R. *In Vitro Cellular & Developmental Biology Plant* **2009**, *45*, 189–192.

(68) Mika, T.; Cse, E.; Ne, A. **2018**, DOI: 10.1021/acs.chemrev.7b00395.

(69) Dusselier, M.; Mascal, M.; Sels, B. F. In *Selective Catalysis for Renewable Feedstocks and Chemicals*, Nicholas, K. M., Ed.; Springer International Publishing: Cham, 2014, pp 1–40.

(70) Bozell, J. J.; Petersen, G. R. *Green Chemistry* **2010**, *12*, 539–55.

(71) Ponnusamy, V. K.; Nguyen, D. D.; Dharmaraja, J.; Shobana, S.; Banu, J. R.; Saratale, R. G.; Chang, S. W.; Kumar, G. *Bioresource Technology* **2019**, *271*, 462–472.

(72) Calvo-Flores, F. G.; Dobado, J. A. *ChemSusChem* **2010**, *3*, 1227–1235.

(73) Shuai, L.; Amiri, M. T.; Questell-Santiago, Y. M.; Héroguel, F.; Li, Y.; Kim, H.; Meilan, R.; Chapple, C.; Ralph, J.; Luterbacher, J. S. *Science* **2016**, *354*, 329–333.

(74) Alonso, D. M.; Bond, J. Q.; Dumesic, J. A. *Green Chem.* **2010**, *12*, 1493–1513.

(75) Shuai, L.; Luterbacher, J. *ChemSusChem* **2016**, *9*, 133–155.

(76) Lynd, L. R.; Van Zyl, W. H.; McBride, J. E.; Laser, M. *Current Opinion in Biotechnology* **2005**, *16*, 577–583.

(77) Yang, B.; Dai, Z.; Ding, S.-y.; Wyman, C. E. *Biofuels* **2011**, *2*, 421–450.

(78) Karimi, K.; Taherzadeh, M. J. *Bioresource Technology* **2016**, *203*, 348–356.

(79) Zeng, M.; Gao, H. N.; Wu, Y. Q.; Fan, L. R.; Zheng, T. H.; Zhou, D. F. *Journal of Macromolecular Science Part a-Pure and Applied Chemistry* **2010**, *47*, 1042–1049.

(80) Li, H.; Qu, Y.; Yang, Y.; Chang, S.; Xu, J. *Bioresource Technology* **2016**, *199*, 34–41.

(81) Sun, Y.; Cheng, J. J. *Bioresource Technology* **2005**, *96*, 1599–1606.

(82) Nair, R. B.; Lundin, M.; Brandberg, T.; Lennartsson, P. R.; Taherzadeh, M. J. *Industrial Crops and Products* **2015**, *69*, 314–323.

(83) Jeong, S.-Y.; Lee, J.-W. *Industrial Crops and Products* **2016**, *79*, 1–6.

(84) Li, H.; Pu, Y.; Kumar, R.; Ragauskas, A. J.; Wyman, C. E. *Biotechnology and Bioengineering* **2014**, *111*, 485–492.

(85) AndSridhar Viamajala, M. J. S.; Decker, S. R.; Tucker, M. P.; Himmel, M. E.; Vinzant, T. B. *Biotechnology Progress* **2007**, *23*, 1333–1339.

(86) Shinde, S. D.; Meng, X.; Kumar, R.; Ragauskas, A. J. *Green Chemistry* **2018**, *20*, 2192–2205.

(87) Jönsson, L. J.; Martín, C. *Bioresource Technology* **2016**, *199*, 103–112.

(88) Zhao, X.; Cheng, K.; Liu, D. *Applied Microbiology and Biotechnology* **2009**, *82*, 815–827.

(89) Lee, S. H.; Doherty, T. V.; Linhardt, R. J.; Dordick, J. S., *102*, 1368–1376.

(90) Shen, X. J.; Wen, J. L.; Mei, Q. Q.; Chen, X.; Sun, D.; Yuan, T. Q.; Sun, R. C. *Green Chemistry* **2019**, *21*, 275–283.

(91) Payne, C. M.; Knott, B. C.; Mayes, H. B.; Hansson, H.; Himmel, M. E.; Sandgren, M.; Ståhlberg, J.; Beckham, G. T. *Chemical Reviews* **2015**, *115*, 1308–1448.

