How molecular evolution technologies can provide bespoke industrial enzymes : application to biofuels.

Laurent Fourage, Jean-Marie Sonet, Frédéric Monot, Gilles Ravot, Antoine Margeot

D o s s i e r

Second and Third Generation Biofuels: Towards Sustainbility and Competitiveness
Seconde et troisième génération de biocarburants : développement durable et compétitivité

How Molecular Evolution Technologies can Provide Bespoke Industrial Enzymes: Application to Biofuels

L. Fourage[1], J.-M. Sonet[2]*, F. Monot[3], G. Ravot[4] and A. Margeot[3]

[1] TOTAL SA, Immeuble Lafayette, 2 place des Vosges, 92051 Paris La Défense 1 - France
[2] Protéus SA, 70 allée Graham Bell, Parc scientifique Georges Besse, 30035 Nîmes Cedex 1 - France
[3] IFP Energies nouvelles, 1-4 avenue de Bois-Préau, 92852 Rueil-Malmaison Cedex - France
[4] SAS PIVERT, Parc technologique des rives de l'Oise, rue des rives de l'Oise, BP 50149, Venette, 60201 Compiègne Cedex - France
e-mail: laurent.fourage@total.com - jmsonet@proteus.fr - frederic.monot@ifpen.fr - gravot@cegetel.net - antoine.margeot@ifpen.fr

* Corresponding author

Résumé — **Comment les technologies d'évolution moléculaire peuvent fournir des enzymes industrielles sur mesure : application aux biocarburants** — L'hydrolyse enzymatique de la lignocellulose est l'un des principaux goulets d'étranglement dans le développement de la conversion biologique de la biomasse lignocellulosique en biocarburants. L'un des organismes les plus efficaces pour la production d'enzymes cellulolytiques est le champignon *Trichoderma reesei*, principalement grâce à sa capacité importante de sécrétion. La conversion de la cellulose en glucose implique trois types de cellulases travaillant en synergie : les endoglucanases (EC 3.2.1.4) clivant de façon aléatoire les liaisons glycosidiques en β-1,4, les cellobiohydrolases (EC 3.2.1.91) attaquant la chaîne de cellulose aux deux extrémités afin de produire le cellobiose, dimère qui sera converti en glucose par l'action des β-glucosidases (EC 3.2.1.21). De façon inattendue, la quantité de β-glucosidase (BGL1) sécrétée par les souches de *T. reesei* représente un très faible pourcentage de la quantité totale des protéines sécrétées qui en fait donc une activité limitante du cocktail. Cette faible activité limite d'autant plus les performances du cocktail que le cellobiose représente le principal inhibiteur de la réaction cellulolyse par les cellobiohydrolases. Ce goulot d'étranglement peut être atténué soit par une surexpression de la β-glucosidase chez *T. reesei,* soit par une amélioration de son activité spécifique. Après un bref aperçu des principales technologies existantes, cet exemple sera utilisé dans cette revue pour illustrer le potentiel des technologies d'évolution dirigée pour développer des enzymes répondant aux besoins de l'industrie des biotechnologies. Nous décrivons comment la mise en œuvre d'une stratégie d'évolution dirigée par le L-Shuffling™ avec trois gènes parentaux provenant de la biodiversité microbienne permet d'obtenir des activités β-glucosidases très améliorées par rapport à la Cel3a β-glucosidase de *T. reesei* (activité spécifique 242 fois plus élevée pour le substrat pNPGlc). Cette amélioration de l'activité glucosidasique, après expression du gène codant pour la β-glucosidase améliorée chez *T. reesei* et sécrétion du nouveau cocktail, permet une diminution d'un facteur 4 de la quantité de cocktail nécessaire à la saccharification d'une biomasse industrielle prétraitée (paille de blé).

Abstract — ***How Molecular Evolution Technologies can Provide Bespoke Industrial Enzymes: Application to Biofuels*** — *Enzymatic hydrolysis of lignocellulose is one of the major bottlenecks in the development of biological conversion of lignocellulosic biomass to biofuels. One of the most efficient organisms for the production of cellulolytic enzymes is the fungus* Trichoderma reesei,

mainly thanks to its high secretion capacity. The conversion of cellulose to glucose involves three types of cellulases working in synergy: endoglucanases (EC 3.2.1.4) randomly cleave β-1,4 glyco-sidic linkages of cellulose, cellobiohydrolases (EC 3.2.1.91) attack cellulose chain ends to produce cellobiose dimers which are converted into glucose by the β-glucosidases (EC 3.2.1 21). Unexpect-edly, the amount of β-glucosidase (BGL1) from T. reesei *hyperproducing strains represents a very low percentage of the total secreted proteins. A suboptimal content of this enzyme limits the perfor-mance of commercial cellulase preparations as cellobiose represents the main inhibitor of the cellul-olysis reaction by cellobiohydrolases. This bottleneck can be alleviated either by overexpressing the β-glucosidase in* T. reesei *or optimized its specific activity. After giving a brief overview of the main available technologies, this example will be used to illustrate the potential of directed evolution tech-nologies to devolop enzymes tailored to fit industrial needs. We describe the L-Shuffling*^{TM} *strategy implemented with three parental genes originating from microbial biodiversity leading to identifica-tion of an efficient β-glucosidase showing a 242-fold increase in specific activity for the pNPGlc sub-strate compared to WT (Wild Type) Cel3a beta-glucosidase of* T. reesei. *After expression of the best improved β-glucosidase in* T. reesei *and secretion of a new enzymatic cocktail, improvement of the glucosidase activity allows a 4-fold decrease of cellulase loading for the saccharification of an industrial pretreated biomass compared to the parental cocktail.*

INTRODUCTION

In the emerging low carbon economy, enzymes have become key tools to develop sustainable and clean indus-trial processes in major sectors such as bioenergy and chemicals manufacturing. This article opens a room for exploiting molecular evolution approaches in order to design ideal cocktails of enzymes for the development of biological conversion of lignocellulosic biomass to bio-fuels. Indeed, lignocellulosic biomass is an abundant and cost-efficient substrate but its breakdown into ferment-able sugars remains one of the major challenges for the development of a sustainable biofuel production process.

