Designing novel cellulase systems through agent-based modeling and global sensitivity analysis

Advait A Apte¹, Ryan S Senger¹, and Stephen S Fong²,*

¹Department of Biological Systems Engineering; Virginia Tech; Blacksburg, VA USA; ²Department of Chemical and Life Science Engineering; Virginia Commonwealth University; Richmond, VA USA

Keywords: agent based modeling, global sensitivity analysis, cellulose hydrolysis, cellulose degradation strategies, enzyme engineering, metabolic engineering

Abbreviations: ABM, agent based modeling; GSA, global sensitivity analysis; CBP, consolidated bioprocess; SA, sensitivity analysis

Experimental techniques allow engineering of biological systems to modify functionality; however, there still remains a need to develop tools to prioritize targets for modification. In this study, agent-based modeling (ABM) was used to build stochastic models of complexed and non-complexed cellulose hydrolysis, including enzymatic mechanisms for endoglucanase, exogluconase, and β-glucosidase activity. Modeling results were consistent with experimental observations of higher efficiency in complexed systems than non-complexed systems and established relationships between specific cellulolytic mechanisms and overall efficiency. Global sensitivity analysis (GSA) of model results identified key parameters for improving overall cellulose hydrolysis efficiency including: (1) the cellulase half-life, (2) the exogluconase activity, and (3) the cellulase composition. Overall, the following parameters were found to significantly influence cellulose consumption in a consolidated bioprocess (CBP): (1) the glucose uptake rate of the culture, (2) the bacterial cell concentration, and (3) the nature of the cellulase enzyme system (complexed or non-complexed). Broadly, these results demonstrate the utility of combining modeling and sensitivity analysis to identify key parameters and/or targets for experimental improvement.

Introduction

Parallel advances in numerous fields (e.g., genomics, metagenomics, bioprospecting, synthetic biology) have rapidly increased the knowledge and experimental tools available for biological research. When used in combination, fundamental knowledge can direct experiments to improve or modify native biological function.1 Practically, this typically involves iterations between analyzing the known information, experimentally implementing a best guess, and adding to the knowledge pool through the generation of experimental results. For complex systems with a large number of variables, it is often difficult to even formulate what the initial best guess is to explore for the system. This is true for cellulose hydrolysis where native cellulolytic systems are highly variable between different cellulolytic organisms.2 Conceptually, this study focuses on developing and demonstrating a computational approach to aid in the analysis of cellulolytic systems to identify key parameters that influence the efficiency of cellulose hydrolysis.

Identification of cellulase expression systems

Cellulolytic bacteria express and secrete cellulase enzymes to hydrolyze cellulose into consumable sugars.3,7 Two distinctly different enzyme systems have been identified for the production and secretion of cellulases for extracellular cellulose degradation: (1) complexed and (2) non-complexed.4,5 Organisms with a complexed hydrolysis system form a cell wall-bound extracellular structure called a cellulosome. Enzymes required for cellulose hydrolysis such as (1) endoglucanases, (2) exoglucanases, and (3) β-glucosidase are localized in the cellulosome.5,8,5 These enzymes have different hydrolysis mechanisms and act at specific sites within cellulose chains. Endoglucanases cut the amorphous sites in the middle while exoglucanases act on the crystalline reducing and non-reducing ends. Beta-glucosidases hydrolyze smaller cellulose chains (e.g., cellobiose) into individual monosaccharides.6,7 Under the non-complexed mechanism, cellulases are secreted into the outer cellular matrix (unbound to a cellulosome structure) to facilitate cellulose hydrolysis. Usually, aerobic organisms such as Acidothermus cellulolyticus (ATCC 43068), Cellulomonas flavigena (ATCC 482), and Thermohibida fusca (ATCC 27730) (among others) exhibit the non-complexed hydrolysis mechanism while anaerobic organisms such as Clostridium thermocellum (ATCC 27405), C. cellulolyticum (ATCC 35319), and Fibrobacter succinogenes (ATCC 19169) employ the complexed cellulose hydrolysis strategy. Recent research has suggested that organisms having a complexed cellulose hydrolysis enzyme system can degrade cellulose more efficiently than organisms...
with the non-complexed system. This is primarily attributed to the synergistic mechanisms of different cellulases inside the cellulosome structure. The close physicality demonstrated by tightly packed enzyme clusters of the cellulosome may enable organisms with a complexed system to achieve better efficiency of degrading cellulose. A third type of cellulase enzyme system also exists and is termed a “hybrid” system. In a hybrid system, both the complexed and non-complexed cellulose hydrolysis mechanisms are active simultaneously. For example, C. thermocellum ATCC 27405, has been shown to secrete two non-cellulosome bound cellulases in addition to expressing a cellulosome.

Advances in recombinant DNA technology have enabled the development of efficient and scalable expression systems to produce cellulase enzymes using more traditional industrial organisms such as Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Pichia pastoris, Streptomyces lividans, and species of Aspergillus and Trichoderma. Cellulases originating from these organisms have a wide range of applications in the biotechnology industry. While some of the most important applications of cellulolytic enzymes from microorganisms are in food, brewing and wine making, animal feed, textile and laundry, pulp and paper industries, and agriculture, ultimately a major goal of industrial biotechnology involves using renewable cellulose as the primary carbon source for fuels and chemicals.