(92) Zhang, Y. H. P.; Lynd, L. R. *Biotechnology and Bioengineering* **2004**, *88*, 797–824.

(93) Lynd, L. R.; Weimer, P. J.; Zyl, W. H. V.; Isak, S. *Microbiology and Molecular Biology Reviews* **2002**, *66*, 506–577.

(94) Li, X.; Chang, S. H.; Liu, R., *Industrial applications of cellulases and hemicellulases*, 2018, pp 267–282.

(95) Sørensen, A.; Lübeck, M.; Lübeck, P. S.; Ahring, B. K. *Biomolecules* **2013**, *3*, 612–631.

(96) Andberg, M.; Penttilä, M.; Saloheimo, M. *Bioresource Technology* **2015**, *181*, 105–113.

(97) Kubicek, C. P.; Kubicek, E. M. *Current Opinion in Chemical Biology* **2016**, *35*, Energy Mechanistic Biology, 51–57.

(98) Miettinen-Oinonen, A.; Suominen, P. *Applied and Environmental Microbiology* **2002**, *68*, 3956–3964.

(99) Igarashi, K.; Ishida, T.; Hori, C.; Samejima, M. *Applied and Environmental Microbiology* **2008**, *74*, 5628–5634.

(100) Demirel, Y., *Biofuels*, 2018; Vol. 1-5, pp 875–908.

(101) Song, B.; Li, B.; Wang, X.; Shen, W.; Park, S.; Collings, C.; Feng, A.; Smith, S. J.; Walton, J. D.; Ding, S. Y. *Biotechnology for Biofuels* **2018**, *11*, 1–11.

(102) Pihlajaniemi, V.; Sipponen, M. H.; Liimatainen, H.; Sirvio, J. A.; Nyyssola, A.; Laakso, S. *Green Chem.* **2015**, 21–23.

(103) Zoghlami, A.; Paes, G. *Frontiers in Chemistry* **2019**, *7*, 874.

(104) Jeoh, T.; Ishizawa, C. I.; Davis, M. F.; Himmel, M. E.; Adney, W. S.; Johnson, D. K. *Biotechnology and Bioengineering*, *98*, 112–122.

(105) Arantes, V.; Saddler, J. N. *Biotechnology for Biofuels* **2011**, *4*, 1–16.

(106) Bansal, P.; Vowell, B. J.; Hall, M.; Realff, M. J.; Lee, J. H.; Bommarius, A. S. *Bioresource Technology* **2012**, *107*, 243–250.

(107) HE, G. *Nature Biotechnology* **1985**, *3*, 155–160.

(108) Oyedeji, O.; Gitman, P.; Qu, J.; Webb, E. *ACS Sustainable Chemistry and Engineering* **2020**, *8*, 2327–2343.

(109) Yeh, A. I.; Huang, Y. C.; Chen, S. H. *Carbohydrate Polymers* **2010**, *79*, 192–199.

(110) Fougere, J. D.; Lynch, M.; Zhao, J.; Zheng, Y.; Li, K. *Energy and Fuels* **2014**, *28*, 2645–2653.

(111) Chang, V. S.; Holtzapple, M. T. Fundamental factors affecting biomass enzymatic reactivity., 2000.

(112) Pihlajaniemi, V.; Sipponen, M. H.; Kallioinen, A.; Nyyssölä, A.; Laakso, S. *Biotechnology for Biofuels* **2016**, *9*, 18–18.

(113) Wu, N.; Hubbe, M. A.; Rojas, O. J.; Park, S. *Bioresources* **2009**, *4*, 1222–1262.

(114) Ishizawa, C. I.; Jeoh, T.; Adney, W. S.; Himmel, M. E.; Johnson, D. K.; Davis, M. F. *Cellulose* **2009**, *16*, 677–686.

(115) Herbaut, M.; Zoghlami, A.; Habrant, A.; Falourd, X.; Foucat, L.; Chabbert, B.; Paës, G. *Biotechnology for Biofuels* **2018**, *11*, 1–17.

