Linking Hydrolysis Performance to Trichoderma reesei Cellulolytic Enzyme Profile

Lehmann, Linda Olkjær; Petersen, Nanna; I. Jørgensen, Christian; Olsson, Lisbeth; Stocks, Stuart M.; Jørgensen, Henrik S.; Hobley, Timothy John

Published in:
Biotechnology and Bioengineering

Link to article, DOI:
10.1002/bit.25871

Publication date:
2016

Document Version
Peer reviewed version

Biofuels and Environmental Biotechnology **Biotechnology and Bioengineering**
DOI 10.1002/bit.25871

Linking hydrolysis performance to *Trichoderma reesei* cellulolytic enzyme profile †

Linda Lehmann[a*1], Nanna Petersen[b2], Christian I. Jørgensen[c], Lisbeth Olsson[d], Stuart M. Stocks[c], Henrik S. Jørgensen[c] and Timothy Hobley[a3].
[a]Center for Microbial Biotechnology, Department of Systems Biology, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[b]Department of Chemical and Biochemical Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[c]Novozymes A/S, Krogshøjvej 36, 2880 Bagsværd, Denmark
[d]Industrial Biotechnology, Department of Chemical and Biological Engineering, Chalmers University of Technology, 412 96 Göteborg, Sweden

*Corresponding author

Email: lle@novozymes.com

[1] Present address: Novozymes A/S, Krogshøjvej 36, 2880 Bagsværd, Denmark
[2] Present address: Novo Nordisk A/S, Novo allé, 2880 Bagsværd, Denmark
[3] Present address: Division of Industrial Food Research, National Food Institute, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

†This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1002/bit.25871]

Additional Supporting Information may be found in the online version of this article.

Received May 09, 2015; Revision Received October 22, 2015; Accepted October 26, 2015

Abstract

Trichoderma reesei expresses a large number of enzymes involved in lignocellulose hydrolysis and the mechanism of how these enzymes work together is too complex to study by traditional methods, e.g. by spiking with single enzymes and monitoring hydrolysis performance. In this study a multivariate approach, partial least squares regression, was used to see if it could help explain the correlation between enzyme profile and hydrolysis performance. Diverse enzyme mixtures were produced by *Trichoderma reesei* Rut-C30 by exploiting various fermentation conditions and used for hydrolysis of washed pretreated corn stover as a measure of enzyme performance. In addition, the enzyme mixtures were analyzed by liquid chromatography - tandem mass spectrometry to identify and quantify the different proteins. A multivariate model was applied for prediction of enzyme performance based on the combination of different proteins present in an enzyme mixture. The multivariate model was used for identification of candidate proteins that are correlated to enzyme performance on pretreated corn stover. A very large variation in hydrolysis performance was observed and this was clearly caused by the difference in fermentation conditions. Besides β-glucosidase, the multivariate model identified several xylanases, Cip1 and Cip2 as relevant proteins to study further.

Key Words: Trichoderma reesei, proteomics, liquid chromatography tanden mass spectrometry, mathematical modelling, cellulase

Introduction

Lignocellulose is a highly complex structure composed mainly of cellulose, hemicellulose, and lignin. It has a great potential as a cheap and renewable feed stock, e.g. for bioethanol production, but because of the heterogeneous nature of lignocellulose, a mixture of different enzymes is necessary to hydrolyze the cellulose and hemicellulose to its monosaccharides. In addition there is a need for synergy between the enzymes indicating that there exist an optimal ratio between the cellulolytic enzymes and also that they need to act in the right order (Mansfield et al., 1999, Fägerstam and Pettersson, 1980; Chanzy and Henrissat, 1985; Zhang and Lynd, 2004). The filamentous fungus *T. reesei* produces at least two cellobiohydrolases (Cel7A and Cel6A), five endoglucanases (Cel7B, Cel5A, Cel12A, Cel61A, and Cel45A), and two β-glucosidases (Cel3A and Cel1A) each with different specificities and activities (Zhang and Lynd, 2004). Similarly, hemicellulolytic enzymes can also be produced in multiple forms with various activities. Because of the heterogeneous nature of the enzyme cocktail and the synergy between the enzymes, the exact role of each individual enzyme component in cellulose (and hemicellulose) hydrolysis is not fully understood yet. The major components of the *T.reesei* secretome (Cel6A, Cel7A, Cel5A, and Cel12A) constitute about 85% of the total secreted protein (Goyal et al., 1991; Tolan and Foody, 1999). It can be suspected that the high production level of these four enzymes is a reflection of their importance in the hydrolysis of cellulose. As is the case in most fungi, enzyme production changes depending on the conditions used for cultivating the fungi e.g. choice of carbon source, pH, temperature, agitation etc. It is unlikely that production of all the enzymes in *T. reesei* is affected to the same degree and thus the cellulolytic enzyme *profile* will be expected to change depending on the choice of cultivation conditions. Within recent years technology has allowed for a more large-scale examination of these fungal proteins e.g. by 2D gel electrophoresis coupled to liquid chromatography tandem mass spectrometry (LC-MS/MS) (Vinzant et al., 2001; Herpoel-Gimbert et al., 2008; Nagendran et al., 2009). From such studies it has become evident that the profile of secreted proteins in *T. reesei* contains many different proteins of cellulolytic and non-cellulolytic activity. However, there has been little attempt to couple the information about the entire protein profile to how this affects hydrolysis performance on lignocellulosic substrates. This is likely due to the fact that the usual profile of secreted proteins in *T. reesei* contains more than just a few enzymes; up to 82 proteins have been reported (Nagendran et al., 2009). In such large data sets it can be difficult to manually determine which proteins have a determining effect on the total enzymatic activity and which proteins are less important.

Partial least squares (PLS) regression is a powerful tool that can handle large data sets, such as the output of LC-MS/MS analysis. PLS is especially useful when handling data matrices containing more variables (e.g. the number of proteins identified by mass spectrometry) than observations (e.g. the number of samples used for mass spectrometry analysis). Furthermore, PLS can handle both noisy and highly collinear data (Eriksson et al., 2004). This has made PLS a standard tool within chemometrics, for example for the analysis of spectroscopic data (Wold et al., 2001). For a more thorough introduction to PLS regression the reader is referred to Martens and Næs (1989).

In this study we investigated the usefulness of a PLS approach for modelling the hydrolysis performance on pretreated corn stover (PCS) based on the enzyme profile determined by mass spectrum. This was done by growing *T. reesei* RutC30 in batch cultivations on crystalline cellulose (Avicel) using different growth conditions to obtain various profiles of secreted proteins. One-dimensional gel electrophoresis (SDS-PAGE) and LC-MS/MS was applied to identify the proteins secreted during the different cultivations. Additionally, the filtered fermentation broth containing the secreted proteins was used to determine hydrolysis performance on PCS. PLS was applied for modeling the relationship between the different protein profiles and the measured performances on PCS. Proteins variations most correlated to hydrolysis performance were subsequently identified, demonstrating that the PLS approach is useful where data is too complex for a traditional hypothesis driven approach

Materials and methods

Strain and spore propagation

In all cultivations *T. reesei* strain RutC30 obtained from Novozymes A/S, Bagsværd, Denmark, was used. The strain was maintained as frozen spore suspensions at -80˚C in 20 % (v/v) glycerol. Spores were propagated on plates containing potato dextrose agar (PDA) for 7 days at 30˚C before inoculation of fermentors. The PDA medium consisted of potato dextrose agar (Difco) 39 g, 1 ml of trace element solution, double distilled water (ddH_2O) until 1000 ml. The trace element solution contained $ZnSO_4*7H_2O$

1.0 g, $CuSO_4*5H_2O$ 0.5 g in 100 ml ddH_2O. Fermentors were inoculated to a final concentration of $1*10^9$ spores/L.

