Thomas Zahel, Bsc

In Silico Visualistation: Modelling Cellulase Activity on a Nanoscale

MASTERARBEIT

zur Erlangung des akademischen Grades eines

Diplom-Ingenieur

in

Biotechnologie

[gemeinsames Studium im Rahmen von NAWI Graz]

eingereicht an der

Technischen Universität Graz

Betreuer:

Univ.-Prof. Dipl.-Ing. Dr.techn., Bernd Nidetzky

Institut für Biotechnologie und Bioprozesstechnik

Graz, Juli 2013

Abstract

An object orientated model of enzymatic cellulose hydrolysis has been developed. The behaviour of single enzymes (EG, CBH I and CBH II), irreversibly bound to the cellulose surface, was modelled on a nanoscale, applying enzyme velocities recently measured by atomic force microscopy (AFM). Cellulose was implemented as an amorphous–crystalline mixed substrate containing microcrystals embedded by an amorphous matrix. For the first time, it was possible to compare a nanoscale bottom–up approach to cellulose degradation patterns on the nanometre scale. Moreover, it was possible to reproduce reported cellulase synergism factors and specific enzyme activities. This was feasible due to recently available AFM data paired with reported hydrolysis results from literature. In detail, it was possible to explain the slow emerging and fast degradation of microcrystals out of the amorphous matrix, recently discovered by AFM measurements. This phenomenon could be ascribed to the relatively inefficient microcrystal degradation of CBH I due to high crowding of CBH I on the crystalline parts. The subsequent uncovered amorphous region underneath the crystal could be rapidly degraded by EG and CBH II due to high specific surface exposure of this amorphous part. Moreover, the model could reproduce synergistic work of exo– and endoglucanases, where it was found that the binary combination of CBH I and CBH II gave the largest synergism compared to all other binary enzyme combinations. The modelled synergism factors in respect to the relative mixing ratios of the binary combinations gave excellent agreement to reported literature data. Furthermore, the bottom–up approach revealed reasonable mean specific activities for surface acting cellulases.

Kurzfassung

Es wurde ein objektorientiertes Modell zur Beschreibung der enzymatischen Cellulose–Hydrolyse entwickelt. Das Verhalten von einzelnen Enzymen (EG, CBH I und CBH II), die irreversibel an die Celluloseoberfläche gebunden sind, wurde auf Nano–Ebene simuliert, wobei Enzymgeschwindigkeiten durch Resultate von AFM–Studien (atomic force microscopy) zugänglich gemacht wurden. Cellulose wurde als Substrat mit amorphen und kristallinen Teilbereichen implementiert. Zum ersten Mal war es möglich, ein phänomenologisches "bottom–up"–Modell sowohl mit Abbaumustern auf Nano–Ebene, enzymatischen Synergismen und biochemischen Parametern abzugleichen. Dies wurde durch den Vergleich von AFM–Daten und Hydrolysedaten aus verschiedenen Literaturquellen mit dem Modell möglich gemacht. Auf Nano–Ebene war es möglich, das in AFM–Studien beobachtete Heraustreten und den raschen Abbau von mikrokristallinen Strukturen, im Vergleich zur umgebenden amorphen Matrix, zu erklären. Dieses Phänomen konnte durch den relativ ineffizienten Abbau von kristallinen Strukturen durch die CBH I erklärt werden, da die CBH I durch Zusammenstöße während der prozessiven Arbeit in ihrer Aktivität limitiert wird. Die unter dem Kristall liegende amorphe Cellulose wird rasch durch EG und CBH II abgebaut, da sie eine relativ hohe spezifische Oberfläche aufweist. Darüber hinaus konnte das Modell den Synergismus zwischen Zweierkombinationen der oberflächenaktiven Enzyme reproduzieren, wobei die Kombination von CBH I und CBH II den größten Synergismus zeigte. Die Synergismus–Faktoren der Zweierkombinationen in Bezug auf das Mischungsverhältnis beider Enzymkomponenten ergaben ausgezeichnete Übereinstimmung mit früheren experimentellen Studien. Des Weiteren zeigte das Modell realistische durchschnittliche spezifische Enzymaktivitäten für die oberflächenaktiven Enzyme.

Danksagung

Ein großes Danke zuerst an Professor Bernd Nidetzky, der mich durch diese Arbeit geleitet hat. Der zeitliche Druck war auch für ihn sicher nicht immer einfach.

Immer ein offenes Ohr fand ich bei Manuel, der mich bei der Erstellung der Arbeit großartig unterstützt hat und immer Zeit gefunden hat auf meine Fragen einzugehen.

Einen sehr aufmerksamen Zuhörer habe ich in Dr. Winfried Kernbichler gefunden, der immer für ein außerordentlich herzliches Gesprächsklima sorgt.

Danke auch an meine Eltern, die mir mein Studium überhaupt ermöglicht und mich immer emotional unterstützt haben, egal, wie schwierig die Lage erschienen ist. Ganz besonders möchte ich meinem Vater dafür danken, dass er die Kraft gefunden hat, mich während meiner Diplomarbeit in Ruhe arbeiten zu lassen.

Der größte Dank geht an Christa, die der wichtigste Mensch und Ansprechpartner in dieser schwierigen Zeit war und mich immer wieder mit ihrer Liebe gestärkt hat.

Davon ganz abgesehen wäre ohne Christa meine Arbneit wahrscheinlih um einige rechtschreibfeler Reicher und einige Beistriche ärmer.

Table of Contents

1. Introduction 1
2. Biological and Computational Fundamentals 5
2.1 Cellulases 5
2.2 Object oriented programming 6
2.3 Atomic force microscopy 7
3. Model Development 8
3.1 Modelling cellulose 8
3.2 Modelling cellulases 13
4. Methods 21
4.1 Computational methods 21
4.2 Parameter sensitivity analysis (PSA) 21
4.3 AFM data 22
4.4 Virtual synergism experiments 23
5. Results 25
5.1 Reproducing AFM observations – validation on the nanoscale 25
5.2 Increasing enzyme surface coverage leads to reduced specific microcrystal degradation 31
5.3 Derive biochemical parameters from a bottom–up–approach 33
5.4 Elucidating influencing input parameters 37
6. Discussion 44
7. Conclusion 49
8. Outlook 51
9. References 53
Attachment 57

1. Introduction

In times of global energy shortage, cellulose based energy carriers seem to be ideal for satisfying our demand for renewable non–oil based energy. Cellulose is the most abundant carbon source on earth, and enzymatic degradation of cellulose can be used for the production of second generation biofuels, containing mainly ethanol as an energy source. Only by improving the enzymatic hydrolysis, which is the key bottleneck of ethanol production from cellulose, it will be possible to compete economically with petroleum gasoline production (Lynd et al., 2008). Cost competitive technology can be achieved by improving the enzymatic machinery and also by making the substrate more susceptible to hydrolysis. In order to reach this goal, it is necessary to understand the complex process of different enzymes working on the surface of a heterogeneous substrate, and to identify and quantify the impact of different system variables on the hydrolysis process (Bansal et al., 2009).

Cellulose is a linear homopolymer of β–1–4 linked D-glucose units. The degree of polymerisation (DP ~ 10000 – 15000 glucose units) depends on the cellulose source material and can be dramatically reduced by mechanical, physical or chemical pre-treatment (DP < 1000) (O'sullivan, 1997; Zhang and Lynd, 2004). Due to intrachain hydrogen bonding between the hydroxyl group (3') and oxygen (5') of the neighbouring glucose molecule, the cellulose chain configuration becomes rather linear (Moon et al., 2011). Interchain hydrogen bonds of glucose molecules of different chains promote stacking of multiple cellulose chains and the formation of elementary fibrils, subsequently forming larger microfibrils (5 - 50 nm diameter). Moreover, these intra- and interchain hydrogen bonds result in various ordered crystalline arrangements of microfibrils with different crystal unit cells (allomorphs) and crystal packing. Thereby many β–glycosidic bonds are hidden within the crystal and not accessible to enzymatic hydrolysis. Naturally occurring crystalline cellulose consists mainly of the allomorph cellulose I, where cellulose chains are stacked in parallel. Besides crystalline cellulose, amorphous cellulose – which is disordered and generally more easily degradable – is present in nearly all natural cellulose sources. Crystalline parts can be transformed into amorphous parts by pre–treatment methods, such as phosphoric acid or ionic liquid treatment. Due to re–crystallisation processes, from amorphous to crystalline cellulose, the allomorph cellulose II can be formed, where cellulose chains are stacked anti–parallel. Due to the altering content

of amorphous parts, crystalline parts, crystalline allomorphs, accessible β glycosidic bonds and various degrees of polymerisation, cellulose is regarded as a heterogeneous substrate (O'sullivan, 1997). As a consequence of this heterogeneity, cellulose is recalcitrant to hydrolysis because β–glycosidic bonds have to be made accessible prior to hydrolysis.

Enzymes acting on the cellulose surface can be divided into three main groups according to their mode of action. Endoglucanases (EG) are known to randomly cut β–1-4 glycosidic bonds of a cellulose chain, cellobiohydrolases I and II (CBH I and CBH II) are known to degrade cellulose from its reducing and non reducing ends, respectively (Bansal et al., 2009; Himmel et al., 2007; Zhang and Lynd, 2004). Therefore, the enzyme efficiency is strongly dependent on the local availability of appropriate reaction sites. Moreover, cellobiohydrolases differ from EG, since they perform a processive movement along a cellulose chain, which enables them to cleave multiple cellobiose molecules, adjoining each other, without detaching from the chain. In contrast to that, EG are able to cleave one cellulose chain into two shorter chains and thereby create one additional reducing and non–reducing end. As a consequence of this mode of action, EG generate appropriate reaction sites for cellobiohydrolases, resulting in the so called "exo–endo synergism". The binding of the enzymes onto the cellulose surface occurs via a cellulose binding module (CBM), which is connected by a linker to a catalytic domain (CD). Besides its binding function, the CBM has a loosening activity onto the hydrogen bonding network, resulting in a disruption of elementary fibrils, called amorphogenesis (Arantes and Saddler, 2010; Boraston et al., 2004).

Previous experimental work was able to shed light on several different aspects of the hydrolysis, such as enzyme adsorption (Linder and Teeri, 1996; Maurer et al., 2012; Moran-Mirabal et al., 2011; Nidetzky et al., 1994a), mechanistic and structural information (Abuja et al., 1988; Divne et al., 1994; Varrot et al., 2003), mode of enzymatic action (Himmel et al., 2007; Horn et al., 2012), up to nanoscale observations of the degradation (Bubner et al., 2012; Bubner et al., 2013; Ganner et al., 2012; Igarashi et al., 2011). Intensive recent research was able to increase the knowledge about the enzymatic mode of action and even show the movement of single cellobiohydrolases along a cellulose chain by atomic force microscopy (AFM) revealing single enzyme velocities (Igarashi et al., 2009; Igarashi et al., 2011).

Despite the extensive experimental research over the past four decades, many questions remain unresolved, most prominently the reason for the drop of the enzymatic hydrolysis rates at higher degrees of conversion, called rate retardation. Therefore, new modelling concepts are regarded as a key tool for elucidating rate retarding factors (Bansal et al., 2009).

Previous modelling attempts focused on the one hand on the hydrolytic reaction as a holistic process, and on the other hand on the enzyme–substrate interaction on a atomic scale (Beckham et al., 2011; Beckham et al., 2010; Chundawat et al., 2011; Dowd et al., 1992; Matthews et al., 2006; Mazeau and Heux, 2003). Giving an example for the latter one, the Gibbs free energy stages of CBMs were investigated using force field measurements (Beckham et al., 2010). For capturing the whole hydrolytic process, empirical models (Scheiding et al., 1984; Suga et al., 1975), which do not give deeper insight, and mechanistic models were applied. Mechanistic approaches can be further divided into two subgroups: (i) deterministic models using differential equations taking chain length into account, which results in a huge amount of ordinary differential equations (Levine et al., 2010; Zhou et al., 2009), and (ii) phenomenological–statistic models using cellular automata or object oriented programming (Kumar and Murthy, 2013; Warden et al., 2011). A major drawback of deterministic models is that local arrangements of crystalline and amorphous parts cannot be taken into account, because global concentration terms for cellulose chain lengths and enzymes are applied. The lack of a description of this heterogeneous substrate emphasises the need for a more detailed modelling of the hydrolysis process, which should be achieved by covering more substrate and enzyme related properties, such as different degrees of crystallinity and diffusive velocities of enzymes on the surface, respectively (Bansal et al., 2009; Zhang and Lynd, 2004). Object oriented models provide a solution for this task because they are able to reconstruct spatial separated crystalline and amorphous substrate parts paired with multi enzyme interactions, such as the process of enzymes bumping into each other.

Object oriented programming was recently successfully applied to virtually create a three dimensional substrate containing properties describing its heterogeneity on the nanometre scale (Kumar and Murthy, 2013; Warden et al., 2011). Moreover, enzymes (objects of the system) were implemented by assigning

mechanistic, kinetic and diffusion features. The previous object oriented reaction–diffusion systems were validated against hydrolysis studies from literature. However, the output of hydrolysis studies, a two dimensional plot of produced reducing sugars versus time, conceals a lot of the actual information present in object oriented systems. Thereby the complex three dimensional model output is lost and only a small fraction of the prevalent information can be used for validation. Therefore, object oriented models with nanoscale resolution require visualisation methods capturing nanoscale processes for validation.

In situ liquid AFM is a perfect analytical tool for capturing cellulase action in nano dimensions on insoluble cellulose surfaces (Bubner et al., 2012; Bubner et al., 2013; Ganner et al., 2012; Igarashi et al., 2011). This opens up a completely new possibility of comparing a three dimensional model output to observed enzyme velocities and nanoscale degradation patterns. This reveals not only mathematical descriptions which fit observations, like Michaelis–Menten concepts, but also gives a phenomenological insight into enzymes and substrates, which could only be indirectly addressed in previous models by comparing model outputs to hydrolysis experiments.

As a novelty of this study, we use AFM surface images for validating the visual model output. Therefore, a special object oriented set up was chosen, where one can describe the substrate and all its related features (accurate description of cellulose chains in amorphous cellulose, detailed crystal structure and spatial orientation), as well as the enzymatic properties (single enzyme description, enzyme diffusion, different mode of action of CBHs and EGs). Thereby it is possible to assign physical (Abuja et al., 1988) and biochemical properties (Igarashi et al., 2011; Jervis et al., 1997; Moran-Mirabal et al., 2013) of the enzyme and the substrate (Ganner et al., 2012) to certain classes of the system.

2. Biological and Computational Fundamentals

2.1 Cellulases

Cellulases are hydrolytic enzymes, which degrade cellulose into its monomer, D–glucose. Due to the heterogeneity of the substrate, different enzymes with varying modes of actions are required to degrade cellulose in a concerted manner. Common sources of cellulases are several fungi (most prominently *Trichoderma sp.*) and anaerobic bacteria (*Clostridium sp.*). Fungal cellulases are single non–complexed enzymes, whereas bacterial cellulases are organized in complexed form called cellulosomes. Biotechnological applications for cellulases range from food, textile, and renewable energy biotechnology (Bhat, 2000). In this study, I will only focus on the fungal non–complexed cellulases. The potential power of cellulases to degrade cellulose to glucose, which is a substrate for bio–ethanol production, has been investigated over the past four decades. Regarding their modes of action and substrate specificity, cellulases working on the surface can be divided into three main groups.

Endoglucanases (EG)

EG are able to cut β–1–4 glycosidic bonds within or at the end of a chain. Cuts in the cellulose chain produce one reducing and one non-reducing end (Himmel et al., 2007). EG do not necessarily release a cellobiose molecule by a single cutting event, unless the chain produced is shorter in length than soluble cellulose chains (DP < 4 - 7 glucose units) (Klemm et al., 2004; Stålbrand et al., 1998; Zhang and Lynd, 2005). This makes it rather difficult to determine the reaction duration of a single cutting event. Most EG perform a non processive movement, which means that a cutting event is followed by a random diffusion on the cellulose surface until the next cutting event occurs. EG show a predominant activity on amorphous cellulose and only minor activity on highly crystalline cellulosic substrates (Hoshino et al., 1997; Zhang and Lynd, 2004).

Cellobiohydrolases I and II (CBH I and CBH II)

The molecular structure of cellobiohydrolases differs from the structure of EG, since CBH I and CBH II have a molecular tunnel (formed by four loops in the case of CBH I and two loops in the case of CBH II), through which the cellulose chain can be

threaded through (Divne et al., 1994; Zhang and Lynd, 2004). The tunnel leads to the active site, where in general after each second glucose molecule a cutting event is initiated. The binding of a cellulose chain via the tunnel leads to a strong adsorption onto the cellulose chain and allows cellobiohydrolases to perform a sliding (processive movement) along the chain after a cellobiose molecule is cleaved off (Horn et al., 2012; Moran-Mirabal et al., 2011). Moreover, this enables cellobiohydrolases to release multiple cellobiose molecules until they detach from the cellulose chain. The strong binding caused by the processive movement also shows drawbacks, since it is suggested that cellobiohydrolases are stuck to the cellulose chain after crashing into an obstacle (another cellulose chain or other enzymes). CBH I attacks cellulose from its reducing end, whereas CBH II degrades from the non–reducing end. Especially CBH I, but also CBH II, shows higher activity on highly crystalline cellulose (Avicel) compared to EG. However, recent AFM measurements revealed that CBH II has no direct hydrolysis effect on purely crystalline cellulose (*Valonia*) but seems to prepare the crystalline parts for CBH I attack (Ganner et al., 2012; Igarashi et al., 2011).

2.2 Object oriented programming

Object oriented programming consists of classes, to which certain properties and functions/methods can be assigned. Subclasses of parent classes can inherit these functions and properties, and moreover, can add their own specific functions and properties. In the case of cellulases, one can imagine a parent class "enzyme" and several subclasses, such as EG, CBH I and CBH II. Furthermore, cellobiose molecules can be regarded as another class with completely different properties and functions. Objects (instances) of a class have the same functions and properties but may differ in the specific values of their function and properties. Data encapsulation is another important aspect of object oriented programming, where some properties are defined as "private". Thereby it is impossible for other objects to read or edit these properties from outside. One "private" property for example is the reaction duration, the time span until the reaction of an enzyme is finished. Only the enzyme itself "knows" how long it still has to wait until the reaction is finished and it can continue diffusing over the cellulose surface. This leads to a very detailed and realistic representation, which comes along with several limitations in computational time since each molecule has to be calculated separately. Furthermore, object oriented

programming can be divided into discrete approaches, applying discrete positions for objects in space, or continuous approaches. These approaches can also be mixed with some discrete features and some continuous features. Moreover, object oriented programming approaches can be divided according to their spatial resolution. Studies investigating atomic interactions (atomic scale models) differ from studies considering whole molecules (mesoscale models). This terminology arises from a modelling point of view. In contrast to this, microscopy, which is used in this study for data generation, uses the term "nanoscale", indicating spatial resolution on the nanometre scale, and "mesoscale", indicating spatial resolution of several nanometres (~ 50 - 100 nm). Since we are interested in the movement and reaction of whole molecules (cellobiose MW ~ 180 and cellulases ~ 55 kDa) and not atoms, which have dimensions in the nanometre scale (cellobiose ~ 1 nm and cellulases ~ 5 – 7 nm), it is a mesoscale model with nanoscale resolution. However, I want to state that these two terms come from two different research fields, namely modelling and microscopy. For clarity reasons and due to the fact that the model output is compared to microscopy data, only the microscopy terminology will be used from now on.

2.3 Atomic force microscopy

AFM is a powerful tool for investigating height differences on a surface. A sharp tip with a width of several nanometres rasters over the sample surface measuring the height profile. As common for biological samples, tapping mode as the standard AFM mode (García, 2010) is used, which ensures that the sample is not mechanically damaged. High resolution imaging can provide spatial resolution of 1 nm (Bubner et al., 2013). Moreover, time resolutions of 1 min with standard AFM and 300 ms with high speed (HS) AFM can be achieved (Igarashi et al., 2009)).

3. Model Development

Object oriented programming was used for implementing cellulose and cellulases in a bottom–up–approach in Matlab as described in the following section.

3.1 Modelling cellulose

Due to the fact that time resolved AFM height degradation profiles from the study of Ganner et al. (2012) are used for validation of the model, it is necessary to virtually reconstruct this substrate. It consists of amorphous and crystalline parts, which can be further divided into larger (up to 10 µm) and smaller crystals (10 nm x 100 nm), called microcrystals. A schematic representation superimposed with a real AFM image from the study of Ganner et al. (2012) is shown in Figure 1.

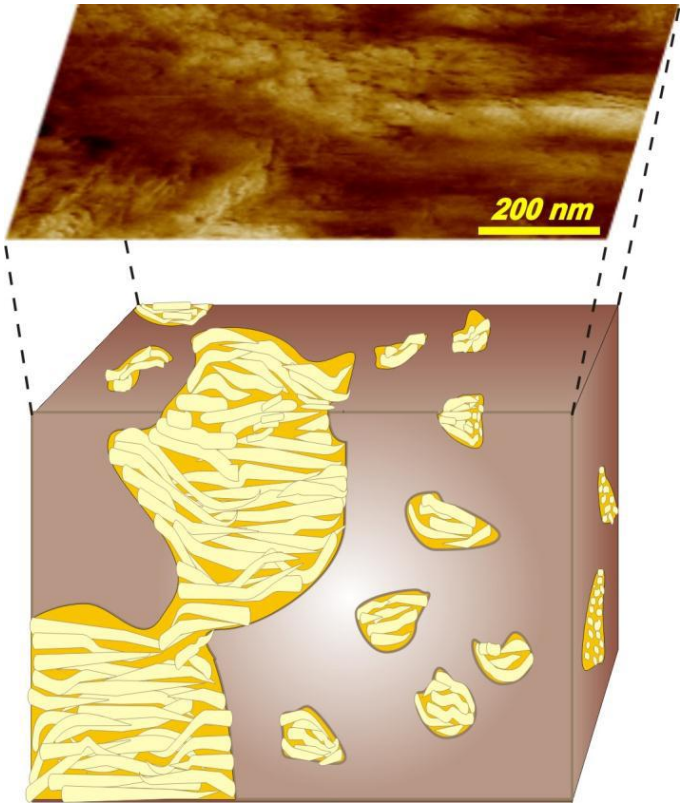

Figure 1: Schematic representation of the cellulosic amorphous–crystalline mixed substrate of Ganner et al. (2012). The substrate contains larger crystals (up to 10 µm) and smaller microcrystals (10 nm x 100 nm). Picture is taken from the supplemental information of Ganner et al. (2012).

