Expression of Highly Toxic Genes in *E. coli*: Special Strategies and Genetic Tools

F. Saïda[1,*], M. Uzan[3], B. Odaert[2] and F. Bontems[2]

[1]*Center for Molecular Genetics, University of California San Diego, 9500 Gilman Drive, La Jolla CA 92093-0634, USA;* [2]*ICSN-RMN, Institut de Chimie des Substances Naturelles, CNRS de Gif-sur-Yvette, 91190 Gif-sur-Yvette et Ecole Polytechnique, 91128 Palaiseau, France;* [3]*Institut Jacques Monod, CNRS et Universités Paris VI et Paris VII, 2 Place Jussieu, 75251 Paris Cedex 05, France*

Abstract: *Escherichia coli* (*E. coli*) remains the most efficient widely-used host for recombinant protein production. Well-known genetics, high transformation efficiency, cultivation simplicity, rapidity and inexpensiveness are the main factors that contribute to the selection of this host. With the advent of the post-genomic era has come the need to express in this bacterium a growing number of genes originating from different organisms. Unfortunately, many of these genes severely interfere with the survival of *E. coli* cells. They lead to bacteria death or cause significant defects in bacteria growth that dramatically decrease expression capabilities. In this paper, we review special strategies and genetics tools successfully used to express, in *E. coli,* highly toxic genes. Suppression of basal expression from leaky inducible promoters, suppression of read-through transcription from cryptic promoters, tight control of plasmids copy numbers and proteins production as inactive (but reversible) forms are among the solutions presented and discussed. Special expression vectors and modified *E. coli* strains are listed and their effectiveness illustrated with key examples, some of which are related to our study of the highly toxic phage T4 restriction endoribonuclease RegB. We mainly selected those strategies and tools that permit *E. coli* normal growth until the very moment of highly toxic gene induction. Expression then occurs efficiently before cells die. Because they do not target a particular toxic effect, these strategies and tools can be used to express a wide variety of highly toxic genes.

Keywords: Highly toxic, expression, *Escherichia coli*, tight control, RegB.

INTRODUCTION

Recombinant protein production in *E. coli* takes advantages of the extensive knowledge on the genetics and the molecular biology of the bacterium. Although general strategies for achieving high-level expression of genes in this organism have been heavily reviewed [1-5], a review exclusively dedicated to the expression of highly toxic genes is still missing.

This paper presents and discusses updated strategies and genetic tools used to express in *E. coli* highly toxic genes. We define a highly toxic gene as a gene that, when introduced into an *E. coli* expression strain, causes cell death or severe cultivation and maintenance defects during the growth phase (prior to induction). Genes that exhibit toxic effects only when they are overproduced (after induction) are not the subject of this paper.

The main goal of the solutions presented in this paper is to enable *E. coli* cells to tolerate the highly toxic gene during the growth phase. After induction, efficient expression signals will ensure rapid and quantitative production of the highly toxic protein before cells die. For clarity purposes, we structured the paper in five main sections: manipulation of highly toxic genes transcriptional and translational control elements, manipulation of the coding sequence, manipulation of the copy number, addition of stabilizing sequences and empirical selection of *E. coli* strains. A short sixth section, named "Special expression protocols", is summarized in Table **3**.

All expression vectors and *E. coli* strains described in this paper are listed in Table **1** and Table **2**, respectively. We choose not to include "periplasmic secretion" in this review because of the heavily reviewed topic in the literature and because of the low expression-purification yields of this method.

I- MANIPULATION OF HIGHLY TOXIC GENES TRANSCRIPTIONAL AND TRANSLATIONAL CONTROL ELEMENTS

Tight control of highly toxic genes transcription and translation is the first key for a successful expression in *E. coli*. Hereafter, we present and discuss transcriptional strategies that reduce the amounts of highly toxic genes mRNAs during the growth phase. These strategies include the phage-mediated delivery of high specificity RNA polymerase, the use of an intrinsically tightly regulated promoter, the use of antisense, competitive, invertible and hybrid promoters, and the judicious positioning of strong transcription terminators. We will also discuss the use of "mRNA switch" mechanism to tightly control the translatability of highly toxic genes

*Address correspondence to this author at the Center for Molecular Genetics, University of California San Diego, 9500 Gilman Drive, La Jolla CA 92037-0634, USA; Tel: (858) 534-2451; Fax: (858) 534-7073; E-mail: fsaida@biomail.ucsd.edu.

