

Section 1: Cover Page

DOE award number:
DE-PS02-09ER09-09

Name of Recipient:
Chesapeake-Protein Expression and Recovery Lab (C-PERL)

Project Title:
Identification, Production and Characterization of Novel Lignase Proteins from Termites for Depolymerization of Lignocellulose.

Project Director/Principle Investigator:
Jeffrey Slack

Authorized Recipient Official:
John Lee Compton

Consortium/Team Members:
Jeffrey Slack
George Buchman
Elena Kovaleva
Michael Scharf
Amit Sethi

Section 2: Cover Page Limitations Notice

Any references to commercial virus platforms used at C-PERL for use in making recombinant baculoviruses are proprietary.

Purdue University is pursuing patents on aldo-keto reductase (AKR), GHF11-1, beta-glucosidase, laccase, catalase, superoxide dismutase and glutathione peroxidase.

Section 3: Executive Summary

Wood is a potential source for biofuels such as ethanol if it can be digested into sugars and fermented by yeast. Biomass derived from wood is a challenging substrate for ethanol production since it is made of lignin and cellulose which cannot be broken down easily into fermentable sugars. Some insects, and termites in particular, are specialized at using enzymes in their guts to digest wood into sugars. If termite gut enzymes could be made abundantly by a recombinant protein expression vector system, they could be applied to an industrial process to make biofuels from wood. In this study, a large cDNA library of relevant termite genes was made using termites fed a normal diet, or a diet with added lignin. A subtracted library yielded genes that were overexpressed in the presence of lignin.

Termite gut enzyme genes were identified and cloned into recombinant insect viruses called baculoviruses. Using our PERLXpress system for protein expression, these termite gene recombinant baculoviruses were prepared and used to infect insect larvae, which then expressed abundant recombinant termite enzymes. Many of these expressed enzymes were prepared to very high purity, and the activities were studied in conjunction with collaborators at Purdue University. Recombinant termite enzymes expressed in caterpillars were shown to be able to release sugars from wood. Mixing different combinations of these enzymes increased the amount of sugars released from a model woody biomass substrate.

The most economical, fastest and energy conserving way to prepare termite enzymes expressed by recombinant baculoviruses in caterpillars was by making crude liquid homogenates. Making enzymes stable in homogenates therefore was a priority. During the course of these studies, improvements were made to the recombinant baculovirus expression platform so that caterpillar-derived homogenates containing expressed termite enzymes would be more stable. These improvements in the baculoviruses included significantly reducing proteases and preventing blackening immune reactions that occur when caterpillars are homogenized. Proteases may degrade enzymes and immune reaction blackening may inactivate enzymes thus compromising the ability of these crude recombinant expressed termite enzyme preparations to release sugars.

Commercial preparations of fungal enzymes currently are used to digest wood for ethanol production. We demonstrated in this study that termite enzymes could improve the efficiency of fungal enzyme cocktails. Although the economic feasibility of using caterpillar expressed termite enzymes alone to treat wood was not proven, this work points to the potential to combine C-PERLXpressed insect enzymes with industrial enzyme cocktails to boost their efficiency at treating wood for biofuels.

Section 4: Comparison of Objectives with Accomplishments

Objective 1: Continue characterization of currently expressed and purified enzymes (singly and in combination) and develop new functional biochemistry methodologies for Phase II.

a) C-PERL will work with enzymes previously discovered and expressed in Phase 1 of this SBIR grant and develop methods to optimize and scale purification.

1. PREVIOUSLY DISCOVERED β -GLUCOSIDASE, CELL-1 AND LACCASE: These enzymes were not further optimized for expression or scale purification as they were already easily purified and expressed at high levels.

2. PREVIOUSLY DISCOVERED CELL-3: The GH7 endosymbiont cellulase Cell-3 was poorly expressed and was not easily purified during Phase I of the grant. Significant efforts were made in Phase II to purify Cell-3 with only limited success. Crude homogenates from larvae expressing Cell-3 did have cellulase activity and were included in crude homogenate type assays. Five other endosymbiont GH7 type cellulases were identified and expressed in PERLXpress. Work was done to optimize their expression and purification.

b) Develop assay technology (e.g., catalases, esterases, feedstock). Purdue to develop testing methods for easily characterizing the enzymatic activity on both model and feedstock substrates. The feedstock assay will be used to analyze future enzymes and also the eventual cocktail of enzymes envisioned as the final commercialization product.

1. GLYCOSYL HYDROLASE FAMILY 7 (GHF7): Specific enzymatic activity of three recombinant GHF7 enzymes (GHF7-3, GHF7-5, GHF7-6) was compared in endo- and exoglucanase activity assays with nine model substrates in HEPES buffer (100mM, pH 7.0) at 30°C for 60 min. These nine model substrates included carboxymethyl cellulose (2%; w/v), microcrystalline cellulose (2%), 4-nitrophenyl β -D-cellobiside (6mM), 4-nitrophenyl β -D-glucopyranoside (6mM), 4-nitrophenyl β -D-fucopyranoside (6mM), 4-nitrophenyl β -D-galactopyranoside (6mM), 4-nitrophenyl β -D-xylopyranoside (6mM), 4-nitrophenyl β -D-mannopyranoside (6mM), and 4-nitrophenyl β -D-arabinopyranoside (6mM). Absorbance of the mixture was read at 420 nm over a period of 90 min in the case of all the model substrates except for carboxymethyl cellulose and microcrystalline cellulose, where absorbance was read as an endpoint at 540 nm relative to a glucose standard curve. Standard curves were prepared as seven serial dilutions in assay buffer with 4 mM EDTA from 5 mM downward to 0.078125 mM. The eighth well that contained only buffer served as the blank control.

To investigate potential effects of calcium as a cofactor, duplicate reactions were performed with the model substrates in the presence and absence of 10mM calcium chloride. All measurements were replicated three times. The linear portion of the curve was used to estimate the reaction rate.

2. CATALASE (CAT): Catalases are known to possess three different kinds of activities: a) catalytic activity (conversion of hydrogen peroxide into water and oxygen); b) peroxidative activity (oxidation of low molecular weight alcohols into aldehydes in the presences of low concentrations of hydrogen peroxide); and b) polyphenol oxidase activity (oxidation of phenolic compounds into quinones in the absence of hydrogen peroxide).

Catalytic activity of the recombinant catalase was tested using 10mM hydrogen peroxide as a substrate in 50 mM phosphate buffer (pH 7.0) with 0.2 % BSA. Absorbance of the mixture was read at 246 nm over a

period of 5 min. The linear portion of the curve was used to estimate the reaction rate. All measurements were replicated three times.

Peroxidative activity of the recombinant catalase was tested using 5.9 M methanol and 4.2 mM hydrogen peroxide in 100 mM phosphate buffer (pH 7.0). Formaldehyde (75 – 0 μ m) was used to prepare the standard curve. Endpoint absorbance of the mixture was read at 540 nm. All measurements were replicated three times.

Polyphenol oxidase activity of the recombinant catalase was tested using 100mM catechol in 100mM phosphate buffer (pH 7.0). Absorbance of the mixture was read at 420 nm over a period of 30 min. The linear portion of the curve was used to estimate the reaction rate. All measurements were replicated three times.

3. ALDO-KETO REDUCTASE (AKR): Specific enzyme activity of aldo-keto reductase was tested against seven model substrates in 100mM phosphate buffer (ph 7.0). These seven model substrates included p-nitrobenzaldehyde (2mM); DL-glyceraldehyde (2mM); 9,10 phenanthrene (2mM); trans-2-hexen-1-al (10mM); D-glucuronic acid (10mM); D-xylose (10mM); and D-glucose (10mM). NADPH (0.2mM) was used as a co-factor. Absorbance of the mixture was read at 340 nm over a period of 10 min. The linear portion of the curve was used to estimate the reaction rate. All measurements were replicated three times.