The crystalline part was modelled as microcrystals, which were observed in the AFM measurements of Ganner et al. (2012). Microcrystals have a defined shape of about 10 nm in width and 100-300nm in length and are thought to be a recrystallisation product (cellulose II) from the substrate preparation method used in this study (Ganner et al., 2012). This implies the anti parallel orientation of crystalline cellulose

chains, which was accounted for in the model. This was done by orientating all crystalline cellulose chains of one z plane in the opposite direction of the cellulose chain of the above z plane. The modelled cellulose block was implemented as two huge three dimensional matrices. The "matter matrix" consists of values of "1" representing amorphous cellobiose molecules and "0" representing bulk liquid. Values higher than 1 indicate a microcrystalline organized cellobiose molecule (Figure 2). The spacing of the matrix is set to 1 nm in each dimension. Although cellobiose molecules have a rectangular two dimensional shape (1.04 nm in length and 0.53 nm in width), the gap between cellulose chains is rather large due to hydrogen bridges, which makes the representation of a cellobiose molecule with 1nm x 1nm realistic (Zhang and Lynd, 2004). Due to the fact that cellulose chains are spatially orientated, a second matrix describing this substrate feature is necessary ("Orientation matrix" in Figure 2). This matrix has the same size as the "matter matrix" containing values of "1" and "-1" for cellulose chains aligned in x and -x direction, respectively and "2" and "-2" for cellulose chain aligned in y and -y direction, respectively.

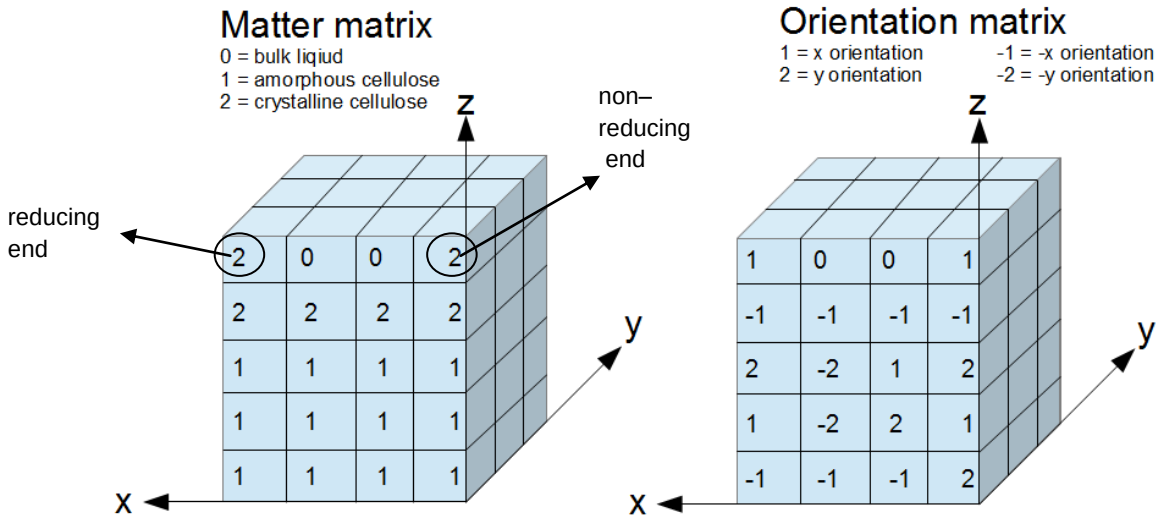

Figure 2: Two matrices describing properties of the cellulose substrate. The scheme shows a two layered crystal with only two cellobiose units left in the top layer resulting in one reducing and one non reducing end. Beneath this crystal, amorphous cellulose is present as shown in the "Matter matrix". The first layer of the crystal is orientated in x direction, the second layer in -x direction as shown in the "Orientation matrix". This implements the anti parallel crystal conformation. The amorphous cellulose is randomly orientated (details described in the model development section and Figure 5).

The internal structure of microcrystal is not homogenous because Avicel has an average degree of polymerisation (DP) of about 300 glucose units (about 150 nm) as reported in Ogeda et al. (2012), and microcrystals observed in AFM experiments are up to 300 nm in length. This means that larger microcrystalline structures in the substrate are built of some overlaid chains, which create internal chain ends, and the ends of the microcrystal may not be flat but rather frayed as shown in Figure 3. Microcrystals used for all simulations, unless stated otherwise, were implemented cylindrical with a radius of 10 nm and a length of 100 nm, which were standard dimensions for microcrystals reported in Ganner et al. (2012). The common microcrystal dimension used for the simulations is shown in Figure 4. The crystal was completely embedded by amorphous cellulose (up to 90 nm z level of Figure 4). This calculates to 15 % crystalline cellulose in the first 20 nm of vertical degradation and is in line with the predominant amorphous substrate structure of the amorphous–crystalline mixed substrate used in the studies of Ganner et al. (2012) and Bubner et al. (2012).

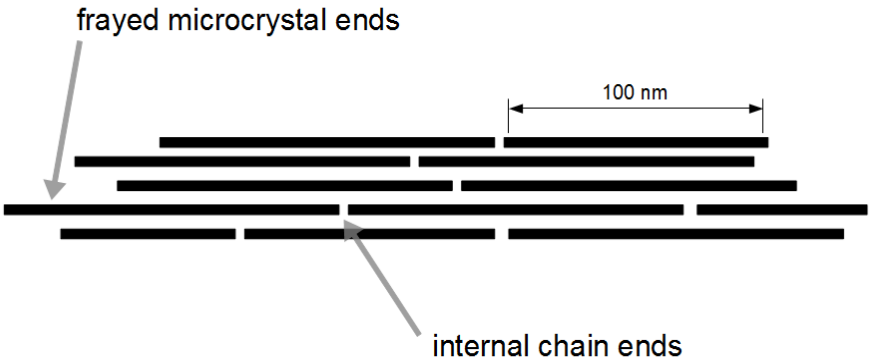

Figure 3: Microcrystal representation. Frayed ends and internal chain ends are shown.

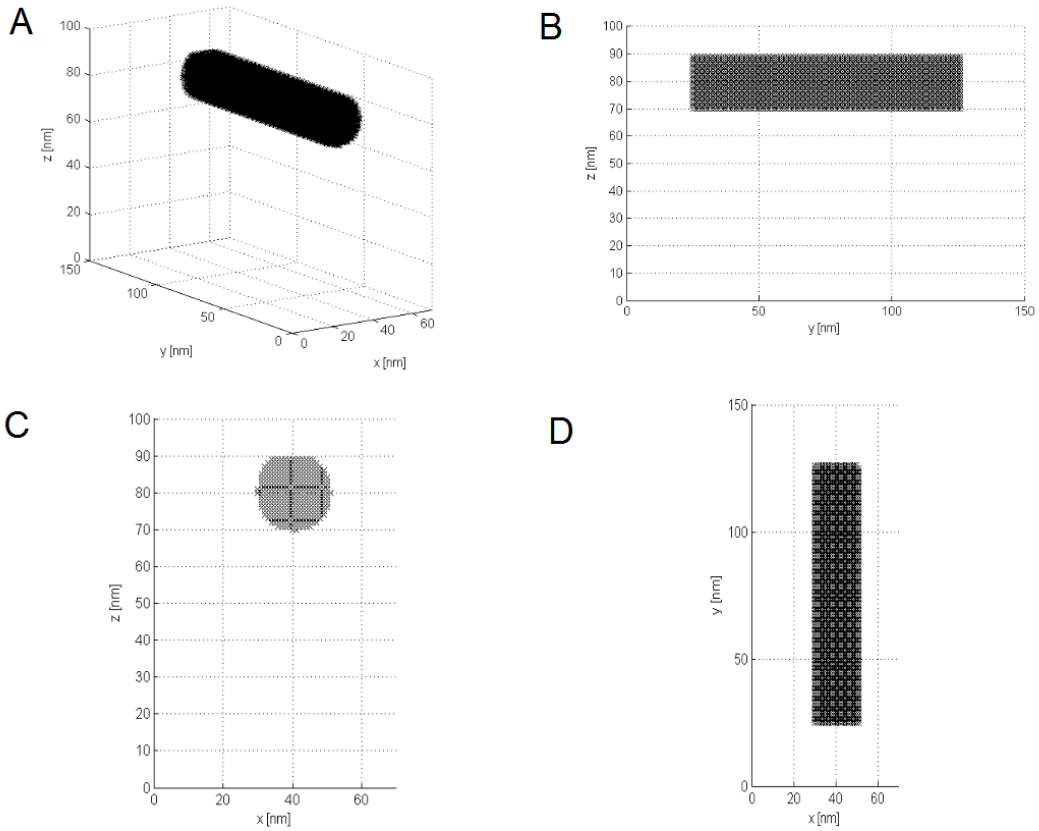

Figure 4: Microcrystal dimension used for simulations of all result sections unless stated otherwise. The surrounding of the crystal was filled with amorphous cellulose up to 90 nm. A: 3D representation of the microcrystal, B: y–z plane, C: x–z plane D: x–y plane. Frayed crystal ends are not shown for clarity reasons. For all simulations the crystal was completely embedded by amorphous cellulose.

For amorphous cellulose, a random direction of cellulose chains is proposed. Considering one z plane, 1/3 of the amorphous cellulose chains point into x direction, 1/3 into y direction and 1/3 into z direction. Although the length of an amorphous cellulose chain segment is not known until it changes its direction and cannot be fully reproduced by molecular modelling, it was assumed that statistically five cellobiose molecules were aligned consecutively (Dowd et al., 1992; Mazeau and Heux, 2003). A parameter sensitivity analysis, performed prior to choosing this value, revealed that the actual number of cellobiose molecules following each other did not affect the model output significantly (see Figure 24). Therefore, each single z plane was filled with cellulose chain segments of five cellobiose units pointing into the same direction (1/3 in x and -x, 1/3 in y and -y direction). The ends of these segments were initially not regarded as free ends for enzymatic attack, and can only be attacked by CBH II

after EG degrades a cellobiose molecule next to the segment end. The residual 1/3 sites were filled with single cellobiose pieces pointing in either x, -x, y or -y direction, representing the amorphous cellulose chains pointing in z direction. Avicel, the original source of the applied amorphous–crystalline mixed substrate, has a degree of polymerisation of about 150 cellobiose molecules (Ogeda et al., 2012). This leads to chain ends within the amorphous matrix, which can be implemented giving each amorphous cellobiose molecule a chance of 1/150 to be a reducing or non reducing end, respectively. A representation of one z plane of the amorphous cellulose is shown in Figure 5.

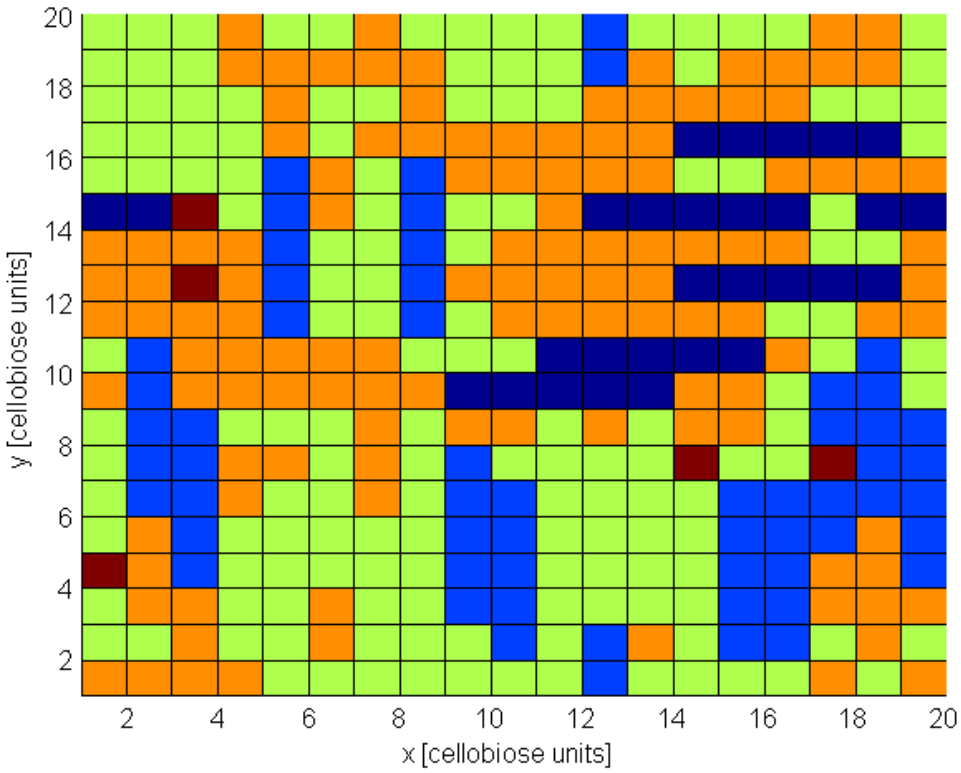

Figure 5: Amorphous matrix representation of one z plane (20x20 cellobiose shown). Orange fields: amorphous cellulose oriented in x direction. Dark blue fields: amorphous cellulose oriented in -x direction (exchanged reducing and non reducing ends). Green fields: amorphous cellulose oriented in y direction. Bright blue fields: amorphous cellulose oriented in- y direction (exchanged reducing and non reducing ends). Dark red fields (only 5 fields visible): free cellulose chain ends due to the degree of polymerisation of Avicel of 150 cellobiose units.

In general, the representation of amorphous and crystalline cellulose made it possible to design each cellobiose molecule separately concerning its degree of crystallinity and its orientation, which was a limiting factor in former modelling approaches. A typical size of the whole matrices used for simulations is 70 nm x 150 nm.

3.2 Modelling cellulases

In line with previous modelling studies (Jalak and Väljamäe, 2010; Kumar and Murthy, 2013; Levine et al., 2010; Warden et al., 2011; Zhou et al., 2009), we have considered three main surface acting enzymes produced by the fungal systems of *Trichoderma sp.*: endoglucanases (EG, e.g. *Tr*Cel7B), cellobiohydrolases I (CBH I, e.g. *Tr*Cel7A) and cellobiohydrolases II (CBH II, e.g. *Tr*Cel6B). This set of the cellulases of the fungal system of *Trichoderma sp.* was chosen because model validation was performed on data gained by Ganner et al. (2012), who used the same set of enzymes.

In the program, a class of enzymes is defined with different properties and methods (Figure 6). The parent class "enzyme" can be divided into three subclasses according to the three main enzymes acting on the cellulose surface (CBH I, CBH II and EG). A set of enzyme properties and function is defined (Figure 6). Properties are the position of the enzyme ([x, y, z]), the size of the whole enzyme, the size of the CBM, the mode of enzymatic action (processivity or endo–activity), the residual waiting duration (time until an enzyme collision is dissolved), the residual reaction time (time until the reaction is finished), the moving threshold (number of cellobiose molecules, which can be hiked through), and the residual processive period (number of cellobiose molecules, which can be processed until the processive motion is stopped). The functions of the enzymes can read and modify all properties. The functions *slide*, which contains a random walk algorithm, and *react*, which reduces the residual reaction duration by one Δt, are defined for all enzymes. In contrast to this, the function *processive_motion_CBH,* which ensures that cellobiohydrolases always degrade one cellobiose molecule within one reaction duration, is only defined for cellobiohydrolases. A detailed description of the functions is given in section 4.1 Computational methods) and the appendix. The initiation of the functions is handled in the *main* script of the program, which initiates the surface diffusion, reaction or the processive motion of the enzymes when a suitable reaction partner is present. I want to state that for family 6 cellobiohydrolases (CBH II or *Tr*Cel6B) endo–activity was proposed, because of the observation that the molecular tunnel responsible for the processivity occasionally opens up (Varrot et al., 1999; Varrot et al., 2003). This endo–activity was neglected for the first modelling attempt since a general predominant exo–activity is assumed in various publications (Bansal et al., 2009; Himmel et al., 2007; Zhang and Lynd, 2004). In the model, EG cut a single cellobiose

molecule out of the surface, which is not entirely realistic since it is known that EG only produce a chemical bound cleavage between two glucose molecules (Bansal et al., 2009; Zhang and Lynd, 2004). However, this assumption was not regarded as limiting, since the action of EG (especially the modelled *Tr*Cel7B) leads mainly to soluble cellobiose in the supernatant, and the main purpose of EG is the generation of free chain ends, which is also performed by this modelling approach (Bansal et al., 2009; Himmel et al., 2007; Karlsson et al., 2002). By the action of either processive (CBH I or CBH II) or non processive enzymes (EG) cellulose chains with a DP < 6 glucose molecules might be generated. These chains are regarded as soluble, which was shown experimentally (Klemm et al., 2004; Stålbrand et al., 1998; Zhang and Lynd, 2005). For all simulations, enzyme ratios of 60 % CBH I, 20 % CBH II and 20 % EG of the whole enzyme mix were applied, according to complete fungal cellulase systems (e.g. *Trichoderma sp.*), as used in the work of Ganner et al. (2012), which was used for validation, and many others (Goyal et al., 1991).

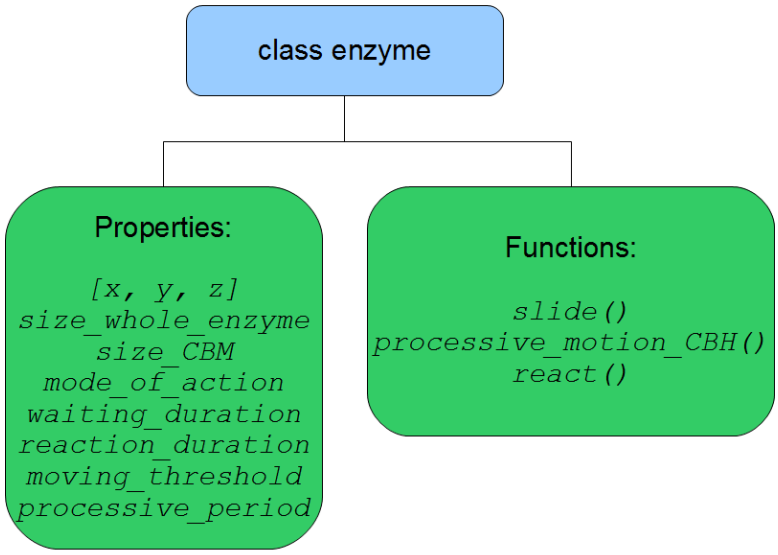

Figure 6: The parent class "enzyme" with its most prominent properties and functions. The code of the functions *slide,* handling the random walk, and *processive_motion_CBH,* required for the processive movement of CBH I and CBH II, are explained in detail with comments in the attachment. Properties: *[x, y, z]* defines the three dimensional position of the enzyme, *size_whole_enzyme* is given as 5 nm x 7 nm x 5 nm, *size_CBM* is the footprint of the CBM (3 nm x 3 nm), *mode_of_action* defines whether the enzyme is endo– or exo–acting. The *waiting_duration* is the duration two collided enzymes stay on the cellobiose chain during their processive motion until they stop their processive motion. The *reaction_duration* defines the time span for degrading one cellobiose molecule. The *moving_treshold* defines the number of cellobiose molecules, which can be hiked through by a random walk (see method section). The *processive_period* defines the maximal number of cellobiose molecules, which can be processed at a stretch.

The general stages of a surface acting enzyme (e.g. CBH I) are shown in Figure 7. Three main steps can be summarised, which will be consecutively elaborated in the next section: (1) the enzyme adsorption onto the cellulose surface, (2) a random walk (surface diffusion) and (3) the reaction (including reaction kinetics) with an appropriate reaction partner.

Adsorption studies of the past suggest irreversible adsorption of solely CBMs (Abuja et al., Jung et al. 2002), partially reversible adsorption (Maurer et al., 2012; Maurer et al., 2013; Moran-Mirabal et al., 2011), to fully reversible adsorption (Linder and Teeri, 1996). Fully reversible adsorption studies were performed on a very short time period (Linder and Teeri 1996), but AFM experiments used for validation of this study lasted for several hours. Partial reversible studies claim that irreversible adsorption is closely connected to the catalytic activity of cellulases. An increase in temperature decreased the reversible bound fraction of processive exo-cellulases, while the endoglucanase fraction stayed constant (Moran-Mirabal et al., 2011). Therefore, processively active cellobiohydrolases can be regarded as tightly bound to the surface. Studies of (Maurer et al., 2012) revealed recently that there is a large fraction of exo- and endo-cellulases irreversibly bound to the surface after a washoff on timescales >1h. A kinetic two step model was proposed in this study and fit to adsorption data resulting in rate constants for irreversible binding. Even after 30 min about 50 % of the enzymes were bound irreversible to the surface and after 1 h about 90 % could not be removed by wash off. In AFM experiments of (Bubner et al., 2012; Ganner et al., 2012)), which gave the data basis of this modelling approach, much longer hydrolysis durations (> 3h) were applied. Therefore a dominant large fraction of cellulases can be regarded as irreversibly bound and an irreversible one step adsorption was formulated for this model. Due to the fact that enzyme adsorption is a rather quick process compared to enzymatic action on an intermediate time scale of several hours, it was assumed that enzymes are already bound to the surface at the beginning of the model simulation (Nidetzky et al., 1994a). Once adsorbed, enzymes can diffuse on the surface according to a random walk and if a suitable reaction site is found, according to the mode of the enzyme, reaction will be initiated.

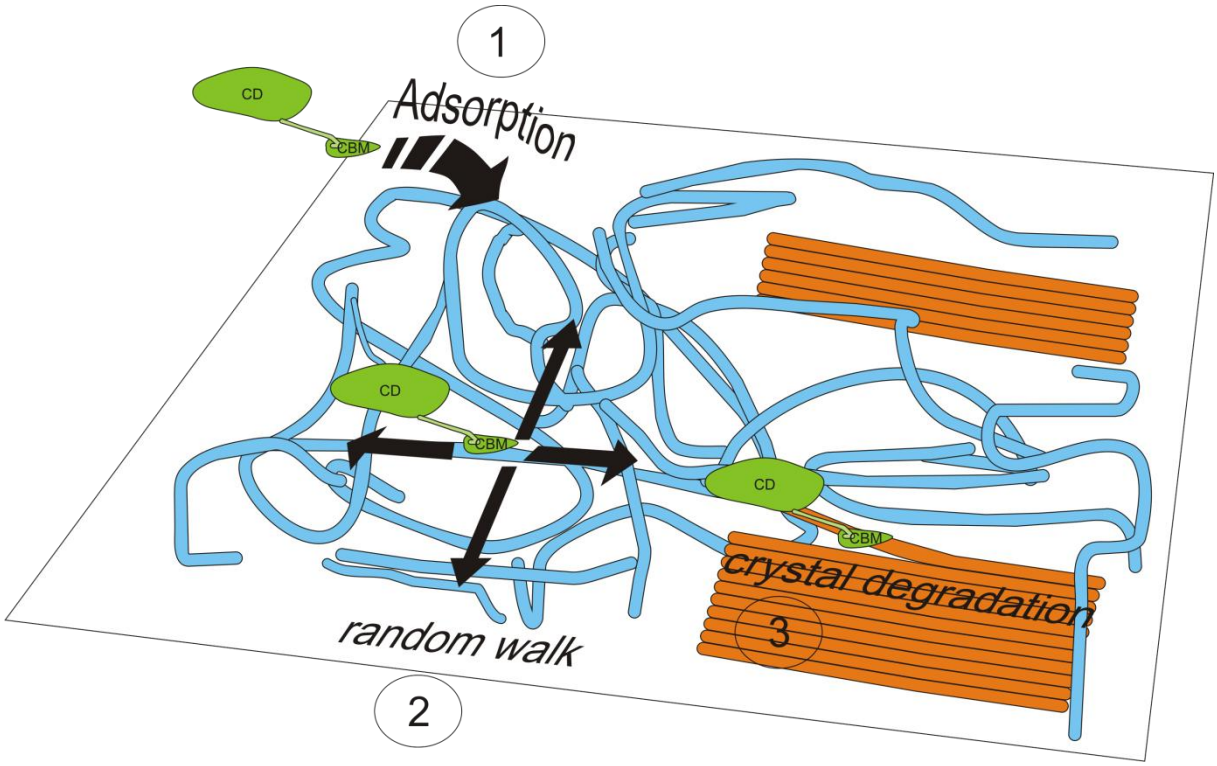

Figure 7: Two dimensional representation of three possible steps in the model for CBH I. Adsorption is followed by a random walk on the amorphous cellulose (blue chains) until a suitable reaction partner (crystalline cellulose – orange chains) is found. CD: catalytic domain, CBM: cellulose binding module of CBH I (green).

