Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Biotechnology 115 (2005) 113–128
JOURNAL OF Biotechnology
www.elsevier.com/locate/jbiotec

# Advanced genetic strategies for recombinant protein expression in *Escherichia coli*

Hans Peter Sørensen, Kim Kusk Mortensen*

*Laboratory of BioDesign, Department of Molecular Biology, Aarhus University, Gustav Wieds Vej 10 C, DK-8000 Aarhus C, Denmark*

Received 29 March 2004; received in revised form 26 August 2004; accepted 30 August 2004

**Abstract**

Preparations enriched by a specific protein are rarely easily obtained from natural host cells. Hence, recombinant protein production is frequently the sole applicable procedure. The ribosomal machinery, located in the cytoplasm is an outstanding catalyst of recombinant protein biosynthesis. *Escherichia coli* facilitates protein expression by its relative simplicity, its inexpensive and fast high-density cultivation, the well-known genetics and the large number of compatible tools available for biotechnology. Especially the variety of available plasmids, recombinant fusion partners and mutant strains have advanced the possibilities with *E. coli*. Although often simple for soluble proteins, major obstacles are encountered in the expression of many heterologous proteins and proteins lacking relevant interaction partners in the *E. coli* cytoplasm. Here we review the current most important strategies for recombinant expression in *E. coli*. Issues addressed include expression systems in general, selection of host strain, mRNA stability, codon bias, inclusion body formation and prevention, fusion protein technology and site-specific proteolysis, compartment directed secretion and finally co-overexpression technology. The macromolecular background for a variety of obstacles and genetic state-of-the-art solutions are presented.


*Keywords:* *Escherichia coli*; Recombinant protein expression systems; Inclusion bodies; Fusion proteins; Rare codon tRNAs

## 1. The modern recombinant expression system

A number of central elements are essential in the design of recombinant expression systems (Baneyx, 1999; Jonasson et al., 2002). Expression is normally induced from a plasmid harboured by a system compatible genetic background. The genetic elements of the expression plasmid include origin of replication (*ori*), an antibiotic resistance marker, transcriptional promoters, translation initiation regions (TIRs) as well as transcriptional and translational terminators.

### 1.1. The replicon

The replicon of plasmids contain the origin of replication and in some cases associated *cis* acting elements (del Solar et al., 1998). Most plasmid vectors used in re-

* Corresponding author. Fax: +45 86 12 31 78.
*E-mail address:* kkm@mb.au.dk (K.K. Mortensen).

0168-1656/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbiotec.2004.08.004

combinant protein expression replicate by the ColE1 or the p15A replicon. Plasmid copy number is controlled by the origin of replication that preferably replicates in a relaxed fashion (Baneyx, 1999). The ColE1 replicon present in modern expression plasmids is derived from the pBR322 (copy number 15–20) or the pUC (copy number 500–700) family of plasmids, whereas the p15A replicon is derived from pACYC184 (copy number 10–12). These multi-copy plasmids are stably replicated and maintained under selective conditions and plasmid free daughter cells are rare (Summers, 1998). Plasmid incompatibility is defined as the inability of two plasmids to be stably maintained in the same cell (Hardy, 1987). Different replicon incompatibility groups and drug resistance markers are required when multiple plasmids are employed for the co-expression of gene products. Derivatives containing ColE1 and p15A replicons are often combined in this context since they are compatible plasmids (Mayer, 1995).

### 1.2. Resistance markers

The most common drug resistance markers in recombinant expression plasmids confer resistance to ampicillin, kanamycin, chloramphenicol or tetracycline. Plasmid mediated resistance to ampicillin is accomplished by expression of β-lactamase from the *bla* gene. This enzyme is secreted to the periplasm, where it catalyse hydrolysis of the β-lactam ring. Ampicillin present in the cultivation medium is especially susceptible to degradation, either by secreted β-lactamase, or acidic conditions in high-density cultures. The latter effect can be alleviated by the less degradation susceptible ampicillin analog, carbenicillin, Kanamycin, chloramphenicol and tetracycline interfere with protein synthesis by binding to critical areas of the ribosome. Kanamycin is inactivated in the periplasm by aminoglycoside phosphotransferases and chloramphenicol by the *cat* gene product, chloramphenicol acetyl transferase. Various genes confer resistance to tetracycline (Connell et al., 2003).

### 1.3. Promoters

Recombinant expression plasmids require a strong transcriptional promoter to control high-level gene expression. Basal transcription in the absence of inducer is minimized through the presence of a suitable repressor. Minimization of basal transcription is especially important when the expression target introduce a cellular stress situation and thereby selects for plasmid loss. Promoter induction is either thermal or chemical and the most common inducer is the sugar molecule isopropyl-beta-D-thiogalactopyranoside (IPTG) (Hannig and Makrides, 1998).

### 1.4. Messenger RNA

Translation initiation from the translation initiation region (TIR) of the transcribed messenger RNA require a ribosomal binding site (RBS) including the Shine–Dalgarno (SD) sequence and a translation initiation codon (Sørensen et al., 2002). The Shine–Dalgarno sequence is located $7 \pm 2$ nucleotides upstream from the initiation codon, which is the canonical AUG in efficient recombinant expression systems (Ringquist et al., 1992). Optimal translation initiation is obtained from mRNAs with the SD sequence UAAGGAGG. The RBS secondary structure is highly important for translation initiation and efficiency is improved by high contents of adenine and thymine (Laursen et al., 2002). Translation initiation efficiency is in particular influenced by the codon following the initiation codon and adenine is abundant in highly expressed genes (Stenstrom et al., 2001).

A transcription terminator placed downstream from the sequence encoding the target gene, serves enhancement of plasmid stability by preventing transcription through the origin of replication and from irrelevant promoters located in the plasmid. Transcription terminators stabilize the mRNA by forming a stem loop at the three prime end (Newbury et al., 1987). Translation termination is preferably mediated by the stop codon UAA in *Escherichia coli*. Increased efficiency of translation termination is achieved by insertion of consecutive stop codons or the prolonged UAAU stop codon (Poole et al., 1995).

### 1.5. Current expression systems

A wealth of expression systems designed for various applications and compatibilities are available. Approximately 80% of the proteins used to solve three-dimensional structures submitted to the protein data bank (PDB) in 2003 were prepared in an *E. coli* ex-