4. GLYCOSYL HYDROLASE FAMILY 11 (GHF 11): Specific enzymatic activity of recombinant GHF11-1 enzyme was tested against eight model substrates in HEPES buffer (100mM, pH 7.0) at 30°C for 60 min. These eight model substrates included beechwood xylan (2%; w/v), arabinoxylan (2%), 4-nitrophenyl β -D-xylopyranoside (6mM), 4-nitrophenyl β -D-arabinopyranoside (6mM), 4-nitrophenyl β -D-glucopyranoside (6mM), 4-nitrophenyl β -D-fucopyranoside (6mM), 4-nitrophenyl β -D-galactopyranoside (6mM), and 4-nitrophenyl β -D-mannopyranoside (6mM). Absorbance of the mixture was read at 420 nm over a period of 90 min in case of all the model substrates except for beechwood xylan and arabinoxylan, where absorbance was read as an endpoint at 540 nm relative to a glucose standard curve. Standard curves were prepared as seven serial dilutions in assay buffer with 4 mM EDTA from 5 mM downward to 0.078125 mM. The 8th well that contained only buffer served as the blank control.

To investigate potential effects of calcium as a cofactor, duplicate reactions were performed with the model substrates in the presence and absence of 10 mM calcium chloride. All measurements were replicated three times. The linear portion of the curve was used to estimate the reaction rate.

Objective 2: Identify potential additional candidate cDNA's (7 to 9 enzymes) from expression libraries for additional development.

a) Continued cDNA library screening and bioinformatic analysis.

Continued cDNA library bioinformatic screening of the original cDNA library sequence led to the identification of cellulases genes representing GH5, GH7, GH10, and GH11 families. Similarly, candidate lignase genes catalase (CAT) and aldoketoreductase (AKR) were identified.

The original cDNA library did not differentiate for genes expressed in the presence of lignin. New termite gut expression cDNA libraries were made from termites feeding on lignin and non-lignin containing diets. These libraries were normalized, and a subtractive library was prepared to enrich for genes expressing activities upregulated in the presence of dietary lignin. Roche 454-titanium pyrosequencing and analysis was done for the two termite cDNA libraries and a total of 9,552 contiguous transcripts (contigs) were assembled.

This data identified 68 candidate lignase gene transcripts including 19 alcohol dehydrogenases, 17 aldo-keto reductases, 15 esterases, 10 glutathione esterases, 7 superoxide dismutases, 3 catalases, and 2 laccases. Of the candidate lignases, 17 were upregulated in the presence of lignin. In addition to lignase candidates, the 454-array data contained 166 contigs for carbohydrate active enzymes including 22 glycosylhydrolases families; GH7, GH45, GH1, GH9, GH2, GH11, GH3, GH5, GH26, GH43, GH28, GH16, GH27, GH53, GH38, GH30, GH10, GH8, GH18, GH30, GH13, GH17. None of the newer families of glycosyl hydrolases (GHs) were complete ORFs and could not be synthetically engineered into recombinant baculoviruses. Future efforts could involve RACE or other methods to identify full-length ORF's.

b) cDNA sequence refinement to identify ORF candidates and reconstruct full ORF.

The original termite cDNA library was used to identify ORFs and synthetic genes were made including GH5-1, GH5-2, GH7-2, GH7-3, GH7-4, GH7-5, GH7-6, GH10, GH11, CAT-1, CAT-2, CAT-3, AKR-1, AKR-2, and AKR-3.

The 454-array based cDNA library provided more gene sequence data in contigs but most contigs did not contain complete ORFs. In the original termite cDNA library, there were bacterial plasmid library clones of cDNAs that could be amplified and sequenced to obtain full length ORFs. Such a plasmid library did not exist for the 454-array based cDNA library. Obtaining full ORFs would have required going back to termites, purifying mRNA and doing 5 prime and 3 prime extension reactions. These types of activities were originally planned but by the time the 454-array data were completely analyzed it was not practical time wise to begin such projects.

However, based on the 454-array data it was possible to identify complete ORFs belonging to smaller genes. Synthetic genes were made of the lignase candidates, superoxide dismutase (SOD) and glutathione peroxidase (GPx). Both SOD and GPx were found to enhance efficiency of cellulases on lignocelluloses substrates.

c) Construct baculoviruses using candidate cDNA. C-PERL will construct a battery of baculoviruses looking for the optimal expression strategy. These various viruses to include different vectors, promoters and/or chaperons.

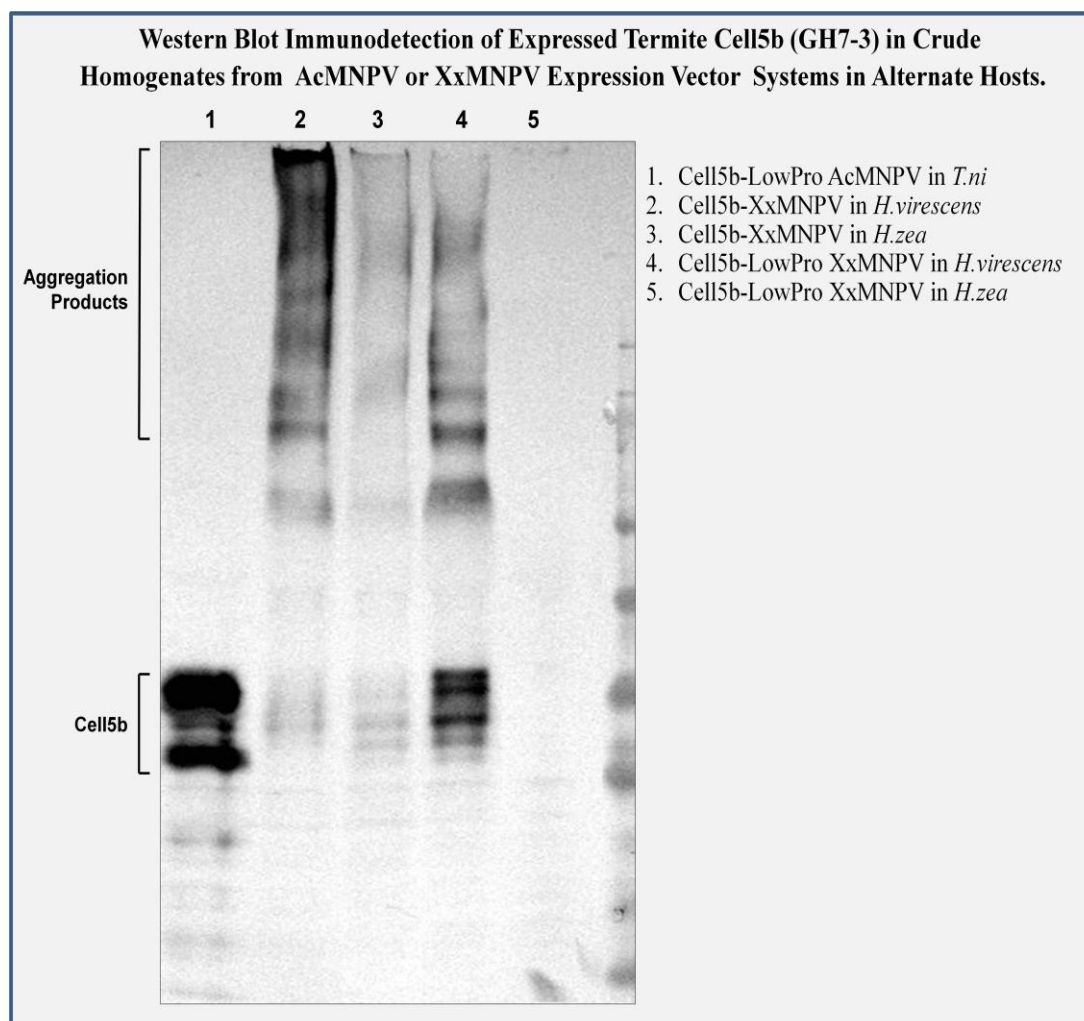
A battery of 67 virus variants expressing termites genes was made including different virus vectors, virus promoters and virus chaperonins.