Recent metabolic engineering approaches

Recent advances in metabolic engineering have uncovered several novel possibilities of engineering cellulase enzymes for use on the industrial scale. Techniques such as (1) high-throughput screening of microorganisms, (2) identification of novel cellulases, (3) enzyme engineering through site-directed mutagenesis, and (4) fermentation process optimization are all suggested to enhance the enzymatic activity and cellular production of desired cellulases. The enzymatic breakdown of cellulosic materials by cellulases for biofuel production is perhaps the most popular application currently being investigated. Conversion of lignocellulosic materials into value-added products (including ethanol) is a multistep process. These processing steps include: (1) pretreatment (mechanical, chemical, or biological), (2) hydrolysis of the complex polymers to produce monosaccharides, (3) bioconversion of hexas and pentose sugars to support microbial growth and production of biochemical products, and (4) the separation and purification of the desired products. Reducing the cost of production, improving cellulase performance, and increasing sugar yields are important considerations for increasing industrial production using microorganisms. Considerable efforts are under way to improve specific cellulase activities for non-complexed cellulase cocktails. These are often implemented through cellulase enzyme engineering based on rational design and/or directed evolution for each cellulase component. But, the identification or engineering of microorganisms capable of performing single-stage cellulose hydrolysis and simultaneous fermentation of liberated hexose and pentose sugars has proven challenging. To improve overall efficiency, enzyme engineering of cellulases is considered a vital step toward engineering recombinant microorganisms suitable for CPB. The heterologous expression of cellulases is central to recombinant cellulolytic strategies, enabling microorganisms to exhibit high product yields and titers. In addition, efforts to detect new enzymes with cellulase activity under adverse conditions are also under way. However, whether these native or engineered cellulase enzymes should be expressed as part of a (1) complexed, (2) non-complexed, or (3) hybrid expression system has not been addressed, until this research through a unique modeling approach.

Agent-based modeling

In an agent-based modeling (ABM) system, an “agent” represents each component of the system that executes different behaviors according to deterministic rules that govern movement and interactions in the simulation matrix. Depending on the biological system being studied, an agent could be (1) a cell, (2) an enzyme or macromolecule, (3) a mitochondria, (4) a tissue, or (5) an entire organ or organism. In a two-dimensional ABM, the simulation matrix can be divided into a number of M x N smaller cells. These cells act as a coordinate system to measure the location of agents at any given instance of the simulation. Each iteration (or step) of the simulation may consider specific unit movement of individual agent to occupy a cell within its von Neumann neighborhood. This is shown in Figure 1. This neighborhood is described as four cells orthogonally surrounding a central cell on a two-dimensional square lattice. In an ABM representing Michaelis-Menten kinetics, enzymatic reactions occur with a pre-determined probability if enzyme and substrate occupy neighboring cells. These local interactions mimic biochemical reactions performed under the close physical proximity present in real biological systems. By simulating the ABM over a large number of iterations, emergent behavior of the system (e.g., diffusion, enzymatic conversion, and cell growth) can be studied as a function of time given a set of physical parameters for each agent. In this research, agents are governed by potentially engineered parameters (e.g., cellulase half-life, expression system, enzyme production rate, among others discussed below). The ABM approach was used in this research to assess the rate of cellulose degradation and consumption in a closed system, given several different sets of physical parameters.

Global sensitivity analysis

Global sensitivity analysis (GSA) uses Monte Carlo simulations to calculate the outputs of a model over the entire range of all input parameters. This variance-based method calculates the contribution of each input parameter to the total variability of the model output. GSA differs significantly from the traditionally used method of partial gradient-based sensitivity analysis (SA). With traditional SA, the change in a model output is calculated by allowing only one parameter to vary, while keeping others constant. This variance in the model output is likely to change if all other parameters are held constant at different values. The GSA approach takes this into account and enables the consideration of multiple parameters simultaneously. To consider only a single parameter is a “first-order” analysis. Two parameters may be considered simultaneously to develop a measure of their interactions. Or, a single parameter can be considered with is interactions with all other parameters in
the model. This is called the “total effect index.”34,35 Thus, the primary advantage of GSA over SA is that GSA accounts for the influence and interactions between input parameters over the entire input space. A simplified tutorial of GSA has been published for a deterministic ammonia emissions model.36 In this research, GSA was performed to identify parameters that most significantly affect the degradation of cellulose and utilization of liberated sugars in a CBP. This study not only identifies previously known areas of significance but also sheds light on some obscure parameters affecting cellulose hydrolysis, providing novel guidelines for future experimental designs.

