Journal of Biotechnology 195 (2015) 1–7

Contents lists available at ScienceDirect

Journal of Biotechnology

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

Gateway-compatible vectors for high-throughput protein expression in pro- and eukaryotic cell-free systems

Dejan Gagoski[a,1], Sergey Mureev[a,1], Nichole Giles[a], Wayne Johnston[a], Mareike Dahmer-Heath[b], Dubravka Škalamera[b], Thomas J. Gonda[c,b], Kirill Alexandrov[a,*]

[a] Institute for Molecular Bioscience, University of Queensland, Australia
[b] University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Australia
[c] School of Pharmacy, University of Queensland, Brisbane, Australia

ARTICLE INFO

Article history:
Received 13 August 2014
Received in revised form 6 December 2014
Accepted 11 December 2014
Available online 19 December 2014

Keywords:
Gateway cloning
Cell-free protein expression
Species Independent Translation Initiation Sequence
Rolling Circle DNA Amplification

ABSTRACT

Although numerous techniques for protein expression and production are available the pace of genome sequencing outstrips our ability to analyze the encoded proteins. To address this bottleneck, we have established a system for parallelized cloning, DNA production and cell-free expression of large numbers of proteins. This system is based on a suite of pCellFree Gateway destination vectors that utilize a Species Independent Translation Initiation Sequence (SITS) that mediates recombinant protein expression in any *in vitro* translation system. These vectors introduce C or N terminal EGFP and mCherry fluorescent and affinity tags, enabling direct analysis and purification of the expressed proteins. To maximize throughput and minimize the cost of protein production we combined Gateway cloning with Rolling Circle DNA Amplification. We demonstrate that as little as 0.1 ng of plasmid DNA is sufficient for template amplification and production of recombinant human protein in *Leishmania tarentolae* and *Escherichia coli* cell-free expression systems. Our experiments indicate that this approach can be applied to large gene libraries as it can be reliably performed in multi-well plates. The resulting protein expression pipeline provides a valuable new tool for applications of the post genomic era.

1. Introduction

The emergence of next generation sequencing technologies has transformed biology by providing rapid and inexpensive access to genomic information. However, the ability to generate gene sequencing data greatly outstrips our ability to experimentally analyze the encoded polypeptides. Currently less than 5% of putative open reading frames have been experimentally studied (UniProt, 2012).

One of the key hurdles to progress in this direction is the recombinant production of proteins. Despite more than 40 years of research, recombinant protein expression is a complex and inherently unpredictable process. It is often difficult or impossible to express functional proteins due to their host toxicity, aggregation, misfolding or degradation. In an effort to address these issues, cell-free protein expression (CFPE) systems have been increasingly employed for recombinant protein production. In CFPE a translation-competent cellular extract is primed with DNA or RNA that encodes the gene of interest (Jermutus et al., 1998; Whittaker, 2013).

The commonly employed *Escherichia coli* and wheat germ cell-free systems have been optimized for production of preparative amounts of "hard-to-express" proteins. Concomitantly there has been an expansion in the species used as lysate sources for CFPE. Common eukaryotic sources are Sf-9 insect cells, mammalian cells, rabbit reticulocytes, and most recently the ciliated protozoan *Leishmania tarentolae* (Mureev et al., 2009). These systems have been used for expression of proteins with unnatural amino acids, production of protein complexes, construction of protein arrays and even expression of membrane proteins (Albayrak and Swartz, 2013; Haberstock et al., 2012; He and Taussig, 2008; Kuruma et al., 2010).

Depending on the CFPE configuration, *in vitro* translation reactions can be primed either with mRNA, or in the case of coupled transcription-translation systems, with the template DNA (Katzen et al., 2005). The coupled CFPE systems are most suitable for multiplexing and high throughput applications due to their single tube/single step format (Sitaraman and Chatterjee, 2009). Such format for instance enables expression of transcriptomes (Goshima

* Corresponding author at: Building 80, 306 Carmody Rd, St. Lucia, 4072 QLD, Australia. Tel.: +61 7 3346 2017.
E-mail address: k.alexandrov@uq.edu.au (K. Alexandrov).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jbiotec.2014.12.006
0168-1656/

et al., 2008), construction of protein microarrays (Zarate and Galbraith, 2014) or directed evolution and library selection experiments (Odegrip et al., 2004). However, rapid and inexpensive cloning of large gene collections requires multiplexing and miniaturization of the process and therefore limits the amount and concentration of resulting template DNA. While methods of template preparation based solely on PCR-based template assembly have been developed, they are less robust in high-throughput format than cloning-based approaches (Ahn et al., 2005). An additional problem is that the 5′ translation enhancing sequences were traditionally unique to each CFPE system, requiring separate template modules for each. We have recently overcome this limitation by developing Species Independent Translation Initiation Sequences (SITS) that allow creation of universal templates for all cell-free systems (Mureev et al., 2009).

Here we present a suite of universal SITS-carrying CFPE vectors suitable for high-throughput cloning based on Gateway recombination technology. We demonstrate efficient vector performance on a selection of Gateway entry clones derived from the Human ORFeome v5.1 collection (Lamesch et al., 2007; Rual et al., 2004) and ImaGenes ORFeome set (Source Bioscience, Lifesciences) arrayed in microwell-plate format designed for robot-assisted handling. In order to streamline the process and to eliminate the requirement for larger scale DNA preparation, we include isothermal Rolling Circle Amplification (RCA) of the resulting templates by using the Phi29 DNA polymerase in (Dean et al., 2001; Fakruddin et al., 2013). We conclude that combination of Gateway cloning and RCA template amplification enables cost efficient and rapid production of *in vitro* expressed proteomes.

