Final Technical Report

Project Title: Addressing the Recalcitrance of Cellulose Degradation through Cellulase Discovery, Nano-scale Elucidation of Molecular Mechanisms, and Kinetic Modeling

Award Number: DE-FG36-08GO18084

Recipient: Cornell University

Project Location(s): Ithaca, NY

Project Period: 7/1/2008-12/31/2010

Date of Report: June 8, 2011

Written by: Larry P. Walker

Program Manager: Larry P. Walker

Principal Investigators: Larry P. Walker, Gary Bergstrom, Stephane Corgie, Harold Craighead, Donna Gibson, David Wilson

Acknowledgment: This material is based upon work supported by the Department of Energy under Award Number DE-FG36-08GO18084.

I. EXECUTIVE SUMMARY

This research project was designed to play a vital role in the development of low cost sugars from cellulosic biomass and contributing to the national effort to displace fossil fuel usage in the USA transportation sector. The goal was to expand the portfolio of cell wall degrading enzymes through innovative research at the nano-scale level, prospecting for novel cellulases and building a kinetic framework for the development of more effective enzymatic conversion processes. More precisely, the goal was to elucidate the molecular mechanisms for some cellulases that are very familiar to members of our research team and to investigate what we hope are novel cellulases or new enzyme combinations from the world of plant pathogenic fungi and bacteria. Hydrolytic activities of various cellulases and cellulase cocktails were monitored at the nanoscale of cellulose fibrils and the microscale of pretreated cellulose particles, and we integrated this insight into a heterogeneous reaction framework. The over-riding approach for this research program was the application of innovative and cutting edge optical and high-throughput screening and analysis techniques for observing how cellulases hydrolyze real substrates. Three major tasks were identified in the original statement of work (SOW):

- **(A)** Expanding the portfolio of cellulases through prospecting for novel cell wall degrading enzymes found in the community of highly virulent plant pathogenic fungi and bacteria.
- **(B**) Elucidating "intrinsic" molecular mechanisms of cellulases at the nanoscale of cellulose fibrils. Total Internal Reflection Fluorescence Microscopy (TIRFM) and Confocal Laser Scanning Microscopy (CLSM) will be used to measure the binding, surface mobility, and activity of single and binary mixtures of fluorescently-labeled cellulases, and to develop mathematical models of these surface phenomena.
- **(C)** Integrating the insights gained from "intrinsic" molecular mechanisms into the mesoscale heterogeneous kinetic framework of pretreated cellulosic particles. A major objective of this research program is to use CLSM to observe the binding and diffusion of cellulases on the surface and into the porous structure of pretreated cellulose particles, and to use a high-throughput binding and hydrolysis system to correlate cellulase cocktail composition and substrate physical and chemical properties with rates and extent of hydrolysis and degree of synergistic effect.

II. OBJECTIVES AND RESULTS

TASK A: Discovery of new cellulases that can be utilized in novel commercial enzyme cocktails for conversion of biomass.

Many plant pathogens rely on cell wall degrading enzymes (CWDEs) to breach the cell-wall barrier and obtain nutrients for their survival while subverting plant defense responses. Our survey of CWDEs from plant pathogens demonstrated that these organisms are highly competent producers of lignocellulolytic enzymes, and their enzyme activity patterns reflect host specificity. More recently as genome data have become available during the course of this study, sequence analyses have revealed large suites of CWDEs in a wide

range of plant pathogens, and particularly necrotrophic fungi. In particular, useful auxiliary enzymes with the potential to improve commercial enzyme cocktails for conversion of various lignocellulose feedstocks should be highly enriched in plant pathogenic fungi.

Objective A: Expand the portfolio of cellulases through prospecting for novel cell wall degrading enzymes found in the community of highly virulent plant pathogenic fungi and bacteria. Our specific objectives were as follows:

Objective A.1. We will refine and employ a high-throughput method to screen a large collection of field-collected virulent plant pathogenic fungi and bacteria for high affinity to native cellulose and high specific activity toward native cellulose. *Trichoderma reesei* RUT-C30 will be used as a positive control. Preliminary screening results have indicated that both cellulase and xylanase activities can be readily detected and several isolates have significantly greater activity than *T. reesei*, supporting our working hypothesis that these isolates will possess novel enzymes for lignocellulosic digestion. We will use several different systems to assay for activity, concentrating primarily on cellulase and xylanase activities and adapting plate assay techniques to a high throughput system using 96 well plates, microplate readers equipped with internal incubation chambers, and robotic platforms for handling multiple plates in continuous operation.

Objective A.2. We will develop and implement fermentation methods to produce crude enzyme extracts that will be tested for activities on different carbon sources. Small-scale fermentation methods will be used to optimize production of enzyme extracts for those microorganisms from Objective 1.1 that have the highest catalytic efficiency for further analysis. We anticipate that approximately 5% of isolates screened will enter into this phase of testing, for a total number of 10-20 isolates for detailed enzyme analysis of crude extracts.

Objective A.3. Activities on the purification and characterization of the most promising isolates will be completed. The protein sequence of the 3-4 promising isolates that were purified by the end of FY2008 will be identified by the end of calendar year 2009.

Objective A.4. Cloning of genes for isolates. Activities on the cloning of the genes for the 3-4 most promising isolates will be completed. By the end of FY2009 the products of PCR vectors from those isolates will be cloned into expression vectors, and their expression evaluated.

TASK A RESULTS

Objective A.1. High-Throughput Screening Platform Development. In order to develop high throughput screening assays for enzyme bioprospecting, a standardized microplate assay was developed for rapid analysis of polysaccharide hydrolysis by fungal extracts, incorporating a number of biomass substrates. Other investigators have shown that activities of hydrolytic enzyme mixtures on relatively simple substrates such as filter paper, Avicel, and CMC do not necessarily correlate with activities on more complex substrates such as biomass, underscoring the importance of developing methodologies that allow assessment of robust enzyme systems.

This objective was accomplished by the development of a high-throughput screening system incorporating multiple substrates using a 96-well format system combined with robotic liquid handling. The results of this work are described in more detail King et. al. 2009. [1] We developed a rapid, quantitative, reproducible, and high-throughput platform for analyzing cell wall degrading enzyme (CWDE) activities of fungal extracts. By choosing appropriate cell wall components as substrates reflective of the types of biomass materials being considered for bioenergy, we have gained insight into the mechanism of hydrolysis for complex and heterogeneous biomass materials.

Objective A.2. Assessment of Media Platform for CWDE Production. To initially test the screening platform, an initial sampling of 12 phytopathogenic fungi was used to test the effect on cell wall degrading enzyme (CWDE) production for three growth media with switchgrass (SG), soybean stem (SS) or Avicel as the primary carbon source. Data were collected for hydrolysis of nine different polysaccharide or biomass substrates. A full-factorial mixed-effect model was built with host (pathogen preference for monocot or dicot tissue), substrate and medium treated as fixed effects and isolate as a random effect. The third order interaction of host*substrate*medium was significant ($P = 0.0135$), as was the second order interaction of medium*substrate ($P = 0.0184$) and the primary effect of substrate ($P = 0.0009$). For all assay substrates, extracts from fungi grown on the Avicel-based medium released either comparable amounts or fewer reducing sugars than cultures grown on SG- or SS-based media, as determined by pairwise *t*-tests of fitted data from the model. The only example where extracts from Avicel medium were noticeably more active than the SG or SS media was the case of dicot pathogens hydrolyzing filter paper (FP). However, this difference was not significant as determined by pairwise *t*-tests ($P = 0.0822$ for dicot pathogens grown on Avicel compared to SG and $P = 0.1097$ for Avicel compared to SS).

When data from cultures grown on Avicel were removed and a mixed-effect model was fitted using standardized data from only biomass-based media (SG and SS), the second order interactions of substrate*host ($P < 0.0001$) and substrate*medium ($P = 0.0091$) were significant. The primary effects of host ($P = 0.1319$) and medium ($P = 0.7287$) were non-significant and the effect of substrate was barely non-significant ($P = 0.0506$). Extracts from cultures grown on SG medium released more sugar than cultures grown on SS medium when tested on the two xylans [arabinoxylan from oat (AXO) and xylan from birch (XY)] and the two grasses [corn stalk (CS) and SG]. The opposite trend was seen for hydrolysis of xyloglucan (XG) and the two legumes [alfalfa (AL) and SS] with cultures grown on SS-based medium releasing more sugar than cultures grown on SG-based medium. Although the model found this substrate*medium effect to be significant, pairwise *t*-tests of both standardized data and values fitted from the model were found to be non-significant. For the host*substrate interaction, extracts from dicot pathogens were significantly more active than monocot pathogens when tested for hydrolysis of the dicot substrates XG ($P = 0.0385$), SS ($P = 0.0017$) and AL ($P = 0.0008$). Other substrate*host interactions were not significant. Therefore, we chose SG medium for use in the large scale screening platform.

Objective A.3. Large-Scale Screening. The large scale screening analysis has been completed and a more complete set of the results in more detail can be found in King et. al.[2] A total of 348 unique isolates was tested for hydrolysis of filter paper (FP), three types of hemicellulose

[arabinoxylan from wheat (AXW), xylan (XY) and xyloglucan(XG)], biomass from two grasses [corn stover (CS) and switchgrass (SG)] and biomass from two legumes [alfalfa (AL) and soybean stem (SS)]. Each isolate was independently replicated 5 times, and each replicate was tested in triplicate on each substrate. Most isolates were from the kingdom Fungi (344, 98.9%), with four (1.1%) isolates from Oomycetes.

The fungi tested showed a broad range of activity on eight substrates. *T. reesei* RUT-C30 was roughly twice as active as the top natural isolates when tested for hydrolysis of FP and SS, but some species exhibited activity greater or equal to *T. reesei* when assayed on either untreated biomass or xylans. Many species were represented by multiple isolates, and we observed that some isolates within a species were highly active, while others showed weaker activities.

Hierarchical clustering was used as a means to organize the moderately large dataset into meaningful groups and to identify coherent patterns. Clustering of the complete dataset (excluding *T. reesei*) identified two major groups of species. The top tier of 86 (55%) species showed moderate or strong hydrolysis of most substrates tested. A bottom tier contained 69 (45%) inactive or very weakly active species. All species tested from the genera *Bipolaris* (4), *Colletotrichum* (7), *Penicillium* (3), *Rhizoctonia* (4), *Sclerotinia* (4) and *Trichoderma* (5) were in the top tier of active isolates. The genus *Fusarium* was represented by 20 species, the greatest number of any genus. Most species of *Fusarium* (17) fell in the top tier of active species. However, three species (*F. decemcellulare*, *F. lateritium* and *F. merismoides*) were average or below average and fell in the bottom tier. Of the nine species tested within the genus *Aspergillus*, four species (*A. lineolatus*, *A. awamori*, *A. fumigatus* and *A. niger*) clustered in the active group, while five species (*A. candidus*, *A. janus*, *A. penicilloides*, *A. peyronelii* and *A. proliferans*) fell in the bottom tier.

Data from the weakly active tier of species was excluded, and the top tier containing 86 species with moderate or strong hydrolytic activities was used for further analysis, revealing a cluster of 27 very highly active species (Figure 1). In particular, the species *F. proliferatum*, *F. oxysporum*, *A. fumigatus*, *Penicillium expansum*, *Mucor hiemalis*, *Rhizoctonia cerealis*, *S. homeocarpa*,

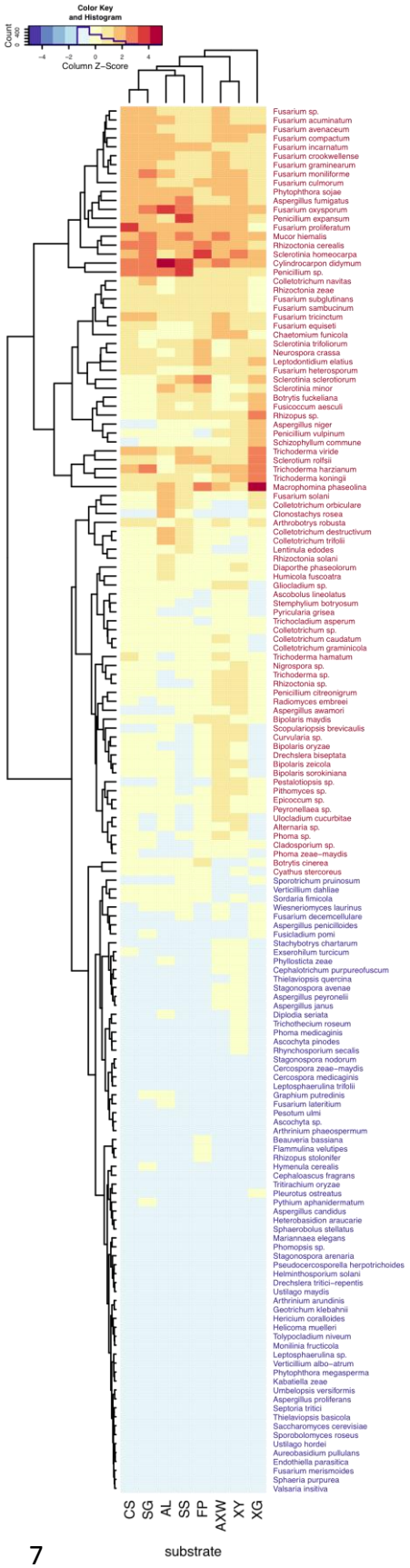

Figure 1. Heat map showing mean activities and clustering of 155 species of plant pathogenic and non-pathogenic fungi when assayed for hydrolysis of eight polysaccharide substrates (XG, xyloglucan (from tamarind); FP, filter paper; AL, alfalfa; SS, soybean stems; SG, switchgrass; CS, corn stalks; AXW, arabinoxylan (from wheat); XY, xylan (from birch)). T. reesei RUT-C30 was excluded from this analysis because of its unusual hydrolytic activities. Negative estimations of reducing sugars were adjusted to zero, and data were standardized within substrates by subtracting the substrate mean and dividing by the standard deviation. Dendrogram and ordering was determined using the distance matrix computation (dist) and hierarchical clustering (hclust) functions in R. Red colors indicate values greater than the substrate mean, while blue colors indicate values less than the mean. Column Z-score and color intensity indicate how many standard deviations the species mean is from the substrate mean.