The article will start with an overview of the context, the social and the economical issues of 2G ethanol pro-duction. Then, we will describe the directed evolution technologies that could be applied to overcome these issues and illustrate such approach by describing a L-Shuffling^{TM} strategy implemented with three parental genes originating from microbial biodiversity leading to identification of an efficient β-glucosidase showing a 242-fold increase in specific activity for pNPGlc sub-strate compared to WT Cel3a beta-glucosidase of *T. reesei*. We will then describe the potential of the new generated secretome after expression of the best improved β-glucosidase in *T. reesei*.

1 CONTEXT, SOCIAL AND ECONOMIC ISSUES

Bioethanol represents a major part of biofuels and is to date nearly exclusively produced from sugar and starch containing crops. However, this first generation bioetha-nol cannot meet all needs and increasing production may lead to competition for arable land and increase pressure on world market prices for cereals and basic crops for food production (Fairley, 2011).

Therefore, growing efforts are undertaken to develop second generation biofuels which use the lignocellulosic, non-edible parts of the plant as substrate for biofuel pro-duction. Using this biomass source is also leading to sub-stantially higher greenhouse gas emission savings than first generation bioethanol (60-80% savings when com-pared to CO_2 emissions from gasoline) (Edwards *et al.*, 2007). The recent new EU directive on renewable energy includes a mandatory market share of 10% for renew-able energy in the transport sector in 2020. The directive includes a minimum saving of 35% GHG (greenhouse gas) emissions for current biofuels and of 60% for pro-duction plants to be installed after 2017. These commit-ments pave the way for second generation biofuels (Directive 2009/28/EC on the promotion of the use of energy from renewable sources).

The classical second generation bioethanol production process comprises four major steps. First, a physico-chemical pretreatment, under acidic or alkaline condi-tions, destroys the rigid structure of the plant cell wall (for review, see Chundawat *et al.*, 2011). This material can then be submitted to enzymatic hydrolysis in order to generate fermentable sugars, mainly glucose. These sugars can be converted into ethanol by yeast fermenta-tion. Finally, the ethanol thus produced is recovered by distillation. The main issue for an industrial application of such process is its high cost compared to first genera-tion bioethanol. It is anticipated that for a competitive production, the 2G ethanol production cost must be divided at least by two.

One of the most important cost-contributors in bio-ethanol production is enzymatic hydrolysis, and many efforts are aimed at decreasing the cost of this step (Margeot *et al.*, 2009). Biotechnology has therefore become one of the key growth drivers for so-called "cellulosic bioethanol", also referred to as "2G bioethanol". Due to its capacity to produce large amounts of cellulose degrading enzymes, *T. reesei* has being extensively studied in various fields of white biotechnology, especially in biofuel production from lignocellulosic biomass (Kubicek *et al.*, 2009). In parallel of studies dedicated to improvement of cellulolytic cocktails, other process schemes allowing for production of 2G ethanol have been proposed in order to lower its cost. In one of these suggested processes, the hydrolysis and fermentation steps are carried out in a single step ("Simultaneous Saccharification and Fermentation", SSF, opposite to "Separate Hydrolysis and Fermentation", SHF) (Olofsson *et al.*, 2008). The advantages of such processes include the consumption by yeasts of sugars that would inhibit cellulases and the reduction of the risks of bacterial contamination during the fermentation step. However, inhibition of cellulases by ethanol and different temperature optima for yeast and cellulases remain problems to be solved. A promising alternative is the consolidated bioprocessing (CBP) which uses only a single organism expressing cellulolytic enzymes and fermenting the released sugars (Olson *et al.*, 2012). The latter concept has yet to reach industrial maturity.

Whatever the process, conversion of lignocellulosic renewable resources into biofuels creates a crucial need for second generation enzymes with optimized characteristics, fine-tuned properties and enhanced performance.

2 MOLECULAR EVOLUTION

Directed evolution of proteins offers different alternatives for optimizing enzyme characteristics and performance. Protein engineering has first been performed by using techniques based on rational design and structural approaches. However, randomized methods rapidly proved to have a much broader scope of application than rational design for rapidly optimizing enzymes and other proteins (Illanes *et al.*, 2012). Actually, randomized directed evolution technologies require no prior knowledge of the tridimensional structure of the target protein. The evolution of the enzymes is solely directed by the assays used for screening for variants with improved characteristics.

Therefore, randomized "directed evolution" technologies have had an increasing impact to address the need

for enzyme optimization. "Directed evolution" mirrors dramatically accelerates what has long been made when breeding plants and animals to produce new crops and livestock in order to fulfill the needs of mankind.

The successful outcome of enzyme directed evolution experiments depends on three major points which include:
– the potential of evolution of the enzyme(s) to be engineered (the so-called "parental" enzymes);
– the method chosen for generating the pool of next-generation enzymes;
– the screening method(s) applied for identifying the enzyme variants having appropriate characteristics).

Until the end of the 90's, mutagenesis was the most widely used tool for generating new variants of an enzyme. The most popular technique for introducing mutations along a gene sequence has long been error-prone PCR (Polymerase Chain Reaction) and is still widely employed (Song *et al.*, 2012). However, it has rapidly been observed that mutagenesis-based technologies suffer severe limits such as the relatively limited sequence space that can be rapidly explored, and a potential of evolution which is limited by the number of beneficial mutations that a gene can cumulate between two selection events. Some of these advanced methods have recently been extensively reviewed elsewhere (Dougherty and Arnold, 2009; Brustad and Arnold, 2011). In order to overcome the limits of classical mutagenesis-based approaches and optimize directed evolution protocols. This review will therefore develop in detail two major technologies developed by Proteus, one called EvoSight[TM] based on random mutagenesis and the other called L-Shuffling[TM] based on gene shuffling.

2.1 Random Mutagenesis

The random mutagenesis EvoSight[TM] method was designed in order to rationalize the directed evolution strategies and make it possible to run enzyme optimization programs within timelines that are compatible with industrial requirements (Fourage *et al.*, 2006; Chodorge *et al.*, 2005).