Cultivation conditions

Batch cultivations were performed in 5 L bioreactors with a working volume of 4 L and the batch cultivation medium consisted of Avicel PH-101 (Sigma) 25 g/L, KH_2PO_4 4 g/L, $(NH_4)_2SO_4$ 13.6 g/L, $CaCl_2*2H_2O$ 0.8 g/L, $MgSO_4*7H_2O$ 0.6 g/L, peptone 6 g/L, 0.2 ml/L Tween 80 and 1 ml/L trace element solution. The trace element solution was made as 1000x concentrated solution and consisted of $FeSO_4*7H_2O$ 10 g/L, $MnSO_4*H_2O$ 3.2 g/L, $ZnSO_4*7H_2O$ 2.8 g/L, $CoCl_2*6H_2O$ 4 g/L. Agitation was achieved with two four-blade disc turbines. The bio-reactor was sparged with air at 0.22 vvm (volume air/fermentation volume/min) and was kept constant at this value in all experiments. The content of CO_2 and O_2 in the exhaust gas was continuously measured in a gas analyzer (Fermentation Monitor Innova 1313).

In total a series of 12 batch cultivations were performed with various settings for pH, temperature and agitation and the chosen values were kept constant during the entire cultivation period, see table 1 for an overview. Temperature and agitation were adjusted prior to inoculation. The pH of the cultivation medium was adjusted immediately before inoculation and kept constant during the cultivation by automatic addition of 2M HCl and 2M NaOH. Foam was controlled by manual addition of sterile antifoam 204 (Sigma) when needed.

Sampling and analysis

Dry cell weight was determined using nitrocellulose filters (pore size 0.45µm, Gelman Sciences). A known weight of cell culture was added to the filter which had been pre-dried in a microwave oven at 150W for 15 minutes, cooled in a desiccator and subsequently weighed. The cell culture was filtered and the residue was washed with 0.9% NaCl and dried on the filter for 15 minutes in a microwave oven at 150W, cooled in a desiccator, and weighed. The cell dry weight, which included mycelium and residual cellulose, was then calculated. The offline pH was measured and online pH probes were adjusted if necessary. Approximately 15 ml of broth was centrifuged at 3000*g* and the supernatant was subsequently filtered (pore size 0.20 µm) and used for analysis of total protein, enzyme activity on filter paper and PCS, respectively, as well as analysis by SDS-PAGE and LC-MS/MS. The amount of total protein in the filtered fermentation broth was determined by using the BCA (bicinchoninic acid) protein assay according to the manufacterer's instructions (product #23227 Pierce Protein Research Products, Thermo Scientific). Filter paper activity assay was performed in microtiter scale according to the 60-µl FPA described by Xiao et al. (2004).

Hydrolysis performance on pretreated corn stover (PCS)

The hydrolysis performance of the enzymes harvested at the end of fermentation was evaluated in a hydrolysis assay, referred to as *performance assay*, using washed PCS as substrate: the PCS, produced by acid steam explosion was kindly provided by the US National Renewable Energy Laboratory (NREL) and treated at Novozymes Inc., Davis, CA, USA by extensive washing and grinding to uniform size. According to NREL, the solid's composition of the washed PCS is 57.5% glucan, 7.0% xylan, 0.9% galactan, 0.7% arabinan, and 27.2% lignin and the hydrolysis assay was performed as the standard hydrolysis assay explained in Harris et al., (2010). Each sample was tested in a concentration of 2, 5, and 8 mg protein/g cellulose respectively, and was performed in triplicates with and without addition of β-glucosidase. The hydrolysis performance of each sample was calculated as the percentage of glucose released after 72 hours at 50°C (average of the triplicate assays) compared to the amount of glucose present as cellulose in the PCS at the beginning of the assay. For samples without β-glucosidase the performance was based on glucose release only, whereas for samples where β-glucosidase was added in the assay the performance was based on glucose and cellobiose released.

Protein separation by SDS-PAGE and extraction

Samples from the different cultivations were prepared for SDS-PAGE by incubating the samples with sample buffer containing dithiothreitol (DTT) for 5 min at 95˚C. After cooling, 10 µl 1 M iodoacetamide in 0.5 M TrisHCl pH 9.2 was added, and the samples were incubated for 20 min at room temperature. Each sample was loaded on a SDS-PAGE gel (Invitrogen NuPage 4-12% Bis-Tris, 10 well, 1 mm, MES running buffer) and run according to Invitrogen guidelines. In each case it was attempted to reach maximum loading capacity on the gel in order to increase the amount of material for the mass spectrometry. Staining with InstantBlue (Expedeon) was performed according to the manufacturer's instructions. Gels were destained with dH_2O.

Each lane of the gel was cut into 8 segments that were digested with trypsin following a standard procedure. In short, the gel segments were washed with 150 µl 50% EtOH/50 mM NH_4HCO_3 for 2x30

min at room temperature in a rotary shaker at 600 rpm. After washing, 50 µl 100% acetonitrile was added to shrink the gel pieces and incubated for 15 min at room temperature. The gel pieces were completely dried in a SpeedVac for approximately 10 min. 15 µl 25 mM NH_4HCO_3 containing protease (trypsin, Roche #11418475001) was added to re-swell the gel pieces and to digest the proteins into peptides. After 10-15 min additional 25 µl 25 mM NH_4HCO_3 buffer (without protease) was added and incubation proceeded overnight at 37˚C. After this point, 50 µl 70% acetonitrile/0.1% trifluoroacetic acid was added and incubated for 15 min at room temperature to extract the peptides. Extraction was repeated and extracts were dried in a SpeedVac and reconstituted in 50 µl 5% formic acid.

Liquid chromatography tandem mass spectrometry

The extracted peptides were analyzed by mass spectrometry (MS) on a LTQ Orbitrap XL LC-MS/MS system (Thermo Scientific) coupled to an Accela LC HPLC system where samples were separated on a 2.1x50 mm ACQUITY UPLC BEH C18 1.7 µm column (Waters) using an acetonitrile gradient in 0.1% formic acid. Both parent peptide ions (MS data) and peptide fragments (MS/MS data) were measured in the high precision Orbitrap. Peptide sequence fragmentation was obtained using the HCD fragmentation of the system.