Cylindrocarpon didymum, *T. viride*, *Macrophoma phaseolina* and *Penicillium* sp. had activities greater than two standard deviations from the mean for at least one of the eight substrates tested and also had high activity on most other substrates. In addition to these broadly active species, *Sclerotium rolfsii* and *Rhizopus* sp. had activity greater than two standard deviations from the mean for hydrolysis of XG and *S. sclerotiorum* had activity greater than two standard deviations from the mean for hydrolysis of FP. However, these three isolates were not extremely active across a broad range of substrates. Several species of *Fusarium* (*Fusarium* sp., *F. acuminatum*, *F. avenaceum*, *F. incarnatum*, *F. graminearum*, *F. crookwellense*, *F. moniliforme*, *F. culmorum*, *F. compactum*, *F. proliferatum* and *F. oxysporum*) clustered together and were highly active on grass cell walls and xylans, as well as being higher than average on most other substrates. Three species of *Trichoderma* (*T. viride*, *T. koningii*, and *T. harzianum*) clustered together and had very high activity on XG and good activity on other substrates. Some species, such as all four species of *Bipolaris*, had activities near or above the mean for AXW and XY, but were average or below average for other substrates. Although negative results from this study do not rule out a species as producing active CWDE, these data identify numerous species that are very capable of hydrolyzing cellulose, hemicellulose and lignocellulosic biomass.

Because the broad-scale screening indicated a number of promising isolates, we recultured a small set of top isolates and tested using a greater number of substrates and hydrolysis times. Data from a subset of isolates with the highest activity for each substrate and time were standardized within each substrate and time. The standardized values were averaged for the two time points, and organized using hierarchical clustering (see Figure 2). Among the top plant pathogenic isolates, *S. homeocarpa* 86-190 showed very good activity on isolated cellulose and xylans, as well as most types of grass cell wall biomass. *F. oxysporum* 85-031 was above average on both untreated and pretreated grasses. The hypercellulolytic mutant, *T. reesei* RUT-C30, shows an unusual pattern of hydrolytic activity. It was highly active on the two pure cellulosic substrates tested [FP and bacterial microcrystalline cellulose (BMCC)] as well as the three pretreated biomass samples [acid pretreated corn stover (PCS), acid pretreated switchgrass (PTSGA) and base pretreated switchgrass (PTSGB)]. However, when compared with the other top isolates, *T. reesei* was the weakest isolate for hydrolysis of the three xylans (AXW, AXO, XY) and five of six types of untreated grass cell walls [SG, eastern gammagrass/switchgrass mix (EGG/SG), big bluestem/switchgrass mix (BBS/SG), tall fescue (TF) and reed canarygrass (RC)]. This clearly illustrates the skewed hydrolytic profile of *T. reesei*, which emphasizes cellulase production that is essential for hydrolysis of pretreated grasses where cellulose is the major component. However, high production of cellulases may result in relatively ineffective hydrolysis of the more complex and heterogeneous untreated plant cell walls in which lignin and hemicellulose may limit cellulose accessibility. For hydrolysis of untreated biomass, high xylanase activity could directly increase the amount of five carbon sugars (mainly xylose and arabinose) as well as stimulate cellulose hydrolysis, perhaps by improving cellulose accessibility.

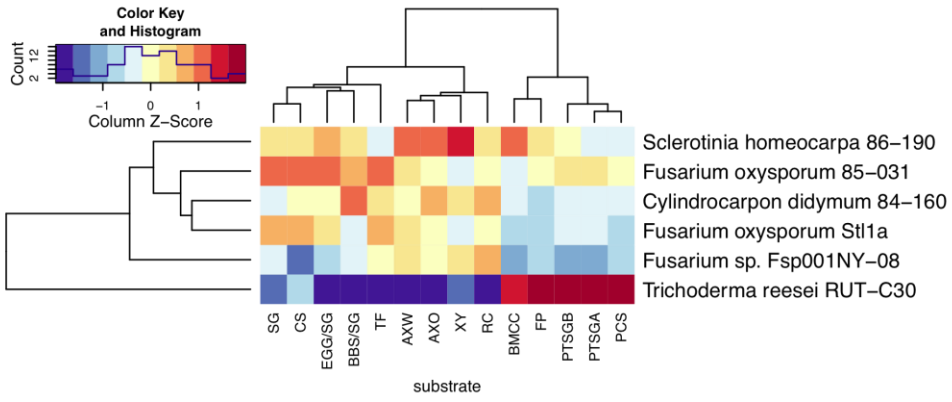

Figure 2. Hierarchical clustering of top isolates and *T. reesei*. Heat map showing mean activities and clustering of the top isolates on each substrate and *T. reesei* when assayed for hydrolysis of 14 polysaccharide substrates (SG - switchgrass; CS - corn stalk; EGG/SG - eastern gammagrass/switchgrass mix; BBS/SG - big bluestem/switchgrass mix; TF - tall fescue; AXW - arabinoxylan (wheat); AXO - arabinoxylan (oat), XY - xylan (birch); RC - reed canary grass; BMCC - bacterial microcrystalline cellulose; FP - filter paper; PTSGB - base pretreated switchgrass; PTSGA - acid pretreated switchgrass; PCS - acid pretreated corn stover). Each substrate was hydrolyzed for two lengths of time, and the mean of those two time points was used for clustering. Selected isolates showed the highest activity on at least one substrate and at least one time point. Measured reducing sugars were standardized within substrates by subtracting the substrate mean and dividing by the standard deviation. Dendrogram and ordering was determined using the distance matrix computation (dist) and hierarchical clustering (hclust) functions in R. Red colors indicate values greater than the substrate mean, while blue colors indicate values less than the mean. Column Z-score and color intensity indicate how many standard deviations the isolate mean is from the substrate mean.

Objective A.4. Cloning and expression of selected plant pathogen-derived CWDEs.

Based on the screening results, many plant pathogens had higher xylanase activity and some highly active isolates had greater activity than *T. reesei* when tested on grass cell walls. Specifically, the species *F. avenaceum*, *F. incarnatum*, *F. graminearum*, *F. crookwellense*, *F. moniliforme*, *F. culmorum*, *F. compactum*, *F. proliferatum*, *F. oxysporum*, *Phytophthora sojae*, *A. fumigatus*, *P. expansum*, *M. hiemalis*, *R. cerealis*, *S. homeocarpa*, *S. sclerotiorum*, *S. trifoliorum*, *C. didymum*, *T. viride*, *T. koningii*, *T. harzianum*, *Chaetomium funicola*, *M. phaseolina*, *S. rolfsii*, and *Leptodontium elatius*, are appeared to be promising candidates in which to discover highly active enzymes in one or more classes of CWDE.

Since a number of the isolates have sequenced genomes, we decided to use the strategy of primer and gene design based on sequencing information for enzyme production. Because of the plethora of *Fusarium* species in the top tier of highly active isolates, we used the publically available *Fusarium* genome, combined with literature analysis of gene and protein expression

data, to identify two endocellulase gene candidates. Genes were synthesized for optimized production in *E. coli* using the vector pET28a (+) for cytoplasmic expression. Constructs were designed to include an inducible promoter, two different secretion systems, and 5xHis and antibiotic tags to aid in clone selection and ease of purification. The cleavable maltose binding (MBP)-His tag was added to each construct for purification of the protein from cell lysates using amylose resin, followed by subsequent digestion with TEV protease to remove the MBP and repurification of the final protein over Ni+ resin for activity testing. Initial expression studies indicated that at least one of the expressed enzymes was active using a Congo red overlay plate assay. Enzyme assays with the initial Cel5A construct indicated that activity was retained as assessed by release of reducing sugar equivalents from CMC cellulose. In enzyme assays, we have detected significant endocellulase activity with the *Fusarium graminearum* Cel5A using several expression systems and fusion partners, and when tested over a range of times and temperatures.

Western blots indicated that expression of the GH61 fusion products were similar to that of Cel5A, and the MBP-GH61 fusion (68 kDa) could be eluted in high purity from the amylose resin. TEV cleavage was rapid (significant cleavage in <30 min on ice), and apparently efficient as determined by disappearance of fusion protein (68 kDa) and appearance of GH61 (26 kDa). At this point, significant losses occurred due to precipitation and degradation of the 26 kDa GH61 product following TEV cleavage and subsequent purification. The protein is expected to contain a disulfide bond, so the fusion was expressed in Origami2 cells and glutathione rather than DTT was used during TEV cleavage. We have also attempted to stabilize the protein and limit precipitation by addition of NaCl; however these strategies have not yet proved successful, although progress has been made. We continued to optimize the process by the inclusion of divalent ions to increase yield and stability. The GH61 fusion construct itself shows little to no endocellulase activity, even upon prolonged incubation with CMC (24 hr at 37 or 50^{o}C). One possibility is that the large fusion partner (or even a small one) could have a serious negative consequence for GH61 binding to cell walls, since it seems to rely on a binding "face." This is not a factor for soluble substrates like CMC that fit into a well defined binding cleft, where we have observed slight but measureable activity. A synergistic effect with other cellulases may be seen with GH61 for the hydrolysis of complex plant cell walls, however until we have stable and purified protein, these types of assays are quite difficult without knowing the precise concentration of GH61 in the cellulase mixture. Furthermore, fusion partners (MBP, 6xHis) may have a more significant effect on binding to the plant cell wall substrates, compared to binding soluble substrates such as CMC. We have determined that there is greater production of the cloned, fusion proteins in *E. coli* by using the Studier auto-induction medium that does not require optimization, and eliminates the need for strict timing for expression analysis. This has resulted in enhanced production of the fusion proteins. Due to the rapid degradation upon cleavage of the fusion site, we have also been evaluating modifications in the protocols so that we retain higher active enzyme.

Although the cloning aspects have experienced some difficulties, there is the opportunity to find "natural variants" of CWDEs that have already been selected by nature through the selection process for pathogenicity. Throughout evolution, the continual biological battles waged by plant pathogens against their plant hosts have led to expanded enzyme arsenals to facilitate the breach

of physical defenses and subsequent utilization of the plant's lignocellulose for nutrients. One interesting aspect of the screening results was the suggestion that xylanase activity was the greatest contributor to hydrolytic efficiency of untreated biomass. Addition of xylanases and xyloglucanases has been shown to enhance overall hydrolysis, especially in cases where pretreatments have not been optimal. The enhanced activity of the extracts produced in this study might also be due to the presence of accessory enzymes that synergize with cellulases and xylanases to increase accessibility. Using data derived from the screening results, plant pathogens appear to display greater activity when hydrolyzing grass biomass than would be predicted from hydrolysis of only cellulose and hemicellulose (Figure 3).

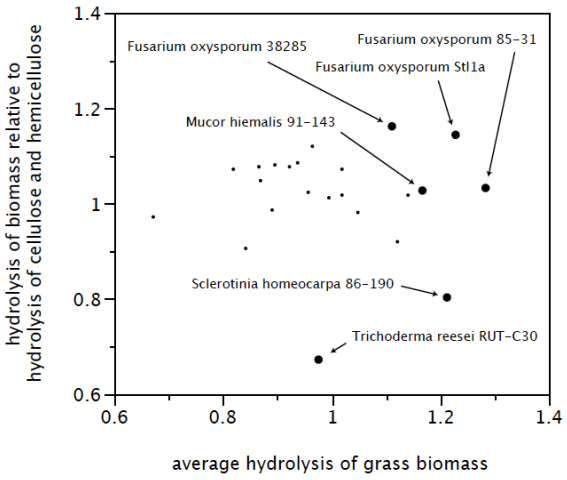

Figure 3: Plot of hydrolysis of grass biomass compared to hydrolysis of cellulose and hemicellulose (y-axis) and hydrolysis of grass biomass (x-axis) for highly hydrolytic isolates of grass biomass (King *et al.*, 2011). Data are based on the 23 highly hydrolytic isolates cultured on switchgrass-based agar, extracted, and assayed on arabinoxylan from wheat for 3 hours (AXW), filter paper for 72 hours (FP}, corn stalk for 72 hours (CS),and switchgrass for 72 hours (SG). Hydrolytic activities for the 23 isolates were standardized to the mean for each substrate. The y-axis illustrates the ratio of hydrolysis of biomass to hydrolysis of purified cellulose and hemicellulose ((CS+SG)/(FP+AXW)). The x-axis represents the average hydrolysis on corn stalks and switchgrass ((CS+SG)/2). All 23 isolates are highly capable of degrading lignocellulosic biomass. Values greater than one on the x-axis indicate isolates especially capable of degrading grass biomass. Values greater than one on the y-axis suggest that hydrolysis of grass biomass is greater than sum of hydrolysis of cellulose and arabinoxylan, implying synergistic involvement of accessory enzymes.

Task B. Elucidating "intrinsic" molecular mechanisms of cellulases at the nanoscale of cellulose fibrils.

Objective B. To use Total Internal Reflection Fluorescence Microscopy (TIRFM) and Confocal Laser Scanning Microscopy (CLSM) to measure the binding, surface mobility, and activity of single and binary mixtures of fluorescently-labeled cellulases, and to develop mathematical models of these surface phenomena.

Objective B.1. Conduct kinetic studies through CLSM. During the FY2009 the surface diffusion and binding characteristics of cellulases to cellulose fibrils and cellulosic structures will be investigated using CLSM. From these experiments we expect to calculate surface diffusion rates which compare to those previously published in the literature.

Objective B. 2. Conduct kinetic studies of cellulases through TIRFM. During FY2009 work will be completed on the tracking of cellulase processivity on cellulose fibrils. By the end of the year cellulase displacement will be evaluated in the context of non-productive surface diffusion and processivity.

Objective B. 3. Carryout characterization of mutants through TIRFM and CLSM. Activities will be initiated on the study of the impact of mutations on cellulase-cellulose interactions through TIRFM and CLSM. During FY2009 the purified fluorescently labeled mutants obtained from FY2008 will be tested for changes in the interactions with cellulose.

Objective B.4. Study of synergism effects through TIRFM and CLSM. On FY2009 activities will be initiated on the study of the effect of cellulase mixtures on cellulase-cellulose interactions. By the end of the year, mixtures of two fluorescently labeled cellulases will have been imaged at the single molecule level, and the changes in activity will be assessed in the context of synergism.