1. COMMERCIAL VIRUS VECTORS/CHAPERONINS: Virus variants included the use of four commercial baculovirus expression vectors. Most success was achieved using a commercial virus vector that made chaperonin proteins. Improved solubility of secreted termite proteins was the major advantage of the chaperonin expressing virus vector. Names of commercial virus vectors used at C-PERL are withheld for proprietary reasons.

2. OTHER C-PERL VIRUS VECTORS AND HOSTS: Work was done to develop an alternate insect and virus platform to express termite enzymes. Although progress was made reducing proteases in crude homogenates using LowPro baculovirus (See Objective 3), there was an undesirable side effect of *Trichoplusia ni* larvae being less easily processed into homogenates when infected with the LowPro baculovirus based baculovirus expression vectors. C-PERL rears a number of insect species including *Helicoverpa zea* and *Heliothis virescens*. These other two species have physiological differences from *Trichoplusia ni* that would make them more easily processed into homogenates after being infected with a LowPro baculovirus. The LowPro baculovirus is based in AcMNPV which does not infect *H.zea* and *H.virescens* by oral route.

An alternate baculovirus expression platform was developed based on a baculovirus that infects *H.zea* and *H.virescens* by oral route. We refer to this virus as XxMNPV. XxMNPV replicates in standard insect cell lines (Sf9, High five). Further, an XxMNPV-based system is less encumbered from an Intellectual Property standpoint, making it more attractive for commercialization. A LowPro modified version of XxMNPV was also available to be made into this new cloning system.

Success was achieved expressing a termite endosymbiont GH7 cellulase using the alternate virus and larval platforms. However the current AcMNPV and *T.ni* platform produced superior amounts of GH7 cellulase products. An example of cellulase expression in the XxMNPV platform is in the following figure.



d) Test viruses in larvae to determine expression rate. C-PERL to evaluate the performance of various viruses in larvae testing expression by Western blot, Coomassie gels and activity.

More than 90 different virus constructs relating to this DOE termite project were expressed in larvae and analyzed by Western blot/Coomassie gels. Preliminary activity assays included carboxymethylcellulose assays and xylanase assays. The enzymes examined expressed over a broad range, with the highest exceeding 10mg/ g tissue (e.g., AKR and GPx)

e) Move the most promising viruses to scaled up expression using C-PERL proprietary production method of oral infection of larval biomass with POV and factory automation.

A total of 36 bulk scaled up expressions of termite-related enzymes were done over the course of this project. Bulk scaled up expression of termite enzymes included hemicellulase (β -glucosidase), cellulases (GHF 45, GHF 9, GHF 7), xylanases (GHF 11, GHF 10, GHF 5), laccases (Lac 6, Lac 12), catalase, aldo-keto reductase, glutathione peroxidase and superoxide dismutase. Most of the larvae were processed to purified termite enzymes for Purdue while some were sent as crude homogenates.

f) Confirm Expression results and provide crude samples to UF/Purdue to test for initial biochemical characterization.

Throughout this collaboration, crude homogenate samples have been sent along side purified samples to Purdue University for activity characterization. Specific examples of crude homogenates being sent to Purdue include Jan 2011 (GH7-2, GH7-3, GH7-3-codon optimized, GH7-4, GH7-5), April 2012 (SOD & PGx1) and Sept 2012 (Lac12, CAT, GH1, GH5, GH7, GH10, GH11, GH45).

g) Purify small samples of enzymes (4 to 6) for use as standards and to further utilize in substrate assays.

Purified enzymes were sent to Purdue continuously throughout this project. Examples of purified enzymes include laccase, AKR, CAT, GH9 (Cell-1), GH1 (β -glucosidase), GH11, GH5, GH7-3, GH7-4, GH7-5, and GH7-6. Other termite gut derived gene products expressed in larvae were not possible to purify to a state that would be suitable for substrate assays. Examples include GH7 (Cell-3), GH10, GH5-1 and GH5-2.

Objective 3: Select and optimize expression of most promising candidates from Year 1 efforts to move toward commercialization.

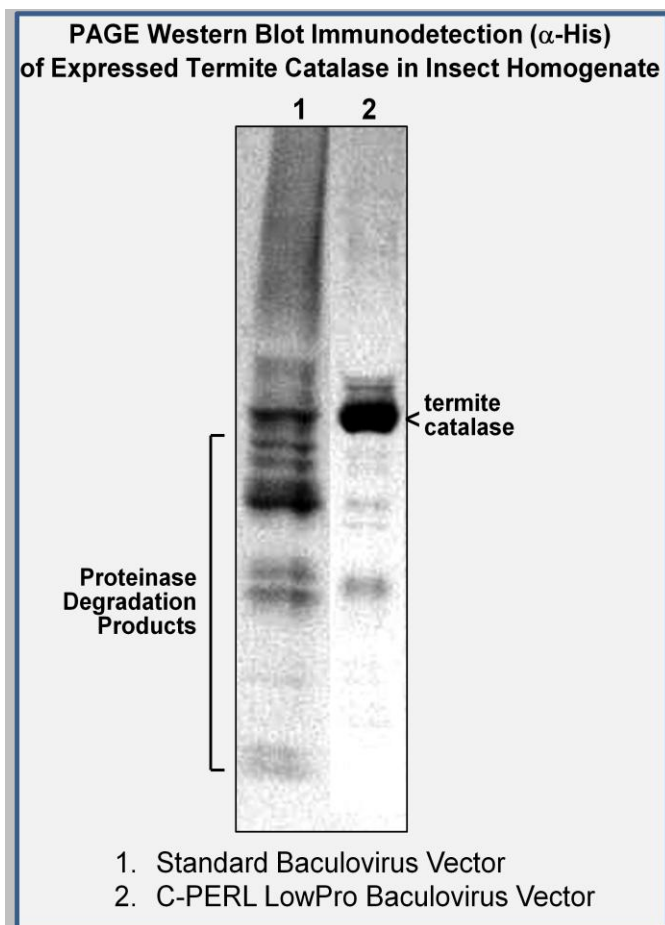
a) Optimize expression and yield. Construct additional baculoviruses using candidate cDNA (2 to 3) discovered in Year 1. C-PERL will construct a battery of baculoviruses looking for optimal expression strategy. These various viruses to include codon optimization, different vectors, promoters and /or helper proteins. Viruses are also to be constructed with and without His tag in order to test His tag effect on activity while facilitating purification.

Virus vectors, promoters and chaperonin proteins were discussed in Objective 2, part c). In the present section, work is described that was done to optimize the expression and stability of termite proteins in larval homogenates. It was recognized from the onset of this study that the most economical way to produce termite enzymes is in a crude unpurified homogenate form. Two major obstacles to using crude homogenates are protease degradation and melanization.

1. **MELANIZATION INHIBITION:** A C-PERL virus was made that expresses an anti-melanization peptide. Melanization is an immune reaction in insect larvae that causes cross linking of proteins and occurs in homogenates that have been prepared of insect larvae. At C-PERL, melanization is prevented by adding a reducing agent such as β -mercaptoethanol. Unfortunately if homogenates are frozen and thawed several times or left above freezing, the β -mercaptoethanol loses effect and the melanization begins to occur. This is detrimental to termite enzymes in homogenates as they are inactivated by cross-linking. Co-expression of anti-melanization peptide in the same virus with termite GH7-3 resulted in homogenates that did not melanize even after several days at room temperature.