EvoSight[TM] is a three-step method. A first step consists in experimentally assessing the "plasticity" (*i.e.*, the ability of a protein to accept mutations with a limited loss of activity) of the specific protein to be engineered, by rapidly evaluating the frequency of functional variants in a series of libraries produced by random mutagenesis under increasing mutational loads. The data generated by this first step are then processed using a mathematical model in order:
– to determine the optimal mutation load for that particular system;

– to statistically determine the minimum number of clones to be screened to ensure that improved variants will be detected.

These results enable to strongly reduce the size of the library to be screened. After screening, the experimental determination of the actual evolution rate of the system (*i.e.* the frequency of improved clones in the screened library) enables to determine the potential of evolution of the specific enzyme/function system considered. The EvoSight[TM] algorithm also enables to determine the number of independent benefic mutations for that particular system and therefore provides guidelines for defining further directed evolution successful strategies.

2.2 Gene Shuffling

Since the late 90's, *in vitro* "sexual" evolution has been developed to allow for poolwise recombination of the parental genes (Tobin *et al.*, 2000). Poolwise recombination enables mutations from many parental genes to recombine in a single progeny, thus dramatically increasing the number of positive mutations that can be cumulated between two selection events and enabling a much broader sequence space to be explored. The term "gene shuffling" was coined to name this approach. It appeared rapidly that recombining related parental genes has dramatically accelerated the rate of evolution. This new scheme achieves a much larger evolutionary distance of the best clones from the parental genes than when using classical random mutagenesis. The first *in vitro* methods for poolwise recombination of parental genes developed, such as DNA shuffling (Stemmer, 1994) or SteP (Zhao *et al.*, 1998), represented a major breakthrough in protein engineering. They were based on polymerase chain reaction-like recombination. Although extremely powerful, these pioneering methods also suffered limitations. The kinetics of hybridization of the megaprimers formed during the PCR-like recombination rounds used to limit the size of the genes that can be efficiently shuffled. Because of the inherent properties of the randomized PCR-like recombination methods, the libraries thus generated frequently contained a significant proportion of non-functional mutants. As a consequence, larger size libraries generally had to be screened so that positive hits are not missed.

To overcome these limitations, a new gene shuffling technique was developed involving no polymerase, called L-Shuffling[TM]. This ligation-based random recombination method (Dupret *et al.*, 2000; Ravot *et al.*, 2005) permits the recombination of parental genes without the use of any polymerase, therefore reducing the risk of unwanted mutations. In this technology, the random recombination is assured by the ligation, by means of a suitable DNA ligase, of the ends of fragments of the parental gene variants hybridized onto an assembling template (*Fig. 1*). This ligation-based process enables randomized recombination, maintaining and combining the DNA information of the parental genes and generating a high proportion of functional variants.

Benefits of L-Shuffling[TM] include the ability of engineering both long and short genes as well as genes with high or limited homology. Most shuffled proteins generated by L-Shuffling are functional, which enables quicker screening and faster release of new products.

The impact of L-Shuffling[TM] for optimizing the characteristics of biocatalysts is well documented (Fourage *et al.*, 2005, 2009) and not limited to the engineering of enzymatic activity. For instance, the technology has been applied for antibody affinity maturation (Chodorge *et al.*, 2008). In this paper, the authors demonstrate that the most improved variant, with a 22-fold affinity gain, emerged from the recombination-based approach and not from a cumulative process such as error-prone PCR. Moreover, an analysis of mutations preferentially selected in the recombined population demonstrated strong cooperative effects when tested in combination with other mutations but small, or even negative, effects on affinity when tested in isolation. This study also demonstrated that the inclusion of an L-Shuffling[TM] recombination step enabled the selection of novel combinations of mutations, and therefore the exploration of a broader sequence space, compared with a parallel strategy which omitted recombination. This result is a strong argument for the combinatorial, poolwise-based approaches to protein evolution when compared to iterative mutagenesis based approaches.
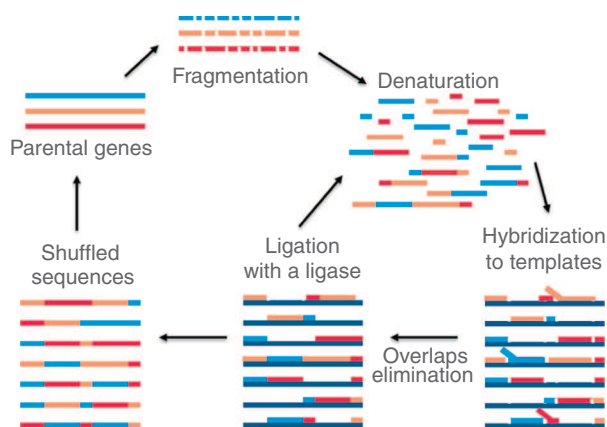

Figure 1

Principle of L-Shuffling[TM].

3 MOLECULAR EVOLUTION OF Cel3a BETA-GLUCOSIDASE

3.1 Experimental Section

3.1.1 Materials

p-nitrophenyl β-D-glucopyranoside (pNPGlc) was purchased from Sigma. Plasmid pET26b + used for creating L-Shuffling libraries was supplied by Novagen.

3.1.2 Construction of Plasmid pET26cay

The cDNA encoding the three parental glucosidases (A, B and C for beta-glucosidases from *Chaetomium globosum*, *Trichoderma reesei* and *Neurospora crassa* respectively) was amplified by PCR with the specific primers

(*see the specific primers below*)

Amplification was performed with 30 cycles of 94°C for 1 min, 60°C for 45 s, and 72°C for 1.5 min, with *Pfu* DNA polymerase (*Promega*) using the following PCR mix:
– 2 µL DNA template at 20 ng/µL;
– 2 µL of primers at 10 pmol/µL;
– 4 µL of dNTPs at 5 mM each;
– 10 µL of 10× buffer;
– 1 µL Pfu at 2.5 U/µL;
– 79 µL dH$_2$O.