As shown by (Moran-Mirabal et al., 2013), the description of randomly sliding enzymes on the surface is close to the experimental findings. The sliding over the surface was modelled as a surface random walk, taking into account that physical barriers such as other cellulose chains and other enzymes cannot be hiked through. The direction of the random walk is the normal vector of a plane, which is checked for obstacles. The size of this wall is between 5 nm x 7 nm or 5 nm x 5 nm depending on the moving direction (see footprint considerations). It is allowed for enzymes to pass this wall unless a certain moving threshold is exceeded. This enables the program to implement the molecular dynamics of the cellulases, which is responsible for dodging a limited amount of cellobiose molecules. The moving threshold was set to 5 cellobiose molecules, which is between 15 and 20 % of the wall.

The minimal distance enzymes keep to each other is defined by their footprint. Due to the fact that sterical hindrances arise from the whole enzyme and not only from the CBM (Sugimoto et al., 2012), a footprint of 7 nm x 5 nm is assumed according to enzyme structure (Abuja et al., 1988). This footprint was also used in

other modelling approaches (Levine et al., 2010). The sterical exclusion due to the footprint is also taken into account for modelling the adsorption process, where newly adsorbed enzymes keep distance from already bound ones. Moreover, the model distinguishes between cellulases, which are processively active and the whole enzyme is bound to the surface, resulting in a footprint of 7 nm x 5 nm, and cellulases diffusing on the surface requiring a footprint of 3 nm x 3 nm (size of CBM). The total height of all cellulases was set to 5 nm, according to the cellulase diameter elucidated by SAXS experiments (Abuja et al., 1988).

Reaction probabilities of enzymes on specific substrate regions were defined according to Ganner et al. (2012), where CBH I has its major activity on microcrystalline cellulose, CBH II shows activity only on amorphous cellulose and EG shows main activity on amorphous cellulose. This is of special interest because the model contains the same amorphous–crystalline mixed substrate as used in the studies of Bubner et al. (2012) and Ganner et al. (2012) and was validated against the AFM observations of these publications.

Enzyme kinetic consideration: The reaction velocity of CBH I was measured recently to 7.1 ± 3.9 nm/s (at 25 ° C) by a well recognised paper of (Igarashi et al., 2011) on purely crystalline substrate (*Valonia* cellulose). This value could be confirmed by measurements of the diffusion speed of reacting enzymes by fluorescence single molecule tracking (Moran-Mirabal et al., 2013). This study revealed an effective diffusion coefficient for intact cellulases of 10^{-13} cm²/s. This calculates to 10 nm/s, which is in good agreement to values extracted by Igarashi et al. from high resolution AFM measurements. In the program the reaction duration (inverse of the reaction velocity) defines the time span between the start and the end of a reaction. This measured reaction velocity for CBH I (7.1 ± 3.9 nm/s) was also applied for CBH II, due to the lack of known molecular velocities of CBH II.

The activity of EG on the amorphous–crystalline mixed substrate (used for validation of this model) and on Avicel, a mainly crystalline cellulosic substrate, was measured by a DNS assay described elsewhere (Xiao et al., 2004). The activity for EG on Avicel resulted in a k_{cat} of 0.29 s^{-1} at 50 ° C (unpublished results). Interestingly, this activity could not be observed in terms of specific degradation of microcrystalline cellulose in AFM experiments (Ganner et al., 2012). This can be explained by the fact that Avicel is not purely crystalline (~ 10 - 30 % amorphous and paracrystalline

cellulose (Ioelovich et al., 2010)) and therefore the measured k_{cat} of 0.29 s^{-1} is the upper boundary for the apparent EG activity on purely crystalline cellulose. Thus, a minor microcrystalline activity was defined in the model for EG as explained in the following section. I want to state that this minor activity of EG on microcrystals has no significant influence on the amorphous degradation and only minor influence (~ 10 % decrease by a 20 fold decrease of EG activity on microcrystals) on the microcrystal degradation, which was elucidated in a PSA shown in Figure 20. In contrast to this, it has a major influence on amorphous degradation when changing the amorphous activity of EG (~ 360 % decrease by a tenfold activity decrease of amorphous EG activity, shown in Figure 18). A similar minor activity for CBH II on microcrystals will be introduced in further model developments, since CBH II also shows activity on Avicel, a highly crystalline substrate.

For EG action on crystalline cellulose an intrinsic reaction velocity ($v_{intrinsic}$ in nm/s) was defined and combined with a probability for starting the reaction (p_R in nm/s) by equation 1:

$$p_R = \frac{1}{\frac{1}{v_{apparent}} - \frac{1}{v_{intrinsic}}} \qquad (1)$$

where $v_{apparent}$ is the measured apparent reaction velocity on crystalline cellulose in nm/s. A k_{cat} of 0.29 s^{-1} at 50 ° C was measured experimentally for EG on Avicel (unpublished results). Similar k_{cat} values for EG on Avicel can be derived from various literature sources (Karlsson et al., 2002; Zhang and Lynd, 2004). Since mainly cellobiose was found as a product and the size of a cellobiose molecule is about 1 nm in length, $v_{apparent}$ could be calculated to 0.29 nm/s. Since the model uses discrete time steps, velocities were introduced into the model as reaction durations, which can be easily derived by taking the inverse of the velocities.

One can derive from equation 1: the higher the intrinsic reaction velocity, the lower the reaction probability. On the whole, by this equation a faster intrinsic reaction velocity is compensated by a lower starting reaction probability leading to the same apparent velocity in the end. In this work, it was elucidated to which extent the selection of a particular pair of reaction probability and intrinsic reaction velocity would influence the overall vertical degradation rate (degradation along the z–axes) on amorphous and crystalline cellulose.

An approximate tenfold increase in reaction activity for EG was assumed on amorphous cellulose in the model, because EG showed a tenfold activity increase on the amorphous–crystalline mixed substrates (used for validation of this model) compared to Avicel (unpublished results). The initial rates on the amorphous–crystalline mixed substrate were measured with a DNS assay as described elsewhere (Xiao et al., 2004). Moreover, an activity increase on mainly amorphous cellulosic substrates was found by several authors (Karlsson et al., 2002) and summarised by (Zhang and Lynd, 2004). I want to state that this tenfold activity increase is the lower boundary of the EG activity on purely amorphous cellulose, since the amorphous–crystalline mixed substrate (used for validation of the model) is not purely amorphous. Therefore, further activity increase on purely amorphous cellulose can be expected. However, this additional increase might be compensated by the fact that the model is simulating room temperature (25 ° C) and not 50 ° C, because all other kinetic values were measured at 25 ° C (Table 1).

For processive active enzymes (CBH I and CBH II), the processivity is a crucial parameter. As reported in recent literature (Horn et al., 2012; Kurašin and Väljamäe, 2011), it was possible to study the internal and apparent processivity of *Tr*Cel7A (CBH I) by fluorescent end–labelled cellulose. The apparent velocity was found to be 64 ± 14, and the intrinsic processivity 4000 ± 570 cutting events. The value for the simulation was chosen to be 100, since it is in the reported range and its influence on the model output was neglectable, as shown by a PSA (Figure 21). The non–influential behaviour can be explained by the high crowding of CBH I molecules on the crystal, which is elaborated in the discussion section in detail.

Enzymatic crowding (jamming): This can only occur for cellobiohydrolases since EG do not perform a processive movement. CBH I and CBH II can collide during their processive motion especially when considering CBH I action on anti–parallel cellulose chains in cellulose II crystals. There CBH I molecules can collide in a head–to–head manner. Therefore, a maximal waiting time of 5 s was defined (Table 1). After this time, the collided enzymes stop their processive motion and start performing a random walk. The value of 5 s was chosen because mean influence on the overall degradation can be expected (see section 5.4 Elucidating influencing input parameters) and Figure 23).

For measuring the vertical degradation rate (degradation along the z–axes) in the model, 900 equidistant measuring spots, each with dimensions of 3 nm x 3 nm, were used to identify the highest point on the projected x–y plane of each spot over time. This is close to the actual mode of action of the AFM, where the tip detects the highest point in the x–y plane. Moreover, these measuring points were separated into two fractions, one fraction of points which are on the projected area of the microcrystal and a second fraction of points solely on amorphous cellulose. Typical model simulations were performed on a total area of 70 nm x 150 nm (10500 nm^2) and spots (900 x 3 nm x 3 nm equals to 8100 nm^2) were capturing the height development of about 77 % of the total surface.

For all simulations, unless stated otherwise, the base case parameters, summarised in Table 1, and the microcrystal configuration, shown in Figure 4, were used.

Table 1: Input parameters for the model (base case parameters). n.d. is short for not defined, due to different kinds of modes of action of the various enzymes.

Enzyme	reaction duration [ms/nm]		processivity [cellobiose molecules]		maximal waiting time due to collision [s]	Fraction on whole cellulase mixture [%]
	amorphous	crystalline	amorphous	crystalline		
EG	350[a]	3500[a]/280[b]	n.d.	n.d.	n.d.	20
CBH I	n.d.	140[c]	n.d.	100[f]	5[g]	60
CBH II	140[d]	n.d.	1-5[e]	n.d.	5[g]	20

[a]apparent velocity (inverse of reacton duration) measured by Eibinger et al. at 50 ° C (manuscript in progress, see text above for calculation)

[b]assumed intrinsic reaction duration (see parameter sensitivity analysis in Figure 19)

[c]measured by (Igarashi et al., 2011) at 25 ° C

[d]assumed to be equal to measured CBH I reaction duration measured by (Igarashi et al., 2011)

[e]dependent on accessible amorphous chain length (see model design of amorphous cellulose)

[f]in the range of reported values of (Kurašin and Väljamäe, 2011) (measured on bacterial cellulose)

[g]assumed as a mean value due to the parameter sensitivity analysis shown in Figure 23

4. Methods

4.1 Computational methods

Matlab version 7.10.0.499 (R2010) was chosen as a programming platform. The architecture of the cellulose is shown in Figure 2. The idea behind the concept is that different properties of the cellulose are stored in separate matrices, all of the same size. Thereby it is possible to store several properties of thousands of substrate objects (cellobiose molecules) in a compact and fast accessible form. The detailed structure of crystalline and amorphous cellulose are explained in the model development section. The object structure of enzymes is depicted in Figure 6, where the most prominent functions and properties are shown. Most of the properties summarised in Figure 6 were elaborated and discussed in respect to their values in the text above. Here I want to give some information on the functions. In short, the function *slide()* was responsible for the surface diffusion of all types of enzymes using a random walk. The random number between 1 and 6, setting the three dimensional moving direction, was generated by a multiplicative congruential algorithm. The walk in the random direction is only performed if there is no obstacle in the chosen direction. The function *processive_motion_CBH()* has no random number generator because the moving direction is set by the orientation of the amorphous or crystalline cellulose chain in advance. The function *processive_motion_CBH(),* as well as the function *slide()*, ensures that enzymes do not hike through obstacles. The function *react()* decreases the internal reaction duration of enzymes when a reaction is finished. The actual deletion of cellobiose molecules of the "matter matrix" (insoluble surface) due to reacting enzymes is handled in the main program outside the functions.

4.2 Parameter sensitivity analysis (PSA)

Input parameters were varied in order to elucidate the differential change in the model in respect to the varied parameter. This technique is a standard method in order to elucidate important system parameters (Hamby, 1994). Except for the varied parameter all other enzyme parameters were applied as the base case parameters from Table 1. The structure of the microcrystal was never changed and had the

dimensions as shown in Figure 4. In order to gain a statistically profound result, five repetitions for each varied parameter were simulated and indicated by error bars.

4.3 AFM data

AFM data was taken from unpublished results of Ganner et al. (2012). For measuring a cellulosic substrate with AFM, a relatively flat surface (roughness of about 200 nm) is a requirement. Furthermore, in order to investigate enzyme activities on crystalline and amorphous parts, it is necessary to create an amorphous–crystalline mixed substrate. Therefore a special substrate preparation method was applied in Ganner et al. (2012), which allowed to fulfil both requirements. In short, the substrate was prepared by dissolving microcrystalline cellulose in 1-Butyl-3-methyl-immadzolonium chloride (ionic liquid) and constant heating and stirring. Subsequent removing of the ionic liquid by ethanol extraction leads to a polymorphic cellulosic model substrate. AFM experiments were conducted by an in situ liquid AFM technique with tapping mode in order to prevent damaging of the sample. For more details on AFM measurements, sources of cellulases and substrate preparation, see Ganner et al. (2012).

The essential results of the study of Ganner et al. (2012), which can be used as input parameters in this modelling work, are summarised here in short. The preferences of single enzymes (EG, CBH I and CBH II) for crystalline or amorphous parts could be qualitatively described. In detail, it was found that CBH II and EG strongly prefer amorphous cellulose for degradation, whereas CBH I shows activity only on small crystals (here called microcrystals). Moreover, it was possible to measure vertical degradation rates (along the z axes) for a complete cellulase mixture. Two distinct vertical degradation velocities of 3.8 ± 0.2 nm/min and 0.7 ± 0.2 nm/min were found and assigned to the crystalline and amorphous degradation, respectively. Although the observation of a faster crystal degradation than amorphous degradation sounds contradictory, it was explained by the fast degradation of microcrystals, which are polished by CBH II and EG prior to their hydrolysis by CBH I. Thereby microcrystals emerge out of the surrounding amorphous matrix until a certain point is reached and a rapid degradation of these microcrystals is initiated. Furthermore, special degradation patterns for microcrystals were described, such as the thinning of

microcrystals starting from the side walls, the shortening of the microcrystals from the tips, and the introduction of defects in the middle of the microcrystals.

4.4 Virtual synergism experiments

In order to investigate the synergistic action of different binary combination of the three main enzymes (EG, CBH I and CBH II), it is necessary to create a substrate which provides amorphous and crystalline chain ends. This is required due to the relatively strict preferences of the cellobiohydrolases in the model, where CBH I is not active on amorphous parts and CBH II is not active on crystalline parts (see model development section and Table 1). For example, a nanoflat surface with neither amorphous nor crystalline ends would result in no activity of the binary combination of CBH I and CBH II. Therefore the model substrate, configured as explained in section 3.1 Modelling cellulose), was pre–incubated with a complete cellulase mixture (EG + CBH I + CBH II), and the reaction was stopped after 500 s. This procedure created a pre–hydrolysed substrate containing amorphous and crystalline chain ends, as shown in Figure 8. For all synergism simulations, this configuration was used in order to obtain reproducible conditions. The pre–hydrolysed substrate had about 15 % crystalline cellulose and 85 % amorphous cellulose. Initial activities, measured by the released cellobiose content of single enzymes and binary combinations, were used in order to obtain synergism factors. The synergism factor was defined as the ratio of the synergistic activity of a binary enzyme combination divided by the sum of the single enzyme activities. All synergism experiments were repeated five times, leading to the error bars shown in Figure 17. In order to compare synergism factors to literature values, it was further assumed, that single enzyme velocities of EG, CBH I and CBH II change in the same range when altering the simulation temperature of 25 ° C to 50 ° C, where most of the experiments used for validation were conducted.

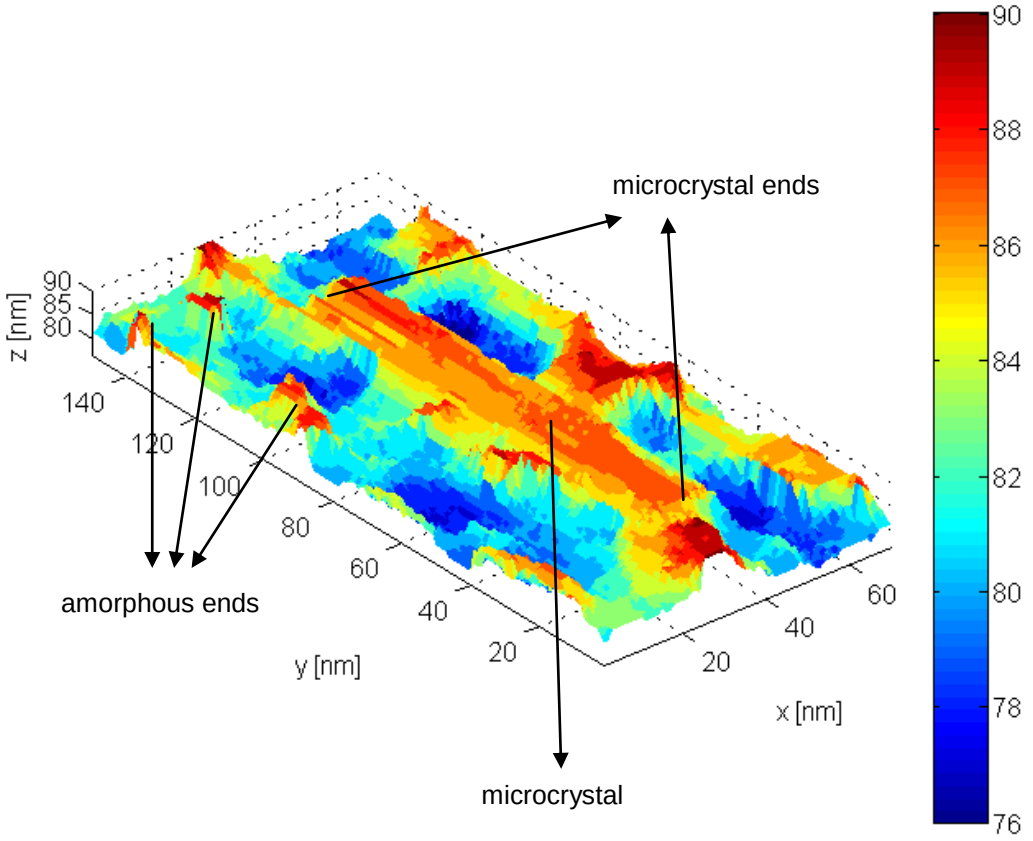

Figure 8: Three dimensional representation of the pre–hydrolysed model substrate, containing amorphous and crystalline chain ends. Colour–bar values are given in nm.

5. Results

As described in the method section, enzymes and cellulose were virtually created by a bottom–up–approach, including physical properties (size) and kinetic data (single enzyme velocities) of cellulases and cellulose according to literature findings of the past. The most outstanding overall result is that only by introducing several known physical and biochemical features and some reasonable assumptions into the model, it is possible to observe nanoscale effects of AFM measurements. Besides these nanoscale degradation profiles, which were the focus of this work, synergism factors of binary enzyme combinations in respect to the binary mixing ratio, and realistic overall hydrolysis kinetics could be reproduced by the model as well. The result section is organized in four distinct parts: Firstly, in order to use the three dimensional nanoscale model output, described as the strength of the simulation, the model was validated against qualitative and quantitative characteristics of AFM measurements. Secondly, the model was validated against literature findings about the degradation efficiency in respect to the adsorbed enzyme concentration. Thirdly, the model was used to describe classical binary synergism experiments between the three surface acting cellulases and classical hydrolysis studies revealing specific activities. Fourthly, the influence of certain input parameters was systematically elucidated on the model output, which represents an internal mathematical validation procedure.

5.1 Reproducing AFM observations – validation on the nanoscale

In order to use the visual information of AFM measurements, data characteristics have to be extracted first. Due to the similar size of the simulation and the dimensions of the microcrystals (10 nm x 100-300 nm), qualitative and quantitative aspects could be extracted out of the experimental microcrystal development. Qualitative degradation characteristics of the work of Ganner et al. (2012), such as the thinning and the separation of microcrystals during hydrolysis, are described in section 4.3 AFM data). As a main quantitative property of microcrystals, the height development (height difference between the top of the crystal and the amorphous level) was extracted out of unpublished results of Ganner et al. (2012). The emerging and subsequent fast degradation of microcrystals was qualitatively described in the work of Ganner et al. (2012) and summarised in section 4.3 AFM data). This feature was chosen for quantitative comparison because it is the best possible way to follow the height development of microcrystals on the nanoscale.

Figure 9 shows qualitatively the thinning (top crystal indicated by the arrow) and the separation of a crystal in two pieces (bottom crystal indicated by red arrow). Moreover, the slow emerging and subsequent fast degradation of the microcrystal can be observed similar to observations of the AFM experiments in the recent publication of Ganner et al. (2012) (see section 4.3 AFM data). Separation can only occur when the microcrystal is thinned enough (red arrow in Figure 9) and EGs are active on the thinned part of the microcrystal. Figure 9 contains screenshots of the video S1 of the supporting information.

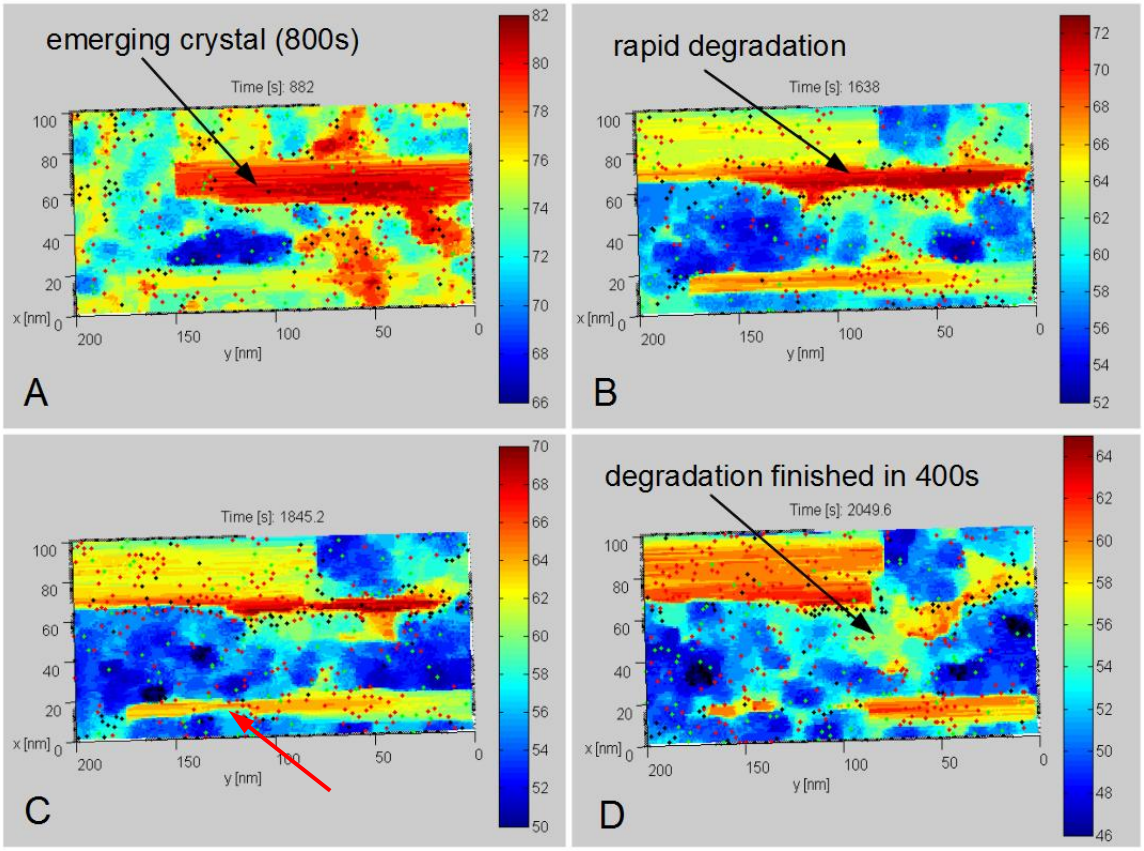

Figure 9: Visual model output of a simulation of a crystalline–amorphous mixed substrate with randomly distributed microcrystals and base case parameters as stated in Table 1. Vertical height is indicated by the colour bar, where values are given in nm. Green dots: EG. Red dots: CBH I. Black dots: CBH II. The crystal development is indicated by arrows, which show the slow emerging and fast degradation of the crystal as well as the thinning of the crystal.

