Protein Expression and Purification 99 (2014) 35–42

Contents lists available at ScienceDirect

Protein Expression and Purification

journal homepage: www.elsevier.com/locate/yprep

High-throughput cloning, expression and purification of glycoside hydrolases using Ligation-Independent Cloning (LIC)

Cesar M. Camilo, Igor Polikarpov *

Departamento de Física e Informática, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador Sãocarlense, 400, 13566-590 São Carlos, SP, Brazil

ARTICLE INFO

Article history:
Received 13 December 2013
and in revised form 25 February 2014
Available online 26 March 2014

Keywords:
High-throughput
Ligation-Independent Cloning
LIC
Solubility screening
Glycoside hydrolases
Thioredoxin

ABSTRACT

Recent advances in DNA sequencing techniques have led to an explosion in the amount of available genome sequencing data and this provided an inexhaustible source of uncharacterized glycoside hydrolases (GH) to be studied both structurally and enzymatically. Ligation-Independent Cloning (LIC), an interesting alternative to traditional, restriction enzyme-based cloning, and commercial recombinatorial cloning, was adopted and optimized successfully for a high throughput cloning, expression and purification pipeline. Using this platform, 130 genes encoding mainly uncharacterized glycoside hydrolases from 13 different organisms were cloned and submitted to a semi-automated protein expression and solubility screening in *Escherichia coli*, resulting in 73 soluble targets. The high throughput approach proved to be a powerful tool for production of recombinant glycoside hydrolases for further structural and biochemical characterization and confirmed that thioredoxin fusion tag (TRX) is a better choice to increase solubility of recombinant glycoside hydrolases expressed in *E. coli*, when compared to His-tag alone.

Introduction

The growth of energy demand promoted by industrialization, urbanization and societal affluence has led the planet to an extremely dependence on petroleum and given the constant increase in oil prices and global warming caused by the greenhouse gases emission, the need for alternative and renewable sources of bioenergy is a growing concern [1]. Among potential alternative bioenergy resources, lignocellulosic biomass has gained particular attention as a rich source of sugars for production of biofuels such as ethanol, as well as other chemicals with high added value. Currently, almost all the production of bioethanol is based on agricultural products such as sugar cane juice and cornstarch [2], but the high production demand and rising concerns over the competition between crops for food and fuel have led to increased focus on lignocellulosic biomass utilization for production of second-generation (or cellulosic) bioethanol.

Plant biomass is the most abundant renewable carbon source on Earth. Lignocellulose, its major constituent, is a highly heterogeneous substrate composed of cellulose (40–50%), hemicellulose (25–35%), and lignin (15–20%) [3]. Cellulose is a linear polymer consisting of glucose units joined by linkages β (1,4) glycosidic bonds. Hemicellulose is a polysaccharide of variable composition, but consists mostly of pentose xylose (5–20%) and arabinose (1–5%) [4,5]. Lignin has a hydrophobic structure, consisted of randomly polymerized phenylpropane monomers, which strengthens its physical properties and protects cellulose and hemicellulose from hydrolytic enzymes [6].

Together, cellulosic and hemicellulosic sugars represent important sources of fermentative material for ethanol production, but the challenge of how to make these sugars available in an economically viable way remains unsolved. Enzymatic saccharification has gained prominence as the most promising approach for cellulose hydrolysis, which is considered a limiting step on lignocellulose utilization process [7–9]. Generally, multiple enzyme activities, including endoglucanase (EC 3.2.1.4), exoglucanase (EC 3.2.1.91) and β-glucosidase (EC 3.2.1.21) are required to release glucose molecules from cellulose [7,10], while xylose, mannose, galactose, rhamnose, and arabinose sugars are depolymerized from hemicellulose by hemicellulases [11]. As the cost of enzymes remains a key economic impediment to commercialization of biofuels, scientific efforts towards elucidation of their catalytic mechanisms, improvement of catalytic activity by enzymatic engineering, directed evolution and site-directed mutagenesis, as well as discovery of new enzymes are fundamental for enabling the cost decrease of enzymatic hydrolysis of biomass.