After digestion using *Nde*I and *Hin*dIII restriction nuclease, the 2.1-kb PCR products were cloned into the pET26b + multicopy vector using the corresponding restriction sites leading to pETcay construct. After validation by automated DNA sequencing with the T7 promoter and terminator primers, the resulting plasmid, named pETcayB, A and C respectively, were used to prepare the three L-Shuffling libraries as described previously (Ravot *et al.*, 2005; Ayrinhac *et al.*, 2011).

3.1.3 Screening of L-Shuffling^{TM} Library

For the first round of evolution, *E. coli* MC1061(DE3) colonies expressing β-glucosidase variants were grown in 96-well microtiterplates at 37°C during 4 hours and 20°C during 20 hours after induction with 100 µM IPTG in 150 µL of Luria-Bertani (LB) medium complemented with 60 mg/mL of kanamycin. For the second and third rounds, cells were grown during 20 hours at 20°C without IPTG induction. After centrifugation (4 min at 4 000 g) and resuspension in 100 µL of 100 mM succinate buffer at pH 5.0 comprising 2.2 mM of pNPGlc (in a presence of 80 g/L of glucose for the third round), cells were incubated for 3 hours at 23°C. After this incubation, activities were estimated spectrophotometrically by measuring the release of p-nitrophenolate ion at 414 nm after adding one volume of Na$_2$CO$_3$. In each plate, the values were compared to the value obtained with cells expressing the WT-glucosidases tested in the same conditions.

3.1.4 Characterization of Glucosidase Variants

Growth conditions used for the characterization of the β-glucosidase variants were the same as the ones used during the screening of the second round of L-Shuffling^{TM} (variants were expressed without IPTG induction. Under such conditions, the level of expression is only driven by the leakage of the promoter and no detectable expression of the variants was observed on SDS-PAGE). After centrifugation (5 min at 8 000 g), *E. coli* MC1061(DE3) clones expressing the improved glucosidase variants were resuspended in 0.8 mL of 100 mM succinate buffer at pH 5.0 and different amounts of resuspended cell pellets were incubated 1.5 h using saturating concentration of pNPGlc at 50°C.

3.2 Results

In order to evolve β-glucosidase activity, a first round of L-Shuffling^{TM} was carried out using two parental sequences sharing 70% of amino acid identity. For microplate expression, induced cultures using 100 µM of IPTG were needed in order to detect β-glucosidase activity of *T. reesei*. It should be noted that activity of the second parental protein (protein A) was not detectable under these screening conditions. Around 20 000 clones were analyzed as described in the experimental section using 2.2 mM of p-nitrophenyl

Specific primers

SgeneA (5'- GGAATTCCATATGCTGGAGGCCGCCGACTG G -3') and
ASgeneA (5'- CCCAAGCTTCTAGGCGGTCAGGCTGCC -3');
SgeneB (5'- GGAATTCCATATGGTTGTACCTCCTGCAGGGAC -3') and
ASgeneB (5'- CCCAAGCTTCTACGCTACCGACAGAGTG -3') and
SgeneC (5'- GGAATTCCATATGGAGACAAGCGAGAAGCAGG -3') and
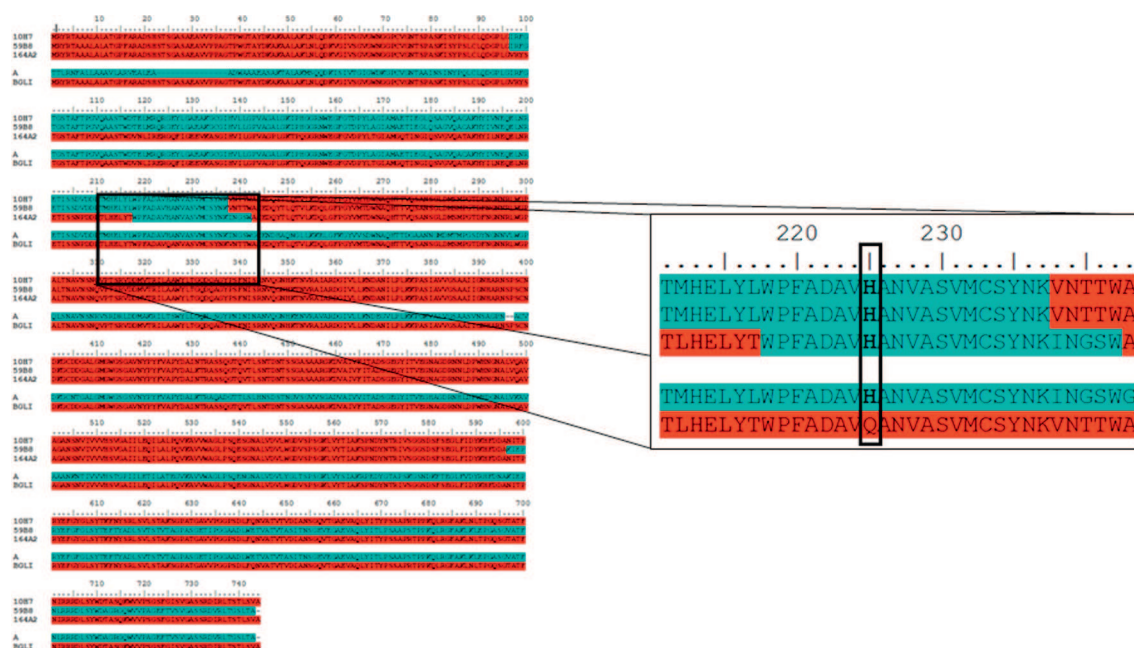ASgeneC (5'- CCCAAGCTTCTAGTATACGTCGAACTTGCC -3')

Figure 2

Sequence alignment of 3 improved variants resulting from the first round of L-Shuffling^TM experiment. Parental genes BGL1 and A are shown below. Red: fragments originating from the reference parental gene encoding protein B; blue: fragments originating from gene encoding protein A.