Database search for protein identification

The combined MS and MS/MS data were searched against the *T. reesei* genome using the Mascot search engine (Mascot server version: 2.2.0, Matrix Science) that is considered a standard software tool for protein identification from combined MS and MS/MS data (Kapp and Schütz, 2007). The segment raw files for each sample were merged and searched with Mascot. The relative amounts of the individual proteins in each sample were measured by emPAI (exponentially modified protein abundance index) values calculated by Mascot. emPAI values give an estimate of relative protein concentrations in mixtures, based on protein coverage by the peptide matches in a database search result, as described in Ishihama et al. (2005).

Processing of data for PLS regression

Based on total protein measurements and the relative distribution of individual proteins measured by LC-MS/MS, the specific amount of each protein in a given cultivation sample was calculated. This absolute protein distribution was structured in an X-matrix shown in figure 3 with rows representing each sample and columns representing each protein identified. In a similar way, a Y-matrix was constructed with the % conversion obtained in the hydrolysis assay for each cultivation. As explained previously, the broth from each of the 12 cultivations was used in 3 different concentrations with and without addition of β-glucosidase meaning that there are 72 rows in the X and Y matrices. The full X and Y matrix can be found in the supplementary material. All values were mean-centered and scaled to unit variance as preprocessing for PLS regression. The modelling was performed in MATLAB using the PLS Toolbox from Eigenvector Research Inc. As with other regression techniques, the independent variables in the X-matrix are used to estimate the dependent variable(s) in the Y-matrix. However, the X-variables are not used directly for estimation of Y. Instead a number of latent variables, which are linear combinations of the original X-variables using different loadings, are identified and used as regressors for Y. The number of latent variables in the PLS models was chosen based on evaluation of the explained variance of Y. A script for the PLS regression can be found in the supplementary material.

Results and discussion

Fermentation results

The various cultivation conditions tested resulted in different titers of protein and filter paper activity (table 2). In general the growth pattern observed showed an increase in total dry weight (dry cell weight including cellulose) during the first 48 hours of cultivation when the fungi were growing on the easily metabolized peptone present in the cultivation medium (figure 1). Afterwards the growth on cellulose began and the dry weight decreased, indicating that the mass lost by hydrolysis of insoluble cellulose to soluble products was larger than the net increase in dry cell weight. While the growth on cellulose occurred, an increase in protein levels and enzyme activity on filter (FPU/ml) paper was observed. The progress of protein production and enzyme activity followed one of three scenarios, 1) both parameters continued to increase throughout the cultivation (as shown in the example in figure 1), 2) both parameters increased during growth on cellulose but reached a constant level before the end of the cultivation, 3) both parameters increased during growth on cellulose but started to decrease again before the end of the cultivation. In each case the sample with the highest content of protein and filter paper

activity and at the latest time point from each cultivation, was chosen for further analysis by LC-MS/MS and the performance assay on PCS.

Enzyme performance on PCS

The performance results from each fermentation listed in table 1 are illustrated in figure 2 showing performance of the cell free enzyme cocktails to the left and to the right are the performance values obtained when pure β-glucosidase was added in excess to the cell free enzyme cocktails. In general the cultivations from pH 3.0 (A, B, and C) resulted in enzyme preparations with poor performance. Addition of β-glucosidase greatly improved the performance, e.g. the performance from cultivation A increased 6-fold when comparing the highest protein dose used. Samples from cultivations D-L corresponding to cultivation pH 4.5 and 6.0, and various temperatures and agitation rates, result in enzyme cocktails with better performance than pH 3.0 cultivations. Enzymes from cultivation D had the overall best performance, both with and without addition of excess β-glucosidase. Interestingly, the cultivation conditions in D are similar to F, except that in cultivation F no Tween 80 was added. The exact role of Tween 80 cannot be concluded from these results but the role of this surfactant could be versatile. In addition to increasing oxygen transfer, stabilization of cellulases (Okino et al., 2013) and protection from unproductive binding to cellulose (Helle et al., 1993) are some of the effects reported in literature. From figure 2 it is clear that enzyme performance can be increased or decreased drastically depending on the cultivation parameters (pH, temperature, and agitation) chosen during the production of the enzymes.

Changes in enzyme profile in response to different cultivation conditions

The same samples that were applied in the corn stover performance assay were further analyzed by LC-MS/MS. By this method, the proteins present in each sample were identified and a semi-quantitative estimation of the relative abundance of each protein was also made. A total number of 60 different *T. reesei* proteins were identified by LC-MS/MS, although not all 60 proteins were present in one single sample, see table 3 and table 4. In the tables proteins are identified by a number which refers to the protein id number at the JGI (joint genome institute) website for *T. reesei* at http://genome.jgi-psf.org/Trire2/Trire2.home.html. Out of the 60 proteins, 42% are already annotated, 38% are not officially annotated but alignments have given clues to the function of the proteins, and the remaining 20% are completely unknown proteins.

The secreted proteins from the cultivations summarized in table 1, which have been assigned a function of glycoside hydrolase (GH), are listed in table 3. The classification of proteins into GH families is based on sequence similarities and was first introduced by Henrissat (1991). A database with the different GH families and also other carbohydrate acting enzymes can be found at www.cazy.org and is continuously updated. It is clear from table 3 that cultivations performed at pH 3.0 (cultivation A-C) resulted in enzyme profiles with a lack of GH family proteins compared to cultivations performed at pH 4.5 and 6.0 (cultivation D-L). Cel7A, Cel6A, and Cel5A which represent the major part of the cellulolytic enzyme system based on protein levels (Goyal et al., 1991) were produced in all the cultivations. Cel7B was only identified in the cultivations from pH 3.0, whereas Cel12A was only produced in cultivations from pH 4.5 and 6.0, not in the cultivations from pH 3.0. Cel61b and β-glucosidase Cel3A were also not produced in the cultivations from pH 3.0 but were found in most of the cultivations from pH 4.5 and 6.0. Xylanase Xyn1 was primarily produced in the cultivations from pH 3.0 and almost never found in any of the other cultivations. In contrast Xyn2 and Xyn3 and xyloglucanase Cel74A were not at all found in the cultivations from pH 3.0 but in most of the cultivations from pH 4.5 and 6.0. β-xylosidase Bxl1 and β-mannanase Man5A were found in the majority of the cultivations performed at various pH values, see table 3. That some proteins are only identified in cultivations from certain pH values could indicate that the corresponding genes are regulated, either directly or indirectly, by pH or that the proteins are less stable at the pH value where they were not detected.

Enzymes classified as belonging to CE (carbohydrate esterase) families are also presented in table 3. This classification system is similar to the GH family system. It is most noteworthy that Cip2 was only produced in cultivations performed at pH 4.5 and 6.0, not at pH 3.0. Cip2 is a glucuronyol esterase hydrolyzing the covalent linkages between lignin and hemicellulose and is known to be expressed during cellulase induced conditions (Li et al., 2007). The Cip2 protein of *T. reesei* contains both a catalytic domain and a cellulose binding module (CBM) belonging to family 1 (Li et al., 2007). The results for Cip2 shown in table 3 indicate that the gene encoding Cip2 (*cip2*) is regulated by pH and thus not expressed at low pH values (pH 3.0). It could also be that the Cip2 protein is highly unstable at low pH values and therefore not detected in cultures grown at this pH value.

