Protein Screening

DOI: 10.1002/anie.201104685

High-Throughput In Vitro Glycoside Hydrolase (HIGH) Screening for Enzyme Discovery**

Tae-Wan Kim, Harshal A. Chokhawala, Matthias Hess, Craig M. Dana, Zachary Baer, Alexander Sczyrba, Edward M. Rubin, Harvey W. Blanch, and Douglas S. Clark**

There has recently been a renewed interest in converting the carbohydrate content (cellulose and hemicellulose) of lignocellulosic biomass to biofuels and biochemicals.[1] This conversion requires depolymerization of the carbohydrate content of biomass to fermentable monosaccharides by using glycoside hydrolases (GHs). Despite the vast number of GHs deposited in publicly available databases,[2] 90% of these enzymes are not characterized and many are unlikely to be suitable for industrial applications because of low activity and stability under processing conditions.[3]

While protein engineering (by using rational strategies as well as directed evolution) has been used to evolve GHs for industrial applications,[4] bioprospecting and metagenomics provide alternative approaches to finding GHs with desired properties.[5,6] The identification of new GHs from large genetic inventories relies on efficient protein expression and rapid high-throughput screening or selection to evaluate large enzyme libraries. However, heterologous expression of GHs (especially cellulases) in microbial hosts is usually very difficult, and often results in misfolded and/or inactive enzymes or non-native enzyme structures.[7] Furthermore, the poor correlation between the activity of many GHs on soluble and insoluble substrates necessitates the high-throughput screening to be carried out on insoluble lignocellulosic substrates.[8] Herein, we present the development of a one-pot high-throughput in vitro glycoside hydrolase (HIGH) expression and screening method by using insoluble lignocellulosic substrates. The HIGH platform can be adapted to screen for cellulases, xylanases, amylases, and β-glucosidases.

[*] Dr. T. W. Kim, Dr. H. A. Chokhawala, C. M. Dana, Z. Baer, Prof. H. W. Blanch, Prof. D. S. Clark
Energy Biosciences Institute, University of California
Berkeley, CA 94720 (USA)
E-mail: blanch@berkeley.edu
clark@berkeley.edu
Dr. M. Hess
School of Molecular Biosciences, Applied Microbial Genomics and Ecology, Washington State University
Richland, WA 99353-1671 (USA)
Dr. A. Sczyrba, Prof. E. M. Rubin
Genomics Division, Lawrence Berkeley National Laboratory
Berkeley, CA 94720 (USA)
and
Department of Energy, Joint Genome Institute
Walnut Creek, CA 94598 (USA)

[**] The research described in this article was funded by the Energy Biosciences Institute.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201104685.

Cell-free protein expression is a simple, high-throughput methodology for synthesizing functional proteins. The cell-free system used in the present work to express GHs is based on *E. coli* crude cell lysate, which contains the translational machinery, and endogenous *E. coli* proteins related to central metabolism, which includes glycolysis,[9] oxidative phosphorylation,[10] and amino acid metabolism.[11] In order to reduce the overall reagent costs, the system employs glucose as an energy source to regenerate adenosine triphosphate (ATP) through glycolysis instead of a substrate-level phosphorylation, which uses expensive high-energy phosphate-based compounds (e.g., creatine phosphate, phosphoenolpyruvate, or acetyl phosphate).[9] Although in vitro expression of *E. coli* is not suitable for all proteins, because the gene cannot be expressed in this procaryotic system and/or posttranslational modification is required, alternative systems for such cases, including wheat germ, are available.