β-D-glucopyranoside (pNPGlc) as substrate. Among the selected improved clones, 16 were sequenced showing large sequence diversity even if a hot-spot was suspected in the N-terminal part of the protein. As expected, due to the chosen L-Shuffling^TM strategy, the skeleton of the improved variants was based on the reference protein. Figure 2 shows a sequence alignment of the 3 best variants together with parental genes BGL1 and A. The figure highlights the fact that comparison of improved variants can lead to identification of potentially important amino acids. Here the Histidine (H) in position 225 from gene A seems more efficient than the glutamine (Q) in the original Bgl1 gene, as this amino acid is found in every improved mutant.

The clone (164A2) selected as the best performer among the three tested ones when expressed in *Escherichia coli* was also analyzed for its β-glucosidase activity improvement on pNPGlc when expressed in *T. reesei* (*Fig. 3*).

Using saturating concentration of substrate, kcat value of the wild-type β-glucosidase was assessed. A strong improvement of the specific activity was observed (*Fig. 4*) whatever the expression system used was. Based on these results, a second round of L-Shuffling^TM was performed using the sixteen sequences shown in Figure 3 and introducing the WT β-glucosidase C as an
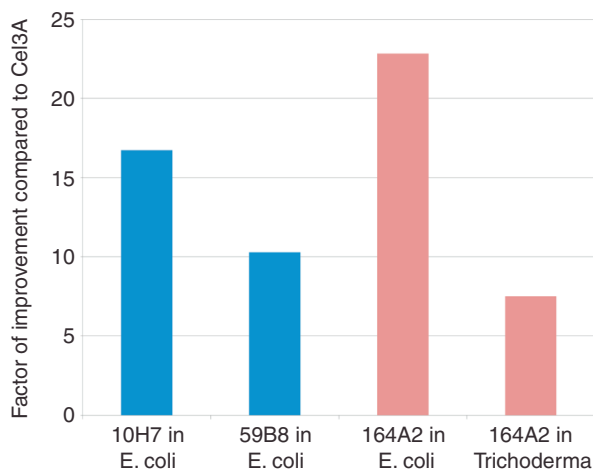

Figure 3

Activity improvement of the best performers from the first round of L-Shufling^TM.

additional parental gene. The same screening strategy was applied except that IPTG induction was no longer necessary due to the strong improvement achieved during the first round. Around 20 000 clones were analyzed as described in the experimental design using 2.2 mM of pNPGlc as substrate. Among the selected improved
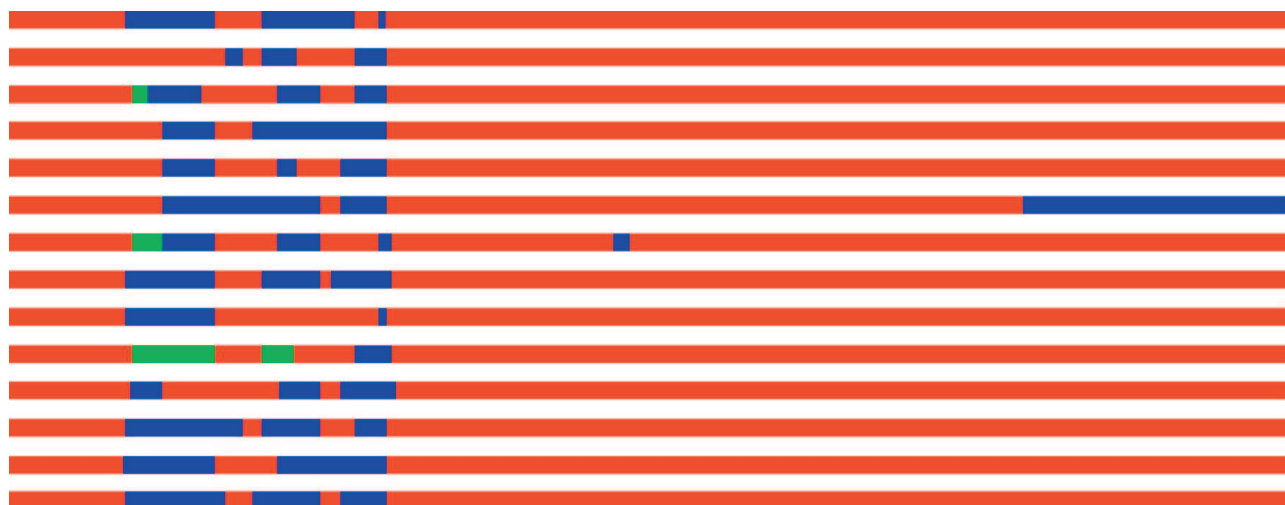
Figure 4

Sequence alignment of improved variants resulting from the second round of L-Shuffling experiment. Red: fragments originating from the reference parental gene encoding protein B; blue: fragments originating from gene encoding protein A; green: fragments originating from gene encoding protein C.
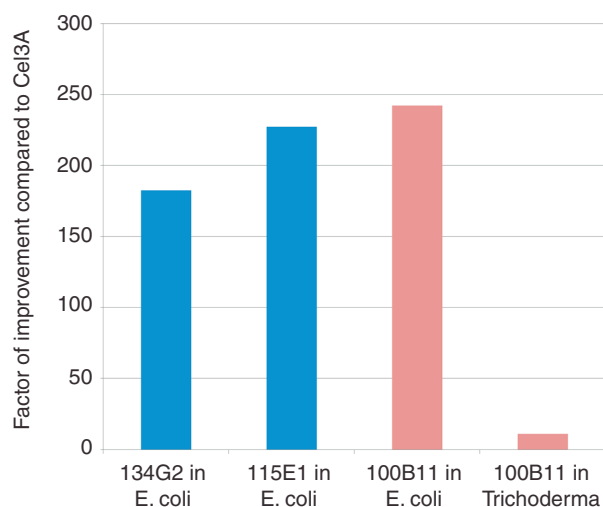

Figure 5

Activity improvement of the best performers from the second round of L-Shuffling^TM.

clones, we determined the DNA sequence of 14 clones. The sequences confirmed the presence of a hot-spot in the N-terminal part of the protein (*Fig. 4*) demonstrating that this protein region is of importance for activity.