(116) Converse, A. O.; Matsuno, R.; Tanaka, M.; Taniguchi, M. *Biotechnology and Bioengineering* **1988**, *32*, 38–45.

(117) Kurašin, M.; Väljamäe, P. *Journal of Biological Chemistry* **2011**, *286*, 169–177.

(118) Gao, D.; Chundawat, S. P. S.; Sethi, A.; Balan, V.; Gnanakaran, S.; Dale, B. E. *Proceedings of the National Academy of Sciences* **2013**, *110*, 10922–10927.

(119) Chen, Y.; Stipanovic, A. J.; Winter, W. T.; Wilson, D. B.; Kim, Y. J. *Cellulose* **2007**, *14*, 283–293.

(120) Arslan, B.; Egerton, K.; Zhang, X.; Abu-Lail, N. I. *Langmuir* **2017**, *33*, 6857–6868.

(121) Kaschuk, J. J.; Frollini, E. *Industrial Crops and Products* **2018**, *115*, 280–289.

(122) Hall, M.; Bansal, P.; Jay H. Lee, M. J. R.; Bommarius, A. S. *FEBS Journal* **2010**, *227*, 1571–1582.

(123) Park, S.; Baker, J. O.; Himmel, M. E.; Parilla, P. A.; Johnson, D. K. *Biotechnology for Biofuels* **2010**, *3*, 1–10.

(124) Hallac, B. B.; Ragauskas, A. J. *Biofuels, Bioproducts and Biorefining* **2011**, *5*, 215–225.

(125) Saini, J. K.; Patel, A. K.; Adsul, M.; Singhania, R. R. *Renewable Energy* **2016**, *98*, 29–42.

(126) Kellock, M.; Rahikainen, J.; Marjamaa, K.; Kruus, K. *Bioresource Technology* **2017**, *232*, 183–191.

(127) Djajadi, D. T.; Pihlajaniemi, V.; Rahikainen, J.; Kruus, K.; Meyer, A. S. *Biotechnology and Bioengineering* **2018**, *115*, 2869–2880.

(128) Cao, G.; Ximenes, E.; Nichols, N. N.; Frazer, S. E.; Kim, D.; Cotta, M. A.; Ladisch, M. *Bioresource Technology* **2015**, *190*, 412–415.

(129) Bhagia, S.; Dhir, R.; Kumar, R.; Wyman, C. E. *Scientific Reports* **2018**, *8*, 1–12.

(130) Igarashi, K.; Uchihashi, T.; Koivula, A.; Wada, M.; Kimura, S.; Okamoto, T.; Penttila, M.; Ando, T.; Samejima, M. *Science* **2011**, *333*, 1279–1282.

(131) Jalak, J.; Väljamäe, P. *Biotechnology and Bioengineering* **2010**, *106*, 871–883.

(132) Hamre, A. G.; Lorentzen, S. B.; Väljamäe, P.; Sørlie, M. *FEBS Letters* **2014**, *588*, 4620–4624.

(133) Hu, J.; Arantes, V.; Pribowo, A.; Saddler, J. N. *Biotechnology for biofuels* **2013**, *6*, 112.

(134) McKee, L. S.; Sunner, H.; Anasontzis, G. E.; Toriz, G.; Gatenholm, P.; Bulone, V.; Vilaplana, F.; Olsson, L. *Biotechnology for Biofuels* **2016**, *9*, 1–13.

(135) Dutta, S. K.; Chakraborty, S. *Bioresource Technology* **2018**, *259*, 276–285.

(136) Du, J.; Cao, Y.; Liu, G.; Zhao, J.; Li, X.; Qu, Y. *Bioresource Technology* **2017**, *229*, 88–95.

(137) Modenbach, A. A.; Nokes, S. E. *Biomass and Bioenergy* **2013**, *56*, 526–544.

(138) Kristensen, J. B.; Felby, C.; Jorgensen, H. *Biotechnology for biofuels* **2009**, *2*, 11–11.

(139) Nguyen, T. Y.; Cai, C. M.; Osman, O.; Kumar, R.; Wyman, C. E. *Green Chemistry* **2016**, *18*, 1581–1589.

(140) Yang, B.; Willies, D. M.; Wyman, C. E. *Biotechnology and Bioengineering* **2006**, *94(6)*, 1122–1128.