The HIGH screening method for detecting glycoside hydrolase activity couples in vitro expression with glycan hydrolysis in one pot. As shown in Figure 1a, cell-free protein synthesis is initiated with a small amount of sugar (e.g., glucose or xylose) as the primary energy source. If the resulting cell-free-synthesized GH enzyme is active, it hydrolyzes the added glycan substrate in the same pot, thus releasing more sugar, which can in turn act as the energy reservoir for ATP regeneration and the subsequent cell-free synthesis of additional enzyme. The sugar, which is released by an active cell-free-synthesized GH, will be used as an additional energy source to extend cell-free protein synthesis (Figure 1b, top), and is also converted into acidic by-products (e.g., lactate, acetate, or formate), which result in acidification of the cell-free mixture (Figure 1b, bottom). Hence, an active enzyme that catalyzes the hydrolysis of the substrate will effect a decrease in pH value, which can be readily detected by using colorimetric pH indicators. The decrease in pH value will also increase the activity of many GHs, which typically have optimal activity between pH 4.0 and 6.5.[6] If the enzyme is inactive against the substrate tested, no additional sugar will be released, and protein production will cease with little change in pH value (Figure 1b, blue lines). Compared with conventional screening methods, which employ cell-based protein expression and soluble substrates, the HIGH screening method provides a simple (single instead of multiple steps), rapid (hours instead of days or weeks), reliable (solid substrate instead of soluble substrate), and universal (various carbohydrates) means to either identify GH enzymes in environmental samples or engineer GH enzymes with altered activity (Figure 1c).