The clone (100B11) selected as the best performer among the three tested ones when expressed in *Escherichia coli* (see *Fig. 5*) was also analyzed for its β-glucosidase activity improvement on pNPGlc when expressed in *Trichoderma reesei*.

As already observed for 164A2 clone, a strong improvement of the specific activity (*Fig. 6*) was observed compared to native Cel3A β-glucosidase, whichever expression system was used. An 11-fold improvement was reached when 100B11 gene was expressed in *T. reesei*. Surprisingly, despite the high β-glucosidase activity of protein C in the tested conditions (data not shown), no DNA fragment of that β-glucosidase C was found in the sequence of the three best performers, although other less improved variants resulting from the same round of shuffling contained fragments originating from β-glucosidase C (*Fig. 7*).

The molecular diversity of the variants selected during the second round of directed evolution was limited, hence suggesting that a plateau of molecular evolution has been reached. In order to further enhance the β-glucosidase activity, the introduction of additional molecular diversity would be required. In the case of the subsequent round of evolution which aimed at improving another parameter using a new screening assay, such diversity can be provided by the parental genes.

As native Cel3A β-glucosidase is inhibited by high concentrations of glucose, which is also a bottleneck for production of 2G ethanol through SHF process, a new screening assay was defined for the third round of L-Shuffling^TM, based on the same conditions than the ones used for the second round, except that 80 g/L of glucose were added during the activity test. In such reaction conditions, around 20 000 clones were analyzed and one clone (149G7, see *Fig. 7* for sequence) was selected
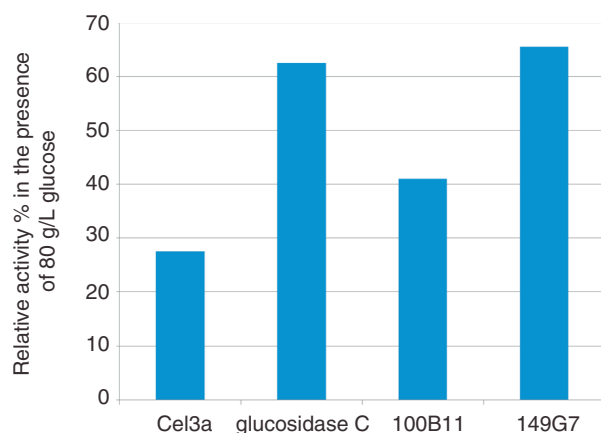
Figure 6

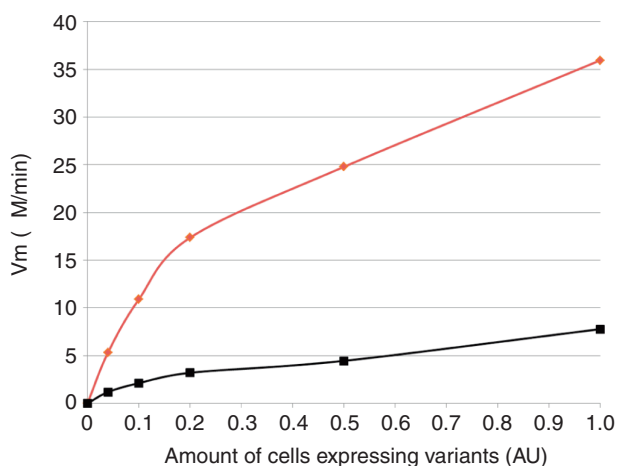Residual activity in the presence of 80 g/L of glucose.

Figure 8

Activity comparison of 149G7 and 100B11 β-glucosidases

Figure 7

Sequence of the improved variant 149G7 resulting from the third round of L-Shuffling[TM] experiment. Red: fragments originating from the reference parental gene encoding protein B; blue: fragments originating from gene encoding protein A; green: fragments originating from gene encoding protein C.

for a better resistance to high glucose concentrations compared to Cel3A (*Fig. 6*).

As shown in Figure 8, an improved residual activity in the presence of 80 g/L glucose was also observed compared to 100B11 (best performer identified during the second round) when 149G7 gene is expressed in *Escherichia coli* and under saturating concentrations of substrate.

However, SDS-PAGE analysis has demonstrated (data not shown) that the observed improvement of activity per amount of cells (see *Fig. 8*) may be due to a better expression of the recombinant glucosidase 149G7.

In conclusion, the protein encoded by clone 149G7 shows a lower inhibition by high glucose concentration than the best variant from the second round of L-Shuffling[TM] (protein 100B11) while conserving a specific activity in the same range.