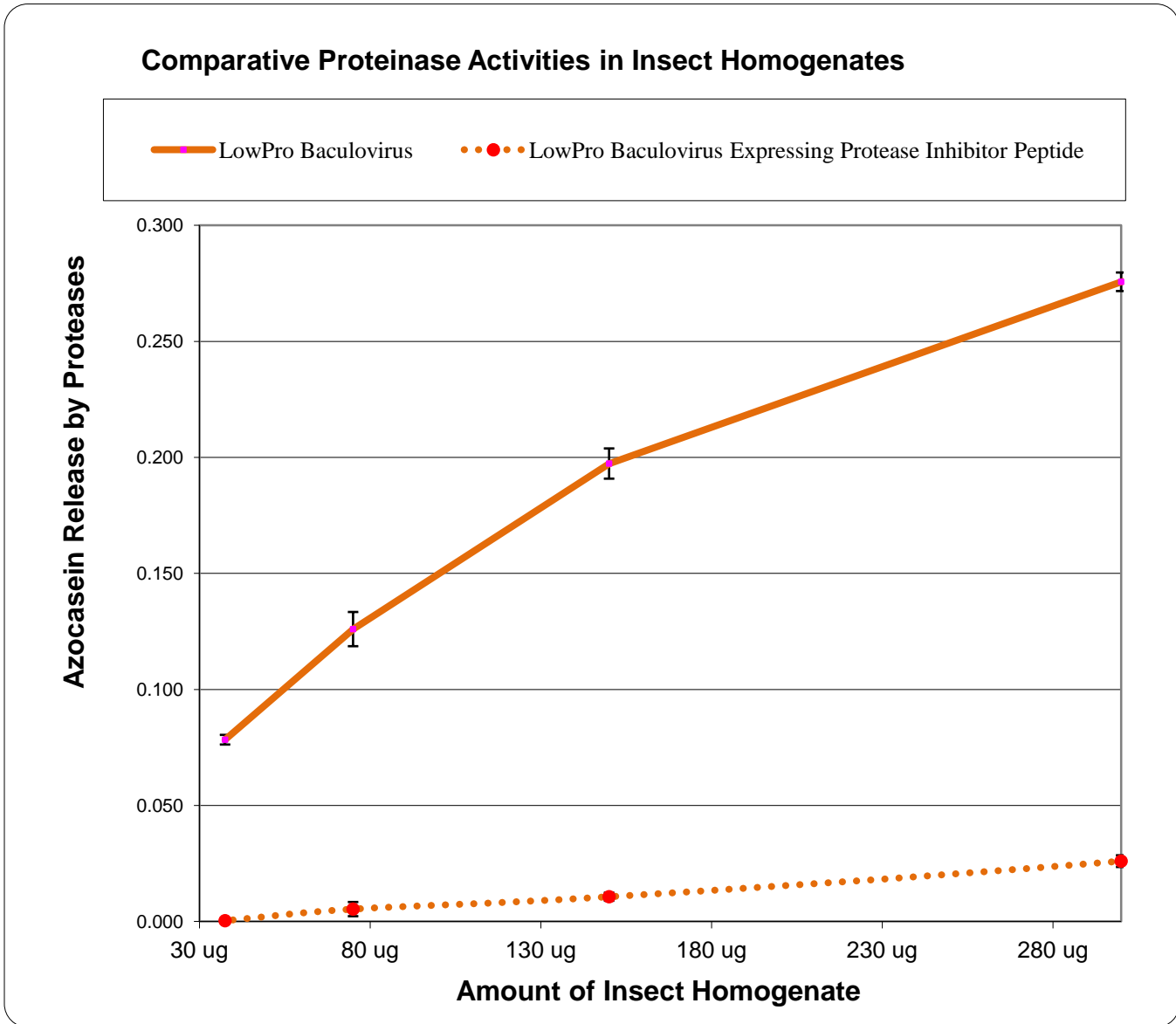
Note: Specific anti-melanization peptide expressed is a trade secret.

2. **OPTIMIZATION EXPRESSION & YIELD/ C-PERL LOWPRO BACULOVIRUS VECTOR:** C-PERL: Termite enzymes were expressed using a low protease (LowPro) baculovirus expression vector which lowered protease levels in larval homogenates. This is a critical development given that homogenates are the most economical form to produce termite enzymes in the C-PERLXpress. Normal virus homogenates contain proteases that degrade expressed cellulases and lignases. The result of this work was an improvement in the stability of expressed termite enzymes in crude homogenates. This is illustrated in the following figure.

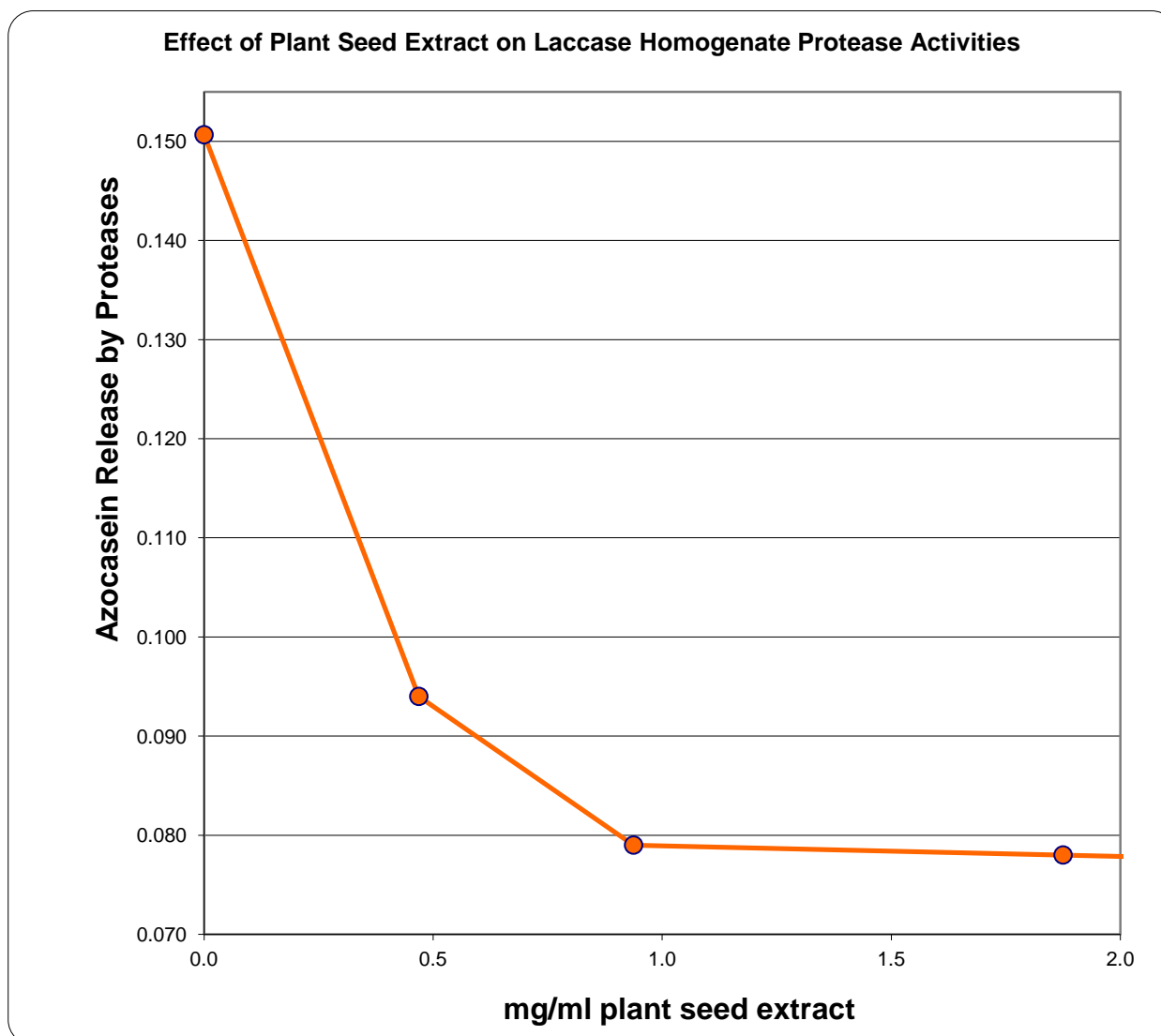


Note: Modifications used to make LowPro baculovirus platform are a trade secret.