(141) Mansfield, S. D.; Mooney, C.; Saddler, J. N. *Biotechnology Progress* **1999**, *15(5)*, 804–816.

(142) Liao, W.; Liu, Y.; Wen, Z.; Frear, C.; Chen, S. *Biotechnology and Bioengineering* **2008**, *101*, 441–451.

(143) Pihlajaniemi, V.; Sipponen, M. H.; Liimatainen, H.; Sirviö, J. A.; Nyyssölä, A.; Laakso, S. *Green Chemistry* **2016**, *18*, 1295–1305.

(144) Bansal, P.; Hall, M.; Realff, M. J.; Lee, J. H.; Bommarius, A. S. *Biotechnology Advances* **2009**, *27*, 833–848.

(145) Jeoh, T.; Cardona, M. J.; Karuna, N.; Mudinoor, A. R.; Nill, J. *Biotechnology and Bioengineering* **2017**, *114*, 1369–1385.

(146) Knott, B. C.; Haddad Momeni, M.; Crowley, M. F.; Mackenzie, L. F.; Gaptz, A. W.; Sandgren, M.; Withers, S. G.; Ståhlberg, J.; Beckham, G. T. *Journal of the American Chemical Society* **2014**, *136*, PMID: 24341799, 321–329.

(147) Ahamed, F.; Song, H. S.; Ooi, C. W.; Ho, Y. K. *Chemical Engineering Science* **2019**, *206*, 118–133.

(148) Nill, J. D.; Jeoh, T. *ACS Sustainable Chemistry and Engineering* **2020**, *8*, 6722–6733.

(149) Chi, C.; Liu, M.; Gong, Y.; Zhang, S.; Zhang, B. *BioResources* **2015**, *10*, 3–5.

(150) Vani, S.; Sukumaran, R. K.; Savithri, S. *Bioresource Technology* **2015**, *188*, 128–135.

(151) Wojtusik, M.; Villar, J. C.; Ladero, M.; Garcia-Ochoa, F. *Bioresource Technology* **2018**, *268*, 592–598.

(152) Lischeske, J. J.; Stickel, J. J. *Biotechnology for Biofuels* **2019**, 1–15.

(153) Luterbacher, J. S.; Parlange, J.-y.; Walker, L. P. *Biotechnology and Bioengineering* **2013**, *110*, 1–9.

(154) Luterbacher, J. S.; Walker, L. P.; Moran-Mirabal, J. M. *Biotechnology and Bioengineering* **2013**, *110*, 108–117.

(155) Nill, J.; Karuna, N.; Jeoh, T. The impact of kinetic parameters on cellulose hydrolysis rates., 2018.

(156) Kari, J.; Andersen, M.; Borch, K.; Westh, P. *ACS Catalysis* **2017**, *7*, 4904–4914.

(157) Kari, J.; Olsen, J. P.; Jensen, K.; Badino, S. F.; Krogh, K. B. R. M.; Borch, K.; Westh, P. *ACS Catalysis* **2018**, *8*, 11966–11972.

(158) Karuna, N.; Jeoh, T. *Biotechnology and Bioengineering* **2017**, *114*, 533–542.

(159) Hwang, H.-a.; Chiao, S.-M. *Chemical Engineering Science* **1995**, *50*, 685–694.

(160) Moszkowicz, P.; Pousin, J.; Sanchez, F. *Journal of Computational and Applied Mathematics* **1996**, *66*, Proceedings of the Sixth International Congress on Computational and Applied Mathematics, 377–389.

(161) Moran-Mirabal, J. M.; Santhanam, N.; Corgie, S. C.; Craighead, H. G.; Walker, L. P. *Biotechnology and Bioengineering* **2008**, *101*, 1129–1141.

(162) Harriott, P., *Chemical reactor design*; Marcel Dekker: New York, NY, 2003.

(163) Stone, J.; Scallan, A. *Cellulose Chemistry and Technology* **1968**, *2*, 343–358.

(164) Bothwell, M.; Daughetee, S.; Chau, G.; Wilson, D.; Walker, L. *Bioresources Technology* **1997**, *60*, 169–178.