2. Materials and methods

2.1. Vector design

First pUC18-based intermediate pLTE EGFP plasmid was constructed by cloning T7 promoter, PolyUUUA-SITS, EGFP and 3′ UTR-coding sequences into the pUC18 plasmid vector between the *AatII-EcoRI* restriction sites, thereby removing the lacZ gene but retaining the lac promoter.

The construction of Gateway destination vectors pCellFree_G01 to pCellFree_G08 was performed by insertion of *de novo* synthesized sequences between the *NcoI* and *NotI* restriction sites of the pLTE EGFP vector. The synthesized sequences were designed based on a core sequence of pDEST14 containing an attR1 recombination site, cddB cassette and chloramphenicol resistance cassette in antisense orientation, followed by an attR2 recombination site (Fig. 1A). Further the core sequence was flanked with affinity or fluorescent tag on the appropriate side of the core (Fig. 1B). A PreScission protease cleavage site sequence was introduced between the core and the fluorescent tags without stop codon and an affinity tag on the given terminus. Stop codons were introduced after the attR2 site for the N-terminal and after the affinity tag for the C-terminal tagging vectors. Each combination of the core sequence and tag where custom synthesized and cloned into the pLTE EGFP construct (GenScript). In vector pCellFree_G09 vector the mCherry sequence tag of pCellFree_G05 was substituted by superfolder GFP (sfGFP). The same procedure was performed on the pCellFree_G08, where the mCherry-his-myc-tag was substituted with sfGFP-his tag resulting in pCellFree_G10. The sequences of the resulting vectors were deposited in the Gene Bank and the corresponding accession numbers are listed in the legend of Fig. 1.

2.2. Cloning of human ORFeome libraries into pCellFree vectors

Human ORF entry clones in pDONOR223 or pENTR201 vectors were picked from a previously generated collection (Skalamera et al., 2012) based on the Human ORFeome collection version 5.1 (Open Biosystems) or the ImaGenes ORFeome set (Source Bioscience, Lifesciences). The open reading frames in these libraries are in frame with the coding elements outside the attL recombination sites and do not contain stop codons. Clones were manually re-arrayed into 96-well plates. The ORFs were transferred into the pCellFree vectors via the Gateway – LR reaction performed in 5 μL volume per well as described previously (Skalamera et al., 2011). Plasmid DNA of the resulting clones was isolated from the transformed α-Select Gold chemically competent *E. coli* cells (Bioline, Australia) using the 5′PRIME Perfectprep plasmid 96 VAC kit (5 Prime) largely according to manufacturer's instructions. The bacterial growth conditions, and final plate elution volumes were adjusted to maximize the DNA yield, as described in Section 3. Clone identity was confirmed by single pass sequencing of arrayed templates at the AGRF facility (Brisbane, Australia), using vector specific primers. Clones in pCellFree_G01, 02, 04, 06, 08 and 10 were sequenced using the SITS forward primer 5′-CCATGACAGTAATGTATAAAGTCTGT-3′, clones in pCellFree_G03 and pCellFree_G09 with the eGFP 3′ end forward primer 5′-CAACGAGAAGCGCGATCAC-3′ and clones in pCellFree_G05 and pCellFree_G07 with the mCherry 3′ end forward primer 5′-AGTACGAACGCGCCGAGG-3′ primers.

2.3. DNA amplification

For medium scale plasmid preparation 200 mL LB medium was inoculated with TOP10 *E. coli* carrying the expression constructs. After an overnight incubation, the cells were harvested and used for plasmid midi-preparation via the NucleoBond® Xtra Midi kit (Macherey-Nagel, Germany).

The PCR amplification was performed with an extended T7 promoter forward primer (5′-TTAGCTGGGACGTCTAATACGACTCACTATAG-3′) and T7 terminator reverse primer (5′-GACGTCTAATACGACTCACTATAG-3′). The reactions were prepared by mixing 1 mM dNTP's (Life Technologies), 5× Phusion® HF Buffer (New England Biolabs), 0.08 units of Phusion® High-Fidelity DNA Polymerase (New England Biolabs), and 25 ng plasmid template. The reaction volume was adjusted with H_2O to 50 μL. The PCR reaction included 30 cycles with 10 s at 98 °C, 15 s at 56 °C and 90 s at 72 °C. The PCR products were purified with Wizard® SV Gel and PCR Clean-Up System.

Isothermal DNA Rolling Circle Amplification (RCA) was performed with the Illustra TempliPhi amplification kit (GE Life Sciences) according to the instructions of the manufacturer. The concentration of the RCA was determined using Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies) and purified DNA (plasmid and PCR) was measured with the Nanodrop ND-1000 spectrophotometer (Thermo Scientific).

2.4. Cell-free expression and analysis of proteins

Proteins were expressed using cell-free expression systems based on *Leishmania tarentolae* and *Eschericha coli*. The extracts (LTE) were prepared as described (Johnston and Alexandrov, 2014; Schwarz et al., 2007). LTE based expression was performed in 20 μL reaction volume typically containing 4 μL LTE-specific feeding solution, 10 μL LTE (11 μL for RCA-product based expressions), 4 μL mix of the appropriately diluted DNA template or 3 μL RCA-product and 2 μL Mg-acetate (35–45 mM for plasmids, 65 mM for PCR-products and 50 mM for RCA-product). The *E. coli* based CFPE reactions were performed by mixing 12 μL of *E. coli*-specific feeding solution, 10 μL *E. coli* S30 extract and 5.8 μL of appropriately diluted DNA (total DNA concentration 30 nM). The reactions were performed at 27 °C either in PCR tubes or in OptiPlate-384 microplates (PerkinElmer). Protein expression was detected either