Table 1. **Vectors Suitable for the Expression of Highly Toxic Genes**

Vector	Features	References, Section/Supplier
pLysS	Constitutive expression of lyzozyme, a natural inhibitor of T7 RNAP	7, I.1/Novagen
pLysE	Expression of more lyozyme than pLysS	7, I.1/Novagen
pBAD series	Tight regulation of gene expression (induction/repression up to 1200)	11, I.2/Invitrogen
pDOC55	Basal expression decreased by the use of a 3' antisense promoter	14, 16, I.3
pNH series	Suppression of basal expression using invertible promoters	17-19, I.4
pACYCT7pmt	Plasmid containing competitors T7 promoters	10, I.5
pLAC11, 22, 33	Use of the highly repressible full-lethgh *lac* promoter-operator region	25, I.6
pPROTet.E series and pPROTLar.A series	Highly respressible hybrid promoters	27, I.7/BD Biosciences
pT7SC series	Blocking read-through transcription by upstream terminators	29 30, I.8/USB
pTYB series	Semisynthesis of highly toxic proteins	44, 45, II.4/New England Biolabs
pOU61	Thermal control of the copy number	47, III.1
pETcoco-1	Arabinose-based control of the copy number	III.2/Novagen
pETKmS1, 2, 3	Addition of stabilizing sequences	52, IV

Table 2. *E. coli* **Strains Suitable for the Expression of Highly Toxic Genes**

E. coli strain	Features	References, Section/Supplier
BL21-AI	T7 RNAP gene under the control of the P_{BAD} promoter	I.2/Invitrogen
PB4144	The strain carries a temperature-sensitive allele of the *supF* suppressor that converts an UAG stop signal into Tyrosine	38, 39, II.2
ECF529, ECF530	The strains carry chromosol copies of the plasmid replication regulator under the control of the P_{BAD} promoter	48, III.2
CopyCutter[TM] EPI400[TM]	Regulation of the copy number of ColE1-type plasmids using a chromosomal copy of the *pcnB* gene under the control of an inducible promoter	49, III.3/Epientre
C41(DE3) and C43(DE3)	Empirically-selected strains with lower basal expression and slower induction rates compared to the original BL21(DE3)	54, V/Avidis
BL21(DE3)NH	Empirically-selected strain that tolerates the expression of instable genes using an unknown mechanism	55, V
ABLE C	The strain reduces 4-fold the copy number of ColE1-derived plasmids	III.3/Stratagene
ABLE K	The strain reduces 10-fold the copy number of ColE1-derived plasmids	III.3/Stratagene

mRNAs. It is important to note that combinations of several strategies are very often needed in order to quantitatively express a highly toxic gene.

I.1. Phage-Mediated Delivery of Highly Selective T7 RNA Polymerase

In their first paper describing the design of an expression system based on the highly selective T7 RNA polymerase (T7 RNAP), Studier F.W. and Moffat B.A. stressed on the fact that small amounts of T7 RNAP provided from a single copy of phage T7 *gene 1* are sufficient to direct high transcription levels from phage T7 gene 10 promoter (T7 promoter) located on a multi-copy plasmid [6]. In the BL21(DE3) expression strain, T7 *gene 1* is under the control of a promoter poorly shut off under repressive conditions (*lac*UV5 promoter). The basal level of T7-dependent transcription in this strain can be reduced by constitutive production of T7 lyzozyme, a natural inhibitor of T7 RNAP, using plasmids pLysS and pLysE [7]. Very often plasmids carrying highly toxic genes cannot be maintained in the BL21(DE3) strain regardless of the presence of lyzozyme-