Identifying new metabolic engineering targets through agent-based modeling

Agent-based modeling was used in this research to effectively compare the efficiencies of (1) complexed, (2) non-complexed, and (3) hybrid cellulase expression systems. This analysis was also accompanied by an analysis of cellulase enzyme composition. Here, the composition of a cellulase enzyme cocktail, consisting of three primary cellulases: (1) endoglucanase, (2) exoglucanase, and (3) β-glucosidase, was varied to determine the impacts on cellulose hydrolysis rates in all three cellulase expression systems studied. This research was performed to investigate the hypothesis that complexed cellulase expression and cellulose hydrolysis mechanism is most efficient when compared with the non-complexed and hybrid systems. The agent-based modeling methods were then combined with GSA in effort to identify new metabolic engineering targets. While considerable research efforts are focused on cellulase enzyme engineering, the results of this research identified additional targets that may be metabolically engineered to improve the overall performance of the CBP.

Results and Discussion

Comparing cellulase enzyme expression systems

In formulating the ABM system for simulations of cellulose hydrolysis, it was found that different concentrations of bacterial cells introduced variability in the model output, making comparisons difficult. This is shown in Figure 2. As the number of bacteria in the system was increased, the number of simulation steps required for cellulose hydrolysis and sugar consumption decreased and had a much smaller variance. This made the model more predictable and required less computational resources for each simulation run. Thus, a fixed starting concentration of at least 60 bacteria cells was required in the system. Above this threshold, the variance of the model reduced to be close to zero, as shown in Figure 3. Using ABM, the following hypotheses were tested: (1) there is no difference between cellulose hydrolysis given complexed and non-complexed cellulase expression systems and (2) there is no difference between cellulose hydrolysis given complexed and hybrid cellulase expression systems. The ABM simulations were run for a total of 200 Monte Carlo iterations for each expression system. The standard student's t test was used to compare the means of the different expression systems, assuming unequal variances and an α of 0.05. The results for t test are presented in Table 2. It is evident from the t test results that differences observed between the simulated hydrolysis efficiencies of the complexed and non-complexed systems are statistically significant. However, those observed between the complexed and hybrid systems was not found to be statistically significant. These results ultimately infer that complexed cellulose hydrolysis is more efficient compared with non-complexed cellulose hydrolysis. According to these calculations, this overall difference in efficiency is about 33%. These results may also indicate that having secreted cellulase enzymes in addition to a complexed cellulose hydrolysis system may not be beneficial to the organism. However, this is still under investigation because hybrid cellulose hydrolysis systems exist in nature (e.g., C. thermocellum). This means the hybrid system must convey a competitive advantage in a particular environment. However, this was not identified in the simulations performed in this research.

Optimizing cellulase ratios

Previously published data37-40 shows that the amount of β-glucosidase in the total cellulase enzyme composition...
(consisting of exoglucanases, endoglucanases, and α-glucosidase) is usually less than 20%. In this research, three different series of ABM simulations were performed by reducing the concentration of α-glucosidase from 20 to 0 in steps of 10. The concentrations of exoglucanases and endoglucanases were modulated. Results are shown in Figure 4. The efficiency of cellulose hydrolysis increased as a result of increasing the concentrations of endoglucanase and decreasing concentrations of exoglucanase, given a constant α-glucosidase concentration. In addition, reducing α-glucosidase content, while maintaining a constant exoglucanase to endoglucanase ratio, showed to increase cellulose hydrolysis efficiency. This indicates a central role for endoglucanases in the cellulose hydrolysis process and suggests that a higher endoglucanase concentration leads to greater efficiency. However, the role of synergy is critical, as simulation results have also shown a cellulase cocktail will outperform a single cellulase. Endoglucanases cut the cellulose chains in the middle making more reducing and non-reducing ends available.

Figure 2. Number of iterations required for cellulose hydrolysis and sugar consumption for 200 independent simulations given bacteria cell density values ranging between 10 and 100 cells per simulation. Large variance was calculated at low cell densities.

Figure 3. Variance of cellulose hydrolysis and sugar consumption over 200 iterations as a function of the bacteria cell density (number of cells) for the three different cellulase enzyme expression systems.
for exoglucanases and smaller chains to β-glucosidase. This explains the increase in efficiency by increasing endoglucanase composition in the cellulase system. It has been shown recently that the endoglucanase is more important to the cellulase mixture when hydrolyzing cellulose with higher degree of polymerization.\(^4\) An upper bound on the concentration of endoglucanase in the enzyme cocktail has been suggested due to the balancing act of exoglucanases as more chain ends become available. The model presented in this research does predict this upper bound on endoglucanase concentration and the balancing act of the endo-exo-synergy is evident in Figure 4.