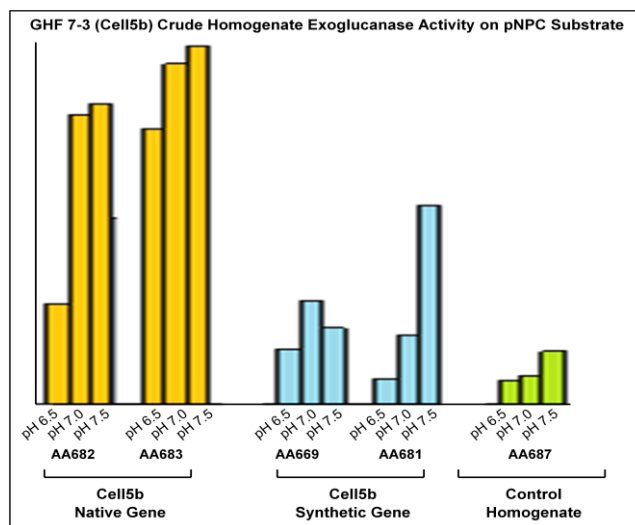
3. OPTIMIZE EXPRESSION & YIELD/ PROTEASE INHIBITORS: A significant portion of the protease activity in larval homogenates is comprised of enzymes from the gut. One strategy pursued during this project involved co-expression of a protease inhibitor peptide in a recombinant baculovirus. Expression of a protease inhibitor peptide in the LowPro virus resulted in homogenates that were further reduced in detectable protease levels. Co-expression of the protease inhibitor peptide with termite enzymes could lead to more stable crude homogenate formulations.



4. OPTIMIZE EXPRESSION & YIELD/ ADD PROTEASE INHIBITORS FROM PLANTS: Termite laccases were abundantly expressed in caterpillars, and were soluble and easily purified. Significant gains in laccase expression were deemed not to be likely given the high expression already achieved. Work was done to improve the crude homogenate formulation of laccase. Protease activities were detected by azocasein assay in laccase homogenates. These protease activities will degrade laccases in homogenates over time. Much of this protease activity was determined to be proteases from the larval host. Synthetic protease inhibitors like ABESF are expensive and are undesirable to have in homogenates being added to lignocellulosic feedstocks. Some plants naturally contain protease inhibitors to protect against insect feeding. We were able to reduce protease activities in laccase homogenates by adding plant seed extracts. The plant seeds used as a source for extracts in this work were from a major agricultural crop and are low in cost. Thus, this is an economical way to stabilize termite enzymes in crude insect homogenates.



5. CODON OPTIMIZATION: A synthetic codon-optimized GH7-3 (Cell5b) gene was made in an attempt to improve expression. This did not significantly enhance the expression of GH7-3 in insect larvae. Comparison of crude homogenates for exoglucanase activity indicated that the native sequences for GH7-3 were more active than the optimized sequences. Improvement in soluble expression by codon optimization is host, algorithm and gene-dependent. Lack of success with GH7-3 is not a predictor for performance of this strategy with other proteins. During the course of this study, we found that termite endosymbiont genes which expressed poorly in insect larvae were difficult to improve by any modifications.



b) Test viruses in larvae to determine expression rate. C-PERL to evaluate the performance of the various viruses in larvae testing by Western analysis, Coomassie-stained PAGE, and activity determination.

Described in Objective 2, part d).

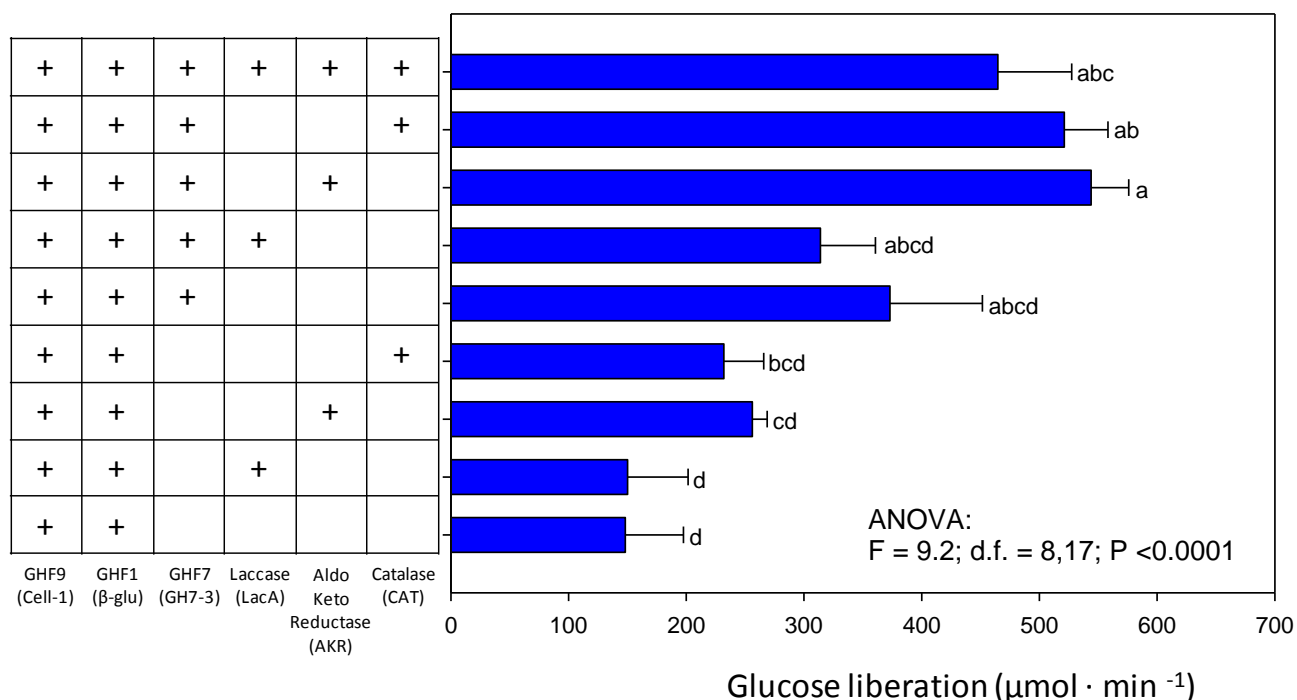
c) Move the more promising viruses into scaled expression using C-PERL proprietary production method of oral infection of larval biomass with POV and factory automation.

Described in Objective 2, part e).

d) UF(Purdue) to focus efforts on trying various combinations of enzymes discovered during the SBIR work, looking for the optimal cocktail for commercialization. This would include both purified and crude (e.g., freeze-dried) forms of enzymes.

Work was not done with freeze dried larvae expressing termite enzymes. The most cost effective form of enzyme production is crude liquid homogenates. Freeze drying was determined to be cost prohibitive as it added labor and energy costs removing water from a material that which could be used directly to treat biomass.

Purdue performed systematic experiments combining purified larval-expressed termite enzymes in various combinations and using them to treat pine sawdust substrate to release sugars. Summary data for one these types of experiments are shown below.