Table 3.   Special Expression Protocols

Method	Principle	Brief Description	Example	References
Plating method	Selection of plasmid-containing cells is more stringent on LB + ampicillin agar plates than in liquid culture media.	After transformation, one single colonie is resuspended in 200 µl of water and plated on LB + ampicillin for 16 hours at 37 °C. The 10^{+3}-10^{+4} obtained colonies are scrapped-off and resuspend in 400 ml of LB + 50 µg/ml ampicillin. When OD reaches 0.4, 50 µg/ml ampicillin is added for the second time and the culture is induced with IPTG for 3 hours at 37°C.	GDNF [1]	56
Use of antibiotic timentin	Timentin is designed to prevent the growth of bacteria using penicillinases secreted by the small number of bacteria that retain the expression plasmid.	Use 75 µg/ml timentin instead of ampicillin. Timentin (Smith-Kline Beecham) is a combinaison of ticarcillin (a semisynthetic derivative of penicillin) and clavulanic acid (a -lactamase inhibitor).	hER [2]	57-58
Induction with very low IPTG concentrations (10 µM)	Cells induced using very low IPTG concentrations have more metabolic control over the toxic effect than those induced with standard IPTG concentrations (0.1- 1.0 mM).	Test IPTG concentrations in the range 0 -10 µM.	CENP-B [3]	59
Expression in M9 + 1% Glucose instead of standard LB	Catabolic repression reduces the basal expression level of T7 RNAP in the BL21(DE3) strain	In BL21(DE3), the T7 RNAP gene is under the control of the *lac*UV5 promoter, a variant of the *lac* promoter with the point mutation (G to A at -9). This mutattion does not abolish its affinity to the cAMP/CAP activation complex. It only decreases it by 5-10 fold. Using the minimal M9 medium in the presence of 1 % Glucose , catabolic repression causes a decrease in the basal expression of T7 RNAP from the *lac*UV5 promoter.	HIV-1 protease	10, 60

[1] Glial cell line-Derived Neurotrophic Factor.
[2] Human Estrogen Receptor
[3] Hamster CENtromere binding Protein B.

encoding plasmids. In some cases, transforming an *E. coli* strain that does not contain the T7 *gene 1* can circumvent the problem of toxicity. Highly toxic gene expression is then induced at mid-log growth phase by delivering the T7 RNAP through infection by phage derivatives containing T7 *gene 1*. Infection by -derivatives DE3 and CE6 has been successfully used to express, respectively, the phage T4 translational repressor RegA and the phage T4 restriction endoribonuclease RegB and transcriptional activator MotA [8-9, Uzan, M., unpublished result]. Infection by an M13 derivative phage containing T7 *gene 1* has been also described for the expression of the human immunodeficiency virus type 1 (HIV-1) protease [10]. Our experience with the phage T4 endoribonuclease RegB shows that the multiplicity of infection (MOI = number of infecting phage particles per *E. coli* cell) is a critical parameter that should be optimized in the range of 5-20 (Saida F., unpublished data).

I.2. Tight Regulation and High-Level Expression using the P$_{BAD}$ Promoter

A ratio of induction over repression of 1200 has been reached with a series of expression vectors containing the P$_{BAD}$ promoter of the *araBAD* arabinose operon and the gene coding the positive and negative regulator of this promoter, *araC* [11] (see Fig. (**1**) for a scheme of the regulation). The basal expression level of these vectors (pBAD, Invitrogen) can be reduced to very low levels in the presence of glucose and to extremely low levels in the presence of glucose-6-

phosphate which catabolite represses the system more efficiently than glucose. Furthermore, the induction level can be controlled over a wide range of inducer (arabinose) concentrations. Despite the 8 to 10-fold reduced efficiency of *E. coli* RNA polymerase as compared to T7 RNAP, the pBAD system has been demonstrated to be very efficient to express highly toxic proteins because of the enrichment of *E. coli* cultures with plasmid-bearing cells during the growth phase. It has been used to express the highly toxic archeal RNase P in amounts compatible with structural studies [12, 13]. Recently, the benefits of the tightly regulated P$_{BAD}$ promoter and the efficient T7 expression system has been combined in the construction of an *E. coli* strain in which the chromosomal copy of the T7 RNAP is under the control of the P$_{BAD}$ promoter. Compared to BL21(DE3)pLysS, the new Arabinose-Induced host BL21-AI (Invitrogen) has a 4 fold lower basal level but a similar expression level after induction with 0.2 % w/v arabinose and 1 mM isopropyl- -D-galactopyranoside (IPTG). When using an expression system under the control of the T7 promoter, BL21-AI strain should be the first strain to investigate if BL21(DE3)pLysS and BL21(DE3)pLysE strains fail to maintain the target gene.