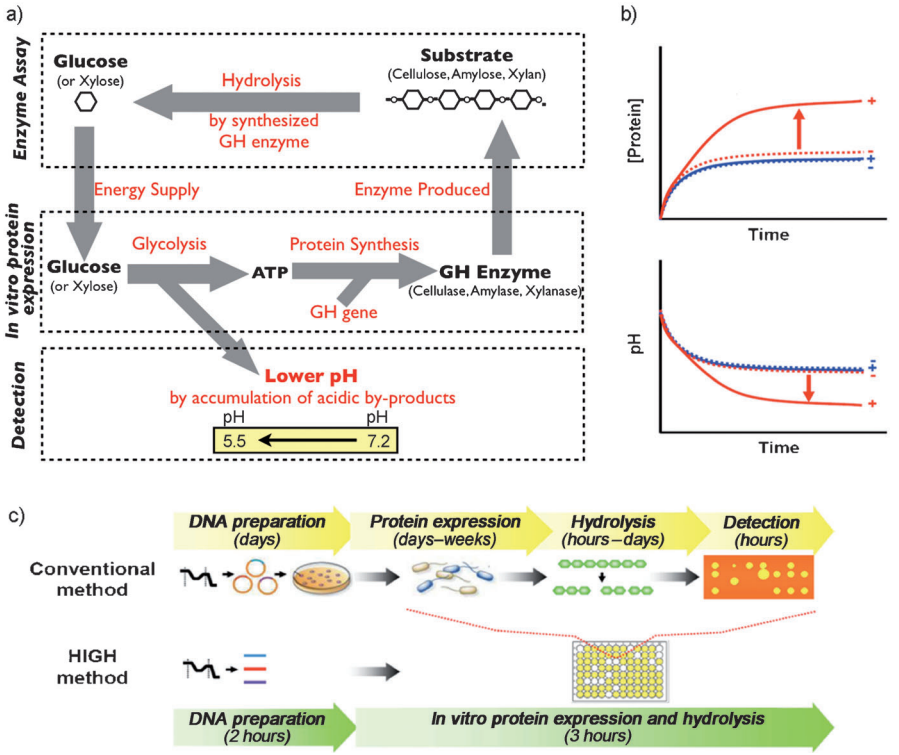

Figure 1. Schematic representation of the high-throughput in vitro glycoside hydrolase (HIGH) screening method. a) A small amount of added sugar (glucose or xylose) regenerates ATP through glycolysis, thus initiating in vitro protein synthesis (in vitro protein expression, middle box). Synthesized glycoside hydrolase (GH) enzyme hydrolyzes the substrate and produces monomeric sugar (enzyme assay, top box). The released sugar is used as an additional energy source to synthesize more GH enzyme and is converted into acidic by-products by central metabolism of *E. coli*, thus resulting in a decrease in the pH value from neutral (yellow) to acidic (colorless; detection, bottom box). b) Representative signal amplification coupled with GH activity. The non-GH (blue) and GH (red) genes were expressed either in the absence (dotted line) or presence (solid line) of substrate in a cell-free system. The yield of synthesized protein increased and the pH value decreased in the reaction that expresses the GH gene with substrate (+) relative to the reaction that expresses GH gene without substrate (−). Reactions expressing a non-GH gene either with substrate (+) or without substrate (−) showed no difference in the yield of protein or pH value. c) Workflow comparison of the conventional screening method and the HIGH screening method. In the conventional screening method, cloning, protein expression, and enzyme hydrolysis are completely separated; HIGH screening method combines all processes in one pot.

Cell extracts that are prepared from cells grown on glucose cannot use xylose as an energy source because of catabolic repression, thus restricting the applicability of the extract to screening for GHs that release glucose or cellobiose as products (cellulases, α/β-glucosidases, and amylases). Furthermore, a cell extract prepared from cells grown on xylose showed only marginal protein-expression ability when xylose was used as the sole energy source (data not shown). Thus, a novel xylose-utilizing cell-free system was developed for xylanase screening on hemicellulose, which is the second largest component in biomass. Xylulose-5-phosphate (Xu-5-P), an intermediate in the pentose phosphate (PP) pathway, was tested to determine if it could serve to regenerate ATP through the PP pathway. Interestingly, Xu-5-P could be used as an energy source to synthesize chloramphenicol acetyltransferase (CAT, from *E. coli*) as a model protein (see Supplementary Figure 1 a in the Supporting Information, red bar). Hence, the conversion of xylose to xylulose-5-phosphate is the limiting step in the use of xylose as an energy source in the current cell-free system. To overcome this limitation, two enzymes, xylose isomerase and xylulose kinase, were added to a cell-free reaction. As shown in Supplementary Figure 1 a (blue bar), the cell-free reaction using xylose produced more CAT protein (950 $\mu g\,mL^{-1}$) relative to a glucose-utilizing reaction (749 $\mu g\,mL^{-1}$), albeit with a lower ATP regeneration efficiency (Supplementary Figure 1 b). In addition, most sugars, such as glucose or xylose, were converted into organic acids, primarily lactic acid, but also formic acid and acetic acid (Supplementary Figure 1 c,d). Therefore, the reduction in pH value that occurs is due to conversion of the sugar into acidic products, and can be measured by using pH indicators.

Of the three pH indicators tested between pH 7.2 and 5.5, which is the range observed in cell-free reactions with 0 to 40 mM glucose or xylose, *p*-nitrophenol showed the highest sensitivity (slope = 1.29, $R^2 = 0.993$) and correlation coefficient (Supplementary Figure 2 a), and can thus be used as a quantitative indicator of sugar released by hydrolysis (Supplementary Figure 2 b). Therefore, a color change from yellow to colorless (decrease in the A_{405nm} value) in a reaction mixture that expresses GH genes is indicative of acid formation and corresponding hydrolytic activity.

Four different GHs were expressed to validate the HIGH screening method for quantifying GH activity. The CAT gene was used as a negative control in all experiments. First, the cell-free reaction was carried out with either the CAT gene or β-glucosidase gene (from *Pyrococcus furiosus*, UniProt, Q51723) in the presence and absence of cellobiose (15 mM). Cell-free protein synthesis of CAT in either the presence or absence of cellobiose, as well as β glucosidase in the absence of cellobiose, ceased after 60 min, whereas β-glucosidase synthesis continued for up to 150 min in the presence of cellobiose (Supplementary Figure 3 a). The absorbance of the reaction that expressed β glucosidase decreased substantially in just 3 h ($\Delta A_{405nm} = 1.267$ for β glucosidase and $\Delta A_{405nm} = 0.557$ for CAT, Supplementary Figure 3 b), thus corresponding to the release of glucose (24 mM) from cellobiose by cell-free-synthesized β glucosidase. More pronounced absorbance differences were obtained by using higher concentrations of substrate (Supplementary Figure 3 c).