### Global sensitivity analysis (first-order indices)

The first-order indices for each parameter are reported in Figure 5 as a percentage of the summation of all first-order index values. The three most significant parameters identified by the first-order index are (1) exoglucanase composition (9.29%), (2) enzyme half-life (11.6%), and (3) exoglucanase activity (12.78%). All three of these parameters are related to cellulases, and two specifically single-out the exoglucanase. In the previous section on optimizing cellulase ratios, it was revealed that increasing exoglucanase composition (at the expense of endoglucanase and β-glucosidase) negatively affects hydrolysis efficiency. The GSA results indicate that exoglucanase composition is responsible for about 9.29% of the cellulose hydrolysis efficiency variation, while exoglucanase activity is responsible for additional 12.78% of the variance. Thus, these results suggest that the exoglucanase composition of the cellulase is an important determinant of cellulose hydrolysis efficiency. Of course, this was confirmed by the previous set of simulations that showed minimizing the exoglucanase composition (while increasing the composition of endoglucanase) was an effective strategy for improving cellulose hydrolysis efficiency. However, exoglucanase activity was shown to account for the most first-order model variance of any parameter. Given a constant composition, increased exoglucanase activity would increase cellulose hydrolysis efficiency. Thus, in addition to optimizing cellulase composition, increasing the activity of the exoglucanase should be seen as a worthwhile enzyme engineering goal. In addition, enzyme engineering to increase the half-lives of all cellulases is also a worthwhile goal, but this result is more intuitive.

### Global sensitivity analysis (total effect indices)

The GSA total effect indices are shown in Figure 6 and identified (1) glucose uptake (7.75%), (2) bacterial cell density (7.61%), and (3) the cellulase expression enzyme system (7.13%) as the three most significant parameters affecting cellulose hydrolysis. When interactions with other parameters are considered, the glucose uptake rate is the most important parameter identified by GSA to impact the rate of cellulose hydrolysis. Thus, when considering optimizing a CBP, it is important to not only hydrolyze cellulose efficiently but also to uptake and convert liberated sugars effectively. Thus, further

---

**Table 1. ABM parameters, bounds, and descriptions**

| Parameter                  | Lower Bound | Upper Bound | Function                                              |
|----------------------------|-------------|-------------|-------------------------------------------------------|
| EnzymeProductionRate       | 1           | 100         | Determines the rate of cellulase enzyme production in the non-complexed system |
| EnzymeProductionRateH      | 0.1         | 10          | Determines the rate of cellulase enzyme production in the hybrid model |
| CompositionExo             | 0           | 100         | Selects the composition of exoglucanase              |
| CompositionEndo            | 0           | 100         | Selects the composition of endoglucanase             |
| CompositionGluco           | 0           | 100         | Selects the composition of β-glucosidase             |
| halflife                   | 100         | 200         | Defines the number of steps a given enzyme stays active |
| PrbExo                     | 1           | 100         | Determines the activity of exoglucanase              |
| PrbEndo                    | 1           | 100         | Determines the activity of endoglucanase             |
| PrbGluco                   | 1           | 100         | Determines the activity of β-glucosidase             |
| EnzymeSystem               | N/A         | N/A         | Selects between (i) non-complexed, (ii) non-complexed, and (iii) hybrid cellulase systems |
| PrbGluIntake               | 1           | 100         | Controls the rate of glucose uptake                  |
| PrbBioIntake               | 1           | 100         | Controls the rate of cellobiose uptake                |
| PrbTrilIntake              | 1           | 100         | Controls the rate of cellotriose uptake               |
| BacteriaConcentration      | 1           | 100         | Determines the number of bacteria in the system       |
| CelluloseChains            | 1           | 10          | Determines the number of cellulose chains in the system |
| CelluloseLength            | 1           | 20          | Selects the length of cellulose chains in the system  |

**Table 2. The t-test results comparing cellulase expression systems**

| Parameter                  | Value  | Parameter       | Value  |
|----------------------------|--------|-----------------|--------|
| Mean Non-complexed         | 280.0  | Mean Complexed  | 188.0  |
| Mean Complexed             | 188.0  | Mean Hybrid     | 190.3  |
| n Non-complexed            | 200    | n Complexed     | 200    |
| n Complexed                | 200    | n Hybrid        | 200    |
| df                         | 392.4  | df              | 381.4  |
| α                          | 0.05   | α               | 0.05   |
| t Calculated               | 13.1   | t Calculated    | -0.3   |
| t α/2 critical             | 2.0    | t α/2 critical  | 2.0    |
optimization of glucose uptake systems can be considered a means of improving the performance of a CBP. The second most important parameter identified by the GSA total effect index was the bacterial cell density. This result is consistent with the initial findings of this study, as shown in Figure 4, that varying the number of bacterial cells in the system can greatly impact the model output. Engineering greater cell densities is an intuitive means of improving the performance of a CBP. Finally, the type of enzyme system itself (complex, non-complexed, or hybrid) was also identified by GSA (correctly) as a source of great variance in the model output. This GSA result reinforced the motivation behind this study that the type of cellulase expression system used is critical to cellulose hydrolysis efficiency. The results of ABM have shown that the complexed system is statistically more efficient than the non-complexed cellulase expression system.

Using ABM and GSA results to guide metabolic engineering efforts

In this study, GSA has revealed important aspects of the cellulose hydrolysis process. Six parameters significantly affecting the hydrolysis were identified and can all be identified as valuable targets for enzyme and metabolic engineering. From these, the exoglucanase enzyme is of particular interest as both its activity and composition significantly affects cellulose hydrolysis efficiency. Modeling results suggest enzyme engineering efforts focused on improving the catalytic action of exoglucanase enzymes would be valuable while metabolic engineering strategies to optimize the exoglucanase to endoglucanase ratio is another area of importance. Few efforts to optimize these ratios are currently under way though more efforts are anticipated in the future. Improving enzyme half-life is another area of great significance. Recent efforts in cellulase engineering have yielded cellulases with higher half-lives.