I.3. The Antisense Strategy

Originally designed to express in *E. coli* the restriction endonuclease EcoRI after inactivation of its cognate methylase [14], this strategy was used to express highly toxic genes with no known endogenous inhibitors. Bovine pancre-
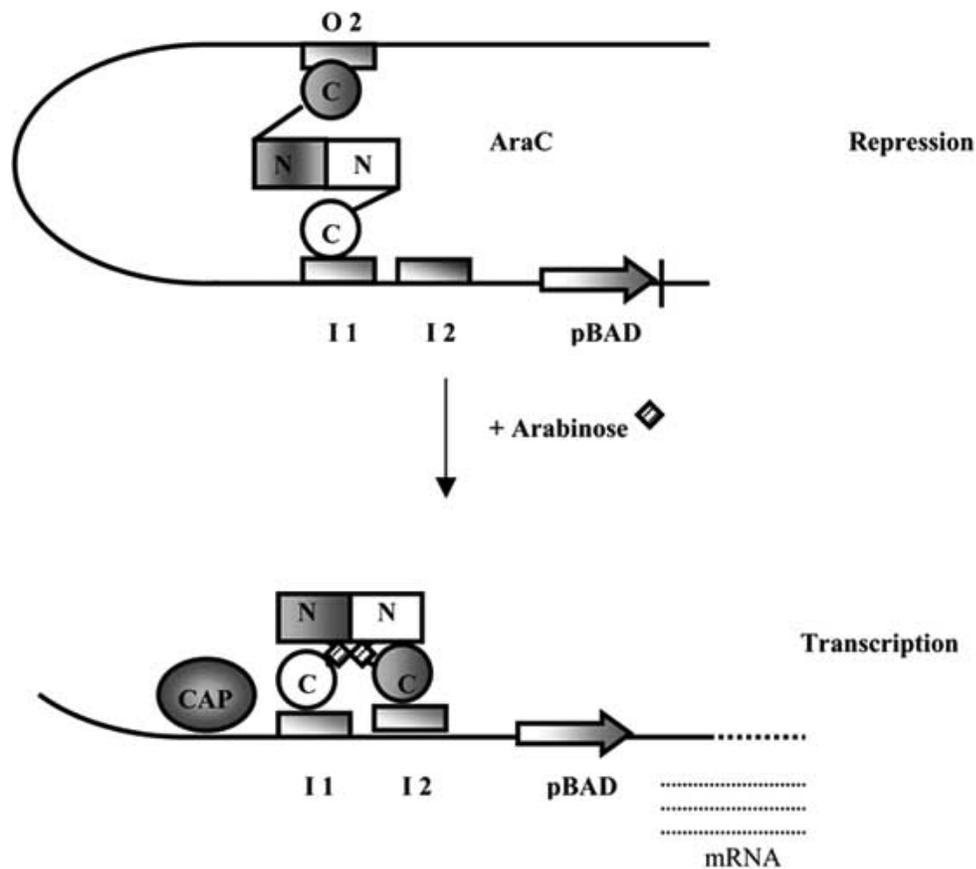
Fig. (1). Control of the P$_{BAD}$ promoter. When arabinose binds to the *araC* repressor, it releases the O2 site and binds to the I2 site. This relaxes the DNA loop and transcription starts. The cAMP activator protein (CAP) stimulates the binding of *araC* to I1 and I2. N and C respectively designate the N-terminus and the C-terminus of the *araC* repressor.

atic DNase I is an endonuclease that cleaves double-stranded DNA to yield 5'-phosphorylated polynucleotides [15]. The lack of success in sub-cloning this enzyme from an M13-derived cloning vector to an expression vector controlled by the *tac* promoter has not been circumvented by the use of the highly repressed P$_L$ promoter [16]. Adding an antisense *lac* promoter downstream of the cloned gene was crucial to minimize the basal level of the highly toxic endonuclease [16]. In the presence of the *lac* inducer IPTG, the *lac* promoter directs the synthesis of antisense mRNAs that form RNA-duplexes preventing the translation of DNase I basal messengers. After derepression of the strong P$_L$ promoter (heat inactivation of the cI857 repressor), significant amounts of DNase I were obtained despite the antisense inactivation of a fraction of the endonuclease mRNAs. The efficiency of this system could be optimized by tuning the induction level of the antisense promoter to the required level that neutralizes the basal expression of the highly toxic gene. An antisense promoter whose induction level depends on the inducer concentration (i.e. P$_{BAD}$ promoter) would be suitable.