TASK B RESULTS

Objective B.1. Conduct kinetic studies through CLSM. Throughout this project we have endeavored to develop methods for the comprehensive study of the binding and diffusion behavior of cellulases on cellulose fibrils, bundles, and mats. Method development focused on immobilization of cellulose onto solid support [3], managing photo-bleaching of fluorphores[4], and developing tracking software for measuring and analyzing the surface diffusion on insoluble cellulose [5]. Our approach has been to use these methods to explore hypotheses regarding cellulases surface diffusion and binding along the cellulose fibrils. Our hypothesis on cellulase displacement was that cellulases associate with the cellulose fibril surface in one of several modes: (a) irreversible binding without motion (immobile fraction), (b) reversible binding without motion, (c) reversible binding with fast micron-length displacement (long range surface diffusion), and (d) reversible binding with slow nanometer-length displacement (short-range surface diffusion). We developed experimental protocols to assess the binding and long-range

surface diffusion of cellulases Cel5A, Cel6B, and Cel9A through CLSM-FRAP and short-range surface diffusion through TIRF-SMT with sub-10nm localization accuracy. Experiments to study binding were conducted in presence of excess of labeled enzymes in free solution, while experiments to test the surface diffusion of cellulases on cellulose surfaces were conducted under continuous buffer flow.

As a result of our successful method development activities we have successfully used confocal microscopy coupled with fluorescence recovery after photobleaching (FRAP) to study the binding/unbinding kinetics of cellulases Cel5A, Cel6B, and Cel9A on bacterial microcrystalline cellulose. The experiments were conducted on individual fibrils and dense cellulose mats at 23, 34, and 45°C as described in previous reports. The FRAP recovery curves obtained for individual bleached areas were averaged over individual experiments and across experiments to obtain curves that are representative of the cellulase behavior under the given experimental conditions. Results for our FRAP experiments are summarized in Figure 4. It can be observed that recovery is slower on dense mats structures than on individual fibrils, a result that supports our previous observation that diffusion hindrance into the porous structure of dense mats is occurring. Qualitatively, the results presented in Figure 4 show that for Cel5A, an increase in temperature from 23 to 34 and from 34 to 45°C, increases the speed of recovery in all structures, while having little effect on the amount of enzymes that can dissociate from the cellulose surface. For Cel6B, only the increase of temperature from 34 to 45°C increases the speed of recovery, while going from 23 to 34°C does not provide a significant increase of recovery for any of the cellulose structures. In addition, the amount of bound Cel6B enzymes that can dissociate from the cellulose matrix is slightly reduced as the temperature is increased. For Cel9A it is hard to qualitatively assess whether the recovery is faster at higher temperatures. However, it is evident that as the temperature is increased the percentage of bound enzymes that could dissociate from cellulose was reduced. Experiments at all temperatures were highly reproducible for Cel5A and Cel6B, giving confidence in the observed behavior. Experiments with Cel9A however, were only reproducible at 23°C, while at 34 and 45°C the experiments varied significantly. Our hypothesis on the observed variability is discussed below.

Faster recovery at higher temperatures was the expected result, because at higher temperatures the added thermal energy allows faster dissociation from the cellulose matrix, leading to a higher dissociation constant and faster turnover of enzyme binding/unbinding. In this sense, the behavior of Cel5A and Cel6B conform to the expected behavior. Conversely, Cel9A does not seem to follow the trend of faster binding/unbinding kinetics with higher temperature. To obtain quantitative results from the experiments, we fitted the recovery curves for fibrils to a FRAP model where recovery is dominated by binding:

$$I_B(t) = \frac{k_1^* C_F \alpha}{k_2}\left(1 - e^{-k_2 t}\right)$$

where k_1^* is the apparent binding rate, C_F is the concentration of fluorescent enzymes in solution, α is the proportionality constant that relates the local enzyme concentration to the fluorescence intensity, k_2 is the unbinding rate, and t is the time in seconds.

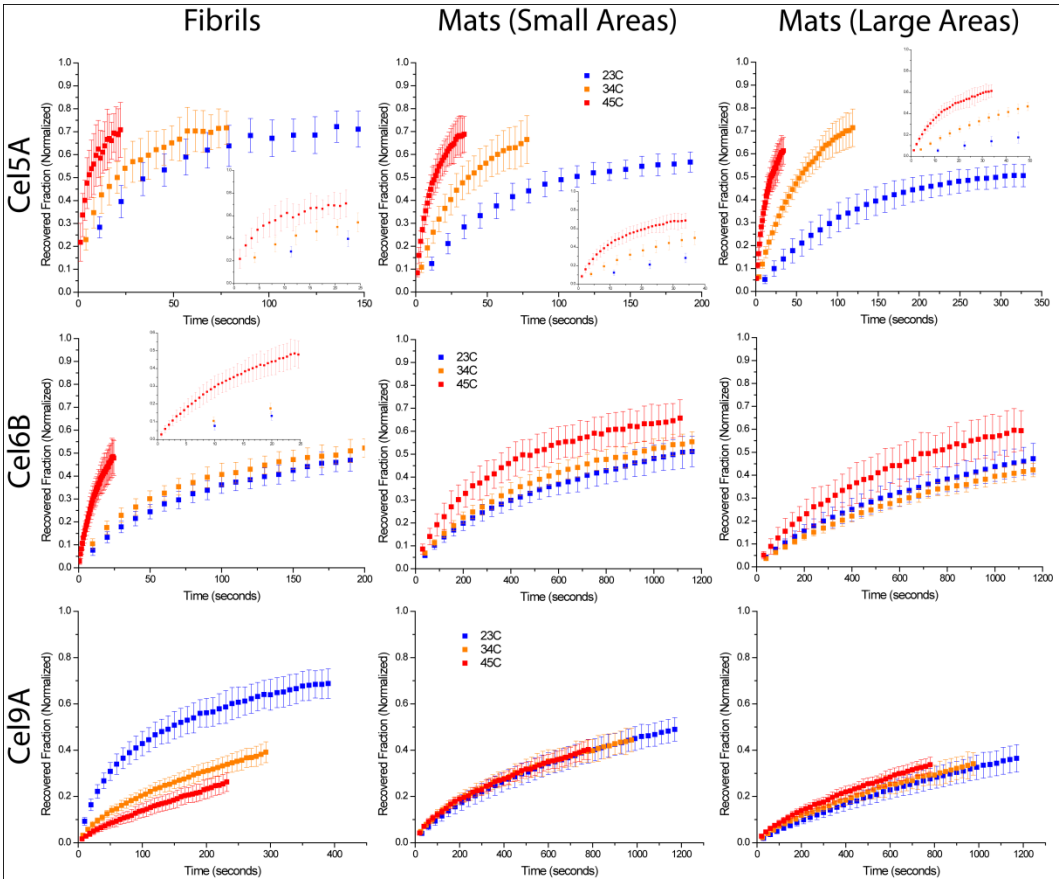

Figure 4. Comparison of the averaged FRAP recovery curves for different temperatures on the different cellulose structures. Curves are averages of recovery curves for individual bleached areas (n = 5-30). Error bars represent the standard deviation of the different recovery curves. Insets present zoomed views of the average recovery curves at 45°C.

FRAP experiments on fibrils were selected for fitting because they allow a simplified interpretation that does not have to take into account the pore entrapment and diffusion hindrance encountered in dense cellulose mats. Fitting to the model was performed on the averaged curves using the known concentration of enzymes in solution and the proportionality constant obtained from calibration measurements, and the values for k_1^* and k_2 were obtained (Figure 5a). The asymptotic maximum of the recovery curves $\left(\frac{k_1^* C_F \alpha}{k_2}\right)$ represents the percentage of the initial fluorescence intensity that is recovered, or the percentage of bound enzymes that can unbind and be replaced by enzymes in solution (mobile fraction). The difference between this number and unity yields the percentage of bleached enzymes that cannot unbind and remain immobilized on the cellulose surface (immobile fraction). From Figure 5c it can be observed that as the temperature is increased, the mobile fraction remains constant for Cel5A, shows a small decrease for Cel6B, and decreases significantly for Cel9A. A decrease in the mobile fraction can only be explained by irreversible binding to the cellulose structure either by denaturation or cellulase stalling on the cellulose fibril. The fact that the decrease is seen on Cel6B and Cel9A suggests a relationship between enzyme processivity and immobile fraction.

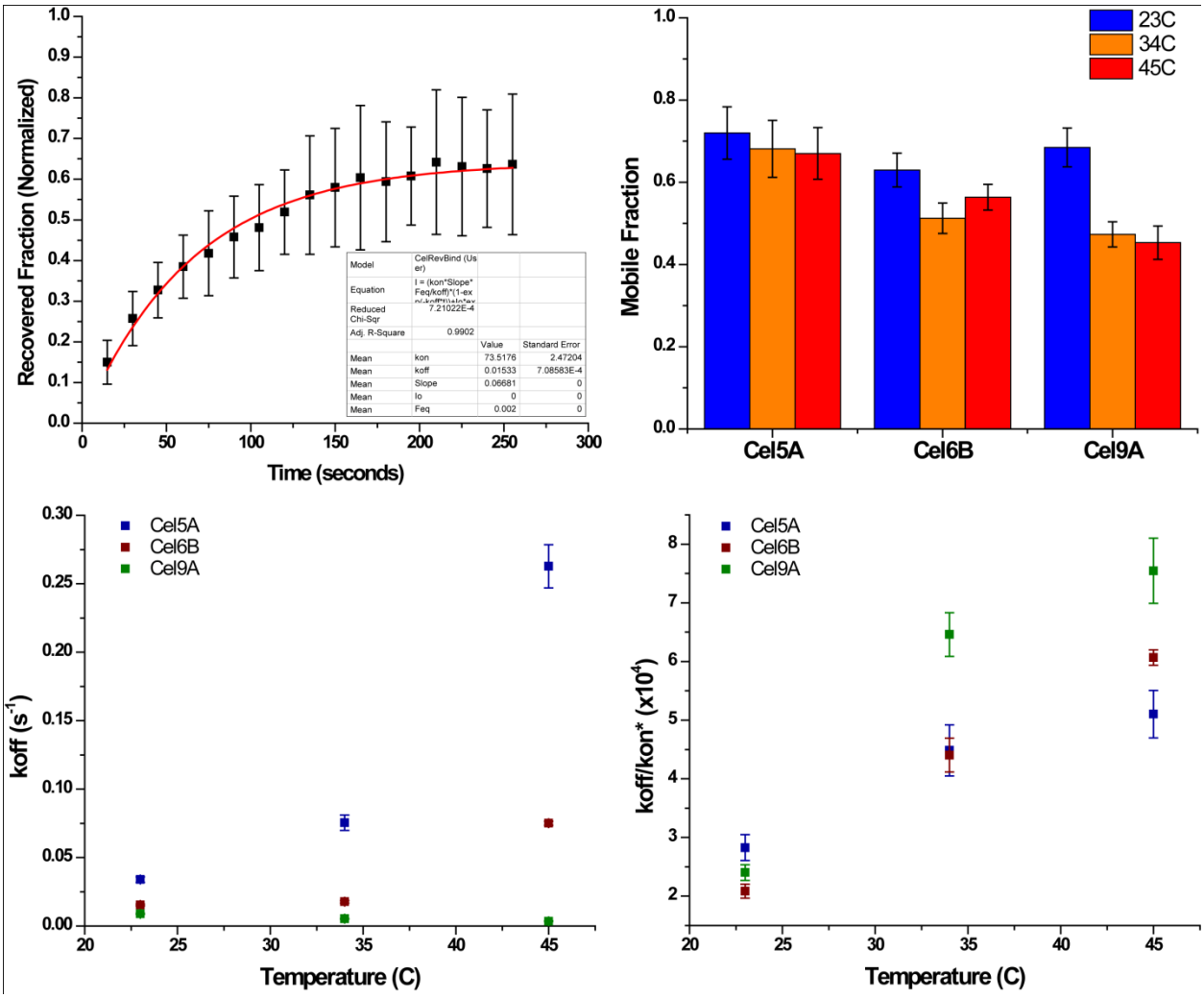

Figure 5. Fitting of recovery curves from FRAP experiments to a binding dominated model allows the calculation of kinetic parameters for all experimental conditions tested for bleaching experiments performed on fibrils. (a) Sample FRAP recovery curve and its corresponding fit to the recovery model. (b) Comparison of the mobile fraction of enzymes for treatments at different temperatures. (c) Unbinding rate extracted from FRAP experiments for the three cellulases at the different temperatures tested. (d) Apparent dissociation constant and its dependence on temperature.

It is also notable that at room temperature, where catalytic activity is minimal, all enzymes show the same percentage of mobile enzymes, suggesting that the fraction of mobile enzymes under conditions with minimal catalytic activity is mainly determined by the CBM binding properties. The fitted k_2 values only depend on the temporal evolution of the fluorescence recovery and represent true unbinding rates (Figure 5c. It is seen that as the temperature is increased, the unbinding rate increases both for Cel5A and Cel6B, indicating faster unbinding or a lower amount of energy required for the enzyme to unbind. Cel9A in contrast shows a slight decrease in its unbinding rate implying tighter binding as the temperature is increased. The ratio $\frac{k_2}{k_1^*}$ is an

apparent dissociation constant which conveys a sense of the impact that temperature has on binding, and its equilibrium of bound concentration. It is not a true dissociation constant because the denominator is an apparent binding rate which has subsumed in it the concentration of binding sites on the cellulose fibrils. Assuming that the concentration of binding sites can be extrapolated from the maximum amount of bound cellulases on the cellulose fibrils, the apparent dissociation constant values would have to be multiplied by a number on the order of 10^{-6}, yielding dissociation constants with values on the order of 0.1-1 nM. This number should serve as an estimate, and not as a definitive value, since errors compounded onto the estimation of $k_1$could make this dissociation constant be under or overestimated by up to a factor of 5.