AFM measurements of Ganner et al. (2012) revealed specific degradation patterns for microcrystals within a polymorphic cellulosic substrate (see section 4.3 AFM data). A slow emerging phase was followed by a rapid degradation of the microcrystal. With the current modelling approach it was also possible to observe these two phases of microcrystal development, which can be distinguished from the

continuous degradation of the surrounding amorphous cellulose (Figure 10). Velocities of emerging and degrading crystals were measured as vertical degradation rates along the z-axis, given in nm/min. The velocity of the emerging phase was 0.6 nm/min, the fast degradation was 3.5 nm/min, and the continuous amorphous degradation 1 nm/min, calculated by linear regression with cellulase surface coverage of 70 % of the total surface, and base case parameters as stated in Table 1. These findings fit to the reported values of Ganner et al. (2012) of 3.8 ± 0.2 nm/min for crystalline degradation and 0.7 ± 0.2 nm/min for amorphous degradation. The error bars shown in Figure 10 represent the height variance within the projected area (amorphous and crystalline). The height profile of the projected area of the microcrystal (shown in red) was more homogenously distributed than the height profile of the amorphous area, since the error bars are much smaller in the microcrystalline area. The fast degradation is initiated just as the degradation of crystalline material finished around 71 nm, because the crystal was implemented from 90 nm - 70 nm in z dimension (Figure 4). Therefore the fast degradation can be ascribed to the action of CBH II and EG, which are the only two enzymes able to degrade amorphous cellulose in the system. This can also be monitored by following time dependent enzyme concentration on the projected crystalline area of the three surface acting cellulases, as shown in Figure 11. Within the time span of 1600s to 2000s, the CBH I concentration is significantly reduced, and the concentration of CBH II and EG is increased on the projected area underneath the crystal. The different time dependent stages (emerging and subsequent fast degradation of the microcrystal) can be watched in the video S2 of the supporting information.

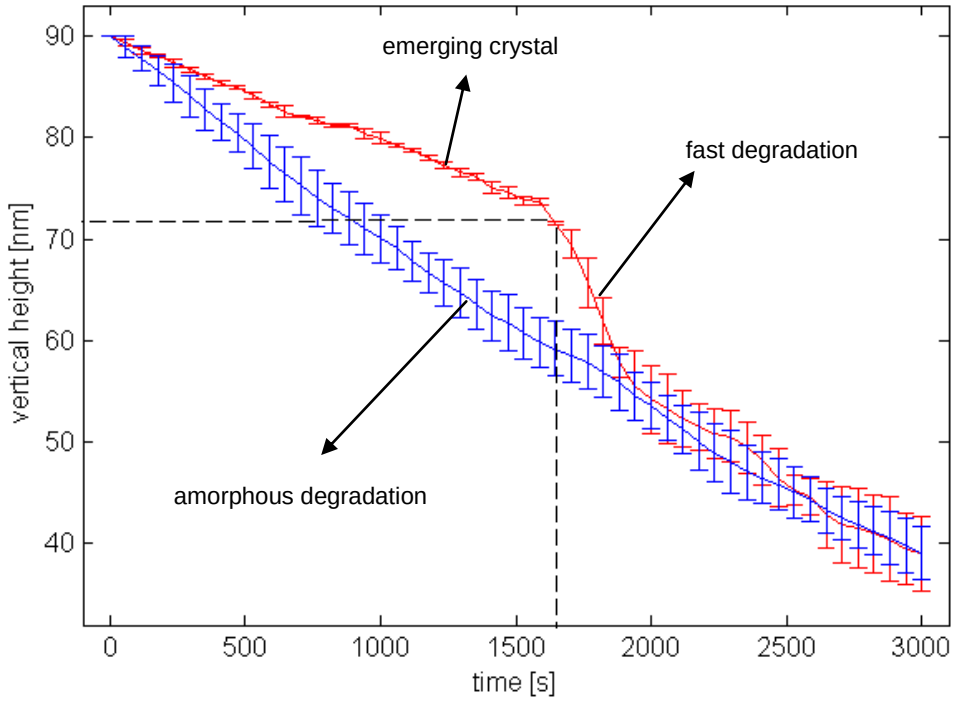

Figure 10: Degradation profile of 900 height measuring spots of a model simulation with 70% surface coverage and one microcrystal (cylindrical: 20 nm diameter, 100 nm length) surrounded by amorphous cellulose. Red solid line: mean value of spots on the projected crystalline area. Blue solid line: mean value of spots on the projected amorphous area. Error bars indicate the height variance within the projected area. Dashed black line indicates the start of the fast degradation of the projected microcrystalline area at ~1600 s.

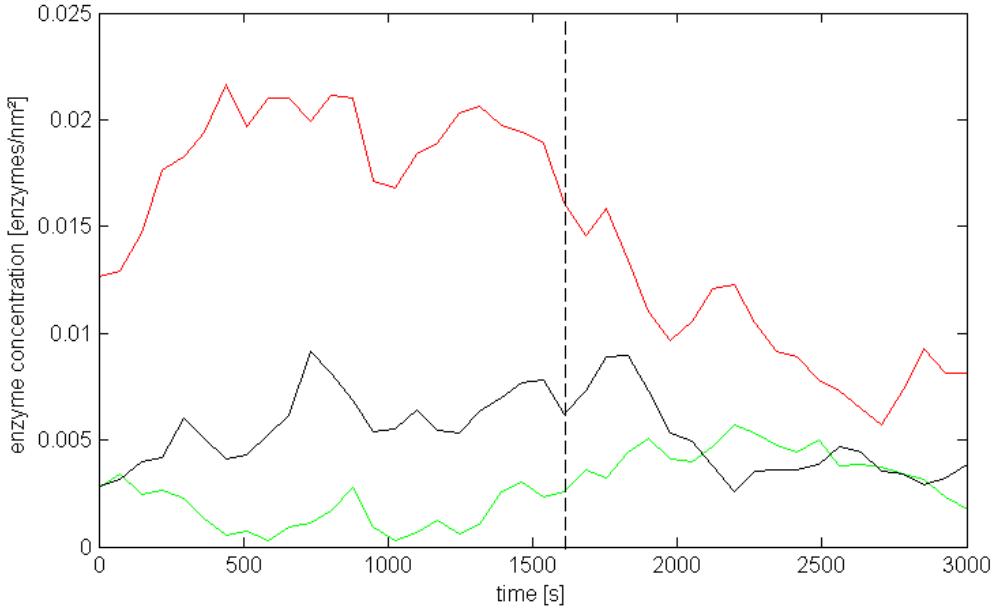

Figure 11: Modelled time resolved enzyme concentrations of EG (green), CBH I (red) and CBH II (black) on projected microcrystalline area. Dashed black line indicates the start of the fast degradation according to Figure 10 at ~1600 s.

Quantitatively, the emerging and degradation of microcrystals could only be experimentally measured in relation to the surrounding amorphous level because no height marker was present in high resolution AFM measurements of Ganner et al. (2012). Usually, height markers consist of a material non degradable by enzymes. This makes it possible to create a reference point for depth measurements. A general data acquisition procedure of an amorphous–crystalline height difference profile of two microcrystals is shown in Figure 12. Figure 13 shows two AFM measurement of an amorphous–crystalline height difference superimposed by the model calculation of two microcrystals. The modelled height profile fits qualitatively and quantitatively to the measured height profile. Moreover, a correlation between microcrystal maximal width (here also called diameter) and the complete degradation time of the microcrystal could be found in the model as well as in AFM experiments: The larger the crystal width, the longer it takes until the crystal is completely degraded. This can be demonstrated on two crystals shown in Figure 13A and Figure 13B. The time for the complete degradation of the crystal experimentally determined in Figure 13A was about 50 min (3000 s). During 50 min the surrounding amorphous cellulose is degraded 35 nm (amorphous degradation rate of 0.7 nm/min reported by Ganner et al. (2012), which is about the same as the maximal width (diameter) of the microcrystal (32 nm). The correlation is also true for smaller crystals as shown in Figure 13B, where the degradation time of 33 min (2000 s) leads to a height degradation of 23 nm (maximal width 20 nm). The same conclusion can be made with the modelled crystal developments shown superimposed in Figure 13A (30 nm diameter) and Figure 13B (20 nm diameter). I want to state that due to the lack of further high resolution AFM measurements, the comparison of modelled and experimental crystal degradation cannot be used statistically for fitting. For the same reason it is not clear to which extent the amorphous material underneath the crystal influences the degradation profile. Nevertheless, a similar general trend of slow emerging crystals, fast crystal degradation and the correlation between degradation time and maximal crystal width can be shown for several microcrystals in the AFM data as well as in the model. Therefore, a statistically profound qualitative conclusion can be drawn.

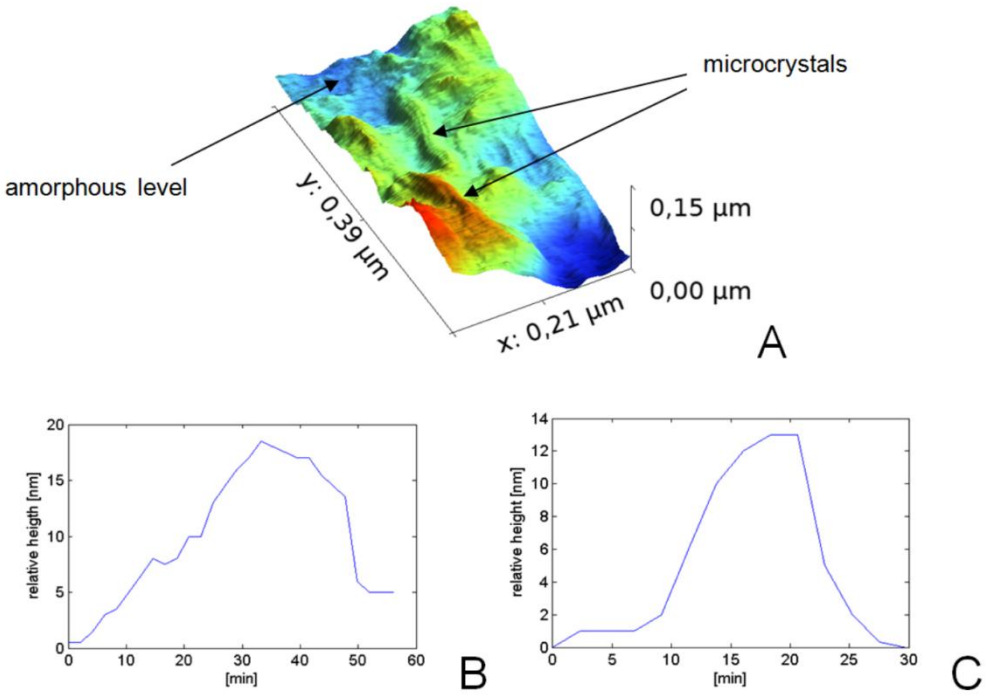

Figure 12: Amorphous–crystalline height difference measurements. A: typical AFM data acquisition consists of evaluating the height of the crystal in respect to its amorphous level. B and C extracted data of two crystals with 32 nm and 18 nm maximal width (=diameter), respectively.

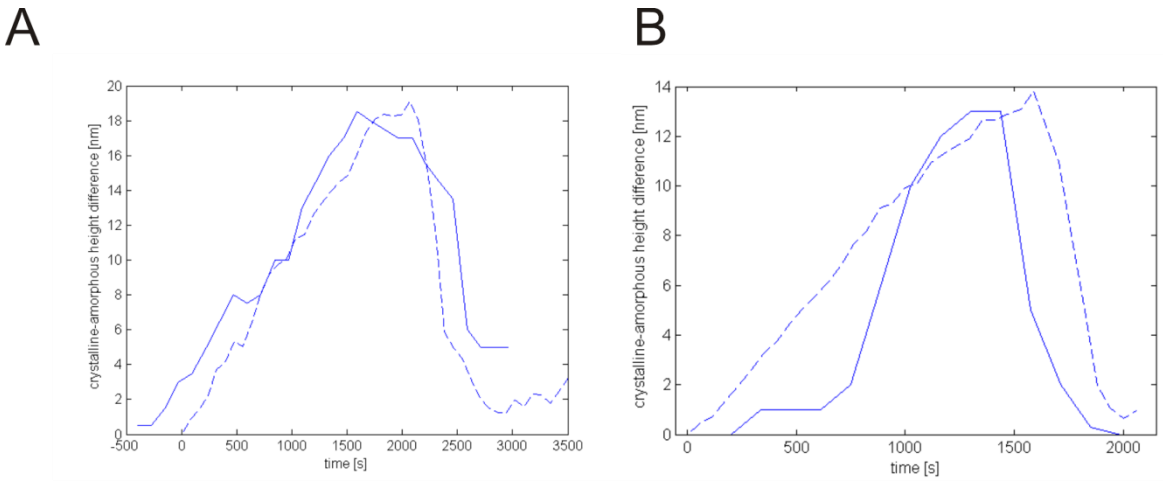

Figure 13: AFM measurements of amorphous–crystalline height differences of a microcrystal (solid lines) superimposed by model simulations (dashed lines). The maximal surface coverage with cellulases in the simulation was 70% of the total surface. All other parameters were applied as stated in Table 1. A: maximal width of AFM measured microcrystal: 32 nm, simulated cylindrical microcrystal with diameter of 30 nm; B: maximal width of AFM measured microcrystal: 19 nm, simulated cylindrical microcrystal with diameter 20 nm.

5.2 Increasing enzyme surface coverage leads to reduced specific microcrystal degradation

In the following sections, parameter sensitivity analyses (PSA) were used to illustrate influences of changed system variables. Each change of the input parameters produces a new simulation output containing all information about the emerging and fast degradation of microcrystals over time, the continuous degradation of amorphous cellulose, and local concentration changes of different cellulases on different parts of the substrate. In order to interpret the influence of a changing input parameter in a compact way, the mean vertical height degradation was defined as an output parameter. Due to the crucial differences of amorphous and crystalline parts of the substrate and the associated specific enzymatic activities, each part (crystalline and amorphous) was analyzed separately by measuring the vertical degradation depth after a certain time period. Thereby it is possible to perceive a compact and holistic, but differentiated picture of the influence of the changed variable.

The degradation rate on microcrystals strongly depends on the surface coverage of the cellulose surface by cellulases. As shown in Figure 14, the specific vertical degradation alleviates on microcrystals when the surface coverage is increased to values higher than 50%. Further surface coverage increase does not increase the vertical degradation velocity of microcrystals significantly. In contrast to this, vertical degradation of amorphous cellulose is linear dependent on the surface coverage. Only at surface coverage >90%, an alleviated dependency is prevalent. In order to give reasons for this phenomenon, the effective mean velocity of all enzymes was investigated for increased surface coverage. As Figure 15 shows, the mean velocity of all enzymes decreases rapidly with increasing surface coverage. This indicates that the mobility of enzymes is reduced due to increased crowding, which is likely to be the reason for inefficient microcrystal degradation at higher values of surface coverage. The amorphous degradation stays linear with increased surface coverage because the activity of CBH II is less, and the activity of EG is not affected by crowding compared to CBH I. The processive movement of CBH II is shorter than the one of CBH I because it is limited by the maximal segment length of an amorphous cellulose chain (see model development section). Thereby, the probability of a collision is reduced, and the CBH II activity stays unaffected until 90 % surface coverage is reached. EG have no processive movement and are able to dig deeper, although crowding increases with higher surface coverage.

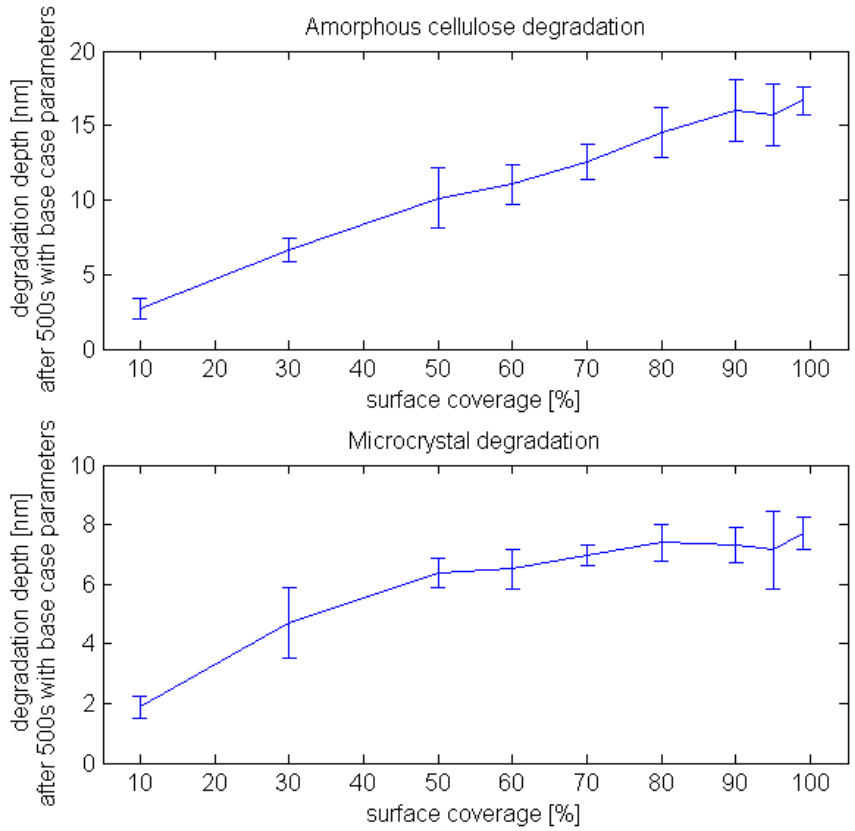

Figure 14: Vertical degradation rate in respect to surface coverage. Influence is shown on amorphous cellulose (top) and microcrystalline cellulose (bottom). Base case parameters of Table 1 and crystal dimensions of Figure 4 were applied.

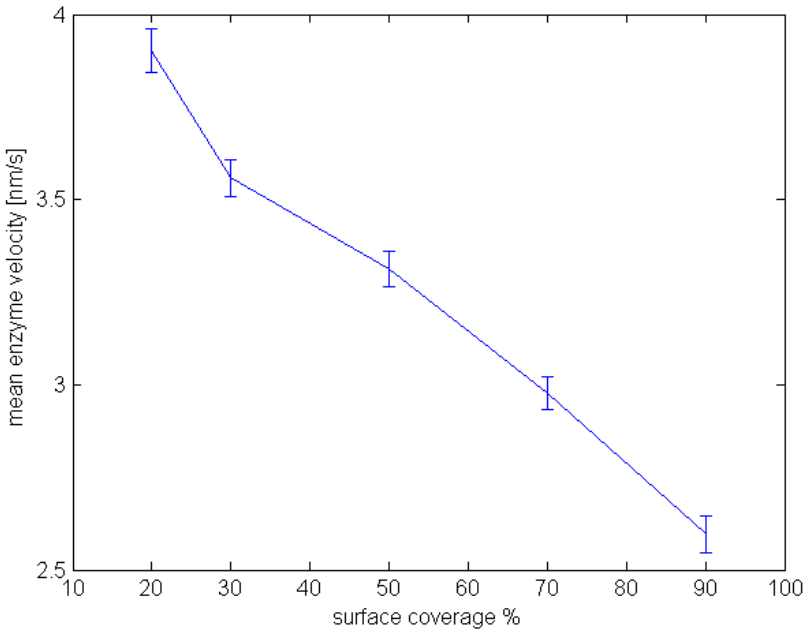

Figure 15: Mean enzyme velocity of all three enzymes (EG, CBH I and CBH II) versus increasing surface coverage

5.3 Derive biochemical parameters from a bottom–up–approach

Classical hydrolysis experiments, conducted by measuring cellobiose in the bulk, were modelled. As a substrate, the amorphous–crystalline cellulose was applied with microcrystal dimension as shown in Figure 4. A result of a modelled hydrolysis experiment is shown in Figure 16. A linear relation between reaction time and cellobiose content was found since the substrate was not limiting, no inhibition by cellobiose and no enzyme deactivation was assumed. Furthermore, some authors suggest a mechanism where cellobiohydrolases are immobilized to the cellulose surface by obstacles, which cannot be hiked through (Jalak and Väljamäe, 2010). This phenomenon was not implemented either. Nevertheless, none of the above mentioned effects is prevalent in the short duration (3–5 h) of AFM measurement and therefore, a linear relation was found experimentally as well (Bubner et al., 2012; Ganner et al., 2012). Linear regression led to a specific activity of 0.6 IU/mg enzyme. This value is a mean value for all three major enzymes (EG, CBH I and CBH II). The specific activity of cellulases on Avicel, which is a highly crystalline substrate (~ 60 - 90 % cyrstallinity) and was the original cellulose source for the applied crystalline–amorphous mixed substrate, is about 0.17 IU/mg for EG1 (*T. reeisei*), about 0.04 IU/mg for CBH I (*T. reeisei*) and about 0.03 IU/mg for CBH II (*T. reeisei*) (Zhang and Lynd, 2004). For purely amorphous cellulose specific activities are reported about 4 - 26 IU/mg for EG1 (*T. reeisei*), about 0.6 IU/mg for CBH I (*T. reeisei*) and about 0.05 IU/mg for CBH II (*T. reeisei*) (Zhang and Lynd, 2004). Since the simulated amorphous–mixed substrate contains a predominant amorphous structure (Bubner et al., 2012) the value of 0.6 IU/mg, as a mean specific activity for all enzymes, seems realistic because it is higher than values reported for highly crystalline Avicel but lower than those for purely amorphous cellulose. Moreover, the simulated value of 0.6 IU/mg is the lower boundary for the specific activity, since all mentioned literature values were measured at 40 – 50 ° C, whereas the model simulates condition at 25 ° C.