The proteins listed in table 4 do not belong to, or are not yet assigned to, any of the CAZy families. Note that proteins Cip1, hydrophobin II (Hfb2) and swollenin were only expressed at pH values 4.5 and 6.0,

not in cultivations performed at pH 3.0. Very little is known about Cip1 except that it has a secretion signal peptide and a cellulose binding module (CBM). Hfb2 is associated with sporulation in *T. reesei* (Askolin et al., 2005) and foaming (Askolin, 2006). The observed expression pattern of Hfb2 fits well with the observation that cultivations performed at pH 4.5 and 6.0 foamed extensively, whereas the foaming in cultivations at pH 3.0 was not significant. In addition, sporulation was observed in cultivations performed at pH 6.0 after a few days of cultivation. Swollenin is an expansin-like protein involved in the amorphogenesis of cellulose (Saloheimo et al., 2002; Arantes and Saddler, 2010). The Bip1 protein was expressed in two out of three cultivations performed at pH 6.0, but not in any of the other cultivations with lower pH value. The folding factor Bip1 is a member of the heat shock 70 protein family (HSP70) of molecular chaperones. In filamentous fungi the gene encoding Bip is overexpressed only under certain conditions (Conesa et al., 2001), which could be related to the response of sporulation at cultivation pH 6.0.

Seven proteins were assigned to proteases and these were primarily detected in the cultivations performed at pH 6.0, see table 4. Especially the cultivation from pH 6.0/23˚C seemed to have produced many different proteases. Only one protease was found in the cultivations performed at pH 3.0 (protein id 82623). In the cultivations performed at pH 4.5 proteases do not seem to have been a significant part of the secreted proteins. A total of 12 proteins identified in the filtered fermentation broth from the different cultivations, were completely unknown with regards to function, see table 4. The majority of these proteins were only expressed at pH 6.0, indicating that a different cellular response occurs at pH 6.0 compared to lower cultivation pH.

In summary it is clear that although *T. reesei* was grown on pure cellulose, hemicellulases were also produced. This is certainly an indication of co-regulation of the genes responsible for the identified enzymes. For example, the genes encoding Cel7A, Cel6A, Cel7B, Cel3A, and Bxl1 (β-xylosidase 1) are regulated by the same transcription factor, Xyr1, which mediates induction of these genes in the presence of inducing carbon sources (Mach-Aigner et al., 2008).

Multivariate data analysis for linking enzyme profile to hydrolysis performance

The enzyme performance measurements showed that changing the cultivation parameters resulted in production of enzymes with a large variation in performance, with values covering the range of 2% conversion to 92% conversion (figure 2). Similarly, the LC-MS/MS analysis showed variation in the profile of secreted proteins during the different cultivations. Many of the secreted proteins are enzymes that are known to act during hydrolysis of cellulose and hemicellulose, see especially table 3, but there are possibly other secreted proteins that also influence hydrolysis performance, e.g. enzymes belonging to table 4. However, in order to determine if there is a correlation between the protein profile and the corresponding performance of the enzymes on PCS, PLS was applied. More specifically, the PLS regression identifies the proteins where the changes in concentration correlate to changes in performance. For the proteins that are present in excess adding more of these will not increase performance.

Using PLS regression it was possible to create a model for prediction of hydrolysis performance on PCS based on the composition of the protein profile. 69.80% of the variance was explained by the first latent variable and only 10.11% was explained by the second latent variable, see figure 4. Therefore only one latent variable was chosen for this model. The proteins (original variables) with the highest positive loading are the proteins where the concentration is most correlated with hydrolysis performance and could be an indication that these proteins are limiting or decisive factors in the PCS hydrolysis. The cutoff value for defining whether a loading is high or low is strictly intuitive and in this case the cutoff was chosen to include 12 proteins shown in table 5.

The function of xylanase (id 69276) is not completely known but the enzyme is suspected to have both endoxylanase and endoglucanase activities. The function of Cip1 is so far unknown and therefore it is difficult to interpret its potential role in hydrolysis, but the model results suggest that this protein is important for hydrolysis of PCS. Cip2 is known to hydrolyze the linkages between lignocellulose and hemicellulose and therefore could also be relevant for PCS. The enzymes Man5A and Cel74A are involved in hemicellulose hydrolysis, although Cel74A has previously been described as an endoglucanase, indicating that this enzyme has activities towards hemicellulose and cellulose. Cellobiohydrolase Cel6A and endoglucanase Cel5A are known to be involved in cellulose hydrolysis and are usually reported to be among the four enzymes secreted in highest levels (Goyal et al., 1991; Tolan and Foody, 1999). The swollenin protein is believed to be involved in amorphogenesis as discussed earlier. Cel3A (β-glucosidase) was already shown to be limiting in the hydrolysis assay, but it is a nice

validation of the model that it also appears as one of the limiting enzymes. The β-glucosidase is important for removing cellobiose, which is inhibiting to cellobiohydrolase and endoglucanase.
In theory, adding more of the high ranking proteins could boost the hydrolysis performance on PCS. Cel6A and Cel5A might be considered candidates for improving hydrolysis performance since their importance in cellulose hydrolysis has already been established in the literature. However, many of the proteins listed in table 5 would not necessarily have been identified as being important for PCS hydrolysis from a standard gel electrophoresis/LC-MS/MS analysis. This is mainly because these proteins are produced in relatively small amounts and therefore it is difficult to determine how small a change in production level is significant for hydrolysis performance. It could be argued that in the case of low levels of enzyme there is the possibility that a significant amount of noise exists in the MS output data. The significance of noise could be examined by performing the SDS-PAGE and LC-MS/MS analysis several times on each sample. Furthermore, the mass spectrometry analysis does not distinguish between active and inactive forms of the same enzyme if they produce fragments with the correct mass. If there exists inactive enzymes in some of the samples this can lead to misleading distribution values in the X matrix and affect the PLS model towards inaccurate or wrong predictions. It is thus recommended that in future work the PLS model(s) should be validated by adding single components of the proteins listed in table 5 and measure how that affects hydrolysis performance on PCS. The production of pure samples of the proteins to be added is a major undertaking and is beyond the scope of this study and should be pursued in further work.
In a different fungal strain, e.g. a strain that produces higher levels of these proteins, it is speculated that other enzymes would then become the limiting factors in efficient hydrolysis of PCS. It is also likely that using a different substrate instead of PCS would identify other proteins as being important for the hydrolysis because the amount of cellulose and hemicellulose varies with the origin of the substrate and the pretreatment method.

Conclusion

It was found not only that protein levels were influenced by the fermentation conditions, but the expression *profile* was influenced as well. This profile had a profound effect on performance of the enzyme mixtures on NREL PCS. Based on the mas spectrum it was not obvious which proteins were specifically responsible for these differences in performance, only with a multivariate approach, PLS, could any sense be made of the mass spectra and it's correlation to performance in hydrolysis. From the PLS model it was possible to identify proteins that most likely are limiting for hydrolysis of PCS. Whether addition of these proteins will indeed increase performance or their exact mechanistic role is a subject for further studies. PLS represents a powerful tool for identification of proteins important for lignocellulosic biomass hydrolysis and could be of great value for industrial development of cellulase mixtures, as well as elucidating the mechanistic roles of various enzyme classes in the recalcitrant problem of lignocellulose hydrolysis.