4 SACCHARIFICATION OF LIGNOCELLULOSIC BIOMASS

Secretomes produced by *T. reesei* strain TR3002 expressing 100B11 gene and *Trichoderma* strain CL847 expressing native Cel3A were used as cellulolytic cocktails for steam-exploded wheat straw hydrolysis (Fourage *et al.*, 2010). Kinetics were determined in shake-flasks and bioreactor (Bio-Laffite) under stirring conditions (500 tr/min). The wheat straw was pretreated by steam explosion (19 bar, 3 min) previously soaked in 0.05 M H_2SO_4 for 15 hours. After filtration, the solid phase mainly containing cellulose and lignin was suspended at 10% Dry Matter (DM) in a total volume of 50 mL for shake-flask experiments and 2 L for bioreactor experiments of 1 M acetate buffer at pH 4.8 and 48°C. Protein concentration was determined using the Folin method (Lowry *et al.*, 1951). Enzymatic cocktails were used at 5, 10 and 20 mg per g of DM and kinetics were carried out for 72 h or 96 h. Shake flasks were done in duplicates. Bioreactor experiments were not replicated but 2 independent samples were taken for each time point. Enzymatic activities were inactivated by thermal denaturation and, after centrifugation and filtration, the released glucose was titrated with a glucose analyzer using glucose oxidase method (*Analox*, UK). The results obtained are shown in Figure 9 (shake-flasks) and Figure 10 (bioreactor). Figure 9 shows that for any enzyme dose TR3002 secretome outperforms the reference CL847. The effect is more visible at high enzymes
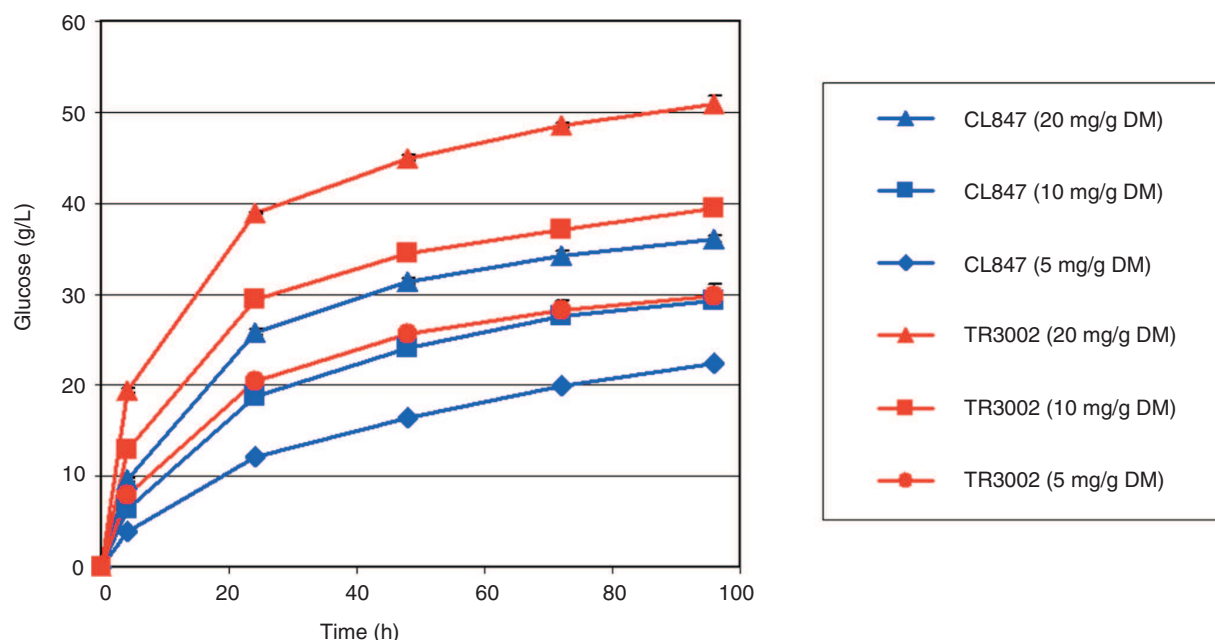
Figure 9

Kinetics of wheat straw saccharification using either improved (red) TR3002 or wild type (blue) CL847 *T. reesei* secretomes in shake-flasks at various enzyme dosages.
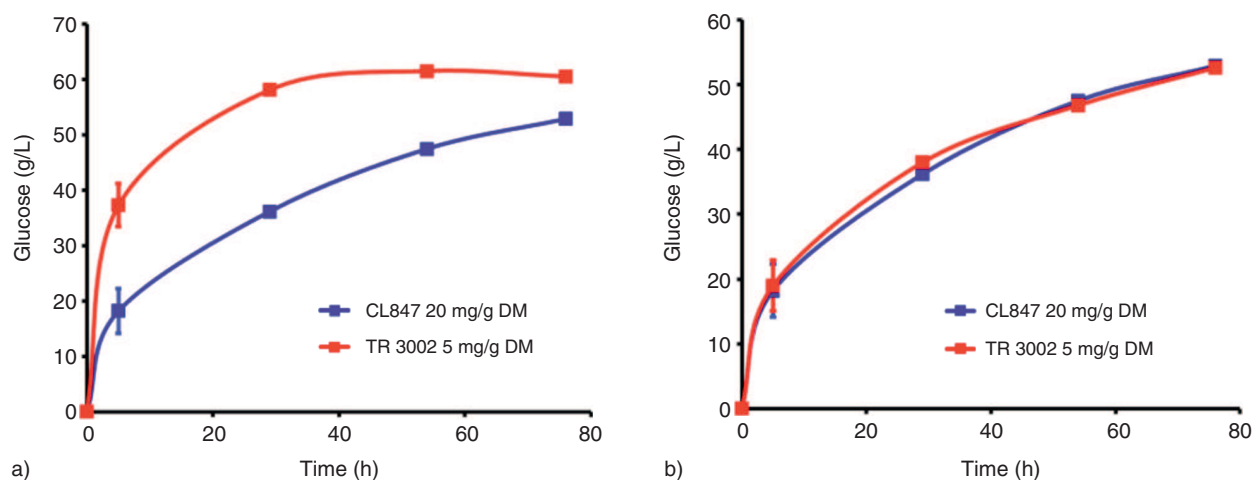

Figure 10

Kinetics of wheat straw saccharification using either improved (red) TR3002 or wild type (blue) CL847 *T. reesei* secretomes in bioreactor. a) Equal amounts of secretome, *i.e.* 20 mg of secretome per g of dry matter; b) 20 mg/g of wild type secretome *vs* 5 mg/g of improved secretome. Figure taken from Ayrinhac *et al.* (2011).

doses, suggesting that at 5 mg/g DM, other activities besides β-glucosidase become critical for hydrolysis. This effect is even more pronounced for doses below 5 mg/g DM (our unpublished results). In bioreactor (*Fig. 10*), steering allows a more efficient mixing and sus-pension of the substrate, resulting in faster kinetics. After 24 h, the maximum yield (60 g/L of glucose) was obtained when TR3002 secretome was used as biocata-lyst, while after 72 h of incubation this maximum yield was not held using the same amount of CL847 secretome

(20 mg of secretome/g of DM, *Fig. 10a*). This major improvement enabled a 4-fold decrease in secretome loading (5 mg of secretome/g of DM) without any loss in activity performance of hydrolysis of a steam exploded wheat straw in bioreactor (*Fig. 10b*).