Objective 4: Perparation for Commericalization.

a) Evaluate enzyme combinations in sawdust and related “real world” models for conversion of various feed stocks to fermentable sugars. These evaluations may use only SBIR-generated proteins, or may also include newly identified proteins in combination with existing recombinant activities.

As mentioned in Objective 4, part d) assays were done at Purdue combining purified expressed termite enzymes with a pine sawdust substrate to release sugars. Work was also done at C-PERL combining crude homogenates with softwood sawdust and corncob dust. Sugars could be liberated but it was clear that more extensive optimization was needed.

1. FEEDSTOCK SACCHARIFICATION ASSAYS USING RECOMBINANT ENZYMES: Each purified recombinant enzyme (Cell-1, β -glu, GHF7-3, AKR, CAT and GHF11-1) was tested at 8 μg per assay except *LacA*, which was tested at 1 μg . In addition to recombinant proteins, assays included 2% w/v substrate in 750 μL HEPES buffer (0.1M, pH7) containing 0.01 M calcium chloride. Each assay was repeated three times. The substrate used was pine sawdust (Nelson Wood Shims; Cohasset, MN) ground to a fine powder using a DCG-20 coffee grinder (Cuisinart; East Windsor, NJ, USA). Assays were incubated in vented 1.5-mL Nalgene tubes at 37°C and 220 rpm for 18 hr, and were stopped with 15 μL of 0.2 M EDTA (final concentration = 4 mM). The addition of EDTA in this manner was found to greatly stabilize color formation after glucose detection (see following sections). Reaction tubes were centrifuged 5 min at 16,000 $\times g$, and the resulting supernatant used for glucose and xylose detection as described below.

2: **GLUCOSE DETECTION:** A commercial glucose-mutarotase (glucose C2) detection reagent was used for glucose detection (Wako Chemical; Richmond, VA, USA). Supernatants from digestion assays were obtained as described above and quantified in 96-well microplates. Each assay technical replicate was divided into three 50- μ L reaction aliquots, which were each used for glucose detection. Reagents for the glucose assay were added at 200- μ L per well, followed by orbital shaking for 5 mins at 37°C. Absorbance was read as an endpoint at 505 nm relative to a glucose standard curve. Each assay plate was run in real-time with its own glucose standard curve. All assays included buffer blanks that contained all reaction components except protein. Standard curves were prepared as seven serial dilutions in assay buffer with 4 mM EDTA from 5 mM downward to 0.078125 mM. The 8th well that contained only buffer served as the blank control.

3. **XYLOSE DETECTION:** Xylose detection was achieved using a commercial D-Xylose assay kit (Megazyme; Wicklow, Ireland). Supernatants from digestion assays were obtained as described above and transferred to microplate wells. Three reaction aliquots were quantified per assay technical replicate using 50 μ L per well. A 50 μ L aliquot of the supernatant was added to 242 μ L of kit mixture (freshly prepared by mixing 160 μ L water, 40 μ L of each solution 1 and solution 2, and 2 μ L of suspension 3) per well, and initial absorbance was read at 340nm after orbital shaking for 5 mins. Next, 5 μ L of solution 4 was added to the reaction mixture and final absorbance was read at 340 nm after orbital shaking for 6 mins. All assay runs included buffer blanks that contained all reaction components except protein. The concentration of D-xylose was calculated using the extinction coefficient of NADH at 340 nm ($6300 \text{ M}^{-1} \text{ cm}^{-1}$).

b) C-PERL to develop freeze-dry larvae and other crude formats that are amenable to a large-scale, low-cost biomass pre-treatment. Samples will be provided to UF (Purdue) for individual activity testing and for incorporation into cocktail.

As mentioned in Objective 2, part f), crude homogenates of larvae expressing termite enzymes were sent to Purdue and evaluated for activity against pine sawdust.

c) Begin early cost modeling and identification of commercialization partner(s).

(Company names are not specified to maintain confidentiality for C-PERL and Purdue so as not to compromise future interactions with these companies)

A meeting was held with representatives from the cellulosic ethanol company company A. There was discussion of using insect crude homogenates to treat lignocellulosic biomass but cost remained an issue. Company A seemed biased towards using fungal and bacterial enzyme products due to the extremely low cost of producing enzymes this way. Work from C-PERL was presented showing the synergy of insect homogenates with fungal enzymes but Company A declined to follow up.

Early discussion with company B did not lead to a collaborative or commercial arrangement.

A meeting was arranged with Company C to discuss use of C-PERLXpress termite enzymes in cellulosic animal feed. Though the application is not biofuel-related, an opportunity exists to characterize novel C-PERL proteins derived from the current work. There was interest in aldo-keto reductase and sample homogenates were sent. There was also interest in esterase and xyloglucanases. Partial esterase genes were identified in the termite 454-array cDNA library but they were not cloned into recombinant baculoviruses.

In conjunction with the Clean Energy Alliance Partnership, C-PERL sought mentors for commercialization assistance. An initial contact was made with a US State Business & Technology Center, but a broader relationship was not developed. Additional candidate mentors were not identified.

There were brief communications with the cellulosic ethanol producer, Company D and a cellulase enzyme testing facility at the provincial the University of E. No strong interest was shown in C-PERLXpress termite enzymes.

It was noted after our interactions, that cellulosic ethanol producers are secretive about the processes and enzyme mixtures that they use. This complicated efforts to find commercialization partners. There is conflict of interest between being willing to evaluate termite enzymes and desire to protect proprietary methods.

Company A is a US company producing cellulosic ethanol using ionic pre-treatment and enzymes.

Company B is a US fortune 500 company with 30 billion in revenues and interests in renewable energy.

Company C is a US animal feed company with 1 billion dollars in revenue.

Company D is a Canadian company producing cellulosic ethanol by steam explosion and enzymes.

University E is in a Canadian province that leads North America producing canola biofuels.

Section 5: Summarize Project Activities

1. ORIGINAL HYPOTHESIS: A full complement of lignocellulolytic enzyme genes will be identified from the termite *R.flavipes*. A series of these termite midgut enzymes will be cloned into recombinant baculoviruses and expressed by C-PERLXpress. Resulting enzyme containing larvae will be homogenized and homogenates will be combined to create an enzyme cocktail. The larval homogenate enzyme cocktail would be used for treating lignocellulose biomass to release sugars. Termites are 95% efficient at releasing sugars from lignocelluloses. Termite enzyme based cocktails could provide the basis for engineering a superior pretreatment process for lignocellulose that can economically surpass current processes involving sugar and starch rich feedstock-based ethanol generation.

2. APPROACHES USED: In brief, several cDNA libraries were made from termite guts; methods included specialized procedures for enriching for target genes, including subtraction and normalization methods. Lignocellulolytic candidate genes were identified by high-throughput 454 DNA sequencing and bioinformatic analysis. Full-length ORFs were engineered into synthetic genes and cloned into recombinant baculoviruses. Recombinant baculoviruses were used to infect *Trichoplusia ni* larvae by PERLXpress technology. Larvae containing termite enzymes were flash frozen at -80C. Later larvae were homogenized. Expressed termite enzymes were purified from homogenates and concentrated. Purified forms of termite enzymes and crude homogenates were subjected to enzymatic assays on artificial and real world substrates. Mixtures of purified enzymes or crude homogenates were assayed for synergistic activities on lignocellulosic substrates. Termite enzymes were also combined with fungal cellulase cocktails to determine synergistic activities with commercial preparations.