Objective B. 2. Conduct Kinetic studies of cellulases through TIRFM. We have completed the analysis of the collection of Single Molecule TIRFM movies obtained from experiments with AlexaFluor647 labeled cellulases Cel5A, Cel6B, and Cel9A at 23 and 45°C. During our analysis of the tracks generated, we have considered only tracks that appear in the movie for a minimum of 30 frames, which guarantees the removal of short, spurious tracks generated due to noise. We have also analyzed the particle tracks for all enzymes at 23 and 45°C with a combination of parameters W = 3, and R = 3 (parameters described in the previous quarterly report), which ensures consistency in the analysis at the expense of eliminating long jumps that could span a distance of more than 3 pixels in a single frame. Qualitatively, the videos at 45°C show more movement from all three cellulases than the ones taken at 23°C. This is an expected result because at higher temperature the enzymes become active and can hydrolyze the glucosidic bonds on the cellulose chains and more thermal energy is available to permit molecular motion. The trend of Cel5A showing larger mobility the Cel6B and Cel9A is maintained at the higher temperature. As discussed in previous reports, an analysis of net displacement (distance between the beginning and the end positions) was conducted for all tracks recorded, and is presented in Figure 6. **The data at 23°C shows that for all three enzymes a significant portion of all tracks have net displacements smaller than a pixel.** With the pixel size being 66.7 nm, these displacements fall within our accuracy of localization (20 nm for enzymes containing one fluorophore and 10 nm for enzymes containing 4 fluorophores). Because of the bimodal distribution observed in the semi-log histograms of Figure 1, we decided to set as a threshold for mobility a net displacement of 1 pixel. This implies that molecules presenting net displacements smaller than 1 pixel are considered immobile and those with displacements larger than one pixel are considered mobile.

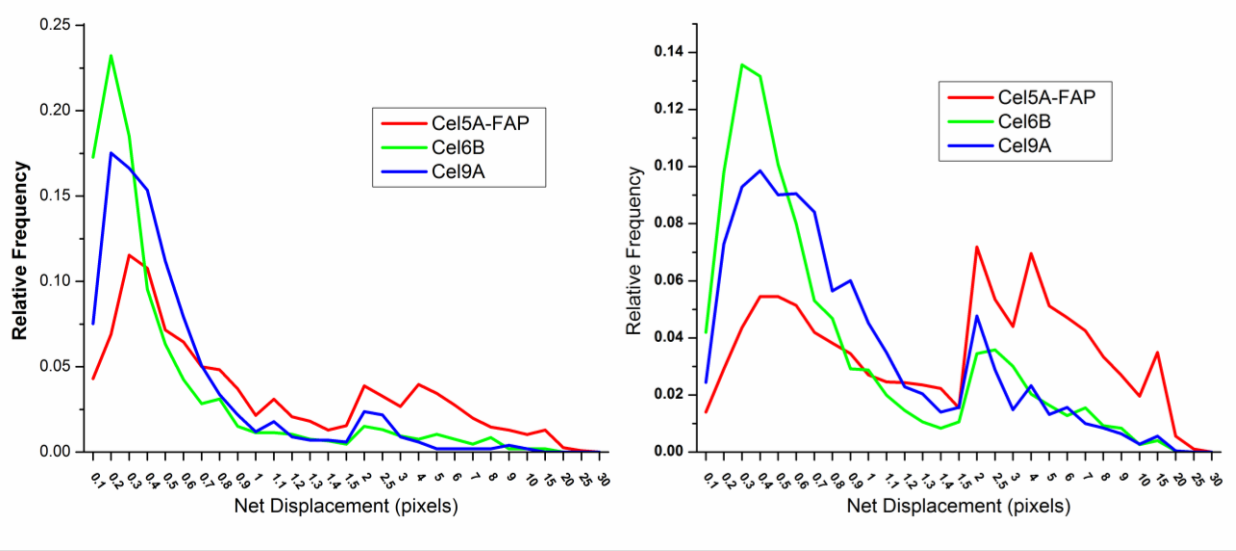

Figure 6. Aggregated data for the net displacement for SM trajectories obtained from *T. fusca* cellulases Cel5A, Cel6B, and Cel9A. Left: Binned net displacements for tracked cellulases at 23°C. Right: Binned net displacements for tracked cellulases at 45°C. From these histograms it can be observed that at both temperatures Cel5A is more mobile than Cel6B or Cel9A, and that this trend continues at 45°C. (pixel size = 67nm)

To further quantify the fraction of enzymes in the mobile versus immobile states it is illustrative to plot the cumulative frequency of net displacements, which is obtained by adding the relative frequencies depicted in Figure 6. Figure 7 shows the cumulative frequency plots. Using 1 pixel as the threshold for immobile enzymes versus enzymes showing displacement from Figure 7 it can be observed that at 23°C, 63% of the Cel5A enzymes are immobile and 37% are mobile; 88% of the Cel6B enzymes are immobile and 12% show some range of motion; and 88% of the Cel9A enzymes are immobile while 12% show some range of motion. As the temperature is increased, these percentages shift, yielding a higher proportion of enzymes in the mobile state. At 45°C, only 38% of the Cel5A enzymes are immobile, while 62% show mobility; 75% of the Cel6B enzymes are immobile and 25% show some displacement; and 71% of Cel9A enzymes are immobile, while 29% show mobility. For all three enzymes, the fraction of mobile enzymes almost doubles from that observed at 23°C. Again it must be stressed that the movements we report are unlikely due to processivity, which would yield much smaller displacements below our resolution limit. Processive movement of enzymes in this case would be included within the fraction that we have considered immobile. The single molecule movies obtained have shown three distinct states that the cellulases exhibit on the cellulose surface: (i) immobile enzymes which don't move or have displacements below our resolution limit, (ii) enzymes which move with small displacement steps over a number of consecutive frames, (iii) enzymes that show fast, long displacements that are short lived. The first state would represent enzymes that are either in the process of catalysis or irreversibly bound to the cellulose surface. The second state represents true surface diffusion along the cellulose surface, with displacements that would resemble Brownian motion. The third state represents the unbinding and local rebinding of cellulases, and should yield diffusion coefficients similar to those observed for cellulases in free solution. This third state is the hardest to characterize, because the large displacements are very short lived, and

happen within one or two frames of the movies. In order to successfully extract all the information from the obtained tracks more complex models are needed, which can account for all three states and the probabilities for the transitions between them. This problem can be approached using Hidden Markov Models, but the development of this modeling approach is beyond the scope of this report.

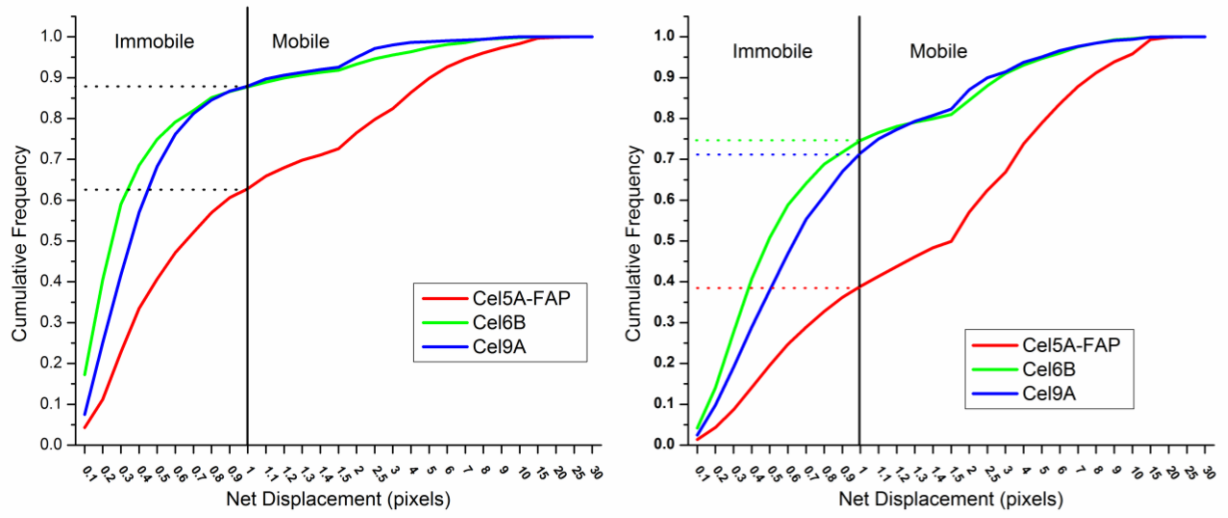

Figure 7. Cumulative frequency for binned net displacements for *T. fusca* cellulases Cel5A, Cel6B, and Cel9A at 23°C (left) and 45°C (right). Histograms from Figure 1 were used to calculate the cumulative frequencies. A threshold of 1 pixel = 67 nm was used for the distinction between mobile and immobile enzymes based on the accuracy of localization for our optical system.

Objective B. 3. Carryout Characterization of mutants through TIRFM and CLSM.

Cel6B mutants S232A and D226A, and Cel6A catalytic domain mutants W16I, and G87A were produced and their activities on CMC, SC, and BMCC tested. From all our results it appears that S232 and D226 act together as the catalytic base by binding a water molecule that carries out hydrolysis. This water can be seen in the X-ray structure of the wild type enzyme and it is missing in the structure of the S232A, D226A double mutant enzyme (unpublished results from Dr. Henrik Hansson, Uppsala University). Additional support for this conclusion is shown in Figure 8 as we see that the double mutant enzyme shows some activity rescue by azide as expected for a mutant that inactivates the catalytic base. There was no azide rescue seen in either of the single mutant enzymes.

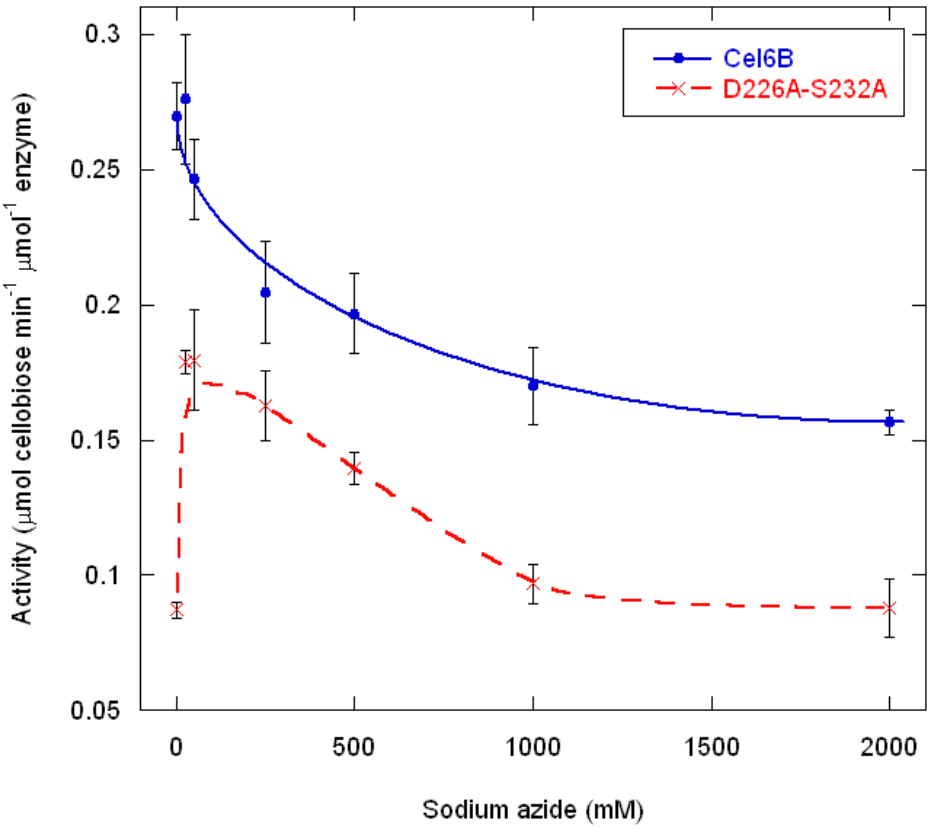

Figure 8. Activity recovery for double mutant enzyme through the addition of sodium azide and comparison with native enzyme activity.

We continue to work on the fluorescently labeling and purification of Cel6B mutants S232A and D226A, and Cel6A catalytic domain mutants. This milestone has been slowed because the student working on the project finished his degree before he could label the enzymes. However, recently Prof. Wilson has just hired a new technician who is producing these labeled enzymes and has had some success. We are anticipating moving these label enzymes into our imaging program this summer.

Objective B.4. Study of synergism effects through TIRFM and CLSM. Understanding the molecular basis of cellulase interactions responsible for synergistic behavior observed between cellulases is critical for rationally designed enzyme cocktails. Factors such as substrate chemical and morphological properties, and/or physical or chemical inhibition can limit substrate hydrolysis. Insight into the interaction of synergistic activity and cellulose morphological structure will clarify these limits. Traditional biochemical techniques that take indirect bulk solution measurements of free protein in the system to determine bound protein concentration limit the spatial and temporal resolution of understanding synergism. Integrating advanced imaging techniques like epi-fluorescent microscopy will provide greater insight into the molecular mechanisms that govern cellulase synergism by allowing real-time observation and direct measurements of bound cellulase concentrations on an insoluble substrate. Binding kinetic data for multiple cellulases with different catalytic actions co-localizing on simple and complex

cellulose morphologies can be determined with a high degree of spatial and temporal resolution that has not previously been possible. This approach could elucidate the heterogenous catalytic interaction between cellulases and cellulose and clarify the role of synergism in cellulose hydrolysis.

Pure fluorescently labeled populations of *T. fusca* endocellulase Cel5A, and exocellulase Cel6B, were applied to immobilized cellulose. Experiments were conducted at room temperature to measure binding without activity, at 45°C to measure binding near peak enzymatic activity and at an intermediate temperature. Binding data, in the form of fluorescent images was recorded using time-lapsed fluorescence microscopy with images taken at defined intervals for a period of four hours to ensure saturating conditions. Kinetic binding curves were established for mixtures of these two cellulases at each temperature to observe the effect of temperature and hydrolytic activity on cellulase binding.

