Biotechnology Advances 30 (2012) 1102–1107

Contents lists available at SciVerse ScienceDirect

Biotechnology Advances

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

Bacterial expression systems for recombinant protein production: *E. coli* and beyond

Rachel Chen *

School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0100, United States

ARTICLE INFO

Available online 24 September 2011

Keywords:
Recombinant protein production
Bacterial expression system
E. coli
Lactoccocus lactis
Pseudomonas expression system
Streptomyces expression system

ABSTRACT

Escherichia coli expression system continues to dominate the bacterial expression systems and remain to be the preferred system for laboratory investigations and initial development in commercial activities or as a useful benchmark for comparison among various expression platforms. Some new developments in overcoming its shortcomings are reviewed in this paper, including antibiotics-free selection plasmids, extracellular production, and posttranslational modifications. The ability for *E. coli* to make mg glycosylated proteins promises even broader applications of the *E. coli* system in the future. Significant progresses have also been made over the past few years in alternative bacterial expression systems. Notably, the *Lactoccocus lactis* system has proven to be a viable choice for membrane proteins. Additionally, several *Pseudomonas* systems were developed and achieved product titers comparable to *E. coli* systems. Other bacterial systems such as *Streptomyces*, coryneform bacteria, and halophilic bacteria offer advantages in some niche areas, providing more choices of bacterial expression systems for recalcitrant proteins.

1. Introduction

Recombinant protein production is an enormous field. There seems no sign that the expansion of this field will abate anytime soon. Impressive progresses in the recombinant protein technology over the past decades have brought hundreds of therapeutic proteins into clinical applications. As there are hundreds more therapeutic proteins in clinical trials, researches aimed to better the technology will continue to speed ahead. The advent of systems biology era is an important driving force that propels the field. The desire to understand the functions of hundreds and thousands of proteins, whose sequences were just recently made available, demands new approaches that deliver each and every single protein (including membrane proteins) in quantities and quality dictated by the structural and biochemical analysis. Moreover, the inherent biodiversity in both peptide sequences and post-translational modifications adds significantly to the complexity of the task to deliver a bioactive protein, because it requires not only the synthesis of peptide backbone but also its authentic modification. Many ingenious strategies were developed over the recent years to meet these challenges. This review hopes to capture some of the newest development.

Understandably, providing a review for such a broad field is difficult, even with the confine of prokaryotic expression systems. To keep the task manageable, this review will cover references from 2006 with only a few exceptions, where continuity demands inclusion of these older references. Many excellent reviews have been published over the recent years (Demain and Vaishnav, 2009; Makino et al., 2011a; Zerbs et al., 2009). In order to make this review a meaningful contribution to the field, I intend to complement the existing reviews by focusing on areas that have received less coverage in other recent reviews. Three bacterial expression systems are reviewed in some details, *Escherichia coli*, *Lactoccocus lactis*, and *Pseudomonas*. Bacillus system is not reviewed here as excellent reviews are published recently (Nijland and Kuipers, 2008; Pohl and Harwood, 2010). Only selected aspects are reviewed for each chosen system. For example, in *E. coli* the focus is on posttranslational modification and trend of developing antibiotic-free selection systems. On *L. lactis*, membrane protein expression is emphasized. Where possible, a comparison to *E. coli* system is made. Other important developments are only briefly mentioned in the review and recent reviews are referenced.

2. *E. coli* systems: new development

E. coli expression system continues to dominate the bacterial expression systems and remain to be the first choice for laboratory investigations and initial development in commercial activities or as a useful benchmark for comparison among various expression platforms. *E. coli* system is also the basis for efforts in protein engineering and high-throughput structural analysis (Gordon et al., 2008; Koehn and Hunt, 2009). Reviews also appeared recently on general aspects of using the *E. coli* system and more specific topics such as single-domain antibody fragments (De Marco, 2011; Koehn and

* Tel.: +1 404 894 1255; fax: +1 404 894 2866.
E-mail address: rchen@chbe.gatech.edu.