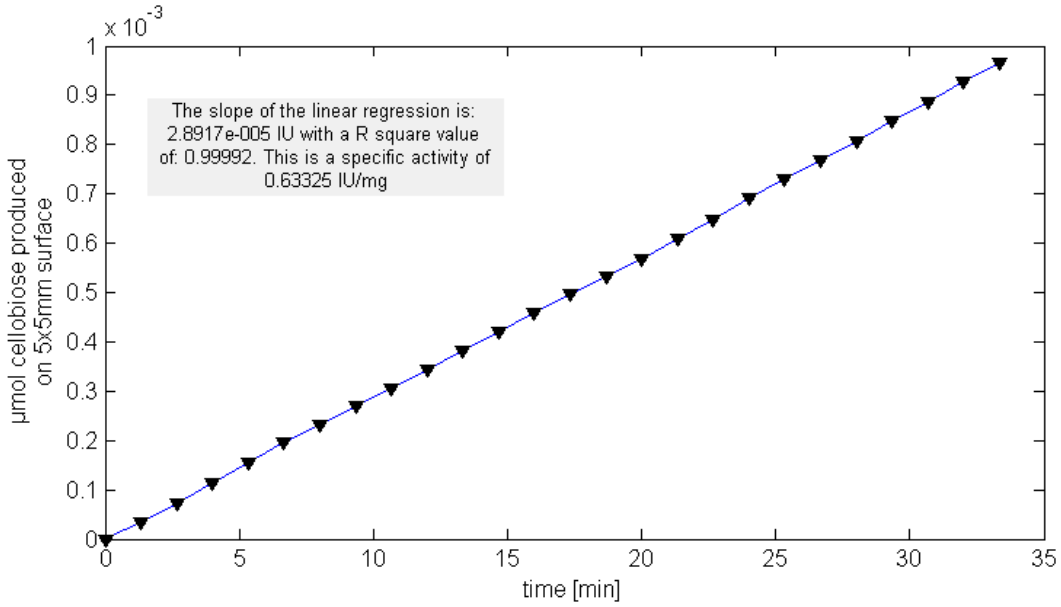

Figure 16: Amount of cellobiose produced on a 5mm x 5mm surface. Resulting mean specific activity of EG, CBH I, CBH II extracted from a bottom–up–approach is shown in the text box.

As a second biochemical validation, the synergism between the three surface acting enzymes was investigated. The effect of synergism between enzymes was measured by applying the concept of the synergism factor (Nidetzky et al., 1993; Nidetzky et al., 1994b). The synergism factor is the ratio of the synergistic activity of a binary enzyme combination divided by the sum of the single enzyme activities. Therefore, single enzyme activities (initial hydrolysis rates) on a substrate offering free amorphous and crystalline ends, as well as accessible β–glycosidic bounds for EG activity, were simulated. In detail, the substrate consisted of 20 % crystalline structures and 80 % amorphous cellulose. Afterwards, the activities (initial hydrolysis rates) of binary combinations of the three surface active enzymes were modelled on the same substrate. Moreover, the ratio between the two applied enzymes of a binary combination was varied with a constant total surface coverage of 50 % (equals to the total amount of adsorbed enzymes). Figure 17 shows synergism experiments for CBH I and CBH II (Figure 17A + Figure 17B), CBH I and EG (Figure 17C + Figure 17D), and CBH II and EG (Figure 17E + Figure 17F). In each experiment, the synergistic activity (shown in green) is higher than the sum of the single activities (shown in blue), a behaviour described in various publications (Irwin et al., 1993; Jalak et al., 2012; Medve et al., 1994; Nidetzky et al., 1993; Nidetzky et al., 1994b). The degree of synergism of CBH I + CBH II was higher than synergism factors of all other binary combinations. The synergism factor of CBH I + CBH II had a maximum

at a level of 30 % CBH II and 70 % CBH I, with a synergism factor of 2.2, which is in excellent quantitative agreement to experimentally measured synergism factors of CBH I and CBH II (Irwin et al., 1993; Medve et al., 1994; Nidetzky et al., 1993; Nidetzky et al., 1994b). Moreover, the qualitative and quantitative synergistic velocities of CBH I and CBH II (Figure 17A) at different mixing ratios are comparable to the values reported by (Igarashi et al., 2011). For binary combinations containing EG (EG + CBH I in Figure 17C and EG + CBH II in Figure 17E) the single enzyme velocity of EG is one magnitude higher than those of the other enzymes (CBH I and CBH II). This can be explained by the simplified mode of EG action in the model, where EG cleave a cellobiose molecule out of the solid surface instead of only breaking a glycosidic bond (see model development section). The synergistic effect of EG + CBH I and EG + CBH II ranged from maximal values of 1.25 for EG + CBH I to 1.4 for EG + CBH II (Figure 17D and Figure 17F, respectively). The maximal values were found for both combinations when the EG fraction was 0.08 and the CBH I/CBH II fraction 0.92. Reported literature values (Nidetzky et al., 1993) for synergism factors of these combinations (EG + CBH I and EG + CBH II) are slightly higher on microcrystalline cellulose (1.6 and 1.5, respectively) but about the same for phosphoric acid swollen microcrystalline cellulose. Nevertheless, the reduced synergistic effect of EG + CBH I and EG + CBH II in comparison to CBH I + CBH II was qualitatively found in experiments (Nidetzky et al., 1993; Nidetzky et al., 1994b). Moreover, the optimal enzyme ratio predicted in the study of (Nidetzky et al., 1994b) for the combination of CBH I + EG was about 90/10, which fits perfectly to the found values in the model (92/8).

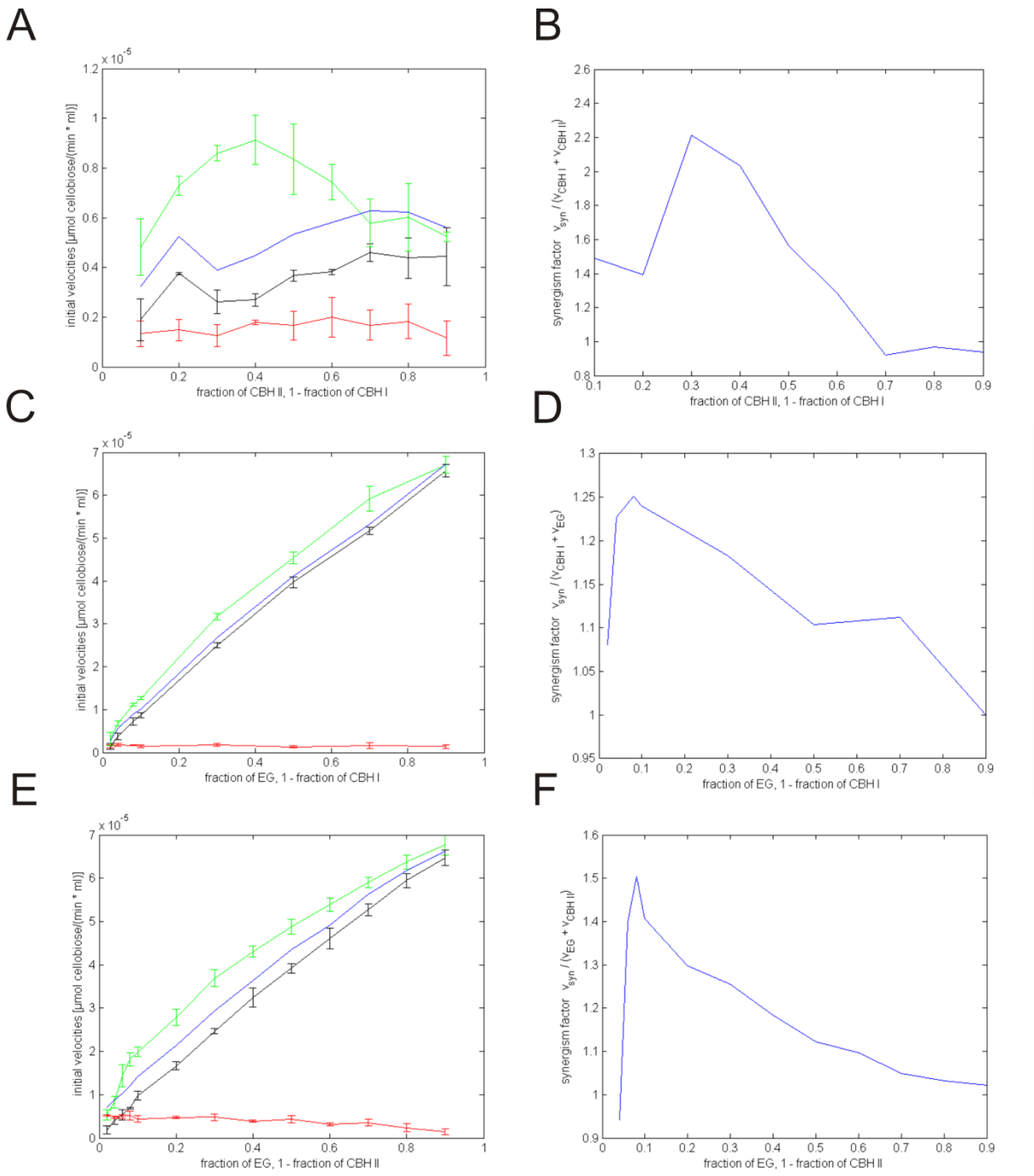

Figure 17: Synergism experiments of CBH I and CBH II (A + B), CBH I and EG (C + D) and CBH II and EG (E + F). Left column: initial hydrolysis rates of single enzymes (black and red lines), sum of single enzymes (blue line) and binary synergism of the enzymes (green line) in respect to different mixing ratios of the binary enzyme combination. Right column: synergism factor (= synergistic activity/sum of single enzyme activities) versus different mixing ratios of binary enzyme combination. A: red line CBH I; black line CBH II. C: red line CBH I; black line EG. E: red line CBH II; black line EG. Error bars indicate standard deviation of five repetitions.

5.4 Elucidating influencing input parameters

In a series of parameter sensitivity analysis experiments, the influence of the amorphous reaction duration of EG (inverse reaction velocity), the intrinsic reaction duration of EG on crystalline cellulose, the length of the processive period of CBH I, and the maximal waiting time of CBH I due to collision was investigated. These PSA experiments were conducted in order to elucidate whether the model shows expected or unexpected behaviour. Thereby it was possible to reveal functionalities of the complex heterogeneous reaction–diffusion system, which could not be intuitively assumed. All PSA experiments were conducted with base case parameters (Table 1) and microcrystal configuration as shown in Figure 4.

An increase in the amorphous reaction duration of EG led to a significant slowdown of amorphous degradation (top of Figure 18), but also the degradation rate of the microcrystalline part was moderately decreased as shown at the bottom of Figure 18.

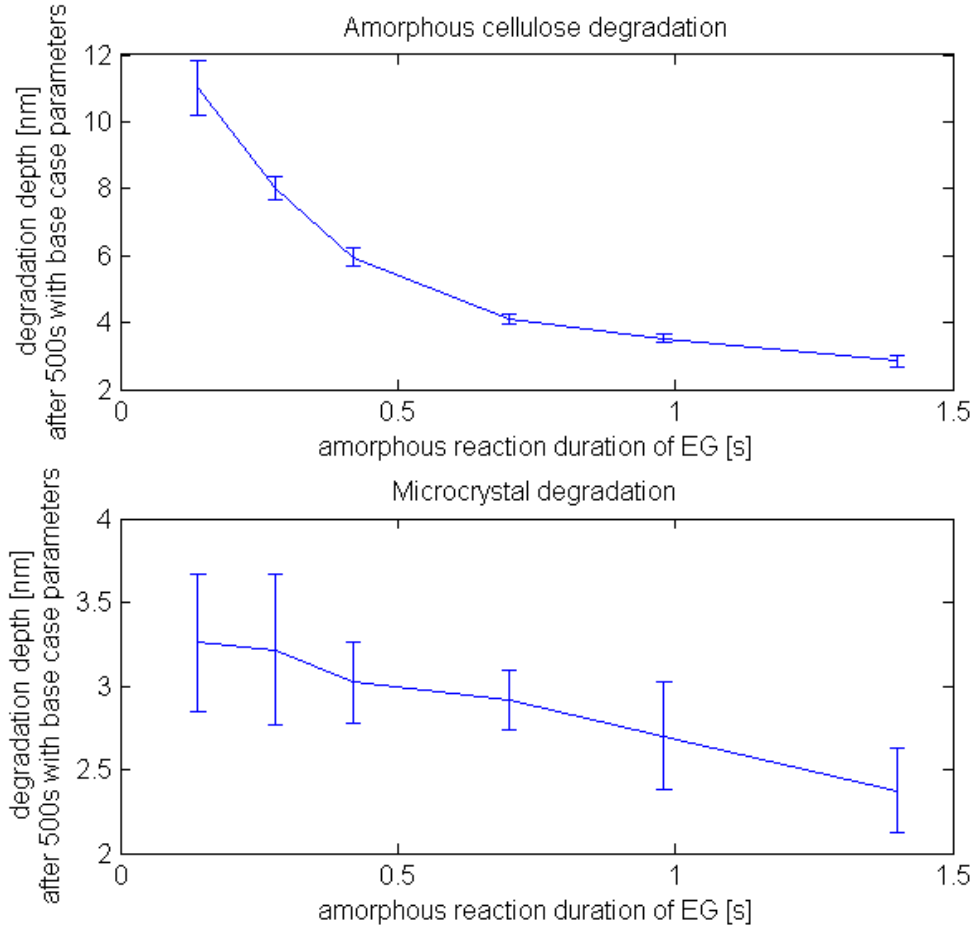

Figure 18: Parameter sensitivity analysis of the amorphous reaction duration of EG on amorphous (top) and crystalline (bottom) parts. Influence is shown on amorphous cellulose (top) and microcrystalline cellulose (bottom).

As described in the model development section, the mode of EG action on crystals is defined by a starting reaction probability together with an intrinsic reaction velocity connected through equation 1. In general, two extreme cases can be imagined: A very high intrinsic velocity leading to a very low reaction probability, or, on the other hand, a very low intrinsic velocity with a high reaction probability. In global terms, this leads to the same apparent reaction velocity, which can be measured by biochemical assays. The question of this section is whether and to which extent the vertical degradation rate of the amorphous and crystalline cellulose is influenced by different degrees of these two extreme modes. The transition between the two extreme modes is possible by varying the intrinsic reaction duration (inverse of intrinsic velocity of equation 1) as shown in Figure 19. The microcrystal degradation (bottom Figure 19) is not affected by the mode of EG action on crystals, which proves that equation 1 produces always the same apparent crystal degradation velocity for different values of the intrinsic velocity. The amorphous degradation (top of Figure 19) seems to decrease slightly when high intrinsic reaction durations (accompanied by a high reaction probability) are applied. In order to stay in a linear range of the vertical degradation velocity compared to the intrinsic reaction velocity the internal reaction duration (inverse of intrinsic reaction velocity) for EG was chosen as 0.28 s/nm for all simulations (see base case parameters in Table 1).

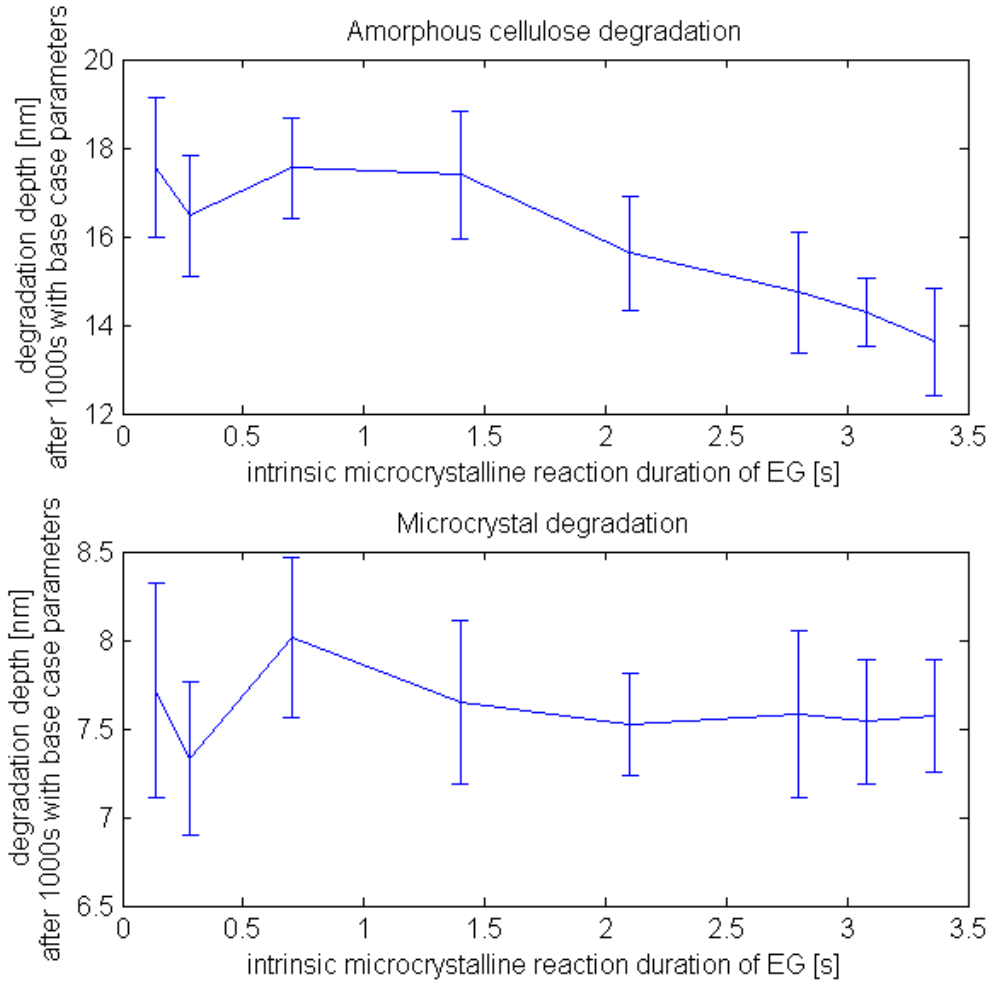

Figure 19: Parameter sensitivity analysis of the intrinsic microcrystalline reaction duration of EG. By varying the intrinsic reaction duration, the starting probability was changed as well according to equation 1. Influence is shown on amorphous cellulose (top) and microcrystalline cellulose (bottom).

The overall influence of the apparent crystalline reaction duration of EG was investigated, shown in Figure 20. No significant (< 5 %) influence on the amorphous degradation and only a minor (~ 10 %) decrease of the microcrystalline degradation rate could be shown, when increasing the reaction duration to the 20 fold of the value used in the base case parameters (3.5 s, see Table 1). This lowers the reaction probability (p_R of equation 1) 20 fold, since the internal reaction duration was kept constant at 280 ms/nm.

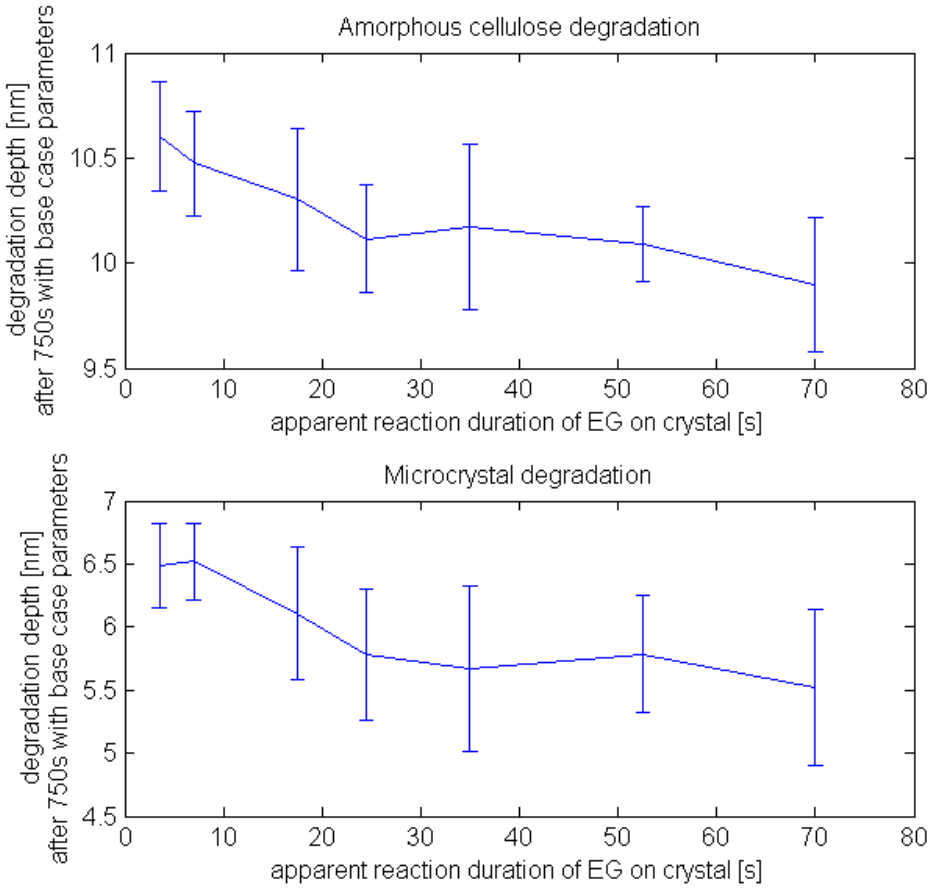

Figure 20: Parameter sensitivity analysis of the apparent reaction duration (inverse of apparent k_{cat}) of EG on crystals. A constant intrinsic reaction duration of 280 ms/nm was applied and thereby solely p_R of equation 1 was varied. Influence is shown on amorphous cellulose (top) and microcrystalline cellulose (bottom).

Figure 21 shows the parameter sensitivity analysis of the length of the processive period of CBH I given in cellobiose molecules. There was no influence of the processive period found neither on the amorphous nor on the crystalline degradation. Due to the fact that enzymes crowd on the crystal, and represent obstacles for each other, the processive movement is finished sooner than the processive length would allow it. The crowding of CBH I was investigated by measuring enzyme velocities on the microcrystal and on the amorphous region, as shown in the histogram of Figure 22. CBH I molecules diffuse with a mean value of 5 nm/s on the amorphous cellulose and 1.8 nm/s on the microcrystal, which results from the high local CBH I concentration of the crystal, as depicted in Figure 11. The enzyme velocity on amorphous cellulose is less than the maximal speed (7 nm/s) because also on the amorphous cellulose crowding occurs between CBH I, CBH II and EG. This model result is in line with the findings of (Kurašin and Väljamäe, 2011), where it was

concluded that the presence of obstacles, leads to a much shorter apparent processivity than it would be possible by the internal processivity.