Acknowledgements
This work was peformed during the PhD of the main author. The work was funded by a PhD stipend from the Novozymes Bioprocess Academy and the authors declare that there are no competing interests. KC McFarland and David Osborn are acknowledged for their excellent technical support for the PCS hydrolysis assay.

References

Arantes V, Saddler J. 2010. Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. Biotechnol. for Biofuels 3:4.
Askolin S. 2006. Characterization of the Trichoderma reesei hydrophobins HFBI and HFBII. Helsinki University of Technology.
Askolin S, Penttilä M, Wösten HA, Nakari-Setälä T. 2005. The Trichoderma reesei hydrophobin genes hfb1 and hfb2 have diverse functions in fungal development. FEMS Microbiol Lett 253(2):281-288.
Chanzy H, Henrissat B. 1985. Undirectional degradation of valonia cellulose microcrystals subjected to cellulase action. FEBS Lett 184(2):285-288.
Conesa A, Punt PJ, van Luijk N, van den Hondel CAMJJ. 2001. The Secretion Pathway in Filamentous Fungi: A Biotechnological View. Fungal Genet Biol 33(3):155-171.
Eriksson L, Antti H, Gottfries J, Holmes E, Johansson E, Lindgren F, Long I, Lundstedt T, Trygg J, Wold, S. 2004. Using chemometrics for navigating in the large data sets of genomics, proteomics, and metabonomics (gpm). Anal Bioanal Chem 380(3):419-429.

Fägerstam LG, Pettersson L. 1980. The 1.4-beta-glucan cellobiohydrolases of Trichoderma reesei QM 9414: A new type of cellulolytic synergism. FEBS Lett 119(1):97-100.
Ghose TK. 1987. Measurement of cellulase activities. Pure Appl Chem 59:257-268.
Goyal A, Ghosh B, Eveleigh D. 1991. Characteristics of fungal cellulases. Bioresour Technol 36(1):37-50.
Harris PV, Welner D, McFarland KC, Re E, Poulsen J-CN, Brown K, Salbo R, Ding H, Vlasenko E, Merino S, Xu F, Cherry J, Larsen S, Leggio LL. 2010. Stimulation of Lignocellulosic Biomass Hydrolysis by Proteins of Glycoside Hydrolase Family 61: Structure and Function of a Large, Enigmatic Family. Biochemistry 49(15):3305-3316.
Helle SS, Duff SJB, Cooper DG. 1993. Effect of surfactants on Cellulose Hydrolysis. Biotechnol Bioeng 42(5):611-617.
Henrissat B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J 280(2):309-316.
Herpoel-Gimbert I, Margeot A, Dolla A, Jan G, Molle D, Lignon S, Mathis H, Sigoillot JC, Monot F, Asther M. 2008. Comparative secretome analyses of two Trichoderma reesei RUT-C30 and CL847 hypersecretory strains. Biotechnol for Biofuels 1(1):18.
Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M. 2005. Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein. Mol and Cell Proteomics 4(9):1265-1272.
Kapp E, Schütz F. 2007. Overview of Tandem Mass Spectrometry (MS/MS) Database Search Algorithms. Curr Protoc in Protein Science 49:25.2.1-25.2.19.
Li X-L, Spániková S, de Vries RP, Biely P. 2007. Identification of genes encoding microbial glucuronoyl esterases. FEBS Lett 581(21):4029-4035.
Mach-Aigner AR, Pucher ME, Steiger MG, Bauer GE, Preis SJ, Mach RL. 2008. Transcriptional Regulation of xyr1, Encoding the Main Regulator of the Xylanolytic and Cellulolytic Enzyme System in Hypocrea jecorina. Appl Environ Microbiol 74(21):6554-6562.
Mansfield SD, Mooney C, Saddler JN. 1999. Substrate and Enzyme Characteristics that Limit Cellulose Hydrolysis. Biotechnol Prog 15(5):804-816.
Martens H, Næs T. 1989. Multivariate Calibration. John Wiley & Sons Ltd.
Nagendran S, Hallen-Adams HE, Paper JM, Aslam N, Walton JD. 2009. Reduced genomic potential for secreted plant cell-wall-degrading enzymes in the ectomycorrhizal fungus Amanita bisporigera, based on the secretome of Trichoderma reesei. Fungal Genet Biol 46(5):427-435.
Okino S, Ikeo M, Ueno Y, Taneda D. 2013. Effects of Tween 80 on cellulase stability under agitated conditions. Bioresour Technol 142:535-539.
Saloheimo M, Paloheimo M, Hakola S, Pere J, Swanson B, Nyyssönen E, Ward M, Penttilä M. 2002. Swollenin, a Trichoderma reesei protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials. FEBS J 269(17):4202-4211.
Tolan J, Foody B.1999. Cellulase from Submerged Fermentation. In: Tsao GT, Brainard A, Bungay H, Cao N, Cen P, Chen Z, editor. Recent Progress in Bioconversion of Lignocellulosics. Springer Berlin / Heidelberg 65:p41-67.
Vinzant, T, Adney W, Decker S, Baker J, Kinter M, Sherman N, Fox JW, Himmel ME. 2001. Fingerprinting Trichoderma reesei hydrolases in a commercial cellulase preparation. Appl Biochem Biotechnol 91-93(1):99-107.
Wold S, Sjöström M, Eriksson L. 2001. PLS-regression: a basic tool of chemometrics. Chemom Intell Lab Syst 58(2):109-130.
Xiao Z, Storms R, Tsang A. 2004. Microplate-based filter paper assay to measure total cellulase activity. Biotechnol Bioeng 88:832-837.
Zhang Y-HP, Lynd LR. 2004. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems. Biotechnol Bioeng 88(7):797-824.

Table 1 –Fermentation conditions

Overview of the settings of pH, temperature and agitation (rpm) in the 12 different batch cultivations. In experiments F-I Tween 80 was not added.

Experiment	pH	˚C	rpm
A	3.0	23	400
B	3.0	28	400
C	3.0	33	400
D	4.5	23	400
E	4.5	28	400
F	4.5	23	400
G	4.5	23	200
H	4.5	23	600
I	4.5	23	800
J	6.0	23	400
K	6.0	28	400
L	6.0	33	400

Table 2 – Fermentation results

Summary of protein concentration and filter paper activity obtained in the 12 different batch cultivations.

Experiment	pH	˚C	rpm	Protein (g/L)	Filter paper activity (FPU/ml)
A	3.0	23	400	4.91	3.05
B	3.0	28	400	4.59	2.08
C	3.0	33	400	1.29	0.14
D	4.5	23	400	8.02	9.45
E	4.5	28	400	3.85	2.47
F	4.5	23	400	14.73	7.36
G	4.5	23	200	2.86	0.39
H	4.5	23	600	9.79	6.67
I	4.5	23	800	11.22	4.96
J	6.0	23	400	3.41	2.03
K	6.0	28	400	9.40	5.93
L	6.0	33	400	3.37	2.10

Table 3 – Proteins assigned to CAZy GH and CE families

Overview of the different proteins assigned to GH or CE families, identified by LC-MS/MS in the filtered fermentation broth from the cultivations given in table 1. Numbers in the table refer to the relative abundance (percentage) of the individual protein in each sample. *T. reesei* protein number refers to the protein id from the JGI website http://genome.jgi-psf.org/Trire2/Trire2.home.html. Where possible, a more detailed function description and the enzyme name(s) have been given.