Recent advances in DNA sequencing techniques have led to an explosion in the amount of available genome sequencing data [12–15] and such phenomenon has provided an inexhaustible

* Corresponding author. Tel.: +55 16 3373 9840; fax: +55 16 3373 8674.
E-mail address: ipolikarpov@ifsc.usp.br (I. Polikarpov).

http://dx.doi.org/10.1016/j.pep.2014.03.008
1046-5928/

source of uncharacterized glycoside hydrolases (GH)[1] to be studied by structural biology. Even though the enzymes can be classified among GH families defined in the CAZy (Carbohydrate-Active EnZymes) database [16], computational methods are unable to distinguish substrate specificities of enzymes belonging to a same family. In this sense, structural and biochemical characterization of a large number of newly identified glycoside hydrolases from different families represent promising tools for understanding molecular features and mechanisms of carbohydrate hydrolysis.

The major bottleneck when working with a large number of targets resides in cloning, expression and solubility screening. Even if a soluble protein expression is achieved, the success of crystallization and 3D structure determination by X-ray diffraction is uncertain. A recent statistic obtained by SGC (Structural Genomics Consortium) showed that, from a target list of 1269 distinct human proteins, 48% of these targets could be expressed soluble in *Escherichia coli* and 46% of them had their structure solved by X-ray crystallography, resulting in an overall success rate of 22% [17]. Such results demonstrate that a large number of cloned targets are required to achieve a reasonable number of solved structures. For this reason, laboratories worldwide are developing and implementing high-throughput cloning and expression pipelines using mainly protein expression screening with *E. coli* as expression host [18–23].

Traditional cloning methods, which involve cleavage of plasmids and target inserts with restriction enzymes followed by ligation with DNA ligase, are almost impractical in a high throughput pipeline because of specific requirements regarding the absence of restriction sites within the targets sequences and also due to the large number of necessary steps to obtain the clone. To serve rapid cloning, many limitations related to generation of multiple expression plasmids have been recently addressed by high-throughput adaptable systems which enable cloning of hundreds of genes and constructs simultaneously. Several worldwide high-throughput facilities have adopted commercial systems such as Gateway® (Life Technologies, USA, CA) [24–26] and In-Fusion™ (Clontech, USA, CA) [27], but despite the overall flexibility of these systems, per-reaction cost can be high because of the dependence on proprietary recombinases. Furthermore, the presence of are combination site as part of the open reading frame can affect protein function and solubility. To avoid this problem, the recombination sites can be placed outside the open reading frame, but it requires large primers with Shine–Dalgarno or Kozak sequences between the recombination site and the gene-specific termini. Alternatively, a protease cleavage site is frequently placed after N-terminal recombination sequence, allowing it to be excluded after protein expression, but large primers and consequently a two-step PCR (polymerase chain reaction) are still necessary. With this, the overall cost of the process associated to reduction in flexibility represent drawbacks when applied to HTP (high throughput) protein expression in an academic environment. Ligation-Independent Cloning (LIC) [28], Restriction site-free cloning (RF-cloning) [29–33], Polymerase Incomplete Primer Extension (PIPE) [34,35] and Enzyme-Free Cloning (EFC) [36] have flexibilities that are comparable to recombinatorial cloning, but the advantage of lower costs due to independence of commercial kits makes them attractive methods for HTP routines. The most common LIC method is based on exonuclease activity of T4 DNA polymerase [37–40], but a number of new methods have been developed to improve versatility and cloning efficiency, such as sequence and ligation independent cloning (SLIC) [41], improved SLIC [42], uracil excision-based cloning [43,44] and Nicking Endonucleases based LIC (NE-LIC) [45].

This paper describes the results of high throughput cloning, expression and purification of glycoside hydrolases from fungal, bacterial and archaean sources using a LIC protocol and a semi-automated solubility screening in *E. coli*.