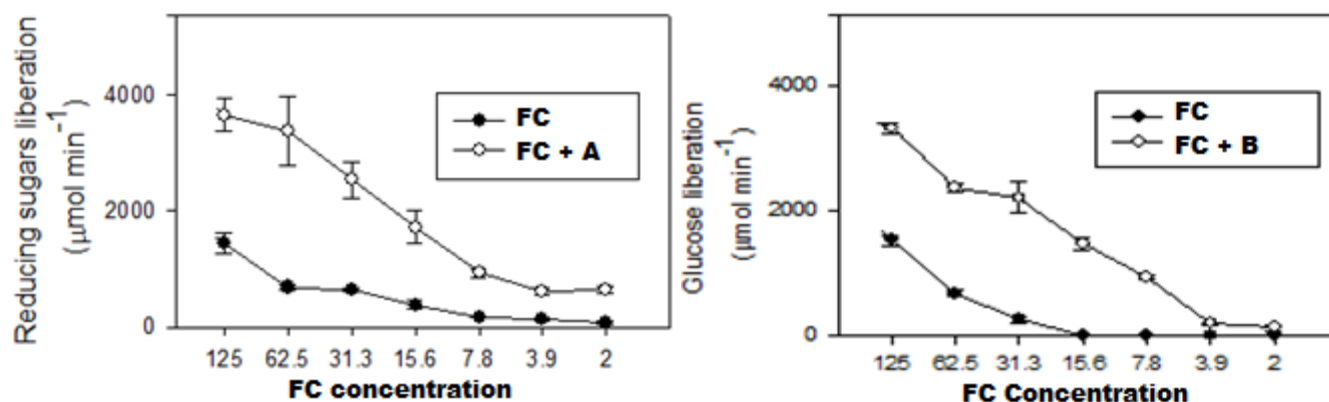
3. PROBLEMS ENCOUNTERED: The termite midgut cDNA libraries were composed of sequences representing genes from termites, endosymbiont protists, fungi and bacteria. Full length gene ORFs from cDNA libraries were rare. Early success in expressing laccase, GH9 (Cell-1), and GH1 (b-glucosidase) was in part due to having complete ORF sequences to clone into recombinant baculoviruses. This was achieved by having full length cDNA plasmid clones and considerable effort doing bioinformatic analysis of contigs. The 454-array termite gut cDNA library provided a large number of new cellulase types and candidate lignases that could be potentially expressed. Unfortunately many contig sequences were incomplete ORFs. It was decided not to pursue obtaining complete ORFs based on incomplete contig sequences from the 454-array as this would have exceeded scope of the project. The incremental effort would have required

obtaining more termite tissues and generating full length individual cDNA's of target gene transcripts. Full-length cDNA would likely not be obtained and reverse transcriptase 5 prime and 3 prime extension methods would need to be applied. Cloned extension products would then need to be DNA sequenced and made into complete synthetic genes. This work for even a small number of genes would have exceeded available resources allotted for this step.

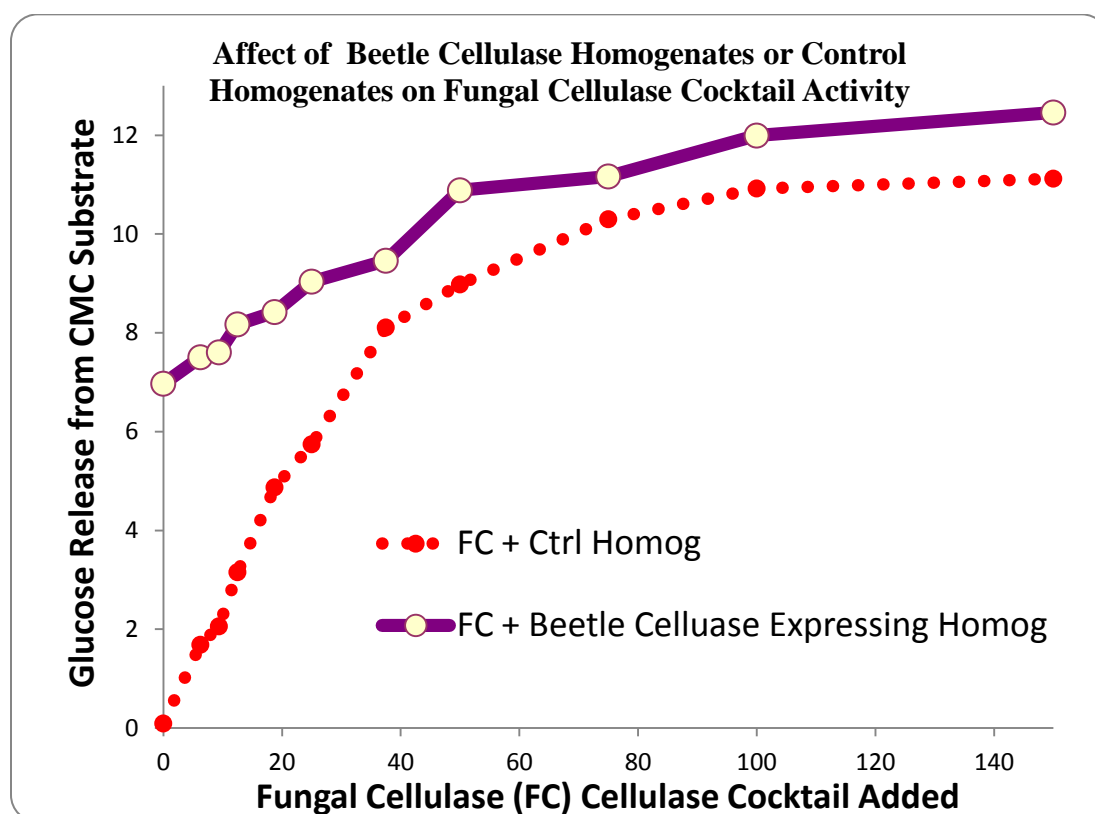
A significant portion of cellulases genes identified in this study were of endosymbiont protist origin based on sequence similarity to known cellulases. In this project, it was observed that most protist cellulases were not abundantly expressed and were not very soluble when expressed in the C-PERLXpress system. A prime example of a difficult to express protist protein is GH7 (Cell-3). Proteins such as this consumed great time and effort in expression and purification. There was little to show for such efforts. Changing of virus platforms, virus promoters, secretion signal peptides, and codon optimization did not significantly improve expression of many endosymbiont protist cellulases. It can anecdotally be concluded the insect-based expression vector system of C-PERL express is better suited to express termite insect origin proteins than endosymbiont origin proteins. The most abundant and soluble proteins expressed were termite origin proteins such as laccase, GH9 (Cell-1), aldo-keto reductase and GH1 (b-glucosidase).

4. DEPARTURE FROM PLANNED METHODOLOGY/ MIXING TERMITE ENZYMES WITH FUNGAL ENZYMES: A departure from methodology concerned the concept of recreating the termite digestive system to digest lignocellulose. This placed heavy reliance on using only termite enzymes. Industrial bioreactors are not likely to be made identical to a termite gut. Companies employing enzymes to make cellulosic ethanol combine enzymes from different sources to make enzyme cocktails. GH7 type fungal and bacterial cellulases are more economically produced than PERLXpressed endosymbiont GH7 cellulases. The termite digestive system provides enzymes to supplement endosymbiont bacterial and protist enzymes. It made sense to add PERLXpressed termite enzymes to fungal enzyme cocktails similar to what are used in industry. The PERLXpressed termite enzymes A, B, C and D were found to be synergistic with commercial fungal cellulase (FC) cocktail on a pine sawdust substrate. Using C-PERL recombinant enzymes to enhance the performance of existing fungal cellulase cocktails is a supplemental to the original concept of using solely termite-derived proteins for this project. Two data plots are shown below illustrating the enhancement of fungal cellulases (FC) by termite A or B.

Note: Enzymes A, B, C and D specifics are withheld so as not to affect patent applications.