				Cultivation type											
				A	B	C	D	E	F	G	H	I	J	K	L
T. reesei protein number	Function	Name(s)	CAZy family	400 rpm	400 rpm	400 rpm	400 rpm	400 rpm	400 rpm	200 rpm	600 rpm	800 rpm	400 rpm	400 rpm	400 rpm
				23˚C	28˚C	33˚C	23˚C	28˚C	23˚C	23˚C	23˚C	23˚C	23˚C	28˚C	33˚C
				pH 3.0	pH 3.0	pH 3.0	pH 4.5	pH 4.5	pH 4.5	pH 4.5	pH 4.5	pH 4.5	pH 6.0	pH 6.0	pH 6.0
123989	Cellobiohydrolase	Cel7A (Cbh1)	GH 7	53.21	49.74	11.99	0.92	1.13	9.42	15.43	0.45	9.23	0.68	1.41	2.51
72567	Cellobiohydrolase	Cel6A (Cbh2)	GH 6	11.85	6.35	0.54	14.34	15.38	10.22	7.43	11.78	10.21	6.93	2.49	0.58
122081	Endoglucanase	Cel7B (Egl1)	GH 7	17.88	18.14	22.02									
120312	Endoglucanase	Cel5A (Egl2)	GH 5	6.72	8.97	1.37	33.9	14.74	23.64	3.61	9.05	19.5	8.03	1.98	15.16
123232	Endoglucanase	Cel12A (Egl3)	GH 12				0.43		5.4	0.46	0.76	1.29	3.34	0.9	3.8
73643	Endoglucanase	Cel61A (Egl4)	GH 61/ AA 9				0.29	1.3						0.08	
49976	Endoglucanase	Cel45A (Egl5)	GH 45	0.68			0.69								
120961	Endoglucanase	Cel61b	GH 61/				0.89		4.82	2.03	0.52		13.05	0.8	

			AA 9												
76672	beta-glucosidase	Cel3A (Bgl1)	GH 3				1.91		2.04	1.54	1.11	0.52	2.62	0.15	
66792	beta-glycosidase		GH 17											0.05	1.14
74223	Xylanase	Xyn1	GH 11	2.14	2.02	14.71		1.66				1.32			
123818	Xylanase	Xyn2	GH 11				0.54	2.1	3.89	4.1	0.99	1.35	3.36	0.38	6.43
120229	Xylanase	Xyn3	GH 10				0.6		2.94				20.03	1.37	
111849	Xylanase	Xyn4	GH 30				3.09			1.23	0.39	1.78	1.33	0.36	0.31
49081	Xyloglucanase	Cel74A (Egl6)	GH 74				4.19	4.46		2.35	1.78	10.93	1.46	0.93	0.69
121127	beta-xylosidase	Bxl1	GH 3	0.83	2.13	15.44	4.85			0.88	0.71	6.05	0.29	0.69	0.39
56996	beta-mannase	Man5A (Man1)	GH 5	0.73	1.12		4.68	2.89	4.63	3.02	1.38	4.39	1.5	0.41	2.13
72526	alpha-glucuronidase	Glr1	GH 67											0.15	
123283	Arabinosidase	Abf1	GH 54										0.28	0.05	
55319	Arabinosidase	Abf1-like	GH 54												1.11
76210	Arabinosidase	Abf2	GH 62						0.82	1.09			0.43		
72704	alpha-galactosidase	Agl3	GH 27		1.03	1.53									
22914	beta-1,3-glucanosyl-transglycosylase		GH 72										0.31		0.28
82235	alpha-glucosidase		GH 31												0.67
124175	beta-1,3-glucanase		GH 64												0.42
121746	beta-1,3-												0.1		

	glucanase		GH 55										7		
65406	Glycoside hydrolase		GH 16					0.21					0.24	0.03	
69276	Xylanase		GH 30				2.82	3.14	4.87	1.05	1.64	4.76	0.79	0.59	3.47
82633	beta-glucanosyltransferase		GH 2		0.44	1.24		0.77							
54219	Acetylxylan esterase		CE5										0.29		
73632	Acetylxylan esterase	Axe1	CE5	0.54		3.16	0.99	1.55		1.23	0.31		0.99		
121418	Acetyl esterase	Aes1	CE16		1.72										2.21
123940	Glucuronoyl esterase	Cip2	CE15				1.04	1.29	0.56	0.25	0.81	1.09	0.56	0.21	0.59

Table 4 – Proteins assigned to non-CAZy families or unknown

Overview of the proteins assigned to non-CAZy families or with unknown function identified by LC-MS/MS in the filtered fermentation broth from the cultivations given in table 1. Numbers in the table refer to the relative abundance (percentage) of the individual protein in each sample. *T. reesei* protein number refers to the protein id from the JGI website http://genome.jgi-psf.org/Trire2/Trire2.home.html. Where possible, a more detailed function description and the enzyme name(s) have been given.

		Cultivation type											
		A	B	C	D	E	F	G	H	I	J	K	L
T. reesei protein number	Function	400 rpm	400 rpm	400 rpm	400 rpm	400 rpm	400 rpm	200 rpm	600 rpm	800 rpm	400 rpm	400 rpm	400 rpm
		23°C	28°C	33°C	23°C	28°C	23°C	23°C	23°C	23°C	23°C	28°C	33°C
		pH 3.0	pH 3.0	pH 3.0	pH 4.5	pH 4.5	pH 4.5	pH 4.5	pH 4.5	pH 4.5	pH 6.0	pH 6.0	pH 6.0
50323	Ooc1 (Cellulose signalling protein)										1.97	0.19	
68067	Cell wall protein			3.42		1.59		1.65				0.12	1.54
73638	Cip1				3.08	2.76	2.17	0.77	0.58	2.93	0.81	0.27	1.59
76155	Acid phosphatase										0.14		
119989	Hfb2 (Hydrophobin class II)				1.38	1.70	15.38		56.00	16.66	2.55	83.55	36.54
122920	Bip1 (Hsp70 family)										0.35		3.33
123967	Hfb3 (Hydrophobin class II)											0.19	
123992	Swo1 (Swollenin, expansin)				19.38	36.80	9.21	13.81	11.06	7.71	11.07	1.57	0.75
22459	Carboxypeptidase A										0.30		
47127	Peptidase M18										0.12		
51365	Peptidase S8 and S53							0.53				0.06	0.79
81004	Peptidase A1, pepsin										1.28	0.11	
81070	Peptidase M28										0.19		
82623	Peptidase S8 and S53	2.14	1.10	1.42						0.29			
123244	Peptidase S8 and S53										2.11	0.05	
55887	Unknown										1.24	0.10	
70840	Unknown	3.28	7.25	22.79		5.97		30.86	0.69				
71092	Unknown												0.20
74807	Unknown										0.88		
80980	Unknown										0.75		0.97
81517	Unknown										0.28	0.06	0.36
104461	Unknown												0.45
119568	Unknown			0.36		0.55							
122127	Unknown							6.70			7.32	0.40	1.60
122374	Unknown										0.56		
123236	Unknown											0.29	
124259	Unknown										1.42		

Table 5 – VIP list of proteins limitng for hydrolysis

Top ranking proteins according to the loading values predicted by the PLS model.