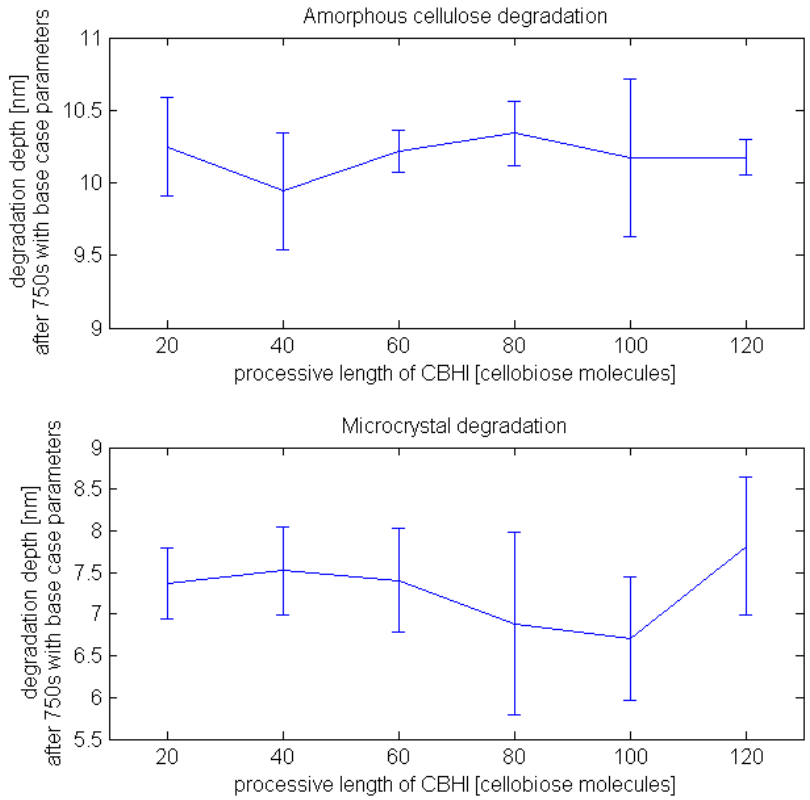

Figure 21: Parameter sensitivity analysis of the length of the processive period of CBH I. Influence is shown on amorphous cellulose (top) and microcrystalline cellulose (bottom).

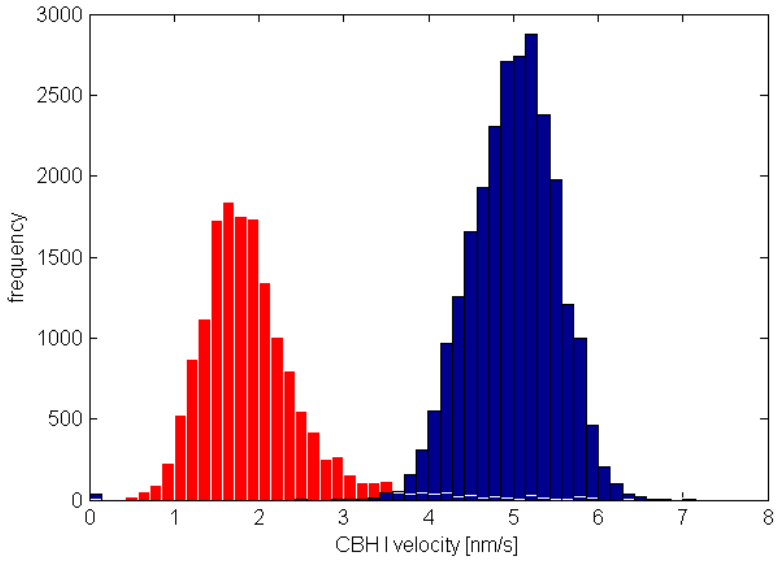

Figure 22: CBH I velocity on amorphous (blue) and crystalline (red) cellulose. Model simulation with base case parameters (Table 1), crystal dimensions as shown in Figure 4 and 70 % surface coverage of all enzymes.

The maximal waiting time of CBH I molecules, which stopped their processive motion due to collision with other cellulases or cellulose obstacles, is a crucial parameter, since its influence on the overall degradation of microcrystalline cellulose is rather high (bottom of Figure 23). In the base case parameters, a maximal waiting time of 5 s was assumed, which is 35 times longer than the processive duration (140 ms) and was regarded as a sufficiently long waiting time. Moreover, this maximal waiting time of 5 s results in a mean vertical crystal degradation (5 nm/ 750 s) compared to maximal waiting time values ranging from 1 to 20 s resulting in vertical degradation rates of 7.5 nm/ 750 s to 3.8 nm/ 750 s, respectively (bottom of Figure 23). The degradation of the amorphous cellulose is rather unaffected (10 % change) by a tenfold increase in the maximal waiting time of CBH II.

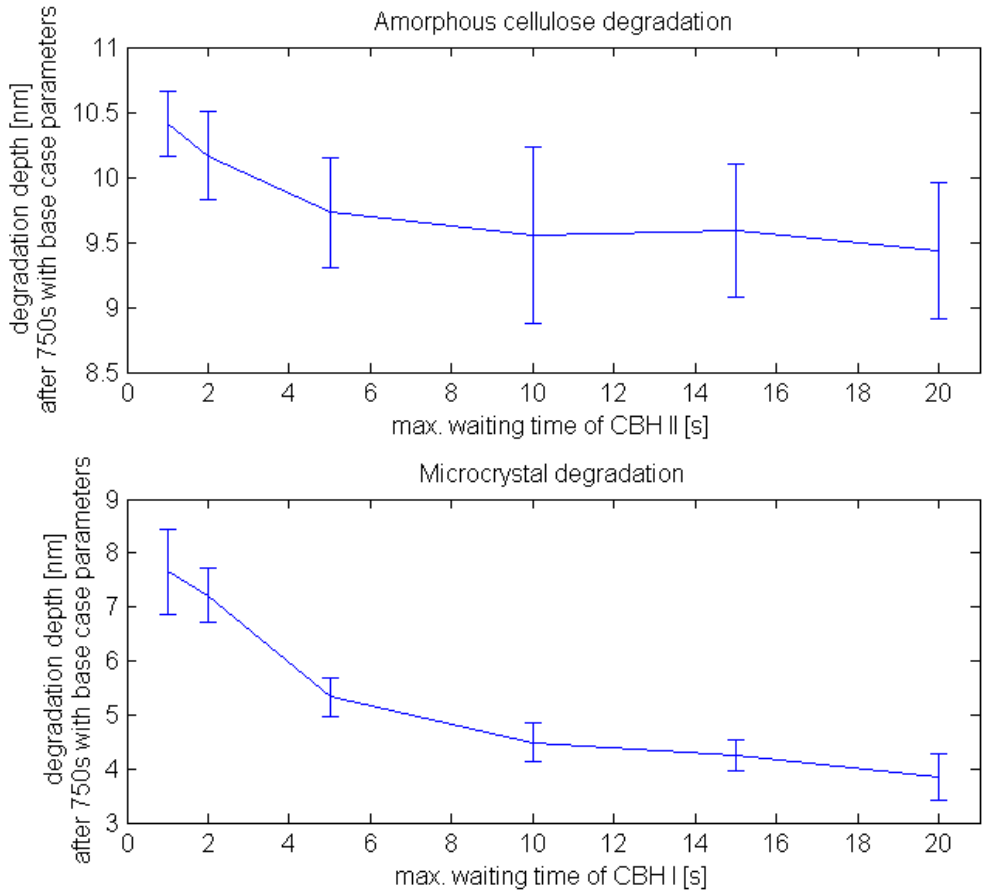

Figure 23: Parameter sensitivity analysis of the maximal waiting time of CBH I due to collision. Influence is shown on amorphous cellulose (top) and microcrystalline cellulose (bottom).

As described in the method section, the amorphous cellulose was designed by cellulose chain segments pointing in x or y direction of a z plane. The chain segments of one z plane have a defined mean segment length of 5 cellobiose molecules in the model (see model development section). In order to elucidate the influence of the segment length, a parameter sensitivity analysis was performed, shown in Figure 24, revealing that the influence on the amorphous degradation is low (less than 2 % change over a range from 2 to 7 cellobiose molecules). The influence on the crystalline degradation is negligible, as expected, since the crystal structure was kept unchanged.

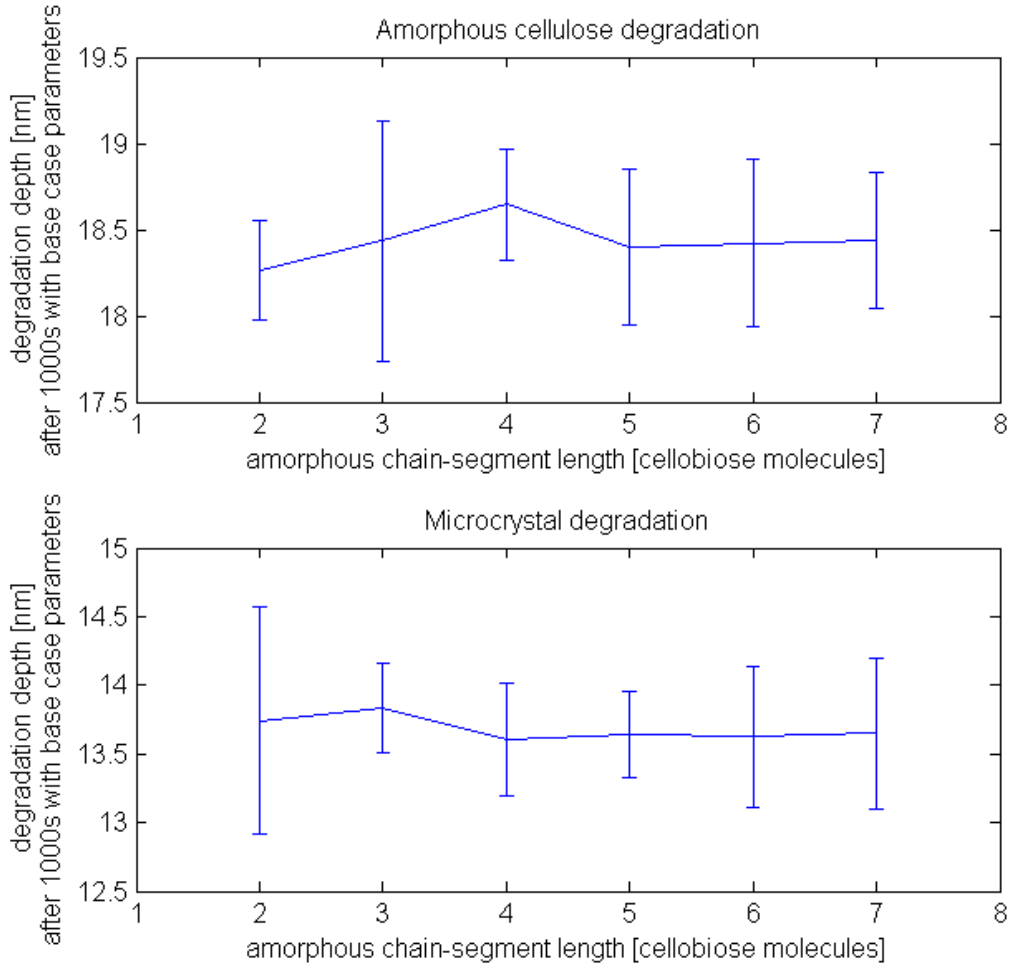

Figure 24: Parameter sensitivity analysis of the amorphous chain–segment length, which is aligned in one direction. Influence is shown on amorphous cellulose (top) and microcrystalline cellulose (bottom).

6. Discussion

The effect of slowly emerging crystals and the subsequent fast degradation can be explained by the synergistic work of different enzymes. Crystalline regions are uncovered from amorphous material by EG and CBH II and free crystalline chain ends are generated for CBH I, because only CBH I has its main activity on microcrystalline parts on the applied substrate as suggested by Ganner et al. (2012). These free chain ends are then slowly degraded by CBH I, although the CBH I concentration is high on the crystal (Figure 11). The reason for the relatively inefficient crystal degradation by CBH I is the high crowding, resulting in a 3.5 fold slower effective CBH I velocity on the crystal (Figure 22). This leads to aggregates of CBH I, which are dissolvable only after the maximal waiting time defined in the base case parameters (Table 1). Therefore, the crystal degradation is 1.5 fold slower than the amorphous degradation (Figure 10). This situation does not change until the crystalline material is fully degraded (~ 1600 s of Figure 10). Exactly in this moment, the fast degradation is initiated because the amorphous part, which is covered underneath the microcrystal (< 70nm of Figure 10), becomes accessible for CBH II and EG. The reason why exactly this amorphous part is subsequently degraded faster than the surrounding amorphous cellulose is the higher accessible surface for enzymatic attack, as shown in Figure 25. The increase in the length of the error bars of the projected crystalline area at points under the microcrystal (< 70nm in Figure 10) arises from the fact that some parts of the projected crystalline area are degraded prior to others. Thereby the height variance increases resulting in longer error bars. On the whole, the observable rapid decrease arises from the synergism of all three enzymes. EG and CBH II "polish" microcrystals and thereby provide reducing crystalline ends for CBH I, which slowly uncovers amorphous parts previously protected by microcrystals. This residual amorphous part is then degraded rapidly because it exposes high specific surface for EG and CBH II attack. The remarkable observation is that this kind of synergism was predicted by the model, solely applying a diffusion reaction system with detailed description of enzymes and the substrate, and no parameter or function specifically describing this phenomenon was implemented. Therefore an object oriented program allows for studying interactions between all acting components, which cannot be explained intuitively.

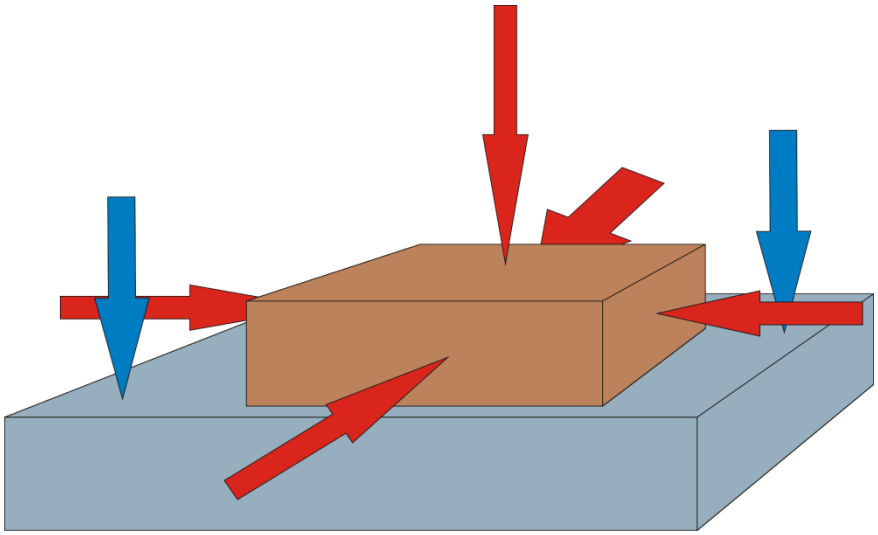

Figure 25: Specific surface of amorphous parts (brown box), uncovered from crystalline cellulose. Blue box represents amorphous cellulose, which was continuously degraded and never beneath any crystal. Red arrows indicate the areas which are accessible for enzymatic attack of the residual amorphous material after crystal degradation, blue arrows indicate the areas of possible enzymatic attack of amorphous material never covered by crystalline material.

As found in a recent publication of (Igarashi et al., 2006), the specific activity of cellulases decreases with increased surface density (here surface coverage). As shown by the model (Figure 14), a decrease of specific activity when increasing surface coverage is prevalent especially on crystalline cellulose, which fits to the experimental data, since (Igarashi et al., 2006) used highly crystalline cellulose. So far, it was not possible to give an experimental evidence for the reduced enzyme velocity with increased surface coverage. This could be investigated with high resolution (HS) AFM, which was performed by Igarashi et al. (2011). In this work a stop–and–go behaviour of CBH I was characterised. In accordance with the current results, an increase in the "stop" periods should be visible when increasing the surface coverage.

In general, parameter sensitivity analysis show rather high error bars, arising from repeated simulations. However, this can be explained by the randomness of the random walk, and the standard deviations, shown by the error bars, do not exceed 10% of the original value.

Numerous input parameters have a crucial effect on the model output, and most of them are known thanks to biochemical and AFM studies of the past. For example, the reaction velocity of CBH I was measured by AFM and the reaction rate of EG was measured experimentally by reducing sugar assays. On the other hand, some of the

parameters do not influence the model output significantly, such as the internal reaction velocity of EG on microcrystalline cellulose or the processive length of CBH I. Unfortunately, some of the crucial parameters are not elucidated yet and can only be assumed, such as the maximal waiting time of CBH I due to collision. However, I want to state that the overall trend and observations on the nanoscale would remain unchanged when considering different values, for example for the maximal waiting time of CBH I. This can be illustrated by the fact, that a decrease in the maximal waiting of CBH I would increase the degradation rate of CBH I only continuously. This is shown in Figure 26, where the time dependent degradation profile of different maximal waiting times was investigated. Therefore the slow emerging and fast degradation of microcrystals cannot be achieved by altering the maximal waiting time of CBH I in any direction.

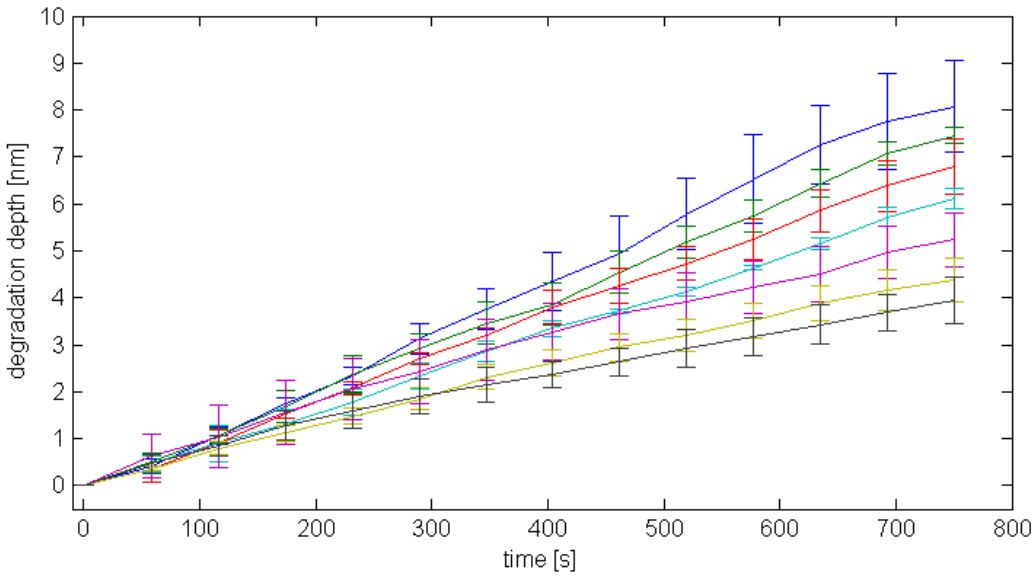

Figure 26: Time dependent degradation profile of different maximal waiting times ranging from 1 to 20 s (1, 2, 5, 7, 10, 15 and 20 s). Different maximal waiting times are indicated by different colours of the lines. The smaller the slope, the higher the maximal waiting time.

The attenuating effect of an increased amorphous reaction duration of EG on the crystal degradation might sound contradictory, since the kinetic parameters were kept unchanged for the crystal degradation (Figure 18). This can be explained by the fact that an increased reaction duration on amorphous cellulose decreases the mobility of enzymes on the amorphous cellulose. Thereby EGs accumulate on the amorphous part and a decreased EG concentration on the crystal is prevalent (Figure 27). Therefore, the crystalline degradation is decreased although the kinetic parameters did not change for this part. This is a good example of how close the diffusion and

reaction aspects are entangled with each other. Furthermore, the slowdown of EG activity on amorphous cellulose limits the generation of free crystal ends, which are initiation points for CBH I. Thereby the substrate accessibility for CBH I is reduced, decreasing the crystalline degradation as well.

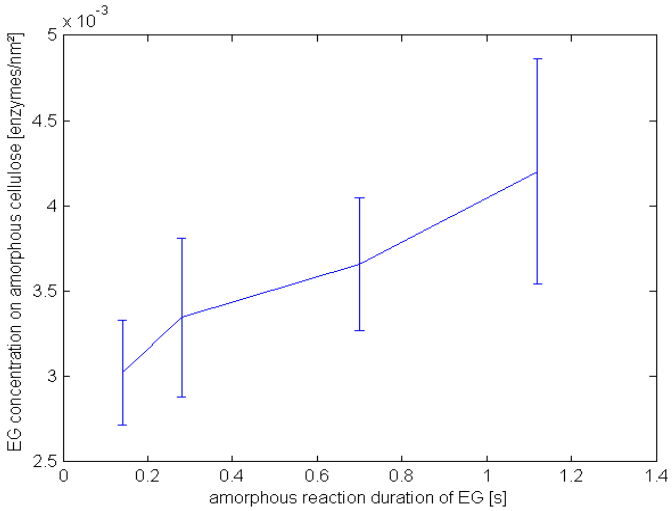

Figure 27: EG concentration on amorphous cellulose for different values of amorphous reaction duration of EG.

The described phenomenon for the increased amorphous reaction duration of EG is true for the increased intrinsic reaction duration of EG on microcrystalline cellulose the other way round. Due to a long intrinsic reaction duration the EG is more immobile on the crystal, which results in a high EG concentration on the crystal and therefore a lower EG concentration on the amorphous substrate part (Figure 28). This explains the attenuated degradation of the amorphous cellulose when the intrinsic reaction duration of EG is increased to higher levels (> 1.5 s).

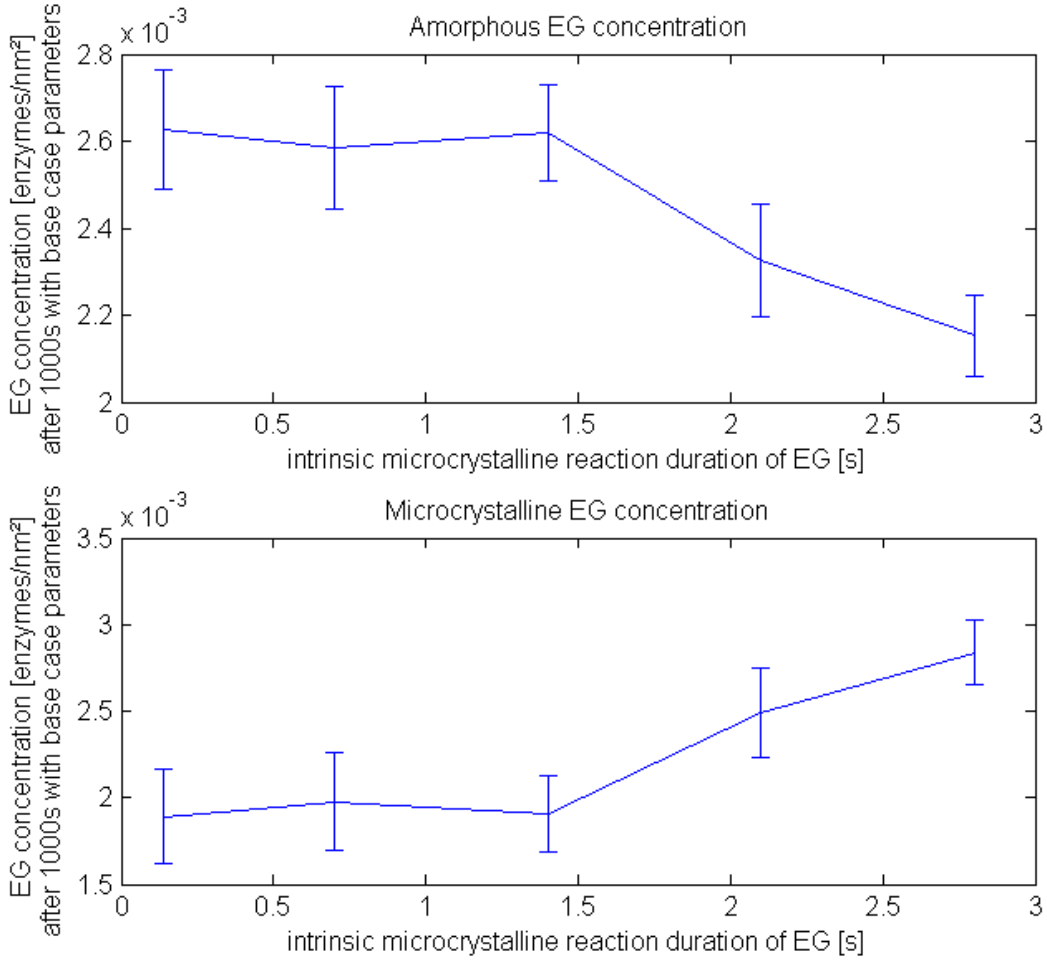

Figure 28: EG concentration on amorphous (top) and crystalline (bottom) cellulose for different values of the intrinsic microcrystalline reaction duration of EG (inverse of reaction velocity $v_{intrinsic}$ of equation1)

With the current settings, the length of the amorphous cellulose chain segment pointing into the same direction was not influential on the model output (Figure 24). This can be explained by the random action of EG creating free cellulose chain ends homogenously distributed over the whole cellulose surface. This leads to rather short chain fragments independent of the previously set segment length.