I.4. The Use of Invertible Promoters

This strategy was initially described by Podhajska A.J. [17] and Hasan N. [18]. The promoter, flanked by two attachment sites (att), is cloned in a direction opposite to the gene to be expressed and inverted by inducing site-specific genetic recombination using integrase. This system was successfully used to express variable domains of a T-cell receptor (TCR) [19]. The expression vector contained all the relevant features of a *lac*-based expression system in addition to a *lac* promoter put within an invertible module. To deliver the integrase, the authors engineered a heat-inducible vector harboring the integrase gene under the tight control of both the P$_L$/cI857 system and the antisense *ompA* promoter. Compared to the previously described strategies, the use of invertible promoters permits a completely free choice of the promoter system that controls the expression of the highly toxic gene. Surprisingly, this strategy has not been adapted yet to the very efficient T7 RNAP-driven expression system.

I.5. The Use of Competitive Promoters

Introducing supplementary T7 promoters into *E. coli* BL21(DE3) cells would compete with the T7 promoter controlling the expression of the highly toxic gene and hence, would significantly decrease its basal expression. This idea was used to express the human immunodeficiency virus type 1 (HIV-1) protease [10]. The authors constructed a competitor plasmid by introducing the T7 promoter into the EcoRV site of plasmid pACYC184 [20]. The recombinant plasmid (pACYCT7pmt) was introduced into BL21(DE3) cells and maintained under the selection pressure of 30 µg/ml chloramphenicol. The cells were then transformed with the expression vector carrying the HIV-1 protease gene under the

control of the same T7 promoter (parental plasmid pAR3040, selection marker: ampicillin resistance [21]). Control cells (without the competitor plasmid) were also transformed with the expression vector. Only *E. coli* cells that contained the competitor plasmid yielded a large number of transformants in the presence of 100 µg/ml ampicillin. Furthermore, these double-transformed cells were able to express about 10 mg of active and soluble HIV-1 protease per liter of induced cells. The authors showed, by plasmid isolation techniques, that the copy number of the competitive plasmid was approximately twice as high as the copy number of the expression vector. This 2 to 1 ratio of T7 promoters (competitor to target) was apparently sufficient to suppress the toxicity of HIV-1 protease during the growth phase. Other highly toxic proteins, with a lower allowed basal expression level, may require higher ratios. The copy number of the competitor plasmid is therefore a parameter that should be optimized.

I.6. The Use of the Full-Length *lac* Promoter/Operator Region

Ratios of induction over repression as high as 1000 have been described, since the 1970s, for the full-length wild-type *lac* promoter/operator region [22]. This tight regulation is due to the interplay of three components: the catabolic repression mediated by the c-AMP-activated protein CAP, the binding of the *lac* repressor to the *lac* operator O1 site and the formation of stable loop structures due to the interaction of the *lac* repressor with one of the two auxiliary operators sites (O2 downstream of the coding region or O3 upstream of the CAP binding site, see (Fig. (**2**) top) and detailed reviews in [23-24]). pET vectors contain the very strong T7 promoter associated with only the *lac* operator O1 site. They lack both the CAP binding site and the O3 auxiliary operator (see Fig. (**2**) bottom). This probably explains why it is not possible to completely shut off the expression of genes cloned in these vectors. Warren J.W. and coworkers designed a new expression vector (pLAC11) containing the entire *lac* control region between the start of the O3 auxiliary operator site to the end of the O1 site [25]. Using the -galactosidase

reporter system, they found that under repressed conditions plasmid pLAC11 exhibits a 84 fold lower expression level than the pET-21(+)/pLysS system whereas under induced conditions pLAC11 has only a 2 fold lower expression level. The ratio of induction over repression for pLAC11 was 590, close to the ratios described for the wild-type *lac* promoter/operator. The nearly complete shut off of gene expression under repressed conditions described here is very suitable for the expression of highly toxic genes because it significantly increases the number of *E. coli* cells surviving the growth phase and able to express the target gene. This enrichment significantly enhances the overall expression yield and shadows the 2 fold lost of induction level when compared to the pET expression system.