Part of the benefit of a comprehensive computational analysis is to equally consider the effects of all possible variables within a system. There have been a number of studies seeking to improve cellulolytic efficiency, but much of this research is heavily skewed to focus on enzyme engineering. A PUBMED literature search (shown in Table 3) was performed to illustrate that far more research to-date has been conducted in enzyme engineering (41 210 articles published to date) than in other identified areas (glucose transport, cell concentration) that can improve cellulolytic efficiency. Thus, the combination of ABM and GSA has been used in this case to identify critical potential metabolic engineering targets for optimizing a CBP that are relatively under-researched compared with cellulase enzyme engineering efforts.

Conclusions

The focus of this computational research was to compare the efficiency of cellulose hydrolysis given different cellulase expression systems. The use of ABM and GSA also allowed the identification of the system parameters that significantly impact cellulose hydrolysis efficiency in a CBP. The randomness introduced by a sparsely populated modeling environment was addressed by assuming sufficient concentrations of bacterial cells. This variability was attributed to physical distances between interacting components of the system at the beginning of the simulation. The ABM predicts that a complexed...
(cellulosome bound) cellulase expression and cellulose hydrolysis mechanism is statistically different from a non-complexed (secreted) cellulase expression system with about a 33% increase in efficiency. The differences observed between the complexed and hybrid systems were not statistically significant, implying that additional cellulosome-free secreted cellulases may not be advantageous to a strictly complexed cellulase expression system.

Figure 5. First-order indices of the GSA for parameters listed in Table 1.

Figure 6. Total effect indices of the GSA for parameters listed in Table 1.
From the ABM results of testing different cellulase ratios, it was concluded that the composition of endoglucanase in the system is directly proportional to the efficiency of hydrolysis while exoglucanase and β-glucosidase concentrations are inversely proportional. This can be explained by the fact that by randomly hydrolyzing the amorphous sites in the middle of a cellulose chain, endoglucanases are producing a greater number of shorter cellulose chains and exposing more reducing and non-reducing cellulose ends in the process, contributing to a higher cellulose hydrolysis efficiency.

The results of the GSA study revealed the importance of the several parameters affecting the cellulose hydrolysis process. Enzyme engineering was found to be a significant aspect of cellulose hydrolysis, as indicated by first-order indices of the GSA. However, while increasing the half-life of cellulase enzymes is an intuitive finding, the importance of the exoglucanase activity is less intuitive. These results correlate well with the amount of research currently focused in these areas. Important targets for metabolic engineering were identified as well. First, the production of optimized ratios of endoglucanases to exoglucanases was found significant according to first-order results. The glucose uptake rate was found to be a highly significant aspect of cellulose hydrolysis when total effect indices of the GSA were considered. Research determining the effects of substrate uptake on hydrolysis efficiency has received significantly less attention in the recent literature compared with cellulase enzyme engineering. Thus, the approach of ABM and GSA provides a valuable tool for the identification of new engineering targets. In the case of the CBP, all model parameters showed to contribute to the variance in cellulose hydrolysis efficiency. While these are all important and can yield an improved process if engineered, those parameters with higher sensitivities can yield greater returns if optimized and may represent better engineering targets for the near term.

Overall, the combined modeling approach of agent-based modeling and global sensitivity analysis was able to simulate the mechanisms and physicality of the cellulose hydrolysis process. Key variables affecting the efficiency of cellulose hydrolysis were identified to suggest future avenues for experimental engineering to improve cellulolytic systems. Additionally, the modeling approach taken appears well-suited for studying biological processes, particularly for its ability to account for stochastic processes and physical biological constraints. AGM and GSA may be applied to ranges of problems from molecular binding events to macroscopic community dynamics (e.g., growth and crashes in large-scale algae ponds or microbial consortia).

### Methods

**Agent-based model construction**

Agent-based modeling was used to study cellulose hydrolysis. The model was built using Netlogo software, version 4.1RC5 (October 13, 2009). The agents of the system included: (1) bacteria, (2) cellulose, and (3) cellulase enzymes (produced by the bacteria). The individual parameters that defined the agents were varied and are given in Table 1. These parameters consisted of: (1) the cellulase enzyme production rate (by the bacteria); (2) the composition of exoglucanases, endoglucanases, and β-glucosidase; (3) the cellulase enzymatic activity and half-life; (4) the cellulase enzyme expression system; (5) the cellular uptake rates of glucose, cellobiose, and cellobiose sugars; (6) the concentration of bacteria cells present in the system; and (7) the concentration and average length of the cellulose chains present. These parameters are discussed in detail below.