HIGH screening for cellulases was tested toward soluble (cellohexaose) and insoluble (phosphoric acid swollen cellulose, PASC) substrates by using an endoglucanase from *Butyrivibrio fibrisolvens* (Uni*P*rot, P20847). A small amount of β glucosidase and purified *T. emersonii* Cel7A (see methods section in the Supporting Information) was added to help complete the hydrolysis of cellobiose, which is the main hydrolysis product of the endoglucanase reaction, into glucose in the cell-free reaction. This reaction, which contained *B. fibrisolvens* endoglucanase and cellohexaose (5 mM), resulted in a 65% decrease in absorbance in 3.5 h (12.5 mM glucose released) relative to the control reaction with CAT (Supplementary Figure 4a), thereby validating the cellulase activity of this enzyme.[12] Cell-free synthesis of the endoglucanase was carried out in the presence of 1% PASC (w/v; Supplementary Figure 4b), and resulted in the release of 17 mM glucose by PASC hydrolysis and a 42% decrease in absorbance over 3 h.

Finally, α amylase from *Bacillus amyloliquefaciens* (Uni-Prot, P00692) and xylanase from *Thermobifida fusca* (Uni-Prot, Q47 L48) were used as model enzymes with the solid substrates amylose (2%, w/v) and xylan (from birchwood, 2%, w/v), respectively. α Glucosidase (from *Bacillus stearothermophilis*) and β xylosidase (from *Bacillus pumilus*) were added to the reaction mixture for amylase and xylanase screening, respectively, to complete the hydrolysis of each substrate to monomeric sugars. After 3 h of incubation, the absorbances in reactions that express amylase and xylanase were 67% and 42% of the negative controls (Supplementary Figure 4c,d), which are equivalent to the generation of 14 mM glucose and 26 mM xylose, respectively. The protein concentration was measured to confirm the amplification of protein production. The protein concentrations increased 1.7-, 1.6-, and 3.9-fold for cellulase-PASC, amylase-amylose, and xylanase-xylan, respectively, relative to the reaction without substrates (Supplementary Figure 4e). The results validate the application of HIGH screening for different types of GH activities, including β-glucosidases, cellulases (endoglucanase), amylases, and xylanases with both soluble and insoluble substrates.

The HIGH screening method was also used to identify different GHs from a metagenomic library constructed from the cow rumen microbiome.[13] The hydrolytic activity of 82 randomly selected putative GH genes was tested. These genes were assembled from the metagenome of cow rumen and were predicted to contain a domain that belongs to one of the following glycoside hydrolase families: 3, 5, 8, 9, 10, 13, or 26 (Supplementary Table 1). These GH families are known to contain β-glucosidases (EC 3.2.1.21), endoglucanases (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91), amylases (EC 3.2.1.1), and endoxylanases (EC 3.2.1.8). HIGH screening, which uses PASC (0.6%, w/v) as the substrate, identified 8 cellulose-hydrolyzing enzymes from the library (Figure 2a and Supplementary Table 1). Xylanase screening with this method gave 16 xylan-hydrolyzing enzymes from the library (Figure 2c and Supplementary Table 1). β-Glucosidase and amylase screening of the same library did not give any active GHs against cellobiose or amylose (Supplementary Figure 5a,c), thus emphasizing the necessity and potential of combining computational sequence annotation with functional screening assays to improve current gene-annotation algorithms and enzyme-classification systems.