I.7. The Hybrid Promoter Strategy

The phage P_L promoter is a strong promoter *in vivo* that can be efficiently repressed by the cI repressor. A better repression was achieved by replacing the cI binding sites with sequences encoding the operator 2 (*tetO2*) of the Tn10 tetracycline resistance operon [26]. The resulting promoter-operator sequence called $P_{Ltet-O-1}$ is highly repressible by the Tet repressor and can be induced by adding anhydrotetracycline (aTc) to the culture [27]. aTC exhibits no antibiotic activity at the concentration required for full induction. When paired with ColE1 (50-70 copies), p15A (20-30 copies) and pSC101 (3-4 copies) replication origins, $P_{Ltet-O-1}$ exhibited induction over repression ratios of 2500, 3500 and 5000, respectively [27]. Similarly, a mutant of the *lac* promoter (P_{lac-8A}) whose strength is 3-fold higher than that of the wild type P_{lac} promoter has been combined with the binding site of the *araC* regulator (see section I.2). The new hybrid promoter, called $P_{lac/ara-1}$ [27], can be repressed by growing the cells in the presence of glucose and induced by adding IPTG and arabinose (see Fig. (**3**)). An induction over repression ratio of 1800 was obtained when the promoter was paired with a pSC101 replication origin. It was successfully used to express the highly toxic restriction endonuclease *Cfr9I* in the absence of its cognate methyltransferase. $P_{Ltet-O-1}$ and $P_{lac/ara-1}$ systems are commercially available under the
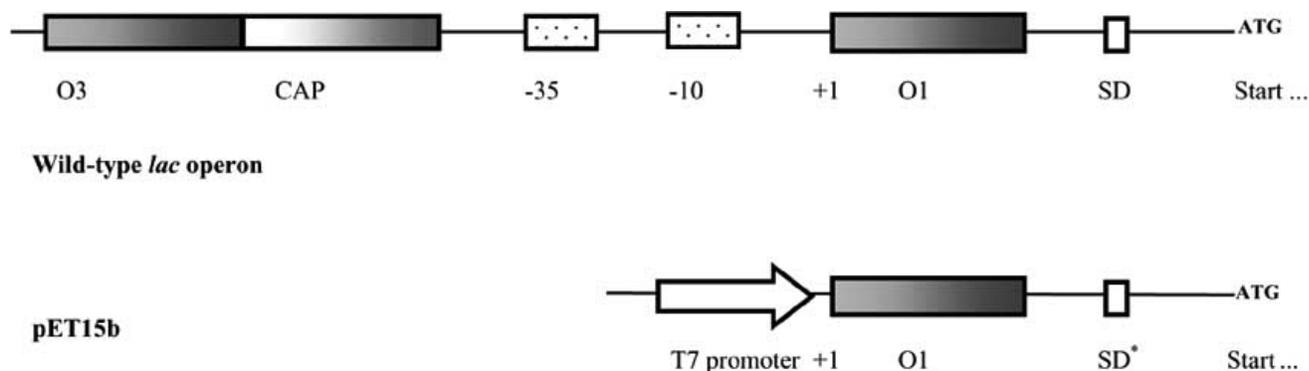
Fig. (2). Promoter-operator regions of the wild-type *lac* operon and the standard pET15b expression vector. +1 indicates the transcription start site. O1 and O3 are *lac* operator sites. Site O2, located in the coding sequence, is not shown. CAP: CAP binding site, SD: Shine-Dalgarno sequence of the *lac* operon, SD*: bacteriophage T7 gene 10 Shine-Dalgarno sequence, T7 promoter: bacteriophage T7 gene 10 promoter. The distance and the sequence between the SD and the ATG translation start site differ in the two systems.
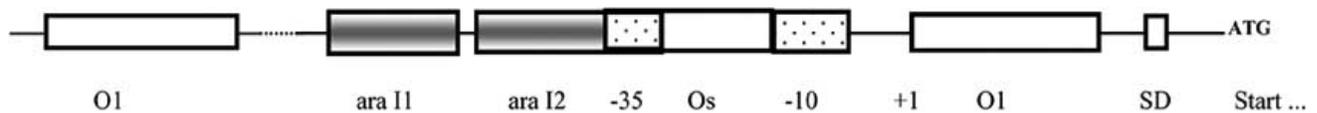
Fig. (3). Architecture of the P$_{lac/ara-1}$ hybrid promoter. +1 indicates the transcription start site. O1 is the *lac* operator site. Note that an additional O1 site is found at the 5' end. Os: a synthetic operator site, ara I1 and ara I2: binding sites of the *araC* repressor, SD: Shine-Dalgarno sequence.

commercial names of pPROTet.E and pPROLar.A respectively (BD Biosciences).