Binary synergistic mixtures. Cellulases are an important class of plant cell-wall degrading enzymes that biochemically convert cellulosic feedstocks to fermentable sugars for bioenergy and bioproducts. These enzymes exhibit different substrate binding characteristics depending on their reactive domains, and while individual cellulases hydrolyze cellulose relatively slowly, mixtures of cellulases with different reactive domains act synergistically to enhance rates and extents of hydrolysis, releasing more sugars. Traditional biochemical techniques that take indirect bulk solution measurements of free protein to determine bound protein concentration are limited in resolving the spatial and temporal patterns of enzyme diffusion and binding that are key to understanding synergism. Integrating advanced imaging techniques like epi-fluorescent microscopy with the existing biochemical data will allow us to overcome these limits and offer the potential for extracting greater insight into the molecular mechanisms that govern cellulase synergism.

Cellulase binding was observed on immobilized cellulose substrates for two *T. fusca* cellulases, Cel5A and Cel6B. Previous studies have shown that cellulases Cel5A and Cel6B have little activity on BMCC at room temperature. Thus, little observed enzyme activity and hydrolysis of the substrate was expected during time course set in the binding experiment. The time course experiment was designed to measure the various stages of binding, including the initial phase, when binding is expected to be linear and time resolution is important, a leveling off phase, when the rate of binding decreases, and finally in the plateau stage, when binding does not appreciably increase. The plateau stage should correspond to a saturation of available binding sites on the cellulose substrate region of interest and equilibrium. All fluorescence data was normalized such that the highest detected signal was set to 1. Due to the heterogeneous nature of the cellulose morphologies and the variations within any one morphological category, normalizing the signal to 1 allowed for comparisons between experiments when the actual structures sampled were of different sizes, with different surface areas, and therefore different binding capacities.

In both the single component and 50/50 mixture experiments, shown in Figure 9 A and B, Cel5A bound rapidly, reaching equilibrium within an hour of the initial binding. Cel6B binds more slowly, approaching the equilibrium stage 4 hours after the initial binding, as shown in figure 1 C inset. Cel5A displayed a different initial binding pattern in mixtures than it had as a single component. This could be a result of competition with Cel6B for easily accessible surface

binding sites. Cel5A reached the plateau stage after 30 minutes in the single component experiments and after 50 minutes in the 50/50 mixture with Cel6B. This difference in time to equilibrium could again be a result of competition for binding sites with Cel6B in the mixture.

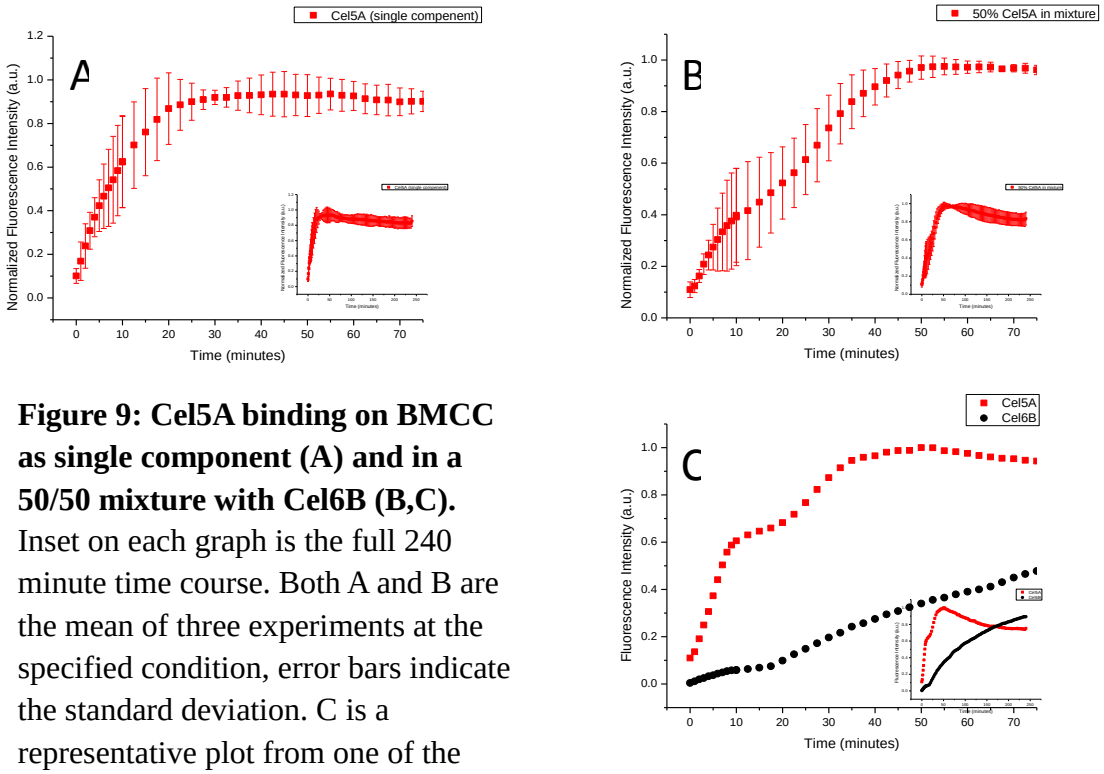

Figure 9: Cel5A binding on BMCC as single component (A) and in a 50/50 mixture with Cel6B (B,C). Inset on each graph is the full 240 minute time course. Both A and B are the mean of three experiments at the specified condition, error bars indicate the standard deviation. C is a representative plot from one of the 50/50 mixture experiments.

TASK C: Integrating the insights gained from "intrinsic" molecular mechanisms into the mesoscale heterogeneous kinetic framework of pretreated cellulosic particles.

Objective C. The overall goal of this task is to utilize infrared technology to chemically characterize cellulosic structures before, during, and after hydrolysis, and to explore the use of confocal laser scanning microscopy (CLSM) to observe the binding and diffusion of cellulases on the surface and into the porous structure of pretreated cellulose.

Objective C.1. *Characterization of cellulase interactions with pretreated cellulose particles.* A major goal of this project has been development of accurate and quantitative analysis using high throughput Fourier Transform Infrared (HT-FTIR). We will conduct a comprehensive set of binding, hydrolysis and synergism experiments using our high-throughput analysis system to correlate cellulase cocktail composition, and substrate physical and chemical properties with rates and extent of hydrolysis, and degree of synergistic effect.

Objective C.2. Develop methods and standardized for evaluation of cellulase binding and cellulose particle hydrolysis on high throughput system

Objective C.3. Evaluation of synergistic effect of cellulase mixtures on hydrolysis of cellulose particles

TASK C RESULTS

Objective C.1. Characterization of cellulase interactions with pretreated cellulose particles.

We have developed methods and standards for the evaluation of the saccharification of bacterial microcrystalline cellulose (BMCC) by *T. fusca* cellulases Cel5A, Cel6B, and Cel9A, and by synergistic mixtures of these cellulases.[6] With these methods the spectral signatures of individual types of bond vibrations have been assigned to peaks observed in the FTIR spectra, background introduced by buffer salts has been removed, and the changes in the spectral signatures of BMCC induced by enzymatic depolymerization have been quantified. This method development lays out the foundation for quantitative analysis of the saccharification of lignocellulose through HT-FTIR technology.

Different methods for quantitative and qualitative standardization of raw HT-FTIR spectra were tested to allow direct comparison between samples and to increase the signal to noise ratio. The first approach was a direct baseline correction followed by a smoothing procedure using Bruker's OPUS software. All the baseline corrected spectra were normalized towards a selected sodium acetate signature in order to standardize the overall spectral data. The second approach consisted in subtracting the signal arising from the acetate buffer (Figure 10). The first approach appeared to be faster and more suitable for the comparison of the datasets if the acetate signals were less intense than the cellulose signal. Standard deviation of 6 replicates (using different ratios of BMCC to acetate, and different loadings) were measured to compare the standardization

of the spectra dataset. The method (ratio, loading and data treatment) yielding the lowest SD was chosen and implemented for the saccharification studies described below.

During this project we have also reviewed and assigned most of the signal peaks found in the mid-infrared spectra of bacterial microcrystalline cellulose (BMCC) as shown in Figure 11. Deconvolution of the spectral peaks, especially in the OH region (3000-3600 cm^{-1}), is necessary to identify the hydrogen bonding signals. Several fundamental peaks were identified and documented as their wavenumbers correspond to their assignment in published literature (+/- 1 cm^{-1}). The wavenumber values reported in the literature were used for the peak fitting procedures in order to determine how many true signals were present and to avoid over-fitting the spectra. Signals were assigned on the entire Mid-IR region (4000 cm^{-1} to 400 cm^{-1}).

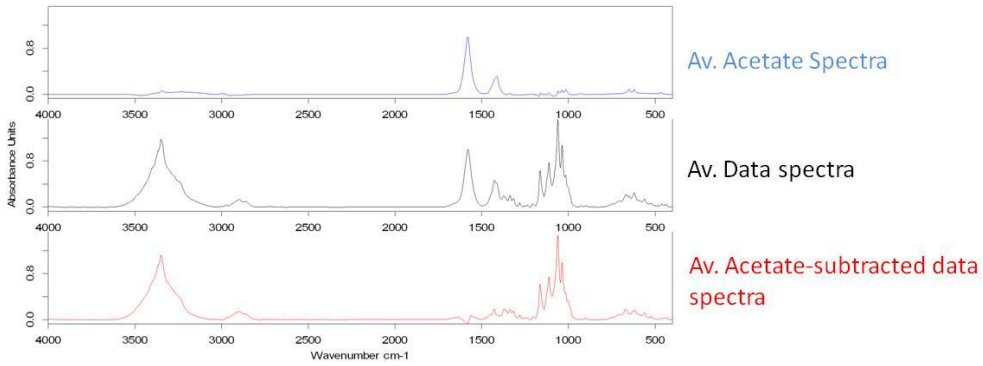

Figure 10. Spectral subtraction of the acetate signals from normalized cellulose in acetate sample (Blue: acetate spectral composite, black: normalized and standardized original spectra, red: subtracted "pure" BMCC spectra).

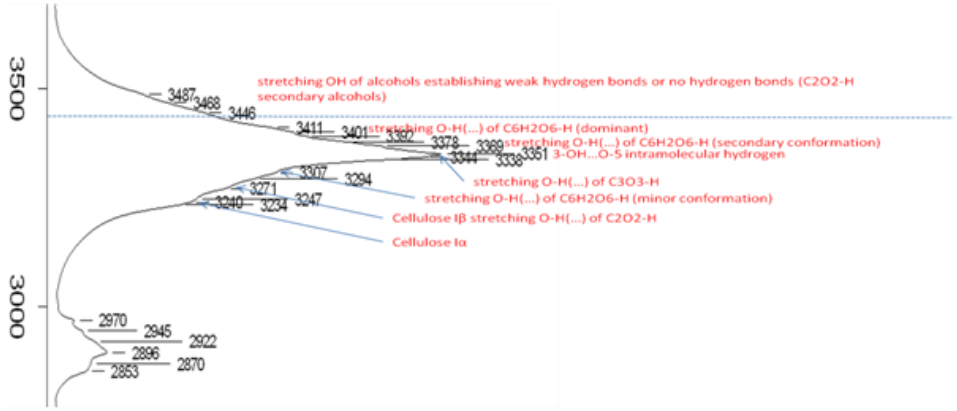

Figure 11. Assignment of the hydrogen bonding of native BMCC in the 3000-3800 cm-1 region).

Using the methodologies described, we have performed a series of BMCC saccharification experiments (sodium acetate buffer, 24hr hydrolysis, 50°C) using Cel9A, Cel5A, Cel6B individually, and synergistic mixtures of Cel5A and Cel6B (1:10 and 1:9 ratio respectively) (see Corgie et. al.[6] The quantity of soluble sugars produced were measured by HPLC and correlated with the data obtained using the discussed analytical parameters from the IR spectra. The concentrations of enzymes and cellulose per reaction were varied to assess the best signal to noise ratio after hydrolysis. Post-dilution of the samples was also tested in order to further achieve the best signal on the analytical wafer.

For the experiments depicted in Figure 12 all the spectra were baseline corrected and normalized towards the acetate peak at 1576 cm^{-1}. Concentrations of glucose, cellobiose, cellotriose and cellotetraose were also measured by HPLC in the supernatant. With these results we demonstrate that the degradation of cellulosic material can be captured using high-throughput FTIR spectrometry, as the degradation measured is well correlated with the sugar measured by HPLC.

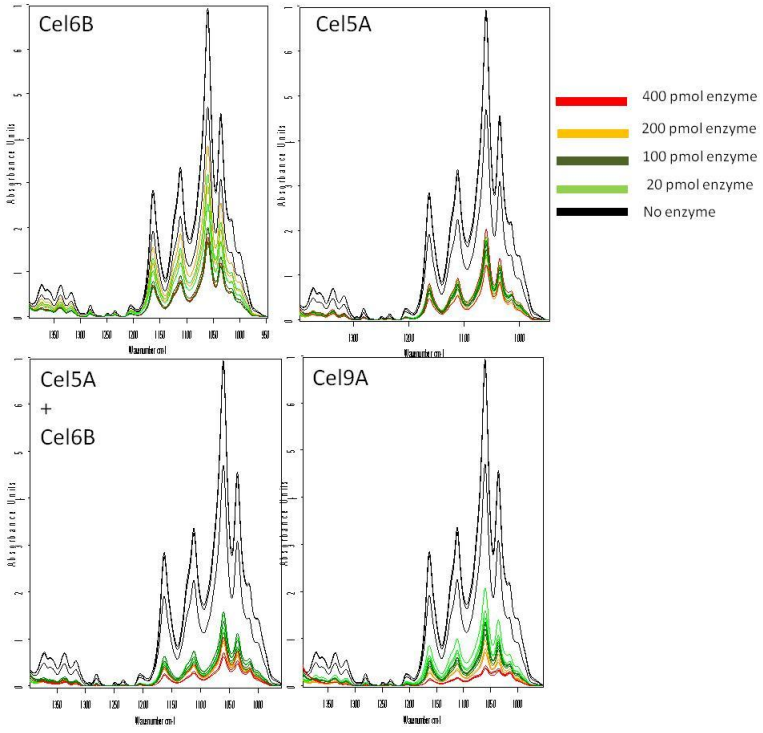

Figure 12. Normalized spectra in the cellulose fingerprint region for the hydrolyzed BMCC with various concentration of cellulases.