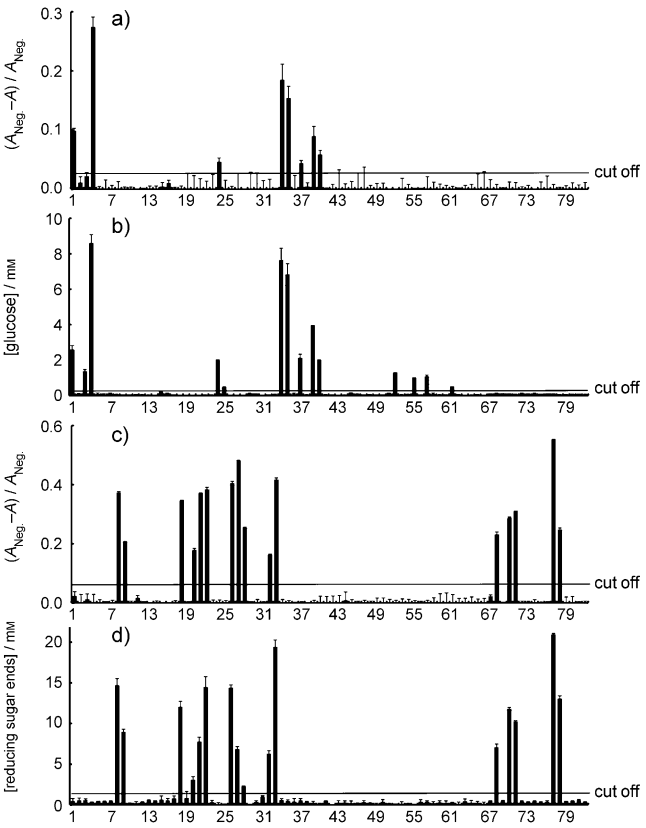

Figure 2. Comparison of a, b) cellulases and c, d) xylanases identified from a cow rumen metagenomic library by using a, c) the HIGH screening method and b, d) the conventional screening method. A 0.6% (w/v; HIGH method) or a 1% (w/v; conventional method) PASC suspension was used for cellulase screening and 2% (w/v) xylan was used for xylanase screening. The PCR products of putative GH genes identified from the cow metagenome library were directly used as the template for cell-free protein expression in a 96-well plate. For the HIGH method, the substrate was introduced into the cell-free mixture and the cell-free reaction was carried out at 37°C for 3 h. The absorbance was directly measured by adding *p*-nitrophenol after removing the solid substrate. The activity is displayed in relative units ($(A_{Neg}-A)/A_{Neg}$), in which A_{Neg} designates the absorbance of the CAT-expressing negative control, and A the absorbance of the sample. The range of activity is between 1 (maximum activity) and 0 (no activity). For the conventional method, cell-free protein expression was carried out without the substrate for 3 h. After 3 h incubation, cell-free-synthesized protein was mixed with the substrate in 50 mM acetate buffer (pH 5.6) for substrate hydrolysis. After 16 h at 37°C, the amount of released sugar was measured by using the glucose oxidase method and DNS assay for cellulase and xylanase activity, respectively. The cut-off value (gray dotted lines) was determined as the mean plus twofold average deviation of the negative controls. All experiments were performed in duplicate. Error bars represent one average deviation around the mean.

To compare the efficiency of the HIGH screening method with conventional screening approaches, the same 82 genes were expressed by using the cell-free expression platform previously described,[13] and assayed against PASC (1%, w/v), xylan (2%, w/v), amylose (2%, w/v), and cellobiose (15 mM),

as previously reported[14] (Figure 2b,d and Supplementary Figure 5b,d). The conventional assay method identified the same xylan-degrading enzymes as the HIGH screening method (Figure 2c). Neither method identified active amylases or β-glucosidases from the library (Supplementary Figure 5b,d). However, screening for cellulase activity by using the conventional method gave 14 active cellulases as compared to 8 found by using the HIGH screening method. The additional 6 cellulases detected were found to have very low activity (< 2 mM sugar released over 16 h by using the glucose oxidase/peroxidase assay), and hence were difficult to distinguish from noncellulolytic enzymes by using the HIGH screening method (Figure 2a). However, as shown in Supplementary Figure 6, the fluorescent pH indicator fluorescein showed greatly improved sensitivity (9 orders of magnitude) for glucose detection in HIGH screening with small amounts of glucose consumed (0.5 to ca. 5 mM). Therefore, addition of more substrate and use of a fluorescent pH indicator[14] should enable the HIGH screening method to better detect enzymes that have very low activity.