7. Conclusion

The aim of this work was to develop an object oriented model describing the action of cellulases on the cellulose surface on a nanometre scale. The outstanding result was that the little amount of physical and biochemical input parameters made it possible to perceive similar degradation patterns on the nanoscale and synergistic effects of cellulases, which lead to accelerated reaction rates. Moreover, the model could reproduce experimental findings, such as the reduced specific activities when increasing the surface coverage, and derive reasonable mean specific activities for the main cellulases from this bottom–up–approach.

Recent work of AFM measurements made it possible to follow the enzymatic action on the nanoscale experimentally. The visual model output reproduces the time dependent emerging and fast degradation of microcrystals as observed by AFM measurements. Moreover, the height profile in respect to the maximal width of the microcrystal was superimposable with the model. The reason for the fast degradation of microcrystals was ascribed to the action of EG and CBH II, due to the fact that the residual amorphous cellulose underneath the crystal exposes high specific surface. Moreover, it was found that high crowding of processive CBH I molecules on the microcrystals made it impossible for CBH I to initiate the fast crystal degradation.

The experimental observation, that increased surface coverage decreases specific enzyme activity, was also found in the model, and could be assigned to high crowding of enzymes, especially reducing the mobility and efficiency on microcrystalline cellulose.

Overall specific activity, gained by the model with virtual hydrolysis studies, was higher than reported literature values for Avicel and lower than values reported for purely amorphous cellulose. This is reasonable, since the modelled substrate consisted of amorphous and crystalline regions. A linear relation between soluble cellobiose content and reaction time was found, since neither an inhibition effect nor a decreased substrate accessibility were implemented. The linear behaviour fits to experimental hydrolysis data of the modelled substrate in time spans applied for AFM measurements (3–5 h). Furthermore, synergism experiments of the binary combinations of CBH I + CBH II showed a maximal synergism factor of 2.2, which is in excellent agreement to reported literature values. Moreover, the optimal mixing ratio of CBH I and CBH II was in perfect comparison to experimental values. The

synergism of CBH I and CBH II was higher than the other binary combinations. Synergism experiments containing EG (EG + CBH I and EG + CBH II) reproduced slightly lower (10 – 20 %) synergism factors than reported ones. However, the simulated optimal mixing ratios, leading to maximal synergism factors, of the binary combinations of EG + CBH I and EG + CBH II were identical to reported values.

8. Outlook

The assumed microcrystalline cross section in the model was spherical, which could not be proven experimentally. A specific experimental design of microcrystals with a defined cross section achieved by a different substrate preparation method may overcome this uncertainty. Subsequent experiments will show whether the fast crystal degradation of the projected area of the microcrystal arises from the fast degradation of amorphous material underneath the crystal (proposed in this work), or from crystalline material itself. The latter one could only be achieved by increasing CBH I activity on microcrystals, since it was found that mainly CBH I is active on this substrate part (Ganner et al., 2012). As shown in this work, the major rate retarding factor of CBH I activity on microcrystals is the high crowding (Figure 11 und Figure 22). Therefore a time dependent rate acceleration of CBH I is only possible by a reducing the crowding, which in turn can only be achieved by loosening the crystal structure; this is called amorphogenesis, and was previously found to be the most prominent overall rate acceleration factor (Arantes and Saddler, 2010).

The implementation of amorphogenesis would require the following steps: The CBMs of all enzymes can increase values of an additional matrix (called "amorphogenesis matrix" analogous to the crystal orientation matrix in Figure 2), when they slide over crystalline stacked cellulose chains. The higher these values, the looser bound the crystalline chain. It is assumed that looser bound cellulose chains, spatially separated from each other, provide more specific surface for enzymatic attack. However, it would be rather difficult to implement spatial reorganisation of cellulose chains. An alternative way of implementing the higher specific accessibility of cellulose chains would be the reduction of the size of the enzymes. This leads to the same phenomenon, where smaller enzymes find more space to attach and move along a cellulose chain. The size of the enzymes should be reduced by the local value of the "amorphogenesis matrix".

By applying this experimental and modelling strategy, it would not only be possible to elucidate whether amorphogenesis is responsible for the fast microcrystalline degradation or not, but also to ascribe specific kinetic values to the action of the CBMs of different enzymes, which equals to the influencing effect of CBMs onto the values of the "amorphogenesis matrix". This will enable us to find parameters influential on the amorphogenesis kinetics, such as enzyme size, size of the CBMs

and diffusion parameters. Moreover, it will be possible to predict how these parameters should be changed in order to achieve an overall effective cellulose degradation.

A minor limitation of the program is that the velocity of surface diffusion is as slow as the actual reaction velocity. This implies that it takes rather long for enzymes to diffuse to a suitable reaction site. A study of (Jervis et al., 1997) showed that effective diffusion coefficients of recombinant CBMs (10^{-11} cm^2/s) are 100 times higher than those reported for catalytically active cellulases (CBM + CD) at 23 – 45 ° C (10^{-13} cm^2/s) (Moran-Mirabal et al., 2013). However, it is not known how often a potential reaction is initiated when a suitable reaction site is prevalent. In the proposed model of this work, reactions are immediately initiated when a appropriate reaction site is prevalent, which compensates the limitation to a certain extent.

9. References

Abuja PM, Schmuck M, Pilz I, Tomme P, Claeyssens M, Esterbauer H. 1988. Structural and functional domains of cellobiohydrolase I from *Trichoderma reesei*. *European Biophysics Journal* **15**:339–342.

Arantes V, Saddler JN. 2010. Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. *Biotechnology for Biofuels* **3**:4.

Bansal P, Hall M, Realff MJ, Lee JH, Bommarius AS. 2009. Modeling cellulase kinetics on lignocellulosic substrates. *Biotechnology Advances* **27**:833–848.

Beckham GT, Bomble YJ, Bayer EA, Himmel ME, Crowley MF. 2011. Applications of computational science for understanding enzymatic deconstruction of cellulose. *Current opinion in biotechnology* **22**:231–238.

Beckham GT, Matthews JF, Bomble YJ, Bu L, Adney WS, Himmel ME, Nimlos MR, Crowley MF. 2010. Identification of amino acids responsible for processivity in a family 1 carbohydrate-binding module from a fungal cellulase. *The Journal of Physical Chemistry B* **114**:1447–1453.

Bhat MK. 2000. Cellulases and related enzymes in biotechnology. *Biotechnology advances* **18**:355–383.

Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochemical Journal* **382**:769–781.

Bubner P, Dohr J, Plank H, Mayrhofer C, Nidetzky B. 2012. Cellulases dig deep. In situ observation of the mesoscopic structural dynamics of enzymatic cellulose degradation. *Journal of Biological Chemistry* **287**:2759–2765.

Bubner P, Plank H, Nidetzky B. 2013. Visualizing cellulase activity. *Biotechnology and Bioengineering* **110**:1529–1549.

Chundawat SPS, Bellesia G, Uppugundla N, da Costa Sousa L, Gao D, Cheh AM, Agarwal UP, Bianchetti CM, Phillips GN, Langan P, Balan V, Gnanakaran S, Dale BE. 2011. Restructuring the crystalline cellulose hydrogen bond network enhances its depolymerization rate. *Journal of the American Chemical Society* **133**:11163–11174.

Divne C, Ståhlberg J, Reinikainen T, Ruohonen L, Pettersson G, Knowles JKC, Teeri TT, Jones TA. 1994. The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science* **265**:524–528.

Dowd MK, French AD, Reilly PJ. 1992. Conformational analysis of the anomeric forms of sophorose, laminarabiose, and cellobiose using MM3. *Carbohydrate Research* **233**:15–34.

Ganner T, Bubner P, Eibinger M, Mayrhofer C, Plank H, Nidetzky B. 2012. Dissecting and reconstructing synergism. In situ visualisation of cooperativity among cellulases. *Journal of Biological Chemistry* **287**:43215–43222.

García R. 2010. Amplitude modulation AFM in liquid. In: . *Amplitude Modulation Atomic Force Microscopy*. Wiley-VCH Verlag GmbH & Co. KGaA.

Goyal A, Ghosh B, Eveleigh D. 1991. Characteristics of fungal cellulases. *Bioresource Technology* **36**:37–50.

Hamby DM. 1994. A review of techniques for parameter sensitivity analysis of environmental models. *Environmental Monitoring and Assessment* **32**:135–154.

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

Horn SJ, Sørlie M, Vårum KM, Väljamäe P, Eijsink VGH. 2012. Chapter five - measuring processivity. In: Harry J. Gilbert, editor. *Methods in Enzymology*. Academic Press, Vol. 510, pp. 69–95.
Hoshino E, Shiroishi M, Amano Y, Nomura M, Kanda T. 1997. Synergistic actions of exo-type cellulases in the hydrolysis of cellulose with different crystallinities. *Journal of Fermentation and Bioengineering* **84**:300–306.
Igarashi K, Uchihashi T, Koivula A, Wada M, Kimura S, Okamoto T, Penttila M, Ando T, Samejima M. 2011. Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface. *Science* **333**:1279–1282.
Igarashi K, Koivula A, Wada M, Kimura S, Penttilä M, Samejima M. 2009. High speed atomic force microscopy visualizes processive movement of *Trichoderma reesei* cellobiohydrolase I on crystalline cellulose. *The Journal of biological chemistry* **284**:36186–90.
Igarashi K, Wada M, Hori R, Samejima M. 2006. Surface density of cellobiohydrolase on crystalline celluloses. *FEBS Journal* **273**:2869–2878.
Ioelovich M, Leykin A, Figovsky O. 2010. Study of cellulose paracrystallinity. *BioResources* **5**:1393–1407.
Irwin DC, Spezio M, Walker LP, Wilson DB. 1993. Activity studies of eight purified cellulases: specificity, synergism, and binding domain effects. *Biotechnology and bioengineering* **42**:1002–1013.
Jalak J, Kurašin M, Teugjas H, Väljamäe P. 2012. Endo-exo synergism in cellulose hydrolysis revisited. *Journal of Biological Chemistry* **287**:28802–28815.
Jalak J, Väljamäe P. 2010. Mechanism of initial rapid rate retardation in cellobiohydrolase catalyzed cellulose hydrolysis. *Biotechnology and Bioengineering* **106**:871–883.
Jervis EJ, Haynes CA, Kilburn DG. 1997. Surface diffusion of cellulases and their isolated binding domains on cellulose. *Journal of Biological Chemistry* **272**:24016–24023.
Karlsson J, Siika-aho M, Tenkanen M, Tjerneld F. 2002. Enzymatic properties of the low molecular mass endoglucanases Cel12A (EG III) and Cel45A (EG V) of *Trichoderma reesei*. *Journal of Biotechnology* **99**:63–78.
Klemm D, Philipp B, Heinze T, Heinze U, Wagenknecht W. 2004. Comprehensive cellulose chemistry: fundamentals and analytical methods. *Comprehensive Cellulose Chemistry*. Wiley-VCH Verlag GmbH & Co. KGaA, pp. 1–7.
Kumar D, Murthy GS. 2013. Stochastic molecular model of enzymatic hydrolysis of cellulose for ethanol production. *Biotechnology for Biofuels* **6**:63.
Kurašin M, Väljamäe P. 2011. Processivity of cellobiohydrolases is limited by the substrate. *Journal of Biological Chemistry* **286**:169–177.
Levine SE, Fox JM, Blanch HW, Clark DS. 2010. A mechanistic model of the enzymatic hydrolysis of cellulose. *Biotechnology and Bioengineering* **107**:37–51.
Linder M, Teeri TT. 1996. The cellulose-binding domain of the major cellobiohydrolase of *Trichoderma reesei* exhibits true reversibility and a high exchange rate on crystalline cellulose. *PNAS* **93**:12251–12255.
Lynd LR, Laser MS, Bransby D, Dale BE, Davison B, Hamilton R, Himmel M, Keller M, McMillan JD, Sheehan J. 2008. How biotech can transform biofuels. *Nature biotechnology* **26**:169–172.
Matthews JF, Skopec CE, Mason PE, Zuccato P, Torget RW, Sugiyama J, Himmel ME, Brady JW. 2006. Computer simulation studies of microcrystalline cellulose Iβ. *Carbohydrate Research* **341**:138–152.

Maurer SA, Brady NW, Fajardo NP, Radke CJ. 2013. Surface kinetics for cooperative fungal cellulase digestion of cellulose from quartz crystal microgravimetry. *Journal of Colloid and Interface Science* **394**:498–508.
Maurer SA, Bedbrook CN, Radke CJ. 2012. Competitive sorption kinetics of inhibited endo- and exoglucanases on a model cellulose substrate. *Langmuir* **28**:14598–14608.
Mazeau K, Heux L. 2003. Molecular dynamics simulations of bulk native crystalline and amorphous structures of cellulose. *The Journal of Physical Chemistry B* **107**:2394–2403.
Medve J, Ståhlberg J, Tjerneld F. 1994. Adsorption and synergism of cellobiohydrolase I and II of *Trichoderma reesei* during hydrolysis of microcrystalline cellulose. *Biotechnology and Bioengineering* **44**:1064–1073.
Moon RJ, Martini A, Nairn J, Simonsen J, Youngblood J. 2011. Cellulose nanomaterials review: structure, properties and nanocomposites. *Chemical Society Reviews* **40**:3941.
Moran-Mirabal JM, Bolewski JC, Walker LP. 2011. Reversibility and binding kinetics of *Thermobifida fusca* cellulases studied through fluorescence recovery after photobleaching microscopy. *Biophysical Chemistry* **155**:20–28.
Moran-Mirabal JM, Bolewski JC, Walker LP. 2013. *Thermobifida fusca* cellulases exhibit limited surface diffusion on bacterial micro-crystalline cellulose. *Biotechnology and Bioengineering* **110**:47–56.
Nidetzky B, Steiner W, Claeyssens M. 1994a. Cellulose hydrolysis by the cellulases from *Trichoderma reesei*: adsorptions of two cellobiohydrolases, two endocellulases and their core proteins on filter paper and their relation to hydrolysis. *Biochem J* **303**:817–823.
Nidetzky B, Hayn M, Macarron R, Steiner W. 1993. Synergism of *Trichoderma reesei* cellulases while degrading different celluloses. *Biotechnology Letters* **15**:71–76.
Nidetzky B, Steiner W, Hayn M, Claeyssens M. 1994b. Cellulose hydrolysis by the cellulases from *Trichoderma reesei*: a new model for synergistic interaction. *Biochemical Journal* **298**:705.
O'sullivan AC. 1997. Cellulose: the structure slowly unravels. *Cellulose* **4**:173–207.
Ogeda TL, Silva IB, Fidale LC, El Seoud O a, Petri DFS. 2012. Effect of cellulose physical characteristics, especially the water sorption value, on the efficiency of its hydrolysis catalyzed by free or immobilized cellulase. *Journal of biotechnology* **157**:246–52.
Scheiding W, Thoma M, Ross A, Schügerl K. 1984. Modelling of the enzymatic hydrolysis of cellobiose and cellulose by a complex enzyme mixture of *Trichoderma reesei* QM 9414. *Applied Microbiology and Biotechnology* **20**:176–182.
Stålbrand H, Mansfield SD, Saddler JN, Kilburn DG, Warren RAJ, Gilkes NR. 1998. Analysis of molecular size distributions of cellulose molecules during hydrolysis of cellulose by recombinant *Cellulomonas fimi* β-1,4-glucanases. *Applied and Environmental Microbiology* **64**:2374–2379.
Suga K, van Dedem G, Moo-Young M. 1975. Degradation of polysaccharides by endo and exo enzymes: A theoretical analysis. *Biotechnology and Bioengineering* **17**:433–439.
Sugimoto N, Igarashi K, Wada M, Samejima M. 2012. Adsorption characteristics of fungal family 1 cellulose-binding domain from *Trichoderma reesei* cellobiohydrolase I on crystalline cellulose: Negative cooperative adsorption via a steric exclusion effect. *Langmuir* **28**:14323–14329.

Varrot A, Frandsen TP, von Ossowski I, Boyer V, Cottaz S, Driguez H, Schülein M, Davies GJ. 2003. Structural basis for ligand binding and processivity in cellobiohydrolase Cel6A from *Humicola insolens*. *Structure* **11**:855–864.
Varrot A, Schülein M, Davies GJ. 1999. Structural changes of the active site tunnel of *Humicola insolens* cellobiohydrolase, Cel6A, upon oligosaccharide binding. *Biochemistry* **38**:8884–8891.
Warden AC, Little BA, Haritos VS. 2011. A cellular automaton model of crystalline cellulose hydrolysis by cellulases. *Biotechnology for biofuels* **4**:39.
Xiao Z, Storms R, Tsang A. 2004. Microplate-based filter paper assay to measure total cellulase activity. *Biotechnology and Bioengineering* **88**:832–837.
Zhang Y-HP, Lynd LR. 2005. Determination of the number-average degree of polymerization of cellodextrins and cellulose with application to enzymatic hydrolysis. *Biomacromolecules* **6**:1510–1515.
Zhang Y-HP, Lynd LR. 2004. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems. *Biotechnology and Bioengineering* **88**:797–824.
Zhou W, Schüttler H-B, Hao Z, Xu Y. 2009. Cellulose hydrolysis in evolving substrate morphologies I: A general modeling formalism. *Biotechnology and Bioengineering* **104**:261–274.

Attachment

In this section, the most decisive functions and code examples from the main script are shown. This comprises the functions *EG_slide*, responsible for the random walk of all enzymes on the surface, and *CBH_processive_motion*, handling the processive motion of CBH I and CBH II. Moreover, two excerpts from the main executing script, handling the chronological action of all enzymes, will be presented by showing the code of how EG and CBH I are processed within one Δt. The handling of CBH II during one Δt is analogous to CBH I, except for the fact that CBH II is solely active on amorphous cellulose and degrades cellulose chains from its non reducing ends.

The source code is presented with comments (green text colour).

1.) EG_slide

```
function [new_x,new_y,new_z ] = EG_slide( intern_pos, E_C_near_ext, Ex, Ey, Ez,
area, intern_treshold, E_radius_sqr)
% EG_SLIDE takes previous enzyme positions and surrounding, and returns new
% enzyme position gained by a random walk algorithm
%   This function is responsible for calculating a random walk for enzymes
%   diffusing on a cellulose surface.

% calculating a random number between [0 and 6)
r = 6*rand;
% assign previous enzyme position to internal variables
intern_x = intern_pos(1);
intern_y = intern_pos(2);
intern_z = intern_pos(3);

% assign internal variables to output variables for cases where no
% diffusion is takes place due to obstacles or hindrance by other enzymes
new_x = intern_x;
new_y = intern_y;
new_z = intern_z;

% if random number is between 0 and 1, the x value of the enzyme will be
% increased
if r >= 0 && r < 1
    x_tmp = intern_x + 1;
    ForwWall = E_C_near_ext(end-1,3:end-2,3:end-2);%z wall starts at 2
    EVol = E_C_near_ext(3:end,2:end-1,2:end-1);

    % unless there is a cellulose wall in front and the enzyme does not try to
    % desorb from the surface, the program may proceed
    if sum(ForwWall(:)) < intern_treshold && sum(EVol(:)) > 0

        % if the enzymes passes the right end of the simulation, it will be set
        % to the outer left end
        if x_tmp > area(1)
            x_tmp = 1;
        end

        % enzymes which are closer to the current enzyme by taking a shortcut
        % through one edge of the simulation have to be edited in their
        % position. This ensures that all enzymes appear with their shortest
        % distance to the currently processed enzyme.
        if x_tmp < area(1)/2
            Ex(Ex - x_tmp>area(1)/2) = Ex(Ex - x_tmp>area(1)/2) - area(1);
        else
            Ex(x_tmp - Ex>area(1)/2) = Ex(x_tmp - Ex>area(1)/2) + area(1);
        end

        if intern_y < area(2)/2
            Ey(Ey - intern_y>area(2)/2) = Ey(Ey - intern_y>area(2)/2) - area(2);
```

```
        else
            Ey(intern_y -Ey>area(2)/2) = Ey(intern_y - Ey>area(2)/2) + area(2);
        end

        if intern_z < area(3)/2
            Ez(Ez - intern_z>area(3)/2) = Ez(Ez - intern_z>area(3)/2) - area(3);
        else
            Ez(intern_z - Ez>area(3)/2) = Ez(intern_z - Ez>area(3)/2) + area(3);
        end

        % calculate the shortest distance of all enzymes to the current one
        dist = (x_tmp - Ex).^2 + (intern_y - Ey).^2 + (intern_z - Ez).^2;

        % unless there is another enzyme within the radius of the enzyme,
        % the temporary position will be assigned to the new position
        if ~any(dist > 1 & dist < E_radius_sqr)
            new_x = x_tmp;
        end
    end
end
% the above section will be repeated for random numbers between 1 and 6
% with a stepsize of 1, leading to an enzyme diffusion in -x, y, -y, z, and
% -z direction, respectively
```

