Metabolic Engineering Communications 2 (2015) 23–29

Contents lists available at ScienceDirect

Metabolic Engineering Communications

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

Identifying promoters for gene expression in *Clostridium thermocellum*

Daniel G. Olson [a,d,*], Marybeth Maloney [a,d], Anthony A. Lanahan [a,d], Shuen Hon [a,d], Loren J. Hauser [b,d], Lee R. Lynd [a,b,c,d,*]

[a] Thayer School of Engineering at Dartmouth College, Hanover, NH 03755, United States
[b] Oak Ridge National Laboratory, Oak Ridge, TN 37830, United States
[c] Mascoma Corporation, NH 37830, United States
[d] BioEnergy Science Center, TN, United States

ARTICLE INFO

Article history:
Received 31 October 2014
Received in revised form
26 February 2015
Accepted 21 March 2015
Available online 30 March 2015

Keywords:
Promoter
Copy number
Structural instability
Rolling circle replication
lacZ
adhB

ABSTRACT

A key tool for metabolic engineering is the ability to express heterologous genes. One obstacle to gene expression in non-model organisms, and especially in relatively uncharacterized bacteria, is the lack of well-characterized promoters. Here we test 17 promoter regions for their ability to drive expression of the reporter genes β-galactosidase (*lacZ*) and NADPH-alcohol dehydrogenase (*adhB*) in *Clostridium thermocellum*, an important bacterium for the production of cellulosic biofuels. Only three promoters have been commonly used for gene expression in *C. thermocellum*, gapDH, cbp and eno. Of the new promoters tested, 2638, 2926, 966 and 815 showed reliable expression. The 2638 promoter showed relatively higher activity when driving *adhB* (compared to *lacZ*), and the 815 promoter showed relatively higher activity when driving *lacZ* (compared to *adhB*).

1. Introduction

Clostridium thermocellum is a promising candidate for the conversion of biomass to ethanol due to its native ability to use cellulose and produce ethanol (Olson et al., 2012). Recently there has been significant progress engineering *C. thermocellum* to improve ethanol production (Argyros et al., 2011; Deng et al., 2013; Van der Veen et al., 2013), however ethanol yield needs further improvement for commercial viability.

To date, most metabolic engineering of *C. thermocellum* has focused on gene deletion but many metabolic engineering strategies require increased or heterologous gene expression, in addition to gene deletion. Only a few genes have been successfully expressed in *C. thermocellum*, including the antibiotic resistance markers *cat* and *neo*; the counter-selectable markers *pyrF*, *hpt* and *tdk* (Olson and Lynd, 2012a); the cellulosome scaffoldin *cipA* (Olson et al., 2013) and the metabolic enzyme pyruvate kinase (Deng et al., 2013).

The most commonly used promoters are from the regions upstream of the *gapDH* (Clo1313_2095), *cbp* (Clo1313_1954) and *eno* (Clo1313_2090) genes in *C. thermocellum*. The *gapDH* gene encodes the glyceraldehyde-3-phosphate dehydrogenase enzyme, which is one of the most highly expressed proteins in the *C. thermocellum* proteome (Olson et al., 2013, 2010; Rydzak et al., 2012). While none of these promoters have been extensively characterized, the *cbp* promoter has been reliably used to drive expression of the *pyrF* (Tripathi et al., 2010) and *tdk* (Argyros et al., 2011) selectable markers. The eno promoter has been used to drive expression of an exogenous pyruvate kinase from *Thermoanaerobacterium saccharolyticum* (Deng et al., 2013). The transcription start sites have not been determined nor is anything known about which sigma factor is responsible for promoter recognition.

In the absence of detailed analysis, it is still useful to search for suitable promoters for expression by scanning those available. It is important to appreciate that any such study based on the assay of a reporter gene activity will also depend on mRNA stability and enzyme efficiency. An ideal promoter for metabolic engineering would have the following characteristics:

1. Low expression of the gene of interest in cloning strains of *Escherichia coli.*
2. Consistently high expression in *C. thermocellum*, independent of genetic context.
3. Low homology to the chromosome (in the case of native promoters, shorter is better).

* Corresponding authors.
E-mail addresses: dan268@gmail.com (D.G. Olson), Lee.R.Lynd@dartmouth.edu (L.R. Lynd).

http://dx.doi.org/10.1016/j.meteno.2015.03.002

The goal of this work is to identify new promoters for gene expression in *C. thermocellum*, and compare them with ones currently in use.

2. Materials and methods

2.1. Plasmid construction

Plasmids were constructed using standard molecular biology techniques (Sambrook and Russell, 2001) and isothermal DNA assembly (Gibson, 2011). Plasmid sequence was confirmed by Sanger sequencing of the promoter region and reporter gene. Plasmids were transformed into *C. thermocellum* DSM 1313 using standard techniques (Olson and Lynd, 2012a). Plasmids were also transformed in to *E. coli* C2566 (New England Biolabs) using standard techniques (Sambrook and Russell, 2001). The thermostable *lacZ* gene was a gift from James Liao (Lin et al., 2014). A list of the promoters and reporter genes in each plasmid is given in Table 1.

