Available online at www.sciencedirect.com
ScienceDirect
Metabolic Engineering 9 (2007) 133–141
METABOLIC ENGINEERING
www.elsevier.com/locate/ymben

Recombinant protein production in an *Escherichia coli* reduced genome strain

Shamik S. Sharma[a,1], Frederick R. Blattner[b], Sarah W. Harcum[c,*]

[a]*Department of Chemical and Biomolecular Engineering, Clemson University, Clemson, SC 29634, USA*
[b]*Department of Genetics, University of Wisconsin, Madison, WI 53706, USA*
[c]*Department of Bioengineering, Clemson University, Clemson, SC 29634, USA*

Received 11 July 2006; received in revised form 4 October 2006; accepted 11 October 2006
Available online 21 October 2006

Abstract

Recently, efforts have been made to improve the properties of *Escherichia coli* as a recombinant host by 'genomic surgery'—deleting large segments of the *E. coli* K12 MG1655 genome without scars. These excised segments included K-islands, which contain a high proportion of transposons, insertion sequences, cryptic phage, damaged, and unknown-function genes. The resulting multiple-deletion strain, designated *E. coli* MDS40, has a 14% (about 700 genes) smaller genome than the parent strain, *E. coli* MG1655. The multiple-deletion and parent *E. coli* strains were cultured in fed-batch fermenters to high cell densities on minimal medium to simulate industrial conditions for evaluating growth and recombinant protein production characteristics. Recombinant protein production and by-product levels were quantified at different controlled growth rates. These results indicate that the multiple-deletion strain's growth behavior and recombinant protein productivity closely matched the parent stain. Thus, the multiple-deletion strain *E. coli* MDS40 provides a suitable foundation for further genomic reduction.

Keywords: Reduced genome; Deletion; Growth rate; Acetate; Transposon; Insertion sequence

1. Introduction

Escherichia coli is one of the most studied and well-understood microorganisms. It naturally occurs in the gut of mammals, which is an anaerobic environment with constantly changing nutrient conditions. *E. coli* is also a commonly used recombinant host for laboratory and industrial recombinant protein production. As a recombinant host, *E. coli* is exposed to only a limited and controlled set of conditions. Specifically, in an industrial fermenter, an aerobic environment is usually desired and maintained, the nutrient concentrations are maintained within narrow ranges, and attachment to the vessel is not desirable. Therefore, the genes required for survival in the gut may not be the same genes required for optimum recombinant protein production. Further, the complete genome sequence of *E. coli* has revealed numerous genes of unknown function and genetic material that has possibly been acquired from other organisms in the recent past (Blattner et al., 1997). In an effort to improve *E. coli* as a recombinant host, many researchers have deleted or added single genes to the genome or modified plasmids to complement the existing genome (Andersen and Krummen, 2002; Andersen et al., 2001; Baneyx and Mujacic, 2004; Bessette et al., 1999; Cebolla et al., 2002; Chen et al., 2003; Chevalet et al., 2000; Chou et al., 1996; Diaz-Ricci et al., 1991; Pecota et al., 1997). These efforts have made progress, but have not truly addressed the global issue of the numerous genes in *E. coli* with unknown function or potentially detrimental function. In this work, a multiple deletion strain—*E. coli* MDS40—was characterized with respect to growth and recombinant protein production. *E. coli* MDS40 was derived from *E. coli* MG1655 by the

*Corresponding author. Fax: +1 864 656 0567.
E-mail address: harcum@clemson.edu (S.W. Harcum).
[1]Current address: New England Biolabs, 240 County Road, Ipswich, MA, USA.

1096-7176/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymben.2006.10.002

excision of many large nucleotide sequences corresponding to potentially non-essential genes, while retaining the backbone *E. coli* genome.