Materials and methods

High throughput Ligation-Independent Cloning (LIC) protocol

LIC cloning, as illustrated in Figs. 1 and 2, was adapted for high-throughput routines using the following protocol:

Vectors preparations

pETTRXA-1a/LIC and pETM11/LIC plasmids were linearized by PCR with Phusion® "High-fidelity DNA Polymerase" (New England Biolabs, USA, MA) using the following primers: Fw- 5′ TGGCGCCCTGAAAATAAAG and Rv- 5′ CCGCGTCGGGTCAC. Briefly, a 50 μl reaction mix containing 2.5 ng vector DNA, 25 pmol of each primer, 0.2 mM dNTP Mix, 1 unit of Phusion polymerase and 1× Phusion polymerase buffer was used in a 3-step PCR reaction: (1) 98 °C for 30 s, 1 cycle; (2) 98 °C for 10 s, 65 °C for 30 s, followed by 72 °C for 105 s; 35 cycles; (3) 72 °C for 10 min, 1 cycle. The PCR product was treated with 20 units of *Dpn*I enzyme (New England Biolabs) for 16 h at 37 °C to remove template DNA and purified by 0.8% (w/v) agarose gel electrophoresis followed by gel extraction with Wizard SV Gel and PCR CleanUp System (Promega, USA, WI). A total of 500 ng of purified vector was treated with3 units of T4 DNA polymerase (Fermentas, USA, MA) in the presence of 1× T4 polymerase buffer, 4 mM DTT and 2.5 mM dTTP in a final volume of 20 μl, incubated for 30 min at 22 °C and subsequently heat inactivated for 20 min at 75 °C.

Template sources

For bacterial, archaean and *Pichia pastoris* targets genomic DNAs were used as templates for PCR amplification. For fungal targets, cDNA libraries were prepared as follows: total RNAs were extracted with Trizol (Life Technologies) and used for first strand cDNA synthesis by "First Strand cDNA Synthesis" kit (Fermentas). *Phanerochaete chrysosporium* was cultivated in submerse fermentation in minimum liquid media as described by Kirk et. al. [46]. For fungi of the genus Aspegillus and Trichoderma,10^6 spores were grow for 3–4 days at 28 °C with stirring in 100 mL of minimal medium: 0.3 g L^{-1} Urea; 1.4 g L^{-1} $(NH_4)_2SO_4$; 1 mL L^{-1} of micronutrients solution 1000X (2.2% $ZnSO_4{\cdot}H_2O$; 1.1% H_3BO_3; 0.5% $MnCl_2{\cdot}4H_2O$; 0.5% $FeSO_4{\cdot}7H_2O$, 0.17% $CoCl_2{\cdot}6H_2O$; 0.16% $CuSO_4{\cdot}5H_2O$; 0.15% $Na_2MoO_4{\cdot}2H_2O$; 5% Na_4EDTA (w/v)); 0,4 g L^{-1} $CaCl_2$; 0.3 g L^{-1} $MgSO_4$; 10 mM sodium citrate pH 5.0; 0.6 g L^{-1} yeast extract. Cultivations were performed using glucose, avicel® (Sigma–Aldrich, USA, MO) or sugarcane bagasse as carbon sources.

Insert preparation

Gene-specific primers were designed using the high throughput primer design tool HTP-OligoDesigner (http://ifsc.usp.br/htpoligo/), with calculated melting temperatures (Tm) ranging from 63 to 65 °C according to thermodynamic data from Breslauer et al. [47]. Signal peptides were removed from coding sequences when identified by SignalP 4.1 Server signal peptide prediction software [48]. To create an insert with complementary overhangs with the LIC vectors, specific 5′ extensions (5′ CAGGGCGCCATG and 5′ GACCCGACGCGGTTA) were added to the forward and reverse

[1] *Abbreviations used:* GH, glycoside hydrolases; LIC, Ligation-Independent Cloning; TRX, thioredoxin fusion tag; CAZy, Carbohydrate-Active EnZymes; SGC, Structural Genomics Consortium; PCR, polymerase chain reaction; RF-cloning, restriction site-free cloning; PIPE, polymerase incomplete primer extension; EFC, enzyme-free cloning; SLIC, sequence and ligation independent cloning; NE-LIC, Nicking Endonucleases based LIC; TEV, tobacco etch virus protease; eGFP, enhanced Green Fluorescent Protein.