2.) *CBH_processive_motion*

```
function [ new_x, new_y, new_z, wait,  processive_period,
possible_reacting_cellulose ] = CBH_processive_motion( pos, dir, E_C_near_ext,
internal_react_pos, wait, processive_period, possible_reacting_cellulose, Ex, Ey,
Ez, treshold, E_radius_sqr, area, max_waiting_time)
%CBH_PROCESSIVE_MOTION responsible for the processive motion of CBH I and
%CBH II
%   Detailed explanation goes here

% assign the previous enzyme position to local variables
x = pos(1);
y = pos(2);
z = pos(3);

% assign internal variables to output variables for cases where no
% diffusion takes place due to obstacles or hindrance by other enzymes
new_x = x;
new_y = y;
new_z = z;

% direction "1" is defined as the processive motion in x-direction
if dir == 1
    x_tmp = x + 1;
    ForwWall = E_C_near_ext(end-1,3:end-2,3:end-2);%z wall starts at 2
    EVol = E_C_near_ext(3:end,2:end-1,2:end-1);

    % if the enzyme passes the right end of the simulation, it will be set
    % to the outer left end
    if x_tmp > area(1)
        x_tmp = 1;
    end

    % enzymes which are closer to the current enzyme by taking a shortcut
    % through one edge of the simulation have to be edited in their
    % position. This ensures that all enzymes appear with their shortest way
    % to the currently processed enzyme.
    if x_tmp < area(1)/2
        Ex(Ex - x_tmp>area(1)/2) = Ex(Ex - x_tmp>area(1)/2) - area(1);
    else
        Ex(x_tmp - Ex>area(1)/2) = Ex(x_tmp - Ex>area(1)/2) + area(1);
    end

    if y < area(2)/2
        Ey(Ey - y>area(2)/2) = Ey(Ey - y>area(2)/2) - area(2);
    else
        Ey(y -Ey>area(2)/2) = Ey(y - Ey>area(2)/2) + area(2);
    end

    if z < area(3)/2
        Ez(Ez - z>area(3)/2) = Ez(Ez - z>area(3)/2) - area(3);
    else
        Ez(z - Ez>area(3)/2) = Ez(z - Ez>area(3)/2) + area(3);
    end

    % calculate the shortest distance of all enzymes to the current one
    dist = (x_tmp - Ex).^2 + (y - Ey).^2 + (z - Ez).^2;

    % unless there is a cellulose wall in front, the enzyme does not try to
    % desorb from the surface, and unless a distance is shorter than the enzyme
    % radius --> the temporary x position can be assigned to the new x
    % position of the enzyme
    if sum(ForwWall(:)) < treshold && sum(EVol(:)) > 0 && ~any(dist > 1 & dist <
E_radius_sqr) ...
            && E_C_near_ext(1 + internal_react_pos(1) + 1,...
            1 +  internal_react_pos(2),...
            1 +  internal_react_pos(3));
        new_x = x_tmp;
```

```
        % the next cellobiose molecules are loaded into the molecular tunnel
        possible_reacting_cellulose(1) = possible_reacting_cellulose(1) + 1;
        if possible_reacting_cellulose(1) > area(1)
            possible_reacting_cellulose(1) = 1;
        end
    else
        % if there is an obstacle, such as another enzyme, the enzyme
        % waits until a maximal waiting time is reached and the processive
        % motion is stopped
        if any(dist > 1 & dist < E_radius_sqr)
            %disp('Wait due to traffic jam')
            wait = wait + 1;
            if wait > max_waiting_time
                processive_period = 0;
                wait = 0;
            end
        else
            % if the enzyme hits a cellulose wall, its processive
            % motion will be stopped immediately
            processive_period = 0;
            %disp('Stopped processive motion due to wall or no cellulose left or no
longer attached')
        end
    end
end
% the above section will be repeated for directions "-1","2" and "-2",
% which represent -x,y and -y direction, respectively
```

3.) Code for EG in the *main* script for on Δt

```
% cycle through all enzymes
for n =1:size(E,1)
    % allocating the position of the n-th enzyme
    Ey = E{n,pos_prop}(2);
    Ex = E{n,pos_prop}(1);
    Ez = E{n,pos_prop}(3);
    % creating a x,y and z vector representing the footprint of the
    % enzyme in x, y and z direction
    y =  Ey - (Eydim-1)/2  - 1: Ey + (Eydim-1)/2 + 1;
    x = Ex - (Exdim-1)/2 - 1: Ex + (Exdim-1)/2 + 1;
    z = Ez - (Ezdim-1)/2 - 1: Ez + (Ezdim-1)/2 + 1;

    % ensuring the infinity of the matrix
    y(y<=0) = y(y<=0) + area(2);
    y(y>=area(2)+1) = y(y>=area(2)+1) - area(2);
    x(x<=0) = x(x<=0) + area(1);
    x(x>= area(1)+1) = x(x>= area(1)+1) - area(1);
    z(z<=0) = z(z<=0) + area(3);
    z(z>= area(3)+1) = z(z>= area(3)+1) - area(3);

    % extracting cellobiose molecules, which can be reached by the
    % enzyme's footprint
    E_C_near = Cellulose(x(2:end-1),y(2:end-1),z(2:end-1));

    % extract all cellobiose molecules, which are potential reaction
    % partners

    possible_reacting_cellulose_IND = find(E_C_near ~= 0);
    if any(possible_reacting_cellulose_IND)
        % randomly take one of these cellobiose molecules as a probable
        % reaction partner
        no_react_cellulose = ...
            round(1+(numel(possible_reacting_cellulose_IND)-1)*rand);
        [xreact,yreact,zreact] = ind2sub(size(E_C_near),...
            possible_reacting_cellulose_IND(no_react_cellulose));
        E{n,possible_reacting_cellulose_prop} = ...
            [x(xreact+1),y(yreact+1),z(zreact+1)]';
    end

    % calculate the amount of cellobiose molecules, to which the enzyme
    % is attached
    E_C_near_ext = Cellulose(x,y,z) ~= 0 & possible_mov;
    E{n,E_C_near_ext_prop} = E_C_near_ext;
    E_C_neighbours(n) = sum(E_C_near_ext(:));
end

rand_react = rand(1,size(E,1));

% create a logical matrix of enzymes still reacting on one cellobiose
% molecule
L_still_react = [E{:,react_time_prop}] > 0;
% decrease the reaction time of these enzymes by dt
E_new_react_time = num2cell(cellfun(@(x) ...
    EG_react(x,dtr),E(L_still_react,react_time_prop)));
[E{L_still_react,react_time_prop}] = E_new_react_time{:};

% determine the enzymes, which do no longer react with cellobiose
% molecules
L_start_react = ~L_still_react;

% create a vector of possible reacting cellobiose molecules of all
% enzymes
react_pos = [E{:,possible_reacting_cellulose_prop}]';

if any(react_pos)
    % derive the index of the reacting cellobiose molecules from x, y
    % and z coordinates
```

```
    react_IND = sub2ind(size(Cellulose),...
        react_pos(:,1),react_pos(:,2),react_pos(:,3))';
    if AE
        % create a logical matrix containing enzymes with higher
        % reaction probability than a threshold. This enables them to
        % start reacting on a crystalline cellobiose molecule
        L_E_active_on_crystal = (Cellulose(react_IND) == 2 | ...
            Cellulose(react_IND) == 3)& activation_energy > rand_react;
        L_start_react = ~L_still_react & (Cellulose(react_IND) == 1 |...
            L_E_active_on_crystal);
    end
    if any(L_start_react)
        % define a logical matrix of enzymes, which do not react and
        % start reacting with an amorphous cellobiose molecule
        L_start_on_Cellulose_1 = Cellulose(react_IND) == 1 & ...
            L_start_react;
        if any(L_start_on_Cellulose_1)
            [E{L_start_on_Cellulose_1,react_time_prop}] = ...
                deal(react_duration_EG_amorph);
            E_new_react_time = num2cell(cellfun(@(x) EG_react(x,dtr),...
                E(L_start_on_Cellulose_1,react_time_prop)));
            [E{L_start_on_Cellulose_1,react_time_prop}] = ...
                E_new_react_time{:};
            Cellulose(react_IND(L_start_on_Cellulose_1)) = 0;
            %free_amorph_ends(react_IND(L_start_on_Cellulose_1)) = -1;
        end
        % create a logical matrix of enzymes with a higher reaction
        % probability than a threshold and start working on a
        % crystalline cellobiose molecule
        L_start_on_crystal = (Cellulose(react_IND) == 2 | ...
            Cellulose(react_IND) == 3) & L_start_react;
        if any(L_start_on_crystal)
            if AE
                % if AE (activation energy model) is "on" the reaction
                % time of the enzymes, which start reacting on crystals
                % is set to the intrinsic reaction time
                [E{L_start_on_crystal,react_time_prop}] = ...
                    deal(react_duration_EG_crystal_intrinsic);
            else
                % not used any more
                [E{L_start_on_crystal,react_time_prop}] = ...
                    deal(react_duration_EG_crystal);
            end
            % reduce the reaction time by one dt (140ms)
            E_new_react_time = num2cell(cellfun(@(x) EG_react(x,dtr),...
                E(L_start_on_crystal,react_time_prop)));
            [E{L_start_on_crystal,react_time_prop}] = ...
                E_new_react_time{:};
            % delete cellulose molecules, which have been degraded by
            % the action of EG
            Cellulose(react_IND(L_start_on_crystal)) = 0;
        end

    end
end

% find enzymes which still react
L_still_react = [E{:,react_time_prop}] > 0;

% all enzymes which do not react any more are allowed to slide over the
% surface
Lslide = ~L_still_react;
Lslide_IND = find(Lslide);

%extract x, y and z data of all EGs, CBH Is, and CBH IIs and
%concatenate them in one variable (Epos)
EGpos = [E{:,pos_prop}]';
CBHpos = [CBH{:,pos_prop}]';
CBHIIpos = [CBHII{:,pos_prop}]';
```

```
Epos = [EGpos;CBHpos;CBHIIpos];

% save the enzymes' positions over time in a cell array
EGpost{1+round(t/dtr)} = EGpos;
CBHpost{1+round(t/dtr)} = CBHpos;
CBHIIpost{1+round(t/dtr)} = CBHIIpos;
%Epost{1+round(t/dtr)} = Epos;

% cycle through all EGs and let them slide according to a random walk
% using the function EG_slide
for k = 1:numel(Lslide_IND)
    [new_x,new_y,new_z] = EG_slide(E{Lslide_IND(k),pos_prop},...
        E{Lslide_IND(k),E_C_near_ext_prop}, Epos(:,1), Epos(:,2), ...
        Epos(:,3), area, EG_moving_treshold, EG_CBM_dim_sqr);
    E{Lslide_IND(k),pos_prop} = [new_x,new_y,new_z]';
end

% display how many enzymes will be deleted in this step due to loosing
% the attachment to the surface
if any(E_C_neighbours == 0)
    disp(['es werden ',num2str(sum(E_C_neighbours==0)),' Enzyme deleted'])
    deleted_EG = deleted_EG + 1;
end
%deletes all enzymes which are no longer attached to the cellulose
%surface
if any(E_C_neighbours == 0)
    E(E_C_neighbours == 0,:) = [];
end
```

4.) CBH I section in the *main* script for Δt

```
% cycle through all enzymes
for n =1:size(CBH,1)
    % allocate the position of the n-th CBH I
    CBHx = CBH{n,pos_prop}(1);
    CBHy = CBH{n,pos_prop}(2);
    CBHz = CBH{n,pos_prop}(3);
    % create a x, y and z vector representing the footprint of the
    % enzyme
    y = CBHy - (Eydim-1)/2 - 1 : CBHy + (Eydim-1)/2 + 1;
    x = CBHx - (Exdim-1)/2 - 1 : CBHx + (Exdim-1)/2 + 1;
    z = CBHz - (Ezdim-1)/2 - 1 : CBHz + (Ezdim-1)/2 + 1;

    % taking edge overlapping into account
    y(y<=0) = y(y<=0) + area(2);
    y(y>=area(2)+1) = y(y>=area(2)+1) - area(2);
    x(x<=0) = x(x<=0) + area(1);
    x(x>= area(1)+1) = x(x>= area(1)+1) - area(1);
    z(z<=0) = z(z<=0) + area(3);
    z(z>= area(3)+1) = z(z>= area(3)+1) - area(3);

    % extract the extended (=footprint + 1 cellobiose in each dimension)
    % cellulose surrounding of the enzyme from the
    % cellulose matter matrix and calculate the number of surrounding
    % cellobiose molecules
    E_C_near_ext = Cellulose(x,y,z) ~= 0 & possible_mov;
    CBH{n,E_C_near_ext_prop} = E_C_near_ext;
    CBH_neighbours(n) = sum(E_C_near_ext(:));

    % unless the CBH I is in a processive motion
    if CBH{n,processive_period_prop} <= 0
        % the internal waiting time is set to 0
        CBH{n,wait_prop} = 0;

        % extract the cellobiose molecules in the close surrounding
        % (under the footprint)
        E_C_near = Cellulose(x(2:end-1),y(2:end-1),z(2:end-1));

        % extract the local crystal orientation from the crystal
        % extension (=crystal orientation) matrix
        local_crystal_extension = ...
            Crystal_extension(x(2:end-1),y(2:end-1),z(2:end-1));

        % search for free ends in y direction
        local_endsy = [Z,diff(logical(E_C_near),1,2)];
        L = local_endsy == -1;
        local_endsy(L(:,2:end)) = -1;
        local_endsy(L) = 0;
        % exchange reducing and non reducing ends, when the crystals is
        % orientated upside down
        L31 = E_C_near == 3 & local_endsy == 1;
        L3m1 = E_C_near == 3 & local_endsy == -1;
        local_endsy(L31) = -1;
        local_endsy(L3m1) = 1;

        % search for free ends in x direction
        local_endsx = zeros(Exdim,Eydim,Ezdim);
        local_endsx(2:end,:,:) = diff(logical(E_C_near),1,1);
        L = local_endsx == -1;
        local_endsx(L(2:end)) = -1;
        local_endsx(L) = 0;
        % exchange reducing and non reducing ends, when the crystals is
        % orientated upside down
        L31 = E_C_near == 3 & local_endsx == 1;
        L3m1 = E_C_near == 3 & local_endsx == -1;
        local_endsx(L31) = -1;
        local_endsx(L3m1) = 1;
```

```
        % local ends can consist of x and y directed ends
        local_ends = local_endsx + 2*local_endsy;
        CBH{n,local_ends_prop} = local_ends;

        % find the position of all ends
        possible_reacting_cellulose_IND = find(( (local_endsx ~= 0 |...
            local_endsy ~= 0) & E_C_near == 1) | (local_endsx == 1 & ...
            local_crystal_extension == 1) | ...
            (local_endsy == 1 & local_crystal_extension == 2));
        if any(possible_reacting_cellulose_IND)
            % define one of the possible ends as a reaction partner
            no_react_cellulose = ...
                round(1+(numel(possible_reacting_cellulose_IND)-1)*rand);
            [xreact,yreact,zreact] = ind2sub(size(E_C_near),...
                possible_reacting_cellulose_IND(no_react_cellulose));
            if (Cellulose(x(xreact+1),y(yreact+1),z(zreact+1)) == 2 ...
                ||Cellulose(x(xreact+1),y(yreact+1),z(zreact+1)) == 3)...
                || rand < p_find_chain_CBHI

                CBH{n,internal_react_pos_prop} = [xreact,yreact,zreact]';
                CBH{n,possible_reacting_cellulose_prop} = ...
                    [x(xreact+1),y(yreact+1),z(zreact+1)]';
                % if a x end and y end are prevalent on the same
                % position and random direction for the processive
                % motion is chosen. Otherwise the direction is chosen
                % according to the crystal orientation
                if local_ends(xreact,yreact,zreact) == 3 ||...
                        local_ends(xreact,yreact,zreact) == -3;
                    if rand < 0.5
                        CBH{n,dir_prop} = 1;% in x direction
                    else
                        CBH{n,dir_prop} = 2;% y direction
                    end
                else
                    CBH{n,dir_prop} = local_ends(xreact,yreact,zreact);
                end
                if Cellulose(x(xreact+1),y(yreact+1),z(zreact+1)) == 2
                    if Crystal_extension(x(xreact+1),...
                            y(yreact+1),z(zreact+1)) == 2
                        CBH{n,dir_prop} = 2;
                    end
                    if Crystal_extension(x(xreact+1),...
                            y(yreact+1),z(zreact+1)) == 1
                        CBH{n,dir_prop} = 1;
                    end
                    CBH{n,processive_period_prop} = ...
                        processive_CBH_length_crystaline;
                end
                if Cellulose(x(xreact+1),y(yreact+1),z(zreact+1)) == 3
                    if Crystal_extension(x(xreact+1),...
                            y(yreact+1),z(zreact+1)) == 2
                        CBH{n,dir_prop} = -2;
                    end
                    if Crystal_extension(x(xreact+1),...
                            y(yreact+1),z(zreact+1)) == 1
                        CBH{n,dir_prop} = -1;
                    end
                    % the processive length is set to the initial value
                    CBH{n,processive_period_prop} = ...
                        processive_CBH_length_crystaline;
                end
                if Cellulose(x(xreact+1),...
                        y(yreact+1),z(zreact+1)) == 1
                    CBH{n,processive_period_prop} = processive_CBH_length;
                end
            end
        end
    else
        % if the processive motion is still ongoing and the reaction on
```

```
        % a single cellobiose molecule is finished
        if CBH{n,react_time_prop} <= 0
            % evaluate all enzymes' position and other properties
            % and save them to local variables
            EGpos = [E{:,pos_prop}]';
            CBHpos = [CBH{:,pos_prop}]';
            CBHIIpos = [CBHII{:,pos_prop}]';
            Epos = [EGpos;CBHpos;CBHIIpos];
            dir = CBH{n,dir_prop};
            E_C_near_ext_n = CBH{n,E_C_near_ext_prop};
            internal_reacting_cellulose_n = CBH{n,internal_react_pos_prop};
            wait_n = CBH{n,wait_prop};
            processive_period_n = CBH{n,processive_period_prop};
            possible_reacting_cellulose_n =
CBH{n,possible_reacting_cellulose_prop};
            % use the function CBH_processive_motion in order to
            % prolong the processive motion in the defined direction.
            % Thereby new enzyme positions and potential abortions are
            % generated and saved to the enzyme properties
            [new_x, new_y, new_z, wait,  processive_period, ...
                possible_reacting_cellulose] =
CBH_processive_motion(CBHpos(n,:),...
                dir,E_C_near_ext_n,internal_reacting_cellulose_n, wait_n,...
                processive_period_n, possible_reacting_cellulose_n, ...
                Epos(:,1), Epos(:,2), Epos(:,3), CBH_moving_treshold,...
                CBH_radius_sqr, area, max_waiting_time);
            CBH{n,pos_prop} = [new_x, new_y, new_z]';
            CBH{n,wait_prop} = wait;
            CBH{n,processive_period_prop} = processive_period;
            CBH{n,possible_reacting_cellulose_prop} = possible_reacting_cellulose;
        end
    end
end
% define a logical vector of CBH I molecules, which are still
% processive active
L_processive_active = [CBH{:,processive_period_prop}] > 0;
% define a logical vector of CBH I molecules, which are still working
% on a cellobiose molecule
L_still_react_CBH = [CBH{:,react_time_prop}] > 0;
% the reaction time of molecules still reacting is reduced
[CBH_new_react_time, L_finish] = cellfun(@(x) CBH_react(x,dtr),...
    CBH(L_still_react_CBH,react_time_prop));
temp_react_time = num2cell(CBH_new_react_time);
temp_finish = num2cell(L_finish);
[CBH{L_still_react_CBH,react_time_prop}] = temp_react_time{:};
[CBH{L_still_react_CBH,finish_prop}] = temp_finish{:};

% concatenate all positions of reacting cellobiose molecules in one
% array. Rows represent the CBH I molecules
react_pos_CBH = [CBH{:,possible_reacting_cellulose_prop}]';

% all CBH I molecules, which found a suitable reaction position
% (reducing end) will be saved in a logical vector.
% Their reaction time will be increased and the specified cellobiose
% molecule will be deleted
if any(react_pos_CBH)
    % convert the reaction position (x,y,z) into a reaction index
    react_IND_CBH = sub2ind(size(Cellulose),react_pos_CBH(:,1),...
        react_pos_CBH(:,2),react_pos_CBH(:,3))';
    L_CBH_react_on_amorph = ...
        Cellulose(react_IND_CBH) == 1 & L_processive_active;
    if any(L_CBH_react_on_amorph)
        [CBH{L_CBH_react_on_amorph,react_time_prop}] =...
            deal(react_duration_CBH_amorph);
        [CBH_new_react_time, L_finish] = cellfun(@(x) CBH_react(x,dtr),...
            CBH(L_CBH_react_on_amorph,react_time_prop));
        temp_react_time = num2cell(CBH_new_react_time);
        temp_finish = num2cell(L_finish);
        [CBH{L_CBH_react_on_amorph,react_time_prop}] = temp_react_time{:};
```

```
        [CBH{L_CBH_react_on_amorph,finish_prop}] = temp_finish{:};
        Cellulose(react_IND_CBH(L_CBH_react_on_amorph)) = 0;
    end

    L_CBH_react_on_crystal = (Cellulose(react_IND_CBH) == 2 |
Cellulose(react_IND_CBH) == 3) & L_processive_active;
    if any(L_CBH_react_on_crystal)
        [CBH{L_CBH_react_on_crystal,react_time_prop}] = deal(react_duration_CBH);
        [CBH_new_react_time, L_finish] = cellfun(@(x)
CBH_react(x,dtr),CBH(L_CBH_react_on_crystal,react_time_prop));
        temp_react_time = num2cell(CBH_new_react_time);
        temp_finish = num2cell(L_finish);
        [CBH{L_CBH_react_on_crystal,react_time_prop}] = temp_react_time{:};
        [CBH{L_CBH_react_on_crystal,finish_prop}] = temp_finish{:};
        Cellulose(react_IND_CBH(L_CBH_react_on_crystal)) = 0;
    end

end

% define CBH I molecules, which still react
L_still_react_CBH = [CBH{:,react_time_prop}] > 0;
% define CBH I molecules, which finished their reaction in this dt
L_finish_reaction = [CBH{:,finish_prop}];

% the processive period of all CBH I molecules, which finished their
% reaction in this dt, will be reduced by 1
if any(L_finish_reaction)
    temp_processive_period =
num2cell([CBH{L_finish_reaction,processive_period_prop}] - 1);
    [CBH{L_finish_reaction,processive_period_prop}] = temp_processive_period{:};
    [CBH{L_finish_reaction,finish_prop}] = deal(false);
end

% all CBH I molecules, which do not react anymore and do not perform
% a processive motion, are allowed to perform a random walk (slide)
Lslide_CBH = ~L_still_react_CBH & ~L_processive_active;
Lslide_IND_CBH = find(Lslide_CBH);
EGpos = [E{:,pos_prop}]';
CBHpos = [CBH{:,pos_prop}]';
CBHIIpos = [CBHII{:,pos_prop}]';
Epos = [EGpos;CBHpos;CBHIIpos];

for k = 1:numel(Lslide_IND_CBH)
    [new_x,new_y,new_z] =
EG_slide(CBH{Lslide_IND_CBH(k),pos_prop},CBH{Lslide_IND_CBH(k),E_C_near_ext_prop},
Epos(:,1), Epos(:,2), Epos(:,3), area, CBH_moving_treshold, CBH_CBM_dim_sqr);
    CBH{Lslide_IND_CBH(k),pos_prop} = [new_x,new_y,new_z]';
end
% the time for a sliding event is equal to the reaction duration
[CBH{Lslide_CBH,react_time_prop}] = deal(react_duration_CBH);

% delete all CBH I molecules, which are no longer attached to the
% surface. This can happen due to the action of other enzymes, which
% erode CBH I and thereby release CBH I into the bulk.
if any(CBH_neighbours == 0)
    disp(['es werden ',num2str(sum(CBH_neighbours==0)),' CBHs deleted'])
    deleted_CBHI = deleted_CBHI + 1;
end
if any(CBH_neighbours == 0)
    CBH(CBH_neighbours == 0,:) = [];% deletes all enzymes which are no longer
                                    % attached to the cellulose surface
end
```