A few research groups have created multiple deletion *E. coli* strains, although not always with the focus of improving recombinant protein production. These multiple deletion methods mainly used random techniques facilitated by transposon libraries (Yu et al., 2002), specialized transposons (Goryshin et al., 2003), and λ phage homologous recombination (Hashimoto et al., 2005). One non-random or targeted deletion method, used sequential plasmid-directed recombination events (Kolisnychenko et al., 2002). Both the random and non-random multiple deletion techniques have generated strains with 8–30% smaller genomes. The random method-derived strains have been observed to have significantly lower growth rates, whereas the reported targeted multiple deletion strains have been observed to have comparable growth rates to the parent strain on the most commonly used laboratory media, including a minimal medium. The targeted deletion method strains were created by excising sections of the *E. coli* MG1655 genome regarded as unnecessary or detrimental, with the aim of improving its properties (Kolisnychenko et al., 2002; Pósfai et al., 2006). The targeted genes were selected based on a comparison of the complete genome sequences of two *E. coli* strains (MG1655 and O157:H7). This analysis revealed a 'backbone' of genetic information, which contained hundreds of strain specific 'islands'. For *E. coli* MG1655, a K12 strain, these islands were designated K-islands. The backbone genome is highly conserved and contained all essential genes. In contrast, the 'islands' contains a disproportionate number of genes with unknown function, toxin genes, transposable elements, and pseudogenes (Blattner et al., 1997; Kolisnychenko et al., 2002). In the first targeted multiple deletion strain described—*E. coli* MDS12—twelve sequential deletions of K-islands were made to *E. coli* MG1655, resulting in a multiple deletion strain with an 8% smaller genome. Twenty-eight further sequential deletions resulted in the strain *E. coli* MDS40, which was used in this study. *E. coli* MDS40 has a 14% smaller genome than *E. coli* MG1655 or roughly 700 fewer genes. Additionally, these deletion strains have improved transformabilities and lower mutation rates (Pósfai et al., 2006).

The focus of this study was to determine if *E. coli* MDS40 was a robust and suitable host for recombinant protein production. Fed-batch fermentations were used to obtain high cell densities under controlled growth conditions in minimal medium typical of industrial processes. Cell yield, byproduct accumulation, and recombinant protein productivity were measured and compared to the parent strain, *E. coli* MG1655. Another aim was to determine the suitability of *E. coli* MDS40 as the primary multiple deletion strain for further targeted deletions. Additional deletions could, for example, target the arabinose utilization genes, such as the *E. coli* HB101 strain, to allow recombinant protein expression using the pBAD vector (Khlebnikov et al., 2002; Sorensen and Mortensen, 2005).

2. Materials and methods

2.1. Strains and expression vector

E. coli MG1655 was obtained from American Type Culture Collection (ATCC). The multiple deletion strain *E. coli* MDS40 was provided by Scarab Genomics (Madison, WI). *E. coli* MDS40 was created by 40 successive targeted deletions from the *E. coli* MG1655 genome i.e., 28 targeted successive deletions from *E. coli* MDS12 (Kolisnychenko et al., 2002; Pósfai et al., 2006). Chloramphenicol acetyltransferase (CAT) was used as the model recombinant protein encoded by the plasmid pPROEXCAT (Invitrogen). The pPROEXCAT plasmid contains a *trc* promoter that controls CAT expression via IPTG-induction, a pBR322 origin of replication, and the β-lactamase gene for ampicillin resistance. *E. coli* MDS40 was transformed with pPROEXCAT via the $CaCl_2$ method (Sambrook et al., 1989). CAT was selected as the initial model recombinant protein for a few reasons. Namely, there is an extensive body of literature available for this protein's expression in *E. coli* MG1655 and related strains. CAT levels in the cell can be easily quantified by an enzymatic assay (Rodriguez and Tait, 1983). Additionally, previous studies have observed that the unusual amino acid composition of CAT (11% phenylalanine) can trigger metabolic burden or stress responses in the host cell (Harcum and Bentley, 1993).

2.2. Shake flask cultivation

For the shake flask experiments, stock cultures stored at $-80\,^{\circ}C$, were thawed and 0.5 ml was used to inoculate 10 ml of defined medium. The medium was as described in Korz et al. (1995), except the batch medium contained 8 g/L KH_2PO_4, 0.4 g/L $MgSO_4$ and 100 μg/ml ampicillin. Precultures grew overnight and were used to inoculate the shake flasks. The volume of preculture was adjusted, such that the shake flasks had equal initial cell densities. Shake flasks were induced with 5 mM IPTG, unless otherwise indicated.

2.3. Fed-batch fermentation

Fermentations were conducted in a 2 L Bioflo 110 fermenter (New Brunswick Scientific, Edison, NJ). Defined batch medium was used (Korz et al., 1995), except the batch medium contained 5 g/L glucose, 8 g/L KH_2PO_4, 0.4 g/L $MgSO_4$ and 100 μg/ml ampicillin. The feed medium contained 481 g/L glucose, 4 g/L $MgSO_4$, 40 mg/L Fe^{III}citrate, and 1.5 × the trace metal concentration of the batch medium. Stock cultures stored at $-80\,^{\circ}C$ were thawed and used to inoculate 100 ml of defined media. Precultures were grown at $37\,^{\circ}C$ to a cell density (OD_{600}) of approximately