Cellulase interactions with pretreated materials. During this project, the evaluation of raw and hydrothermal pretreated biomass was performed using ATR and HT-FTIR (diffuse reflectance and transmittance sampling modes) on mixed hardwood, switchgrass, and mixed perennial grasses. Size and concentration of biomass, as well as deposition and drying methods were investigated to determine the conditions that limit the inter- and intra-variability for HT-FTIR methods. Internal calibration standards were used to validate the methods. Sodium acetate, potassium acetate, boric acid, and ammonium thiocyanate were tested as internal standards (IS) in order to evaluate the interferences with the relevant signals for biomass and pretreated biomass. Optimal ratios of biomass-to-IS were also tested in reflectance and transmittance modes. Results showed that ammonium thiocyanate had the least overlapping vibrational modes with raw and pretreated biomass samples and had a scrap and characteristic peak around 2060 cm^{-1} due to the strong signals of the C≡N bond. Finally, signal assignments were performed on raw and pretreated switchgrass and mixed grasses focusing on the cellulose, hemicellulose, and lignin fingerprints in the 1800-800 cm^{-1} region (see Figure 13).

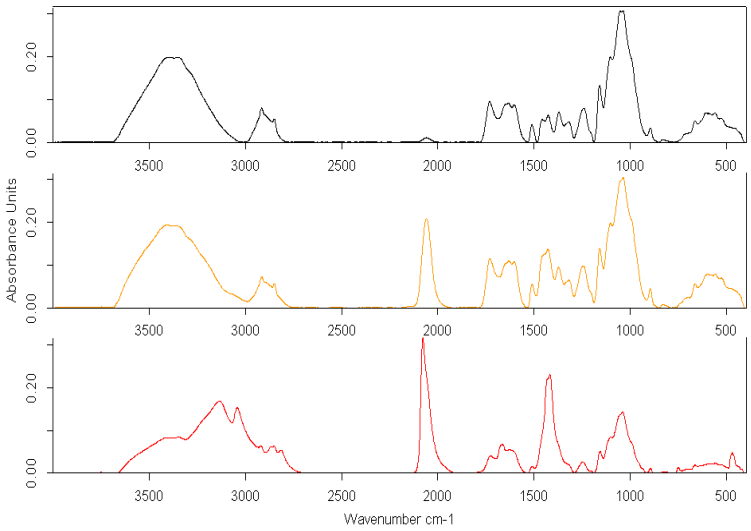

Figure 13. HT-FTIR spectra of raw switchgrass with increasing concentration of ammonium thioisocyanate (characteristic signal around 2060 cm^{-1}, low, medium, high concentrations in black, orange and red respectively).

Quantitative screening of pretreated biomass by QHT-FTIR**.** Based on the insights and analytical parameters described in Corgie et. al. [6], we have developed a subsequent quantitative method for the HT analysis of raw and pretreated biomass samples based around the same FTIR platform (HTS-XT-Vertex 70, Bruker). We tested several internal standards with strong signals that do not interfere with the biomass ones in order to be able to calibrate the samples and reduce the variability between replicates. We also tested the effect of particle size from various grinding methods on the signal quality (see Figure 14). A coarse grinding of the biomass in moist conditions provided the best results in term of particle size for a good signal-to-noise ratio. The quantity of biomass and the biomass-to-standard ratio were optimized. Standard curves were developed for cellulose and lignin monomers (ferulic, coumaric and synaptic acids).

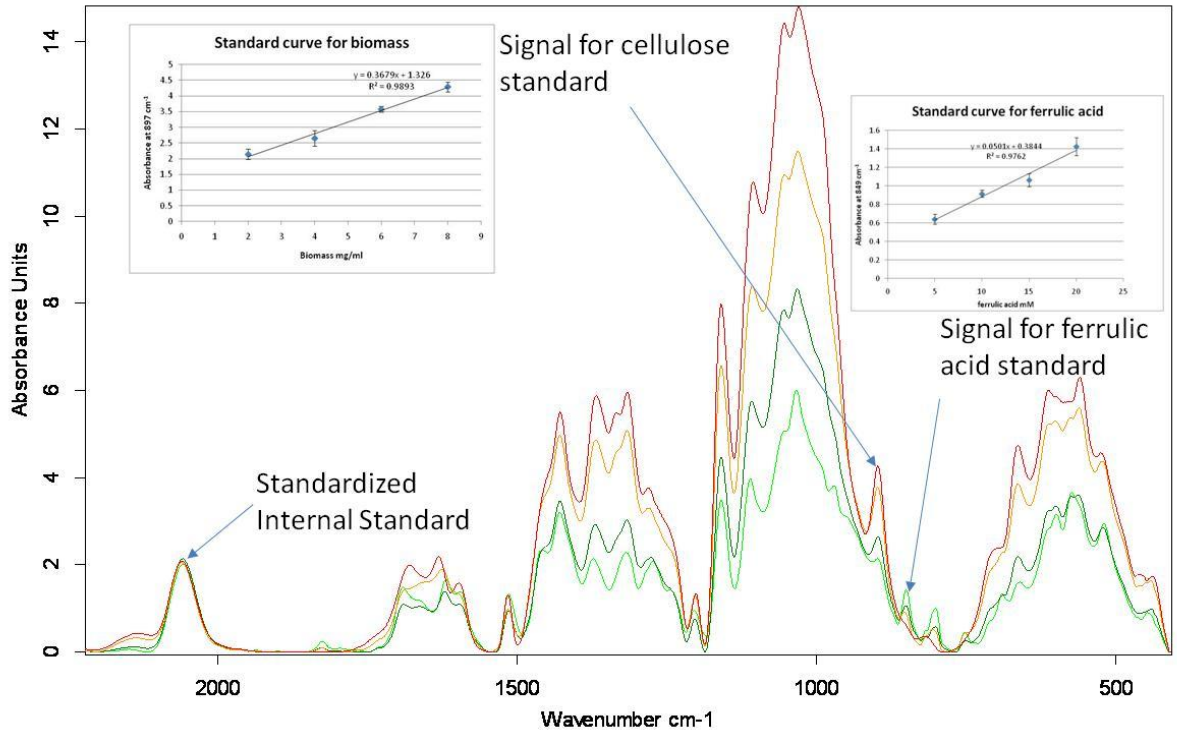

Figure 14. Standard curves for cellulose and ferrulic acid residues. Solkaflock (red, orange, dark green, light green at 2, 4, 6 and 8mg/ml respectively) mixed with a standard for ferrulic acid (red, orange, dark green, light green at 5, 10, 15, 20 mM respectively). Inserts are showing the linear regression using the signal at 897 cm^{-1} for cellulose (β glucosidic linkage) and 849 cm^{-1} for ferrulic acid (Out of plane CH bending of the aromatic ring). The normalized internal standard is shown at 2058 cm^{-1}

Samples coming from our previously developed pretreatment method were grinded and calibrated for particle size and concentration. Wood, Corn Stover and Big Bluestem samples were used. Pretreatment conditions ranged from 150 to 250°C and 40 sec to 1 hour residence time. Pretreated and hydrolyzed samples were also analyzed. The FTIR analysis of the pretreated samples showed a preferential hydrolysis of the hemicelluloses fraction in the range of 150 to 180°C. The overall quantity of lignin residues appeared to be constant in this range of conditions, although a modification of the composition of the aromatic fraction was noticed at higher temperatures. The cellulose fraction was not modified under these conditions but was oxidized above 180°C. Signal enhancement and deconvolution of the regions of interest are underway to determine and quantify the fundamental atomic modification of the cellulose, hemicelluloses and lignin by pretreatment conditions.

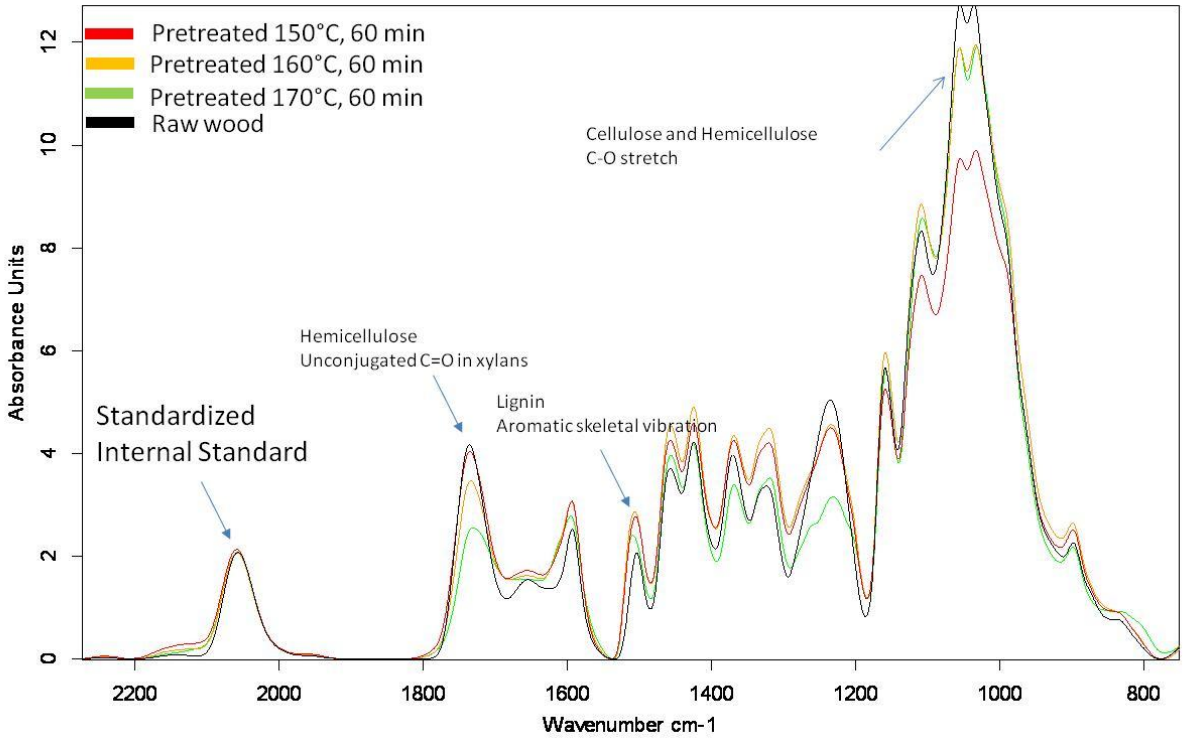

Figure 15. Standardized FTIR spectra of supercritical CO_2 pretreated softwood at 150, 160, and 170°C for 60 min.

Objective C.2. Integrating the insights gained from "intrinsic" molecular mechanisms into the mesoscale heterogeneous kinetic framework of pretreated cellulosic

In this effort, we have used fluorescent probes ranging from FITC and TRITC-labeled Poly Ethylene Glycol (PEG) probes and Dextran probes to assess solute diffusion in pretreated cellulosic materials. We have found significant interaction between PEG probes and cellulose particles, resulting in PEG probes getting attached to the cellulose particles and could not be eluted in subsequent washing process. This is partly explained by the fact that PEG probes appear as string-like molecules in aqueous solutions and have a large surface to interact with adjacent particles. Therefore we chose FITC and TRITC labeled Dextran probes, because Dextran particles appear as solid spheres in aqueous solution, and their sizes closely related to their molecular weight. In addition, Dextran probes do not react chemically with cellulose particles. They have been used in other studies involving diffusion into porous media. The Dextran particles we chose as diffusive probes to investigate porosity of cellulose particles had molecular weight 20kDa, 40kDa and 70kDa. These sizes of particles were selected because they have hydrodynamic radii between 4 and 6 nanometers, close to those of *T. Fusca* cellulases. Whereas 155kDa Dextra-TRITC particles could diffuse into cellulose particles very slowly, and become trapped inside the particles, the time frame for 155kDa Dextra-TRITC diffusion is on the order of days, far slower than we had observed in cellulase diffusion. On the other hand, elusion measurements showed 40kDa FITC-labeled Dextran particles were easily washed out of the cellulose particles when they have been incubated with the particles for several days, indicating their fast diffusion constant within the cellulose particles.

We incubated 20kDa, 40kDa and 70kDa FITC-labeled Dextran with pretreated wood particles for several hours, and analyzed the fluorescence intensity within the cellulose particles over time (see Figure 16). When fluorescent probes are first added to the cellulose particles, the fluorescence in the cellulose-occupied region consists of the particles' autofluorescence alone. The fluorescence due to probes is assumed to be nil, due to exclusion effect of the particles. However, over time the fluorescent probes will diffuse into the cellulose particles via pores and crevices, making the fluorescence within particle's volume increase. By analyzing the increase of the fluorescence, we studied the diffusion behavior of the labeled probes.

From the plot above, we see the diffusion of the small labeled probes into the particles is rapid. In fact the particles are occupied with these probes within an order of tens of minutes. This is true for probes of all three molecular weights used. And there does not seem to be discernible difference in diffusion behavior among the three. We attribute this to the various pore sizes present on the wood particles, which allow both large and small probes to pass through. Due to the diverse morphology of the particles, an exact diffusion constant cannot be calculated. But we do get an order of magnitude estimation of the diffusion behavior, which agrees with what we had obtained before using numerical simulation. In fact, by comparing the measurement results above and our simulation results before, the diffusion constant of the labeled probes within cellulose is between 10^{-16} and 10^{-15} m^2/s.