T. reesei protein number	Function	Ranking number
69276	Xylanase	1
73638	Cip1	2
123940	Glucuronoyl esterase Cip2	3
56996	beta-mannase Man5A (Man1)	4
120312	Endoglucanase Cel5A (Egl2)	5
123818	Xylanase Xyn2	6
123232	Endoglucanase Cel12A (Egl3)	7
123992	Swo1 (Swollenin, expansin)	8
76672	beta-glucosidase Cel3A (Bgl1)	9
72567	Cellobiohydrolase Cel6A (Cbh2)	10
49081	Xyloglucanase Cel74A (Egl6)	11
111849	Xylanase Xyn4	12

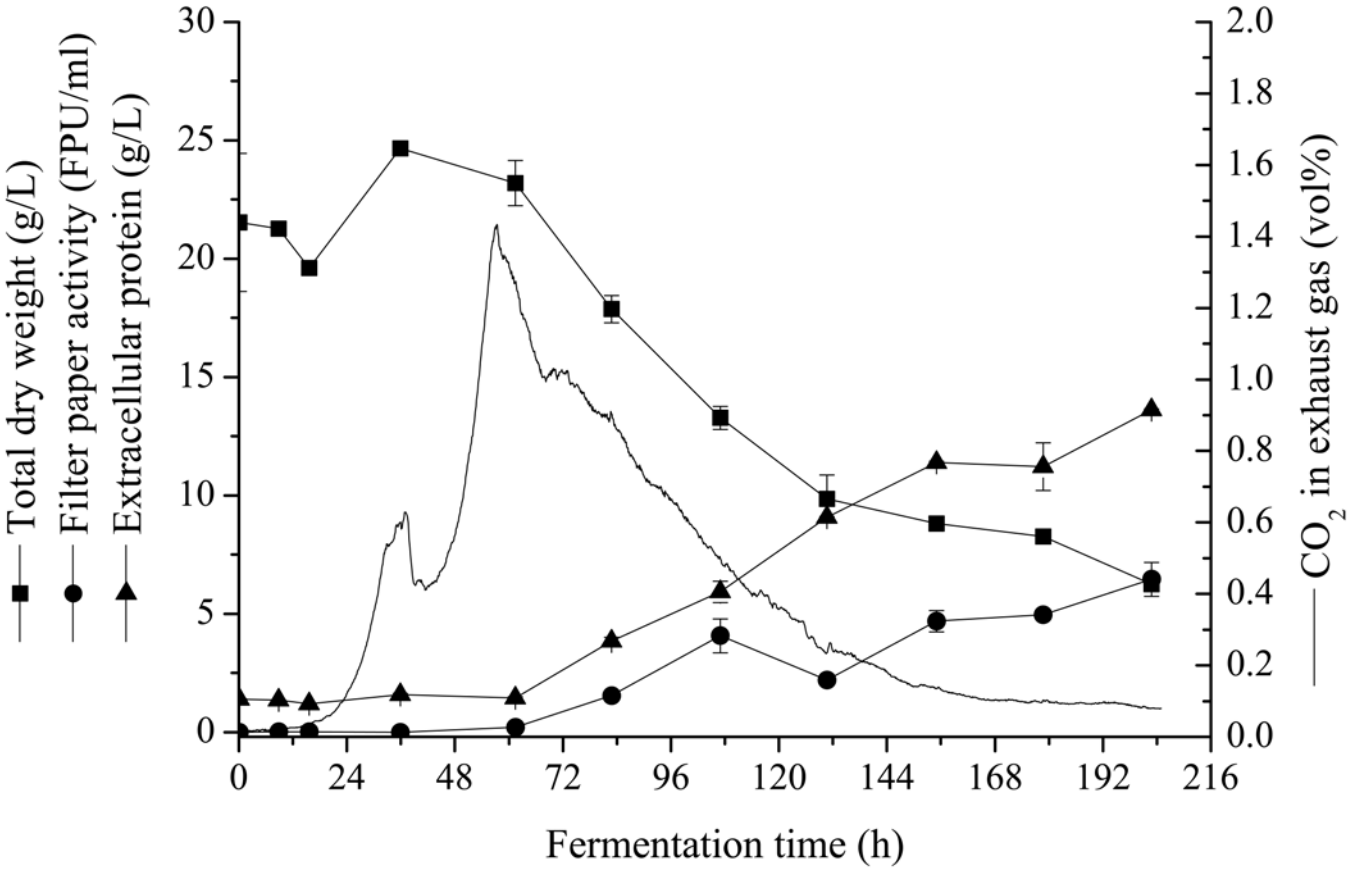

Figure 1

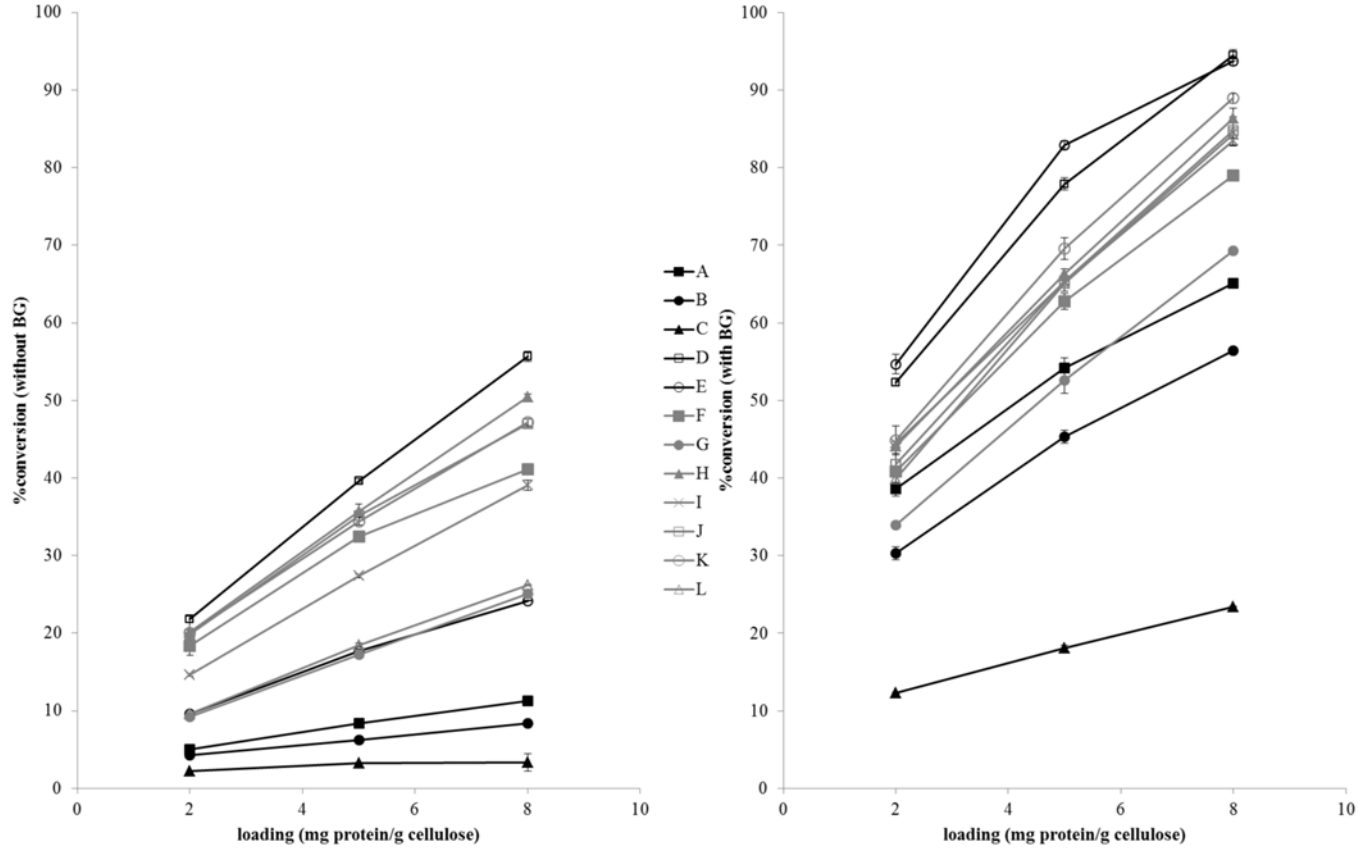

Figure 2

X	Endoglucanase	beta-mannase		Acetylxylan esterase	Xylanase	added beta-glucosidase
A 2mg/g	0.014	0.015	.....	0.011	0.043	0
A 5mg/g	0.034	0.036	.....	0.027	0.107	0
A 8mg/g	0.055	0.058	.....	0.043	0.171	0