0734-9750/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.biotechadv.2011.09.013

Hunt, 2009), stress and stress responses associated with recombinant protein production (Sevastsyanovich et al., 2010).

Extracellular production of proteins is highly desirable as it could greatly reduce the complexity of a bioprocess and improve product quality. The inability of *E. coli* in secretion of protein products to growth medium has long been considered as a major drawback of the system. However, the dogma that *E. coli* secretes no protein has been challenged recently by the recognition of both its natural ability to secrete protein in common laboratory strains and increased ability to secrete proteins in engineered cells. Significant efforts have been made in recent years to address this issue (Ni and Chen, 2009). A survey of recent literature shows that the strategies fall into four categories, 1) engineering dedicated secretion systems that naturally exist in pathogen *E. coli*; 2) use carrier proteins with no known translocation mechanisms; 3) use cell envelope mutants; and 4) co-expression a lysis-promoting protein (Kil) (Ni and Chen, 2009). Efforts in enhancing its ability to secrete proteins to growth medium are significant and some successful technology has begun to impact the production beyond laboratory investigations. Additional reviews on the topic include the review on using autotransporters (Jong et al., 2010).

While potential problems of using antibiotic and antibiotic-resistance markers in large scale recombinant protein production have long been recognized, only recently, efforts in developing alternative methods for selections have born fruits in practical sense. Additionally, engineering glycosylation and other post-translational modification into the system is an exciting development in recent years. These two areas will be reviewed in the following sections.

2.1. Glycosylation and other post-translational modifications

Just a few years ago, it was widely believed that bacteria were incapable of making glycosylated proteins. The discovery of N-linked glycosylation system in a Gram-negative bacterium *Campylobacter jejuni* and subsequent transfer of the system to *E. coli* (Wacker et al., 2002) opened up the exciting possibility to use *E. coli* to synthesize glycoproteins. Since this seminal work, *E. coli* was shown to glycosylate the native PglB (the oligosaccharide transferase from *C. Jejuni*) substrate, AcrA, with various O-antigen polysaccharides native to *E. coli* (Feldman et al., 2005; Wacker et al., 2006). More recently, attempts were made to produce glycosylated proteins other than the native PglB substrate AcrA or other types of glycans non-native to either *E. coli* or *C. jejuni* (Fisher et al., 2011; Ihssen et al., 2010; Lizak et al., 2011; Pandhal et al., 2011). Ihssen et al. used the system to attach *Schigella O* antigens to two carrier proteins derived from *C jejuni* and *Pseudomonas aeruginosa*, respectively. The resulting glycoproteins have the potential to be used as vaccines against shigellosis, a disease causing 1 million deaths per year in developing world, mostly of young children (Ihssen et al., 2010). In this work, the host strain CLM24 carries a chromosomal deletion of the *waaL* gene encoding an O polysaccharide ligase. The genes encoding the necessary components: PglB, carrier proteins, and Shiga O antigen biosynthesis were put on three compatible plasmids. Successful production of glycoproteins was detected by using the anti-shigella O1 antibodies. A ladder of glycoprotein bands were observed for each glycoprotein, due to polymerization of the repeating O-units by the action of the enzymes Wzy and Wzz. This study also identified several challenges of the glycosylation. First challenge is that glycosylation is often incomplete. In fact, glycosylation efficiency is sometimes as low as single digit percentage. When more than one glycosylation sites are present, heterogeneity is also an issue. Second challenge is the growth impairment by one or more glycosylation components and conditions for optimal production of carrier proteins may be different from those leading to optimal glycosylation, necessitating an exhaustive search for best conditions. A fed-batch process established after significant optimization led to a yield of 18–24 mg L^{-1} and productivity of 0.75–1.0 mg L^{-1} h^{-1} (Ihssen et al., 2010). In another work by Lizak et al., antibody fragments, a potential class of therapeutic proteins, were shown to be effectively N-glycosylated with the *C. jejuni* PglB system, improving the stability and solubility of the protein (Lizak et al., 2011). Specifically, the murine single-chain fragment of the anti-His tag antibody 3D5 was engineered to carry two N-glycosylation sites with the optimal sequence DQNAT within the flexible linker region connecting the variable light chain (V_L) with the variable heavy chain (V_H) (Lizak et al., 2011). Production of the diglycosylated 3D5 was successful, although the efficiency was decreased when scaled up to 5 L. Change of antibiotic resistance gene from chloramphenicol to kanamycin was found to be helpful to retain *pglB*-containing plasmid. After column purification, a yield of 2 mg L^{-1} diglycosylated protein was obtained. This compared to 8 mg L^{-1} naked protein under similar conditions. While retaining the same affinity to the target antigen, the diglycosylated 3D5 was found to be more resistant to proteolysis, presumably due to steric hinderance. Additionally, the solubility of the protein was increased 2.5 fold (Lizak et al., 2011). Thus, N-glycosylation could not only be used to produce proteins with the biological activity but also be used as an effective tool to modify physical property of a protein. Therefore, it seems that *E. coli* can be used to produce mg of glycosylated proteins with the desired glycans.