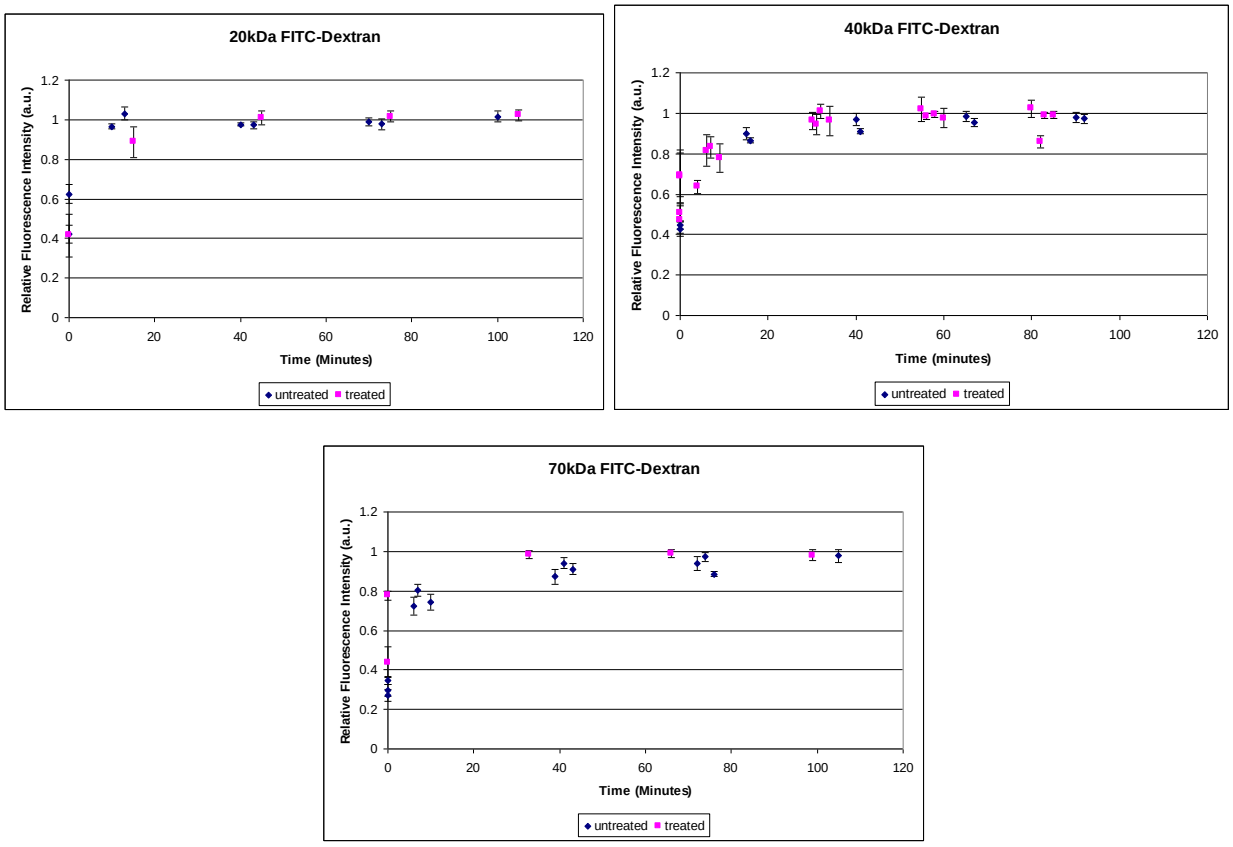

Figure 16. Fluorescence intensity increase inside the volume occupied by cellulose particles. Each data point is an average pixel intensity of five region of interests of a given measurement sequence, with error bar showing standard deviation. Two kinds of particles were used in this experiment. The blue diamonds denote untreated wood particles, whereas purple squares denote treated particles.

Diffusion of fluorescent probes into cellulose particles. During this project we have carried out a comprehensive characterization of the diffusive behavior of fluorescently labeled probes that do not interact with cellulosic materials. Our first goal was to determine the effective hydrodynamic radius of such labeled probes. To this end, we obtained fluorescently-labeled dextrans with molecular weights ranging from 4kDa to 2000kDa, and measured their effective sizes using a Zetasizer (Malvern Instruments), as shown on Figure 17. This apparatus utilizes dynamic light scattering measurements to determine the hydrodynamic size and zeta potential of molecules.

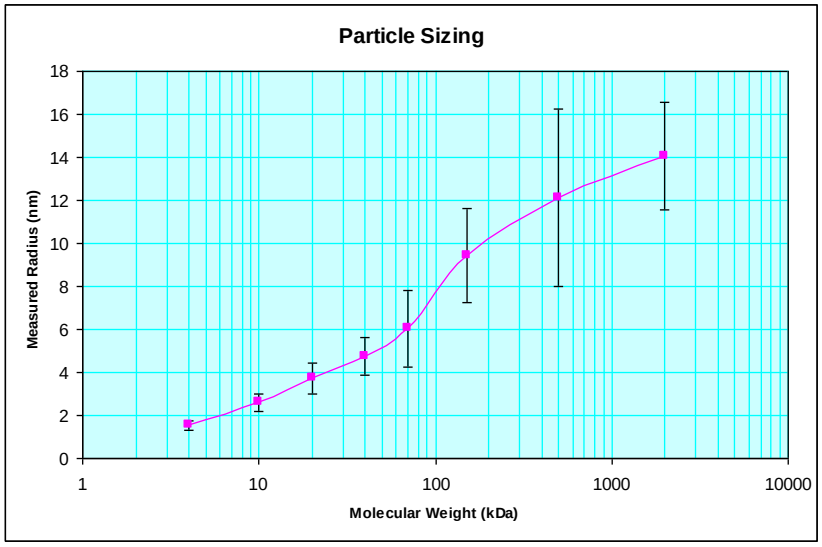

Figure 17. Size distribution of dextran molecules of different molecular sizes as measured through dynamic light scattering using the Zetasizer apparatus. Error bars represent the standard deviation of the measured radius. These variations could be due to the polydispersity of molecular sizes or due to non-specific molecular interactions.

From a theoretical calculation using Einstein's relation, we estimated that the sizes of cellulases are on the order of tens of nanometers. Therefore, to mimic the size of cellulases we focused on the diffusive behavior of fluorescent probes in the 10 kDa to 1000 kDa range. To test the diffusion of fluorescently labeled probes into the porous structures of pretreated wood particles we used a mixture consisting of 1 μM 40 kDa FITC-labeled and 1 μM 155 kDa TRITC-labeled Dextran molecules in 50 mM sodium acetate buffer, the same buffer used in enzyme binding experiments. After incubating the pretreated wood particles with the labeled Dextran solution for two days, we used CLSM to image the particles after gentle washing with pure buffer, to remove unbound fluorescent probes. From the images obtained, we noted there was a definite lack of fluorescence in the green channel from FITC-labeled 40kDa molecules, and clearly visible fluorescence in the red channel from TRITC-labeled 155 kDa Dextran molecules (Figure 18). The same particles were further incubated with a solution containing the same concentration of labeled probes, and were imaged again at seven and 13 days. Again little to no fluorescence was detected from the 40 kDa molecules and increasing fluorescence was observed from the 155 kDa molecules as the incubation progressed. The total increase of fluorescence was quantified and is shown in Figure 19.

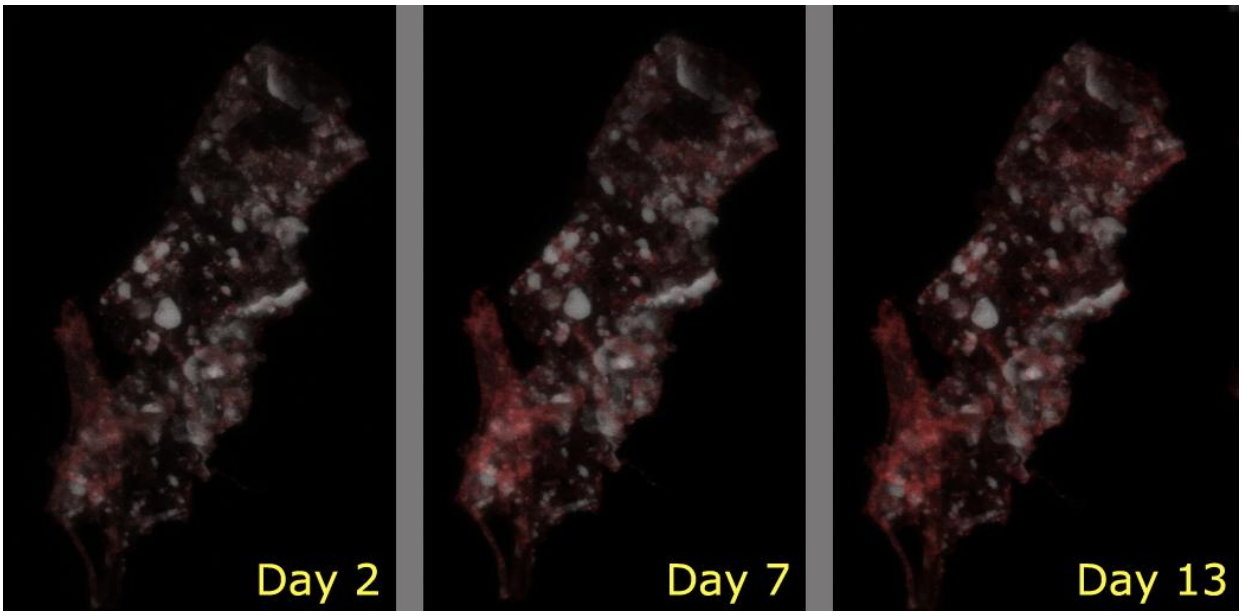

Figure 18. Sample images of one pretreated wood particle that has been incubated in solution containing 1μM 40 kDa Dextran-FITC and 155 kDa Dextran-TRITC for 2, 7, and 13 days. Autofluorescence of the pretreated particle is shown in gray, fluorescence of TRITC is shown in red, and fluorescence of FITC, which is too low to be observable, is shown in green.

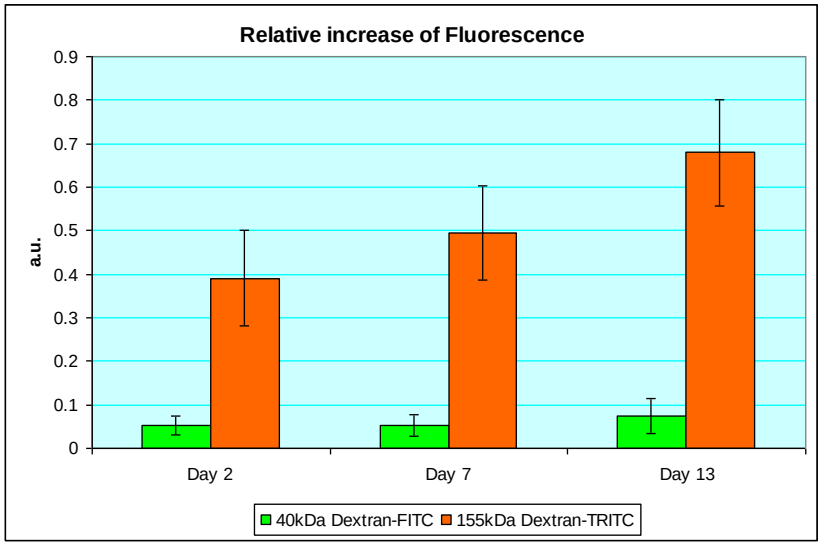

Figure 19. Fluorescence intensity extracted from image series of pretreated wood particles after they have been incubated in solutions containing 1 μM each of 40 kDa-FITC and 155 kDa-TRITC labeled Dextran molecules. We observe an increase in fluorescence derived from TRITC labeled molecules, but little increase in that of FITC, indicating possible pore sizes that are large for 40 kDa Dextran but small enough to trap 155 kDa Dextran.

Since 40 kDa FITC-labeled Dextrans were measured to be approximately 5 nm in radius, and 155 kDa TRITC-labeled Dextrans were measured to be approximately 10 nm in radius, we infer from our imaging results that there are pores within the wood particles with dimensions on the order of 10 nanometers which can trap molecules with radii on the order of 10 nm, but which allow the free diffusion of molecules with 5 nm radii. Thus, 155 kDa Dextran particles with 20 nm diameter will likely diffuse into the pores very slowly, and not get washed away when the solution is replaced with pure buffer. On the other hand, 40 kDa Dextran particles with sizes smaller than this pore size are unlikely to be trapped inside the pores, and can easily be washed away when the solution is replaced. While the same experiment must be conducted as a control with molecules with reversed fluorescent labels, this result points us to a possible porous medium model of the pretreated wood particles.

Cellulase diffusion and binding into cellulose particles. We studied cellulase binding to cellulose particles using time-resolved confocal scanning microscopy. Hydro-thermal pretreated wood particles were immobilized on borosilicate substrates before fluorescently-labeled cellulase solutions at various concentrations were added. Time-lapse CLSM revealed that cellulases Cel5A, Cel6B and Cel9A quickly bound to certain areas of wood particles, slowly diffused into and adsorbed to less accessible areas, but showed little affinity for other areas of the wood. Cellulase-to-substrate association constants were estimated using a transient enzyme binding kinetics model, and were found to be in agreement with published values. In order to accurately account for the fluorescence signal of labeled enzyme mixed with wood autofluorescence, we also developed a spectral deconvolution method to separate signals from multiple fluorochromes. However, order of magnitude estimation can be made and agrees with previously published results. For all three enzymes, the forward binding rates are on the order of $10\mu M^{-1}\cdot min^{-1}$. [7]

Comparing forward binding rate of cellulases with the diffusion constant of labeled probes, we see that binding reaction is much faster than diffusion. Therefore the rate limiting step is diffusion of cellulases into particles' pore structures. The combined effect of fast adsorption with relatively slower diffusion, is an even smaller apparent diffusion constant, with a reduction factor of (1+R), where R is the ratio of bound enzyme to free enzyme.

Comparing the finite element simulation results above with our measurement results obtained before, we see the diffusion rate agrees with the measured results using labeled Dextran probes while the reaction rate was assumed to agree with measured values, i.e., in the order of $10uM^{-1}min^{-1}$ (See Figure 20). Again due to the diverse morphology of the pretreated particles, an exact numerical value cannot be obtained without large variation, using the CLSM method. A bulk measurement would be more appropriate to average out the large variations among pretreated particles.

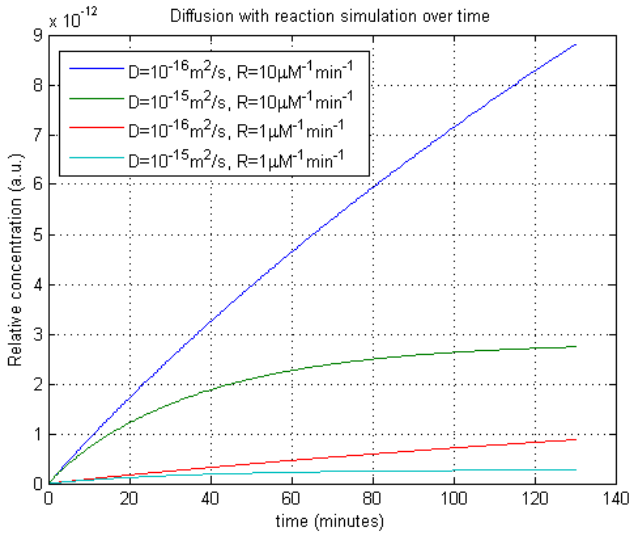

Figure 20. Simulated enzyme concentration inside a block of porous cellulose particle. Diffusion constant was varied between $10^{-15}m^2/s$ and $10^{-16}m^2/s$; Reaction rate was varied between $1uM^{-1}min^{-1}$ and $10uM^{-1}min^{-1}$.