To demonstrate the ABM method, six simulation steps involved in the enzymatic hydrolysis reaction are given in Figure 1. Here, the parameters \(S\), \(P\), \(E\), \(SE\), and \(PE\) stand for substrate, product, enzyme, substrate-enzyme complex, and product-enzyme complex, respectively. The probability of \(S\) and \(E\) occupying one of their neighboring cells is 0.25. When \(S\) and \(E\) occupy neighboring cells, they can interact to form \(SE\). A \(PE\) complex can form in similar way. The transitional probability for the substrate-enzyme complex, \(SE(i)\), to produce product-enzyme complex, \(PE(i)\), is assumed to be 0.5. It can also dissociate into \(S\) and \(E\) with a 0.5 probability, suggesting a reversible reaction. Differences in the transport (by diffusion) of cells, cellulose, and cellulase enzymes in the closed system, arising due to their respective sizes, were given special consideration. Diffusion constants for cellulose chains in aqueous media were set inversely proportional to the length of chain. Thus, a smaller chain was allowed to diffuse faster than a larger chain. Each simulation consisted of complete breakdown of all cellulose chains and cellular uptake of all liberated sugars. The number of steps (see Fig. 1) required to complete each simulation was recorded for each set of parameters tested. Results for each set of parameters were averaged over 200 independent simulations. The average number of steps required for each case ultimately determined the efficiency of cellulose hydrolysis.

**Agent-based modeling parameters**

The following parameters were taken into consideration while building and simulating the model. The upper and lower bounds on the parameters and their functionalities are listed in Table 1. The parameters used in this modeling effort are described in detail below.

1. **Cellulase enzyme production rate (by the bacteria)**

The rate at which enzymes are released outside the bacterial cell was controlled using the \(EnzymeProductionRate\) parameter for the non-complexed cellulase expression system. Similarly, the \(EnzymeProductionRateH\) parameter was used to control enzyme production rate for the hybrid system. These rates were expressed in terms of percentage probability between 1 and 100, with 100 representing the maximum production rate. The only known hybrid system to secret two non-cellulosome bound cellulases is \(C. thermocellum\). Other organisms using a non-complexed
system secrete about 20 different cellulases, on average. So, for this hybrid system, the secreted cellulase enzyme production rate was set to 10% of that of the other non-complexed systems.

(2) Composition of exoglucanases, endoglucanases, and β-glucosidase

The proportion of individual cellulases (exoglucanases, endoglucanases, and β-glucosidases) relative to the total amount of cellulase enzyme present in system was set using the CompositionExo, CompositionEndo, and CompositionGlucO parameters. These parameters fix the percentage of a certain type of cellulase (exoglucanase, endoglucanase, or β-glucosidase) released from bacterial cells, and these values were set between 1 and 100 percent. The summation of individual cellulase levels (defined by the CompositionExo, CompositionEndo, and CompositionGlucO parameters) was always equal to 100.

(3) Cellulase enzymatic activity and half-life

The ability of cellulase enzymes to catalyze biochemical reactions was expressed in terms of percentage probability. This probability dictates how likely a particular enzyme is to catalyze cellulose hydrolysis given the enzyme and cellulose are in physical proximity. The different factors such as (1) temperature, (2) pH, and (3) substrate concentration, which determine rate of a biochemical reaction, were combined into a single interaction probability for cellulases. This probability, expressed as a percentage, was allowed to vary between 1 and 100. By changing the PrbExo, PrbEndo, and PrbGlucO parameters, the activities of the different types of cellulases (exoglucanases, endoglucanases, and β-glucosidases) were controlled. The cellulase enzyme half-life, which decides the decay in its activity over time before denaturation, was also built into the model. The cellulase enzymes were kept active for a specified number of steps during the ABM routine using the HalfLife parameter. This parameter was allowed to vary between 100 and 200 steps. The lower bound was chosen based on preliminary ABM simulations runs (results not shown), and the upper bound was chosen based on the possibility of extending the half-life of a cellulase enzyme by a factor of 2 through enzyme engineering.

(4) Cellulase enzyme expression system

The EnzymeSystem model parameter allowed simulation of either of the three enzyme systems: (1) complexed, (2) non-complexed, or (3) hybrid cellulose hydrolysis.

(5) Cellular uptake rates of glucose, cellobiose, and cellotriose sugars

Different factors such as the proton motive force and the kinetics of the membrane transporters responsible for sugar uptake were combined into a single probability model parameter that controlled the rate of sugar uptake. This probability was also expressed as percentage ranging between 1 and 100, and it dictates the likelihood of a sugar molecule being transported inside a bacterial cell located in the physical proximity. The parameters PrbGlu intake, PrbBio intake, and PrbTri intake determine the probabilities (and ultimately the rate) of (1) glucose, (2) cellobiose, and (3) cellotriose uptake by the cell. The potential range (0 to 100) was identical for each of these parameters, even though the probability (and rate) of glucose uptake exceeds that of cellobiose and cellotriose in observed cultures. However, allowing a full range for each parameter allows the in silico exploration of possibilities of cells engineered for increased cellobiose or cellotriose uptake.

(6) Concentration of bacteria cells present in the system

The number of bacteria in the system was varied through the BacteriaConcentration parameter. A concentration of 60 cells was used for the simulations to maintain a sparsely populated environment to mimic Michaelis-Menten kinetics. This choice was made based on an initial simulation study. The BacteriaConcentration parameter was varied between 1 and 100 cells for the GSA study.