5. DEPARTURE FROM PLANNED METHODOLOGY/ EXPRESSING BEETLE ENZYMES: A departure from methodology was the inclusion of a non-termite insect cellulase. Wood boring beetles are another group of insects that consume wood. The 454-array cDNA library revealed that the *R.flavipes* termite expresses cellulase types expressed by wood boring beetles. Unfortunately, some of these cellulase gene transcripts identified in the termite gut were of bacterial origin and were incomplete ORFs. Bacterial origin genes are better expressed in bacteria. Work was done to clone and express a wood boring beetle cellulase. This was done to provide an additional cellulase of insect origin to the project. In the C-PERLXpress system, the beetle cellulase like termite origin cellulases, was abundantly expressed in larvae, was soluble and had high cellulase activity in crude homogenates. This further supports the premise that insect origin cellulases are ideally suited to being expressed in C-PERLXpress. It was also found that the beetle cellulase could increase the total sugars released from carboxymethyl cellulose substrate by fungal cellulase cocktails. This is illustrated in the figure below.



6. ASSESSMENT OF METHODOLOGY CHANGES IMPACT ON PROJECT RESULTS: By combining C-PERLXpress insect lignocellulolytic enzymes with commercial fungal enzymes, it expands the possibility of using insect enzymes to enhance enzyme processes that are already in use to treat lignocelluloses for biofuel production. This fits better with industrial models and makes it possible to more quickly apply the products of this research.

Future efforts will include supply of proteins to collaborators at Purdue for further study, particularly in bulk format which has greater relevance for a cost-effective cocktail suitable for woody-biomass pre-treatment. C-PERL will continue to seek potential partners with capability to characterize proteins identified and expressed in this study.

Section 6: Products Developed

a) Publications

1: JOURNAL PUBLICATIONS:

- Sethi, A., J. Slack, E. S. Kovaleva, G. W. Buchman, and M. E. Scharf. 2012. Lignin associated metagene expression in a lignocellulose-digesting termite. *Insect Biochemistry and Molecular Biology*. (In press).
- Sethi, A., and M. E. Scharf. 2012. Biofuel: fungal, bacterial and insect lignocellulose degraders. *Encyclopedia of Life Sciences*, John Wiley & Sons, Ltd., UK. (In press).
- Scharf, M.E., Z.J. Karl, A. Sethi, and D.G. Boucias. 2011. Multiple levels of synergistic collaboration in termite lignocellulose digestion. *PLoS One* 6: e21709.
- Scharf, M.E., Z.J. Karl, A. Sethi, R. Sen, R. Raychoudhury, and D.G. Boucias. 2011. Defining host-symbiont collaboration in termite lignocellulose digestion: The view from the tip of the iceberg. *Communicative and Integrative Biology* 4: 761-763.
- Coy, M.R., Salem T.Z., Denton J.S., Kovaleva E., Liu Z., Barber D.S., Campbell J.H., Davis D.C., Buchman G.W., Boucias D.G., Scharf M.E. 2010. Phenol-oxidizing laccases from the termite gut. *Insect Biochemistry and Molecular Biology* 40: 723-732.
- Scharf, M.E., Kovaleva, E.S., Jadhao, S., Campbell, J.H., Buchman, G.W., Boucias, D.G. 2010. Functional and translational analyses of a beta-glucosidase gene (glycosyl hydrolase family 1) isolated from the gut of the lower termite *Reticulitermes flavipes*. *Insect Biochemistry and Molecular Biology*. 40: 611-620.
- Zhou, X., Kovaleva, E.S., Wu-Scharf, D., Campbell, J.H., Buchman, G.W., Boucias, D.G., Scharf, M.E., 2010. Production and characterization of two recombinant beta-1,4-endoglucanases (GHF9) from the termite *Reticulitermes flavipes*. *Archives of Insect Biochemistry and Physiology*. 74: 147-162.
- Sun, J.Z. and Scharf, M.E. 2010. Exploring and integrating cellulolytic systems of insects to advance biofuel technology. *Insect Science* 17: 163-165.
- Scharf, M.E. and D.G. Boucias. 2010. Potential of termite-based pre-treatment strategies for use in bioethanol production. *Insect Science* 17: 166-174.
- Wheeler MM, Tarver MR, Coy MR, Scharf ME. 2010. Characterization of four esterase genes and esterase activity from the gut of the termite *Reticulitermes flavipes*. *Archives of Insect Biochemistry and Physiology*. 73: 30-48.

2: CONFERENCES AND OTHER PUBLIC RELEASES:

- Sethi, A., J. Slack, E. S. Kovaleva, G. W. Buchman, and M. E. Scharf. Lignin associated metagene expression in a lignocellulose-digesting termite. *Entomological Society of America, 2012 [Oral]*.
- Scharf, M. E., and A. Sethi. Termite digestion: From transcriptomics to biofuel technology. *Entomological Society of America, 2012 [Oral]*.
- Sethi, A., J. Slack, E. S. Kovaleva, G. W. Buchman, and M. E. Scharf. The termite gut bioreactor of transcriptional and translational analysis of lignin degradation. *Arthropod Genomics Symposium, 2012 [Poster]*.
- Sethi, A., E. S. Kovaleva, J. Slack, S. Brown, G. W. Buchman and M. E. Scharf. Bioprospecting for novel enzymes from protist communities of *Reticulitermes flavipes* for efficient biomass processing". Annual

Sigma Xi Graduate Student and Postdoctoral Researcher Poster Competition, Purdue University, 2012 [Poster].

Sethi, A., E. S. Kovaleva, J. Slack, S. Brown, G. W. Buchman and M. E. Scharf. *Bioprospecting for novel enzymes from protozoan communities of Reticulitermes flavipes for efficient biomass processing. Entomological Society of America, 2011 [Poster].*

Scharf, M. E. *Hypothesis-driven insights into proximate mechanisms of termite sociality: nutrition and endocrinology. Entomological Society of America Annual Meeting, Reno, NV, Nov. 2011 [Oral].*

Scharf, M. E. *Host-symbiont collaboration in termite lignocellulose digestion. Entomological Society of America Annual Meeting, San Diego, CA, Dec. 2010 [Oral].*

b) Web sites

<http://www.gizmag.com/termite-enzymes-for-biofuel-production/19139/>

<http://www.purdue.edu/newsroom/research/2011/110705ScharfTermites.html>

http://abcnews.go.com/Technology/DyeHard/termite-save-world/story?id=9046512#.UJmhWW_A9p4

<http://www.dailytech.com/Termite+Guts+Could+Lead+to+Biofuel+Production+/article22098.htm>

http://www.energydigital.com/green_technology/termite-may-be-key-to-cellulosic-ethanol

<http://www.c-perl.com/about-us/news-a-events.html>

c) Networks or collaborations fostered

Suganit (communication and meeting with)

Elanco (communication and meeting with)

Iogen (communication with)

Saskatchewan Research Council (communication with)

Entomological Society of America (presentation at meetings)

d) Technologies/Techniques

Technology to engineer termite enzymes into recombinant baculoviruses.

Technology to improve the secretion and solubility of recombinant termite enzymes.

Techniques to assay cellulases on wood sawdust substrates.

Techniques to use termite enzymes to improve fungal cellulases

e) Inventions/Patent, Licensing Agreements

US Provisional Patent No. 61/602,149 “Novel lignases and aldo-keto reductases for biomass conversion”.

Inventors: M.E. Scharf, A. Sethi. File date: Feb 21, 2012.

US Provisional Patent No. 61/728,216 “SOD and GPX”. Inventors: M.E. Scharf, A. Sethi. File date: Feb 21, 2012.

f) Other Products such as databases

Termite gut metatranscriptome sequences obtained after Roche 454-titanium pyrosequencing were deposited under the accession numbers KA650619 - KA659892.