Development of integrated kinetic models. In literature research we came upon work done by F. J. W. Roughton in the modeling and analysis of diffusion and chemical reaction of oxygen and carbon monoxide in thin layers of Haemoglobin. The underlying physics of O_2 or CO diffusing and binding to haemoglobin is similar to that of cellulases diffusing into and binding onto cellulose fibrils with finite porosity and volume. Specifically, both reactions involve an infinite volume of solution of reactive solute, held at constant concentration, surrounding a thin slab of substrate, to which the solute can diffuse in and adsorb. Combining:

Fick's second law of diffusion:

$$\frac{du}{dt} = D(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial x^2}) = D\nabla^2 u$$

,

With enzyme binding reaction:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \quad , \quad \frac{du_b}{dt} = k_1 u_f (S^M - u_b) - k_{-1} u_b$$

Where E is enzyme, S is substrate, S^M is maximum available number of binding sites, k_1 is forward reaction rate, k_{-1} backward reaction rate, u_f is concentration of free enzyme, u_b is concentration of bound enzyme. The combined diffusion and binding model is then:

$$D\nabla^2 u_f = \frac{\partial u_f}{\partial t} + k_1 u_f (S^M - u_b) - k_{-1} u_b$$
$$= \frac{\partial u_f}{\partial t} + k_1 S^M u_f - (k_1 u_f + k_{-1}) u_b$$

To solve the above equation, we make several assumptions. First of all we assume the cellulose particle is a thin slab, with size in one dimension much smaller than the other two dimensions. This effectively changes the problem into a one-dimensional differential equation. From scanning electron images, we note the thickness of most pretreated particles is only micrometers, whereas the width and length of most particles are in the tens to hundreds of micrometers range. Hence our one-dimensional assumption is valid. Secondly, we assume ub is small when t is small, meaning bound enzyme is negligible at the start of the experiment. Thus we have a simplified 1-dimensional binding and diffusion equation:

$$\begin{cases} D\dfrac{\partial^2 u_f}{\partial x^2} = \dfrac{\partial u_f}{\partial t} + \kappa u_f \\ \dfrac{du_b}{dt} = \kappa u_f \end{cases}$$

Where κ is the effective binding constant: $\kappa = k_1 S^M$.

The solution to the above equation is:

$$u_f = u_0 \left[\frac{\cosh(x\sqrt{\frac{\kappa}{D}})}{\cosh(b\sqrt{\frac{\kappa}{D}})} - \frac{2}{b} \sum_{n=0}^{\infty} \frac{(-1)^n \beta_n}{\beta_n^2 + \frac{\kappa}{D}} \cos(\beta_n x) e^{-\beta_n^2 Dt - \kappa t} \right] \text{ where } \beta_n \triangleq \frac{2n+1}{2b}\pi. \text{ And } u_b = \kappa u_f t.$$

The important parameters to extract from the solution are: effective diffusion constant D which describes the porosity of the material and effectiveness of pretreatment, and $\kappa = k_1 S^M$ which describes maximum available binding sites and binding rate.

Integrating u_b and u_f throughout the volume of $-b<x<b$, we should have the model that approximates measured fluorescent data reflecting the total amount of labeled enzymes within the cellulose particle. Fitting this model to experimental data could present challenges, however, because each model is an infinite sum of terms. Therefore, further simplification will be needed in order to utilize numerical fitting routines.

Objective C.3. Evaluation of synergistic effect of cellulase mixtures on hydrolysis of cellulose particles.

Modeling of binary binding of cellulases. During this project we have focused on developing a model for competitive binding of cellulases incubated in mixtures. We have analytically and numerically solved the system of coupled differential equations that describe the simultaneous binding of a binary mixture of cellulases. In a previous report we had shown that there is a temporal evolution to the fractional portion of bound cellulases of each component of the binary mixture even when the mixture is equimolar. Our simulations of the competitive binding indicate

that the temporal evolution depends strongly on the kinetic constants. Figure 21 shows a sample curve which mimics some of the observed temporal evolution of the binding pattern of an equimolar binary cellulase mixture. We propose that fitting the intensity traces obtained from binding of binary mixtures of fluorescently labeled cellulases can yield another approach for estimating the kinetic binding parameters of the cellulases. Furthermore, imaging can provide the temporal evolution due to more complex mixtures and evidence other effects such as synergistic binding with high spatial and temporal resolution.

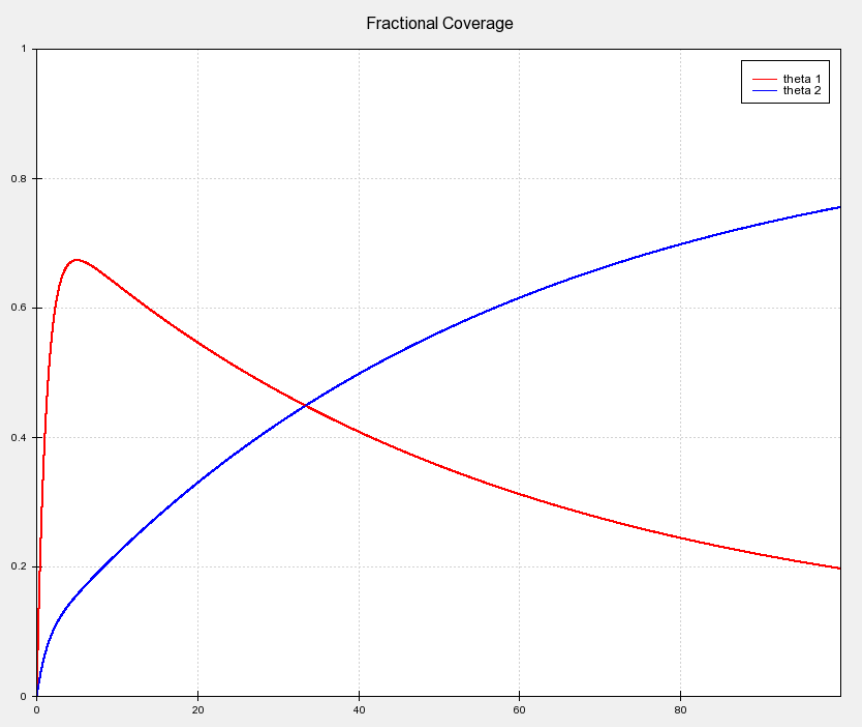

Figure 21. Numerical solution to the coupled differential equations describing simultaneous binding of a binary mixture of cellulases to a cellulose substrate with a maximum number of available binding sites. Conditions for the simulation were chosen to mirror some of the observed behavior from experimental runs.

III. VARIANCE FROM PROJECT OBJECTIVES

For a project with so many complex method development activities and deployment to address the recalcitrance of cellulose degradation there were not as many variances from the objectives one might expect. Yes, there were specific adjustments that we had to make as we moved forward and sometimes these adjustments limited how fast we could deploy methods to address scientific problems. For example, in the second quarter of this program the kinetic studies (Objective B.1. and B.2.) of labeled cellulases on cellulose via TIRFM were slowed down due to a desire to have suitable labeled molecules that could yield meaningful data for single molecule experiments. Nevertheless, we were able to aggressively move this component of the program forward with a better set of data and to completion. Another example is the initial delay that was encountered with the development of an FTIR assay to assess the interaction between cellulases and cellulose particles (Objective C.2.). This was a new approach to study cellulose degradation, and thus requires much more optimization and development to obtain a robust and reliable measurement system for bond cleavage to enzyme binding. This has inevitably delayed the achievement of the goals proposed for Objective C.2.

Cloning some of the new cellulases has been challenging (Objective A.4). For example we did not anticipate that the fusion construction would have a significant effect on enzyme activity. However, we are working to redesign the gene construct to remove tags during purification. We do not anticipate that this redesign will prevent us from proceeding with evaluations of the enzymes in the future. Also, the purification of one enzyme of particular interest, the GH61 construct, proved challenging and has limited our ability to introduce this enzyme into our molecular mechanisms studies. Modification of purification using metals and salts proved to be insufficient to prevent degradation and provide sufficient quantities for analysis for the GH61 protein. Expression with IPTG exhibited mixed results. Eventually experiments were conducted to increase production of the cloned GH61 fusion protein in *E. coli* for enzymatic analysis. We introduced the use of a new auto-induction medium that minimizes optimization and unstable expression, and this seems to help address this problem. Also switching to using magnetic beads to minimize exposure times to TEV protease has been beneficial in the efficient recovery of products.

Progress has been slow in obtaining the *T. fusca* clones. *Cel6B mutants S232A and D226A, and Cel6A catalytic domain mutants W16I, and G87A were produced and their activities on CMC, SC, and BMCC tested*. However, only lately have these clones been moved into the labeling process. Part of this delay is due to the student working on this cloning task finished his degree before he could label the enzymes**.** Prof. Wilson has just hired a new technician who has been successful in producing the labeled enzymes, and we are now using TIRFM to study these mutants.

IV. CURRENT STATE OF THE RESEARCH

Although funding of the project has come to an end, we are continuing to wrap-up some major experimental studies that are important to bringing this project to closure and to lay the foundation for new research exploration focused on addressing the recalcitrant nature of cellulose to enzymatic saccharification. As noted earlier the inventory of CWDEs that we have identified is quite extensive and we continue to clone the genes of individual enzymes for future use in expression systems. As indicated in Objective A.3, each purified enzyme and protein of interest will be subjected to tryptic digestion to isolate peptide fragments that will be sequenced by mass spectrometry. Sequences will be compared to those in existing databases (e.g. CAZy, NCBI, Swiss-Prot) to determine if homologues of full genes from other species are available. If we identify unique sequences not reported in the databases, primers will be designed to code for the peptide fragment(s) from our samples and used to PCR-amplify products from the organism showing the original activity. Our goal is to eventually introduce more of these enzymes into our imaging and high-throughput analytical systems to assess their behavior on model substrates such as BMCC or pretreated wood and grasses.

We are very excited about the imaging and high-throughput analytical methods that we have developed and the insights that we are gaining. We have just submitted a paper on determination of the molecular states of the processive endocellulase *T. fusca* Cel9A during crystalline cellulose depolymerization that builds on methods that we developed under this contract. In addition, we are completing the following studies that we expect to submit for publication this summer:

- Investigation of cellulase synergistic binding and activities on simple and complex cellulose morphological structure using epi-fluorescent microscopy. This study is being performed at room temperature and 45°C.
- An assessment of chemical and structural transformations of pretreated mixed hardwood and switch grass using *T. fusca* Cel5A, Cel6B, CEL9A. This assessment is being accomplished using QHT-FTIR spectroscopy.
- Single molecule kinetics of *T. fusca* cellulases studied through TIRFM at room temperature – reduced activity – and 45°C with high activity.
- We are completing a study using fluorescently labeled cellulase and commercial enzyme cocktails to observe and model cellulase binding and biomass depolymerization. BMCC stained with a fluorescent dye, filter paper, and pretreated mixed hardwood and switch grass are observed in situ during enzymatic depolymerization.

Our goal remains to exploit the spatial and temporal resolutions that we can achieve with the methods developed under this contract to assess and model the heterogeneous catalysis of insoluble substrates by familiar and new CWDEs and enzyme cocktails. We thank the U. S. Department of Energy for providing this opportunity to contribute to this import research and development program for biofuels.

Publications Generated from this Contract:

1. King, B. C., Donnelly, M. K, Bergstrom, G. C., Walker, L. P., and Gibson, D. M. (2009) An optimized microplate assay system for quantitative evaluation of plant cell wall-degrading enzyme activity of fungal culture extracts. *Biotechnology and Bioengineering*. 102: 1033-1044.

2. King, B.C., Waxman, K.D., Nenni, Ni V., Walker, L.P., Bergstrom, G.C., Gibson. D. M. 2011. Arsenal of plant cell wall degrading enzymes reflects host preference among phytopathogenic and saprophytic fungi. *Biotechnology for Biofuels*, **4**:4.

3. Moran-Mirabal, J. M., Santhanam, N., Corgie, S.C., Craighead, H. G., and Walker, L. P. (2008) Immobilization of cellulose fibrils on solid substrates for cellulase binding studies through quantitative fluorescence microscopy. *Biotechnology and Bioengineering*. 101: 1129-1141.

4. Moran-Mirabal J.M., Corgie, S.C., Bolewski, J.C., Smith, H.M., Cipriany, B.R., Craighead, H.G., Walker, L.P. (2009) Labeling and purification of cellulose-binding proteins for high resolution fluorescence applications. *Analytical Chemistry.* 81: 7981-7987.

5. Moran-Mirabal, J.M., Bolewski, J.C., Walker, L.P. 2011. Reversibility and binding kinetics of *Thermobifida fusca* cellulases studied through fluorescence recovery after photobleaching microscopy. *Biophysical Chemistry*, **155:** 20 - 28..

6. Corgie, S.C., Smith, H., Walker, L.P. 2011 Enzymatic transformations of cellulose assessed by Quantitative High-Throughput Fourier Transform Infrared (QHT-FTIR) spectroscopy. Biotechnology and Bioengineering, **108:** 1509 - 1520.

7. Zhu, P. Moran-Mirabal, J.M., Luterbacher, J.S., Walker, L.P., Craighead, H.G. 2011. Observing *Thermobifida fusca* cellulase binding to pretreated wood particles using time-lapse confocal laser scanning microscopy. *Cellulose*, **18**:749–758.

8. Maxim, K., Moran-Mirabal, J.M., Walker, L.P., Wilson, D.B. 2011. Determination of the molecular states of the processive endocellulase *Thermobifida fusca* Cel9A during crystalline cellulose depolymerization. *Biotechnology and Bioengineering*, ***In Press***