(7) Concentration and average length of the cellulose chains present

The number and length of cellulose chains in the system was set using the CelluloseChains and CelluloseLength parameters, respectively. The number of cellulose chains in the system was varied between 1 and 10, and the cellulose chain length was varied between 1 and 20 glucose molecules. While these values are small compared with cellulose concentrations and especially chain length in a physical system, these small values enabled ABM on a realistic time-scale given the very large number of calculations involved. The upper bound on the cellulose chain length was also determined through initial sets of simulations in which longer chains were used at the expense of computational time (data not shown).

Global sensitivity analysis

As shown in Table 1, input parameters of the model are independent of one another. Due to the large number of parameters involved, the GSA was limited to calculating the first-order and total effect indices. The correlations between individual parameters were not calculated in this study. The first-order index measures the sensitivity of the model output to changes in that parameter over its entire range and the entire range of all other parameters. The total effect index measures the sensitivity of the model output for a particular parameter when all interactions of that parameter with all other parameters are considered. A step-by-step procedure for calculating above-mentioned indices has been published. Additional details of the GSA calculation are provided in the Supplemental Materials. It was observed that the values of individual indices changed by increasing number of Monte Carlo simulations performed in the GSA. A sufficiently large number of simulations, 2000 in this case, were used to calculate first-order and total effect indices. This sample size was determined through initial sets of simulations (data not shown).

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

This research was supported by research grants from the Institute for Critical Technologies and Applied Science and the Biodesign and Bioprocessing Research Center at Virginia Tech.

Supplemental Materials

Supplemental materials may be found here: www.landesbioscience.com/journals/bioe/article/29160/
References

1. McArthur GH 4°, Fong SS. Toward engineering synthetic microbial metabolism. J Biomed Biotechnol 2010; 2010:106976; PMID:20497734; http://dx.doi.org/10.1155/2010/490760

2. Gowen CM, Fong SS. Exploring biodiversity for cellulose biofuel production. Chem Biodivers 2010; 7:1086-97; PMID:20491068; http://dx.doi.org/10.1002/cbdv.200900934

3. Bayer EA, Kneisl R, Lamed R. Adherence of Clostridium thermocellum to cellulase. J Bacteriol 1983; 156:818-27; PMID:6630152

4. Bayer EA, Lamed R, Himmel ME. The potential of cellulates and cellolyses for cellulose waste management. Biotechnol Bioeng 2007; 97:253-45; PMID:17462879; http://dx.doi.org/10.1002/bit.20074.004.004

5. Dey AE, Newcomb M, Wu JHD. Cellulase, clostridia, and ethanol. Microbiol Mol Biol Rev 2007; 71:124-54; PMID:17595936; http://dx.doi.org/10.1128/MMBR.69.1.124-154.2005

6. Dien BS, Cotta MA, Jeffries TW. Bacteria engineered for fuel ethanol production: current status. Appl Microbiol Biotechnol 2003; 63:258-66; PMID:15680206; http://dx.doi.org/10.1007/s00253-003-1344-x

7. Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev 2002; 66:506-77; PMID:12209002; http://dx.doi.org/10.1128/MMBR.66.3.506-77.2002

8. Shoham Y, Lamed R, Bayer EA. The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. Trends Microbiol 1999; 7:275-81; PMID:10396657; http://dx.doi.org/10.1016/S0966-842X(01)02153-4

9. Raghotra S, Eberhardt RY, Simpson P, Wigelsworth D, White P, Hazlewood GP, Nagy T, Gilbert HJ, Williamson MP. Characterization of a cellulosome dockerin domain from the anaerobic fungus Piromyces equi. Nat Struct Biol 2001; 8:775-81; PMID:11457512; http://dx.doi.org/10.1038/91333

10. Seoensen A, Teller PJ, Lübke PS, Abrink BG. Onsite enzyme production during bioethanol production from biomass: screening for suitable fungal strains. Appl Biochem Biotechnol 2010; 164:1058-70; PMID:20560902; http://dx.doi.org/10.1007/s12010-011-9194-z

11. Koch A, Weigel CTO, Schulz G. Cloning, sequencing, and heterologous expression of a cellulase-encoding cDNA (ch1) from Penicillium janthinellum. Gene 1993; 124:75-67; PMID:8440481; http://dx.doi.org/10.1016/0378-1119(93)90762-Q

12. Mazzoli R, Lambert C, Pessione E. Engineering new metabolic capabilities in bacteria: Lessons from recombinant cellulolytic strategies. Trends Biotechnol 2012; 30:111-9; PMID:21930521

13. Bonabeau E. Agent-based modeling: methods and techniques for simulating human systems. Proc Natl Acad Sci U S A 2002; 99(Suppl 3):7280-7; PMID:12101147; http://dx.doi.org/10.1073/pnas.082080899

14. Scholl HJ. Agent-based and system dynamics modeling: A call for open research. Syst Sci. 2001; Sciences 2001 Proceedings of the 34th Annual Hawaii International Conference on 2001:1-8 pp.