The HIGH screening method provides a rapid approach to discover active GHs in environmental samples. A large and growing number of GH-candidate genes are included in the Carbohydrate Active Enzyme (CAZy) database; current methods are limited in their ability to express and assay the corresponding enzymes, and less than 10% have been characterized to date.[2] The HIGH screening method affords several advantages over conventional methods to increase the throughput of functional screening for GHs. While cloning, protein expression, and enzyme hydrolysis are separated in the conventional methods, HIGH screening combines these steps in one pot. The entire process from gene expression to activity detection requires only three hours. In addition, HIGH screening provides a universal screening method for GH enzymes by using multiple substrates (cellobiose, cellulose, xylan, and amylose), including solids.

Received: July 6, 2011
Published online: September 16, 2011

Keywords: biomass · enzymes · glycoside hydrolases · high-throughput screening · protein expression

[1] A. J. Ragauskas, C. K. Williams, B. H. Davison, G. Britovsek, J. Cairney, C. A. Eckert, W. J. Frederick, Jr., J. P. Hallet, D. J. Leak, C. L. Liotta, J. R. Mielenz, R. Murphy, R. Templer, T. Tschaplinski, *Science* **2006**, *311*, 484 – 489.
[2] B. L. Cantarel, P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, B. Henrissat, *Nucleic Acids Res.* **2009**, *37*, D233 – 238.
[3] M. E. Himmel, M. F. Ruth, C. E. Wyman, *Curr. Opin. Biotechnol.* **1999**, *10*, 358 – 364.
[4] J. R. Cherry, A. L. Fidantsef, *Curr. Opin. Biotechnol.* **2003**, *14*, 438 – 443.
[5] L. L. Li, . S. R. McCorkle, S. Monchy, S. Taghavi, D. Van der Lelie, *Biotechnol. Biofuels* **2009**, *2*, 10.
[6] C. J. Yeoman, Y. Han, D. Dodd, C. M. Schroeder, R. I. Mackie, I. K. O. Cann, *Adv. Appl. Microbiol.* **2010**, *70*, 1 – 55.
[7] T. Jeoh, W. Michener, M. E. Himmel, S. R. Decker, W. S. Adney, *Biotechnol. Biofuels* **2008**, *1*, 10.
[8] J. K. McCarthy, A. Uzlec, D. F. Davis, D. E. Eveleigh, *J. Biol. Chem.* **2004**, *279*, 11495 – 11502.
[9] K. A. Calhoun, J. R. Swartz, *Biotechnol. Bioeng.* **2005**, *90*, 606 – 613.
[10] M. C. Jewett, K. A. Calhoun, A. Voloshin, J. J. Wuu, J. R. Swartz, *Mol. Syst. Biol.* **2008**, *4*, 220.
[11] K. A. Calhoun, J. R. Swartz, *J. Biotechnol.* **2006**, *123*, 193 – 203.
[12] T. W. Kim, H. A. Chokhawala, D. Nadler, H. W. Blanch, D. S. Clark, *Biotechnol. Bioeng.* **2010**, *107*, 601 – 611.
[13] M. Hess, A. Sczyrba, R. Egan, T. W. Kim, H. A. Chokhawala, G. Schroth, S. J. Luo, D. S. Clark, F. Cheng, T. Zhang, R. I. Mackie, L. A. Pennacchio, S. G. Tringe, A. Visel, T. Woyke, Z. Wang, E. M. Rubin, *Science* **2011**, *331*, 463 – 467.
[14] J. Han, K. Burgess, *Chem. Rev.* **2010**, *110*, 2709 – 2728.