(165) Schiesser, W. E., *The Numerical Method of Lines: Integration of Partial Differential Equations*; Academic Press: San Diego, 1991.

(166) Shampine, L. F.; Reichelt, M. W. *SIAM Journal on Scientific Computing* **1997**, *18*, 1–22.

(167) MATLAB, *version 8.5.0 (R2015a)*; The MathWorks Inc.: Natick, MA, 2015.

(168) Foston, M.; Ragauskas, A. J. *Energy and Fuels* **2010**, *24*, 5677–5685.

(169) Pu, Y.; Hu, F.; Huang, F.; Davison, B. H.; Ragauskas, A. J. *Biotechnology for Biofuels* **2013**, *6*, 1–13.

(170) Behera, S.; Arora, R.; Nandhagopal, N.; Kumar, S. *Renewable and Sustainable Energy Reviews* **2014**, *36*, 91–106.

(171) Van Dyk, J. S.; Pletschke, B. I. *Biotechnology Advances* **2012**, *30*, 1458–1480.

(172) Novy, V.; Nielsen, F.; Olsson, J.; Kevin, A.; Saddler, J. N.; Wallberg, O. **2020**.

(173) Roberts, K.; Lavenson, D.; Tozzi, E.; McCarthy, M.; Jeoh, T. *Cellulose* **2017**, *18*, 759–773.

(174) Kristensen, J. B.; Felby, C.; Jørgensen, H. *Biotechnology for biofuels* **2009**, *2*, 11.

(175) Du, J.; Li, Y.; Zhang, H.; Zheng, H.; Huang, H. *Cellulose* **2014**, *21*, 2409–2417.

(176) Jorgensen, H.; Vibe-Pedersen, J.; Larsen, J.; Felby, C. *Biotechnology and Bioengineering* **2007**, *96*, 862–870.

(177) Arantes, V.; Saddler, J. N. *Biotechnology for Biofuels* **2011**, *4*, 1–16.

(178) Vidal, B. C.; Dien, B. S.; Ting, K. C.; Singh, V. *Applied Biochemistry and Biotechnology* **2011**, *164*, 1405–1421.

(179) Davis, M. E.; Davis, R. J., *Fundamentals of Chemical Reaction Engineering*, 2013, pp 184–239.

(180) Nill, J.; Jeoh, T. *bioRxiv* **2019**, 691071.

(181) Bansal, P.; Hall, M.; Realff, M. J.; Lee, J. H.; Bommarius, A. S. *Biotechnology Advances* **2009**, *27*, 833–848.

(182) Huron, M.; Hudebine, D.; Lopes Ferreira, N.; Lachenal, D. *Biotechnology and Bioengineering* **2016**, *113*, 1011–1023.

(183) Gao, D.; Haarmeyer, C.; Balan, V.; Whitehead, T. A.; Dale, B. E.; Chundawat, S. P. *Biotechnology for Biofuels* **2014**, *7*, 1–13.

(184) Zanchetta, A.; dos Santos, A. C. F.; Ximenes, E.; da Costa Carreira Nunes, C.; Boscolo, M.; Gomes, E.; Ladisch, M. R. *Bioresource Technology* **2018**, *252*, 143–149.

(185) Luterbacher, J. S.; Parlange, J. Y.; Walker, L. P. *Biotechnology and Bioengineering* **2013**, *110*, 127–136.

(186) Seiboth, B.; Verena, C. I.; Seibot, S.-. *Biotechnology for biofuels* **2011**, *6*, 127.

(187) Sluiter, a.; Hames, B.; Ruiz, R.; Scarlata, C.; Sluiter, J.; Templeton, D.; Crocker, D. *Laboratory Analytical Procedure (LAP)* **2012**, 17.

(188) Shuai, L.; Luterbacher, J. *ChemSusChem* **2016**, *9*, 133–155.

(189) M., B. *Anal Biochem* **1976**, *72*, 248–54.

(190) Gama, F. M.; Texeira, J. A.; Mota, M. *Biotechnology and Bioengineering* **1994**, *43*, 381–387.

(191) Neuman, R. P.; Walker, L. P. *Biotechnology and Bioengineering* **1992**, *40*, 218–225.