B 2mg/g	0.000	0.022	.....	0.000	0.040	0
B 5mg/g	0.000	0.056	.....	0.000	0.101	0
B 8mg/g	0.000	0.090	.....	0.000	0.161	0
C 2mg/g	0.000	0.000	.....	0.063	0.294	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
J 8mg/g	0.000	0.120	.....	0.079	0.000	1
K 2mg/g	0.000	0.008	.....	0.000	0.000	1
K 5mg/g	0.000	0.021	.....	0.000	0.000	1
K 8mg/g	0.000	0.033	.....	0.000	0.000	1
L 2mg/g	0.000	0.043	.....	0.000	0.000	1
L 5mg/g	0.000	0.107	.....	0.000	0.000	1
L 8mg/g	0.000	0.171	.....	0.000	0.000	1

Y	% conversion 72h
A 2mg/g	5.03
A 5mg/g	8.43
A 8mg/g	11.31
B 2mg/g	4.30
B 5mg/g	6.27
B 8mg/g	8.42
C 2mg/g	2.24
⋮	⋮
J 8mg/g	84.74
K 2mg/g	44.88
K 5mg/g	69.55
K 8mg/g	88.93
L 2mg/g	39.88
L 5mg/g	65.11
L 8mg/g	84.33

Figure 3

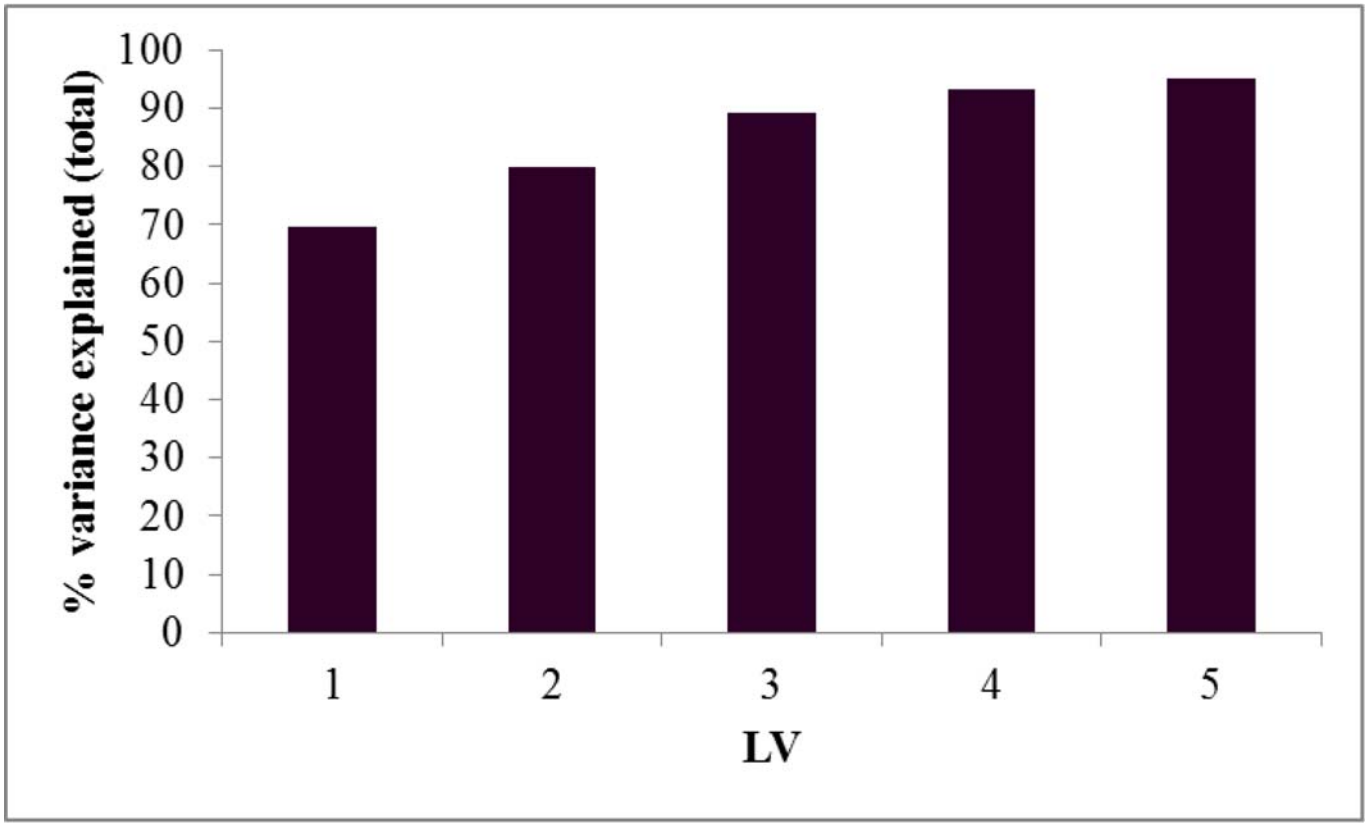

Figure 4