It was reported that in eukaryotic cells, up to 98% proteins are N-terminally acetylated (Johnson et al., 2010). Additionally, N^{ε}-Acetylation of lysine is also common. Functional importance of acetylation probably rivals that of phosphorylation (Newmann et al., 2008). Acetylation, among others, influences DNA replication, repair and recombination, maintenance of genomic stability, cytoskeletal dynamics, metabolism, signal transduction, protein folding and trafficking (Newmann et al., 2008). It is thus important to generate authentic acetylation on the peptide backbone in order to produce bioactive recombinant proteins. In 2005, a fusion strategy for acetylation was reported (Acharya et al., 2005). In this work, Acharya et al. fused a yeast acetyltransferase with the fragment of the tumor suppressor p53 protein (residues 320–356) and found that the lysine residue at 320 of p53 fragment was successfully acetylated (Acharya et al., 2005), suggesting that the yeast acetyltransferase is functional in *E. coli* system. It is not known from this work, however, whether the fusion is absolutely necessary. In a more recent study, a different approach was used. Modification of the lysine residue, acetyllysine, was incorporated into several recombinant proteins as a non-natural amino acid. Accordingly, N^{ε}-acetyllysyl-tRNA synthetase/$tRNA_{CUA}$ pair was generated through protein engineering (Newmann et al., 2008). An endogenous activity deacetylase-CobB, apparently interfered with this acetylation strategy. Thus it was necessary to use an inhibitor, nicotinamide, to suppress the deacetylation activity. As reported, recombinant acetylated manganese superoxide dismutase was successfully produced with the modification at position 44 (Newmann et al., 2008). For amino terminal acetylation, Johnson et al. showed that co-expression of yeast NatB acetylation complex with the substrate proteins is very effective. The complex recognizes three terminal sequences, Met.Asp, Met.Glu, and Met.Asn. Several proteins from both yeast and human origins bearing one of the three sequences were acetylated (Johnson et al., 2010). In another work, Fang et al. discovered that RimJ is an endogenous acetyltransferase in *E. coli*, responsible for the partial acetylation of recombinant peptide, thymosin (Fang et al., 2009). Upon further optimization in a follow up work, they showed that a complete acetylation of the N-terminal serine residue was achieved when RimJ was overexpressed, resulting a high-yield and fully acetylated product (Ren et al., 2011). It would be interesting to see how effective the enzyme works with other substrates varying in sizes as this has only been shown for a 28-residue peptide.

Although not as widely reported, using similar approach as acetylation, phosphorylation of recombinant proteins could also be carried out in *E. coli*. Murata et al. showed that by co-expressing human Jun N-terminal kinase and its cognate substrate (human Jun dimerization protein 2) in *E. coli*, phosphorylated human Jun dimerization protein 2 was produced (Murata et al., 2008). Similarly,