CONCLUSION

The present results demonstrate that molecular evolution technologies combined with an efficient high throughput screening assay enable to rapidly identify strongly improved variants for industrial purposes. In this particular example, we demonstrate how an evolved secretome resulting from *in vitro* recombination processes enables to lower the cost-contribution of the saccharification step in biofuel manufacturing cost. Despite no structural information was available for Cel3A β-glucosidase from *Trichoderma reesei* and no β-glucosidase activity was demonstrated for the two other parental proteins A and C using in this approach, enzyme variants showing up to 242-fold improvement in specific activity and lower inhibition in the presence of high concentration of glucose were identified after three L-Shuffling rounds. After introduction of the improved 100B11 β-glucosidase gene in an industrial *Trichoderma reesei* strain, a good productivity was observed without significant changes of ratio between the different cellulosic activities. The modified *Trichoderma reesei* strain produces a new efficient secretome in which β-glucosidase activity is 11 times higher than in the secretome from the unmodified strain. This new secretome enables a 4-fold decrease in cellulase loading without any loss in hydrolysis performance for degradation of a steam exploded wheat straw in bioreactors.

ACKNOWLEDGMENTS

REFERENCES

Ayrinhac C., Margeot A., Lopes-Ferreira N., Chaabane F.B., Monot F., Ravot G., Sonet J.M., Fourage L. (2011) Improved Saccharification of Wheat Straw for Biofuel Production Using an Engineered Secretome of *Trichoderma reesei*, *Org. Proc. Res. Devel.* **15**, 275-278.

Brustad E.M., Arnold F.H. (2011) Optimizing non-natural protein function with directed evolution, *Curr. Opin. Chem. Biol.* **15**, 2, 201-210.

Chodorge M., Fourage L., Ullmann C., Duvivier V., Masson J.M., Lefevre F. (2005) Rational Strategies for Directed Evolution of Biocatalysts – Application to Candida Antarctica Lipase B (CALB), *Adv. Synth. Catal.* **347**, 1022-1026.

Chodorge M., Fourage L., Ravot G., Jermutus L., Minter R. (2008) *In vitro* DNA recombination by L-Shuffling during ribosome display affinity maturation of an anti-Fas antibody increases the population of improved variants, *Protein Eng. Des. Sel.* **21**, 343-351.

Chundawat S.P., Beckham G.T., Himmel M.E., Dale B.E. (2011) Deconstruction of lignocellulosic biomass to fuels and chemicals, *Annu. Rev. Chem. Biomol. Eng.* **2**, 121-145.

Dougherty M.J., Arnold F.H. (2009) Directed evolution: new parts and optimized function, *Curr. Opin. Biotechnol.* **20**, 4, 486-491.

Dupret D., Masson J.M., Lefevre F. (2000) Method for obtaining *in vitro* recombined polynucleotide sequences, sequence banks and resulting sequences, WO 00/09679.

Edwards R., Larivé J.F., Mahieu V., Rouveirolles P. (2007) Well-To-Wheels analysis of future automotive fuels and power trains in the European context, v.2c. WTW Report 010307, Well-To-Wheel joint study by the research institutes European auto industry (EUCAR), refinery industry (Concawe) and European Union (JRC) (http://ies.jrc.ec.europa.eu/uploads/media/WTW_Report_010307.pdf).

Fairley P. (2011) Next generation biofuels, *Nature* **474**, S2-S5.

Fourage L., Lefevre F., Wieland S., Weber A., Beckers A., Maurer K.H., Kottwitz B. (2005) Alkaline protease variants having improved performance and washing and cleaning agents containing said alkaline protease variants having improved performance, WO05/118793.

Fourage L., Chodorge M., Lefevre F., Masson J.M. (2006) Method of determining the mutational load of a gene library obtained by random mutagenesis of a particular gene and means for implementing same, WO 06/003298.

Fourage L., Wahler D., Bruhlmann F. (2009) Modified 13-hydroperoxide lyases and uses thereof, WO 09/001304.

Fourage L., Lopes-Ferreira N., Margeot A., Mathis H. (2010) Beta-glucosidase variants having improved activity, and uses thereof, WO2010029259.

Illanes A., Cauerhff A., Wilson L., Castro G.R. (2012) Recent trends in biocatalysis engineering, *Bioresour. Technol.* **115**, 48-57.

Kubicek C.P., Mikus M., Schuster A., Schmoll M., Seiboth B. (2009) Metabolic engineering strategies for the improvement of cellulase production by Hypocrea jecorina, *Biotechnol. Biofuels.* **1**, 2, 19.

Lowry O.H., Rosebrough N.J., Farr A.L., Randall R.J. (1951) Protein measurement with the Folin phenol reagent, *J. Biol. Chem.* **193**, 265-275.

Margeot A., Hahn-Hagerdal B., Edlund M., Slade R., Monot F. (2009) New improvements for lignocellulosic ethanol, *Curr. Opin. Biotechnol.* **20**, 3, 372-380.

Olofsson K., Bertilsson M., Lidén G. (2008) A short review on SSF - an interesting process option for ethanol production from lignocellulosic feedstocks, *Biotechnol. Biofuels* **1**, 7.

Olson D.G., McBride J.E., Shaw A.J., Lynd L.R. (2012) Recent progress in consolidated bioprocessing, *Curr. Opin. Biotechnol.* **23**, 3, 396-405.

Ravot R., Masson J.M., Lefevre F. (2005) *Methods in Microbiology: Application of Extremophiles*, Academic Press, London.

Song L., Siguier B., Dumon C., Bozonnet S., O'Donohue M.J. (2012) Engineering better biomass-degrading ability into a GH11 xylanase using a directed evolution strategy, *Biotechnol. Biofuels* **5**, 3.

Stemmer W.P. (1994) Rapid evolution of a protein *in vitro* by DNA shuffling, *Nature* **370**, 389-391.

Tobin M.B., Gustafsson C., Huisman G.W. (2000) Directed evolution: the 'rational' basis for 'irrational' design, *Curr. Opin. Struct. Biol.* **10**, 421-427.

Zhao H., Giver L., Shao Z., Affholter J., Arnold F. (1998) Molecular evolution by staggered extension process (StEP) *in vitro* recombination, *Nat. Biotechnol.* **16**, 258-261.