2.2. Choice of promoter regions

Promoter regions were chosen based on comparison of several sets of published gene expression data. These sets include Riederer et al. (2011), Raman et al. (2011), Gowen and Fong (2010) and Van der Veen et al. (2013). Note that for the last reference, strain characterization is described in the referred paper, but the transcription data is available from the GEO database, accession GSE27046. Promoters were selected on the basis of showing moderate to high expression of the genes they control across several of the data sets. For a given gene, the region upstream of the translation start site of the gene was selected. In general this region was 100–500 bp in length (Table 2). For promoters that are listed twice (i.e. cbp and cbp_2), the first version of the promoter caused problems with cloning, so a slightly different region was selected.

2.3. Making cell-free extracts

To prepare cell-free extracts (CFE), *C. thermocellum* cells were grown to mid-log phase. 10 ml of cells was centrifuged and the pellet was resuspended in B-PER buffer (Thermo Scientific, IL). The cells were lysed by addition of Lysonase according to the manufacturer's directions (EMD Millipore, MA). After centrifugation for 1 min at 15,000*g* to remove cell debris, the resulting supernatant was used for protein and enzyme assays. *E. coli* cells were also lysed using B-PER buffer and Lysonase enzyme. Protein concentration of the CFE was measured using the Coomassie Plus Bradford Assay (Thermo Scientific, IL) using bovine serum albumin as the standard. Absorbance was measured at 595 nm and 450 nm and the ratio of the two was used to determine concentration.

2.4. LacZ activity assay

The activity of the LacZ enzyme was determined by measuring the formation of the yellow cleavage product of o-nitrophenyl-β-D-galactoside (ONPG) as described by Miller (1972) with modifications as described below. The 0.2 ml assay solution contained 100 mM sodium phosphate pH 7.5, 1 mM magnesium chloride, 50 mM beta mercaptoethanol, 0.655 mg/ml ONPG and varying amounts of CFE. The assay solution was incubated at 37 °C for 4 h and the rate of increase in absorbance at 420 nm was measured in a 96-well plate with a BioRad Powerwave XS spectrophotometer. Purified LacZ enzyme (G4155 from Sigma) was used as a standard. Several 2-fold dilutions of CFE were added, and activity was determined in the linear range.

2.5. AdhB activity assay

AdhB is a bifunctional secondary alcohol and aldehyde dehydrogenase from *Thermoanaerobacter pseudethanolicus* ATCC 33223 (Teth39_0218) that is NADPH-dependent (Burdette et al., 1997). The activity of the AdhB enzyme was determined by measuring the disappearance of NADPH by measuring changes in absorbance

Table 1
Plasmids used in this study.

Plasmid	Promoter	Putative SigA/RpoD motif	Predicted conventional RBS Sequence	Predicted TIE strength *(arbitrary units)*[a]	Reporter gene
pDGO-66	gapDH	TTGA(A)A–N17–TA(A)AAT	AGGAGG	–	none
pDGO80	none			369	*adhB*
pDGO81	0544		GGAGG	5190	*adhB*
pDGO83	cbp	TTGA(A)(T)–N17–TATAAT	AGGAGG	244,912	*adhB*
pDGO84	eno	TTGA(A)A–N18–(C)AT(T)AT	GGAG	8763	*adhB*
pDGO86	1194	TTG(T)(T)(T)–N15–TATAAT	AGG(G)GG	19,141	*adhB*
pDGO87	0966	TTG(C)(A)(T)–N15–TGNTATAAT	AGGA	5587	*adhB*
pDGO88	2638	TT(A)A(A)A–N15–TATAAT	AGGAGG	42,645	*adhB*
pDGO89	0815	TT(T)A(A)A–N12–TGNTAT(T)AT	GAGG	19,843	*adhB*
pDGO90	2926		AGGAGG	15,731	*adhB*
pDGO92	0307		GGAG	2764	*adhB*
pDGO95	gapDH	TTGA(A)A–N17–TA(A)AAT	AGGAGG	7372	*lacZ*
pDGO98	none			–	*lacZ*
pDGO99	0544		GGAGG	716	*lacZ*
pDGO100	gapDH_2	TTGA(A)A–N17–TA(A)AAT	AGGAGG	1154	*lacZ*
pDGO102	eno	TTGA(A)A–N18–(C)AT(T)AT	GGAG	32,320	*lacZ*
pDGO104	1194	TTG(T)(T)(T)–N15–TATAAT	AGG(G)GG	18,381	*lacZ*
pDGO105	0966	TTG(C)(A)(T)–N15–TGNTATAAT	AGGA	1734	*lacZ*
pDGO106	2638	TT(A)A(A)A–N15–TATAAT	AGGAGG	6441	*lacZ*
pDGO107	0815	TT(T)A(A)A–N12–TGNTAT(T)AT	GAGG	2091	*lacZ*
pDGO108	2926	TGNTA(A)(T)AT	AGGAGG	7656	*lacZ*
pDGO109	2463		AGGAGG	8067	*lacZ*
pDGO110	0307		GGAG	858	*lacZ*
pDGO111	lac	TT(T)ACA–N18–TAT(G)(T)T	AGGA	390	*lacZ*
pDGO112	lacUV5	TT(T)ACA–N18–TATAAT	AGGA	390	*lacZ*
pDGO117	cbp_2	TTGA(A)(T)–N17–TATAAT	AGGAGG	50,011	*lacZ*
pDGO118	3011_2	TTGAC(T)–N17–TATAAT	AGGAGG	5978	*lacZ*

[a] See Section 2.10 for details about calculation of translation initiation efficiency (TIE).