I.8. Blocking Read-Through Transcription by Upstream Terminators

Several independent reports showed that some highly toxic genes cloned using the pET/T7 expression system cannot be maintained not only in the BL21(DE3) strain but also in *E. coli* strains that do not contain any source of T7 RNAP. Examples include the phage T4 *ndd* gene [28], the thermostable esterase from *Bacillus acidocaldarius* [29], the *S. cerevisiae* DNA polymerase [30] and our own observation with the phage T4 restriction endoribonuclease RegB [31-32]. This phenomenon is, in part, explained by readthrough transcription mediated by *E. coli* RNA polymerase. This transcription initiates at cryptic promoters located upstream of the highly toxic gene promoter and continues through its coding sequence. mRNAs produced by this readthrough transcription are probably low in quantity but could yield critical amounts of the highly toxic proteins when they harbor a very efficient ribosome-binding site (RBS), i.e., the phage T7 gene 10 RBS, found in pET expression vectors. This explanation is supported by our observation that the phage T4 *regB* gene cannot be cloned in the pET15b vector and maintained in a T7 RNAP-free strain unless the efficiency of the pET15b RBS is decreased by mutations that suppress key base-pairing between the target mRNA and the 16S rRNA [31]. To circumvent the read-through transcription problem without altering the efficiency of the RBS, strong transcription terminators upstream of the target gene promoter/coding sequence region are needed. The Rho-independent transcription terminators T1 and T2 from *E. coli rrnB* operon [33] were inserted in tandem upstream of the T7 promoter controlling the expression of DNA polymerase gene [30]. *E. coli rpoC* terminator inserted upstream of the phage T4 *regA* control region was mandatory to quantitatively express this protein [8].

I.9. Tightening the Translation by "mRNA Switch" Mechanism

Cossart and coworkers showed that the mRNA encoding the virulence-activating transcription factor PrfA from *Listeria monocytogenes* behaves as a thermosensor that regulates its own translation in a temperature-dependent manner [34]. At temperatures below 30°C, the ribosome-binding site of the *prfA* mRNA is sequestered into a secondary structure that prevents its binding to the ribosome and facilitates its degradation by intracellular ribonucleases. At temperatures above 37°C, the secondary structure is destabilized, the ribosome-binding site associates with the 16S rRNA and translation is

initiated. The 5' leader region in the *prfA* mRNA was shown to be necessary and sufficient for this "mRNA switch" mechanism [34]. Interestingly, the same authors showed that the system can be reproduced in the gram-negative bacterium *E. coli* by coupling this 5' leader region to a target mRNA. It is therefore conceivable to couple this 5' leader region to the coding sequence of a highly toxic gene, to allow cells to grow at a temperature below 30°C and to induce protein expression by shifting the culture to a temperature above 37°C. For overexpression purposes, enhancement of the efficiency of the Shine-Dalgarno sequence present in the 5' leader region of the gene *prfA*, without compromising its thermodependence, is required. To our best knowledge, this strategy has not been investigated yet.

II. MANIPULATION OF THE CODING SEQUENCE

The strategies presented in this section target the coding sequence of the highly toxic gene. They report the suppression of an amber stop codon at position +1, the suppression of internal stop codons, the use of fusion proteins and the expression of the highly toxic product as two separate polypeptides.