15. Van Dyke Parunak H, Savit R, Rialto A. Agent-based modeling vs. equation-based modeling: A case study and users’ guide multi-agent systems and agent-based simulation. In: Schman J, Conte R, Gilbert N, eds.: Springer Berlin / Heidelberg, 1998:277-283

16. Salehi A, Chan K, Scott M. Sensitivity analysis. Wiley 2000

17. Salehli A, Ratto M, Andre T, Camplong F, Carboni J, Garetli D, Saiamana S, Tarantola S. Global sensitivity analysis: The primer. Wiley 2008

18. Ogejo JA, Sengar RS, Zhang RH. Global sensitivity analysis of a process-based model for ammonia emissions from manure storage and treatment structures. Atmos Environ 2010; 44:3621-9; http://dx.doi.org/10.1016/j.atmosenv.2010.06.053

19. Chen S, Wilson DB. Protocime and transcriptomic analysis of extracellular proteins and mRNA levels in Thermobifida fusca grown on cellulose and glucose. J Bacteriol 2007; 189:6260-5; PMID:17601791; http://dx.doi.org/10.1128/JB.00584-07

20. Dadi AP, Schall CA, Varanasi S. Mitigation of cel- lulase recalcitrance to enzymatic hydrolysis by liquid liquor pretreatment. Appl Biochem Biotechnol 2007; 137:1407-407; PMID:18478405; http://dx.doi.org/10.1007/s12033-007-0968-9

21. Martinez D, Berka RM, Henriass B, Saloimeto M, Arus M, Baker SE, Chapman J, Chertkov O, Courtois PM, Cullen D, et al. Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina). Nat Biotechnol 2008; 26:553-60; PMID:18454138; http://dx.doi.org/10.1038/nbt.1403

22. Raman B, Pare G, Hurst GB, Rodriguez M Jr., McKeown CK, Lankford PK, Samatova NF, Mielentz JR. Impact of pretreated Switchgrass and bio- mass carbohydrates on Clostridium thermocellum ATCC 27405 cellulose consumption: a quantita- tive proteomic analysis. PLoS ONE 2009; 4:e5271; PMID:19384422; http://dx.doi.org/10.1371/jour- nal.pone.0005271

23. Levine SE, Fox JM, Clark DS, Blanch HW. A mechanistic model for rational design of optimal cellu- lase mixtures. Biotechnol Bioeng 2011; 108:2561- 70; PMID:21702033; http://dx.doi.org/10.1002/bit.23249

24. Liao H, Zhang XZ, Rollin JA, Zhang YH. A minimal set of bacterial cellules for consolidated bioprocess- ing of lignocellulose. Biotechnol J 2011; 6:1409- 18; PMID:21751395; http://dx.doi.org/10.1002/biot.201001557

25. Badiane S, Bevan DR, Zhang C. Study and design of stability in GH5 cellulases. Biotechnol Bioeng 2012; 109:31-44; PMID:21869329; http://dx.doi.org/10.1002/bit.23280

26. Binner R, Menath V, Huber H, Thimm M, Bischof F, Schmack D, Reuter M. Comparative study of sta- bility and half-life of enzymes and enzyme aggregates implemented in anaerobic biogas processes. Biomass Conversion and Biorefinery 2011; 1:1-8; http://dx.doi.org/10.1007/s13399-010-0002-2

27. Lin H, Li W, Guo C, Qu S, Ren N. Advances in the study of directed evolution for cellulases. Front Environ Sci Eng China 2011; 5:519-25; http://dx.doi.org/10.1186/1677-1733-5-013-026

28. Lonardo P, Bernal B, Bernal JM, Puchebut M, Vautier M. Stabilized immobilized cellulase by liquidic ions for saccharification of cellulose solutions in 1-butyl-3-methylimidazolium chloride. Green Chem 2011; 13:1406-10; http://dx.doi.org/10.1039/c1gc12529a

29. Myung S, Zhang XZ, Zhang YH, Deppenmeier U. Ultra-stable phos- phoesterase isomerase through immobilization of cel- lulose-binding module-tagged thermophilic enzyme on low-cost high-capacity cellulase adsorbent. Biotechnol Prog 2011; 27:969-75; PMID:21630486; http://dx.doi.org/10.1002/btp.606
48. Park JI, Kent MS, Datta S, Holmes BM, Huang Z, Simmons BA, Sale KL, Sapra R. Enzymatic hydrolysis of cellulose by the cellobiohydrolase domain of CelB from the hyperthermophilic bacterium Caldicellulosiruptor saccharolyticus. Bioresour Technol 2011; 102:5988-94; PMID:21421309; http://dx.doi.org/10.1016/j.biortech.2011.02.036

49. Wilensky U. Netlogo. 1999

50. Apte AA, Cain JW, Bonchev DG, Fong SS. Cellular automata simulation of topological effects on the dynamics of feed-forward motifs. J Biol Eng 2008; 2:2; PMID:18304325; http://dx.doi.org/10.1186/1754-1611-2-2