II.1. Suppression of an Amber Stop Codon at Position +1

E. coli initiator tRNA has been mutated, in its anticodon region, from CAU to CUA [35]. The mutant tRNA has been shown to be amino-acetylated with glutamine instead of methionine. More interestingly, this mutant tRNA is formylated by *E. coli* methionyl-tRNA transformylase and exhibits features known to be important for its direction to the *E. coli* ribosomal site P [36-37]. Using the chloramphenicol acetyltransferase gene in which the initator AUG codon was changed to the amber stop codon UAG, it was shown that the mutant tRNA can initiate protein synthesis at the stop codon both *in vitro* and *in vivo* [35]. Despite the fact that the efficiency of initiation at the UAG stop codon was 60 to 70 % as high as normal initiation at the AUG codon, these findings provide an interesting approach for tight control of highly toxic genes expression. By mutating the AUG initator codon of the target gene to UAG, one can transfer his highly toxic gene to the appropriate expression vector (pET system for example) using very standard cloning techniques and regular *E. coli* strains. Expression of the highly toxic gene can be obtained by using an *E. coli* strain pretransformed with a plasmid containing the mutant tRNA and the glutamine aminoacyl-tRNA synthetase (GlnRS) coding sequences [35]. Although these two genes have been used by the authors of the original work in two different plasmids and expressed constitutively, it would be judicious to trans-

fer the two genes to a single vector and to control the expression, at least, of the mutant tRNA using an inducible promoter.

II.2. Suppression of an Internal Stop Codon

The phage T4 *ndd* gene product is involved in the delocalization and disruption of *E. coli* nucleoid shortly after infection. It was not possible to clone the intact *ndd* gene under the control of tightly repressed promoters [38]. Site-specific mutagenesis was used to convert the tyrosine UAC codon at position 121 to an UAG stop codon. *E. coli* strain PB4144 [39] carrying a temperature-sensitive allele of the *supF* suppressor was used to express the *ndd* gene. Cells grown at 42°C were able to tolerate the *ndd* gene under the control of the *lacZ* promoter. *ndd* expression was induced by shifting cultures growing exponentially from 42°C to 30°C and adding IPTG [38].

II.3. Expression as Fusion Proteins

In this section, a removable fusion partner neutralizes the toxic effect of the target protein. It is important to note that the neutralizing partner is not necessarily a specific antidote that targets the highly toxic protein. A general polypeptide that causes the fusion to strongly interact with new partners, different from regular targets, may lead to a suppression of the toxic effect. The expression of the mammalian apoptosis modulator protein Bax in *E. coli* is an important example illustrating the rationale design of a neutralizing partner. The Bax protein leads to the lysis of *E. coli* cells at concentrations as low as 0.01% of total proteins [40]. Attempts to direct the expression of the protein to the periplasm failed to alleviate the toxicity [41]. When a 17-amino acids leader sequence encoding the GroEL-binding loop of the *E. coli* cochaperone GroES (S-loop) was fused to the N-terminus of Bax, the toxicity did not significantly decrease. However, when the expression is induced under conditions designed to enhance the expression of *E. coli* chaperones (addition of 2% ethanol), the culture expressing the S-loop-Bax fusion continued to grow, albeit more slowly. At the end of the four-hours induction period, the recovered biomass in the presence of 2% ethanol was more than 200% that obtained in the absence of ethanol. [41]. The polarity of the fusion (at the N-terminus) seems to be important. As the N-terminus leader peptide emerges from the ribosome, it directs the fusion immediately to GroEL. This prevents toxic segments of the protein from interacting with vital cellular components. Expression of antimicrobial peptides is another example of the use of neutralizing partners. Buforin II is a 21-amino acids antimicrobial agent from the amphibian *Bufo bufo gargarizans*. This very basic peptide is lethal to *E. coli* but can be produced as a fusion with a neutralizing acidic peptide [42]. The target peptide can be cleaved from its protecting part, at methionine positions, by using cyanogen bromide. In some cases the design of the fusion partner cannot be rationale. Empirical testing of many available fusion partners can be rewarding. *E. coli*-lethal cytokine receptor homology domain (CRH) of granulocyte-colony-stimulating factor (G-CSF) receptor was fused to different proteins including the maltose-binding protein MalE, thioredoxin and glutathione S-transferase [43]. The thioredoxin-fusion permitted the expression of significant amounts of active CRH.

II.4. Semi-Synthesis of Highly Toxic Proteins

Restriction endonuclease from *Haemophilus parainfluenzae* (HpaI) was produced using a semisynthetic approach that utilizes an intein, a protein-splicing element, to generate a reactive thioester at the C-terminus of a first recombinant polypeptide. Nucleophilic attack on this thioester by the N-terminal cysteine of a synthetic peptide led to the ligation of the two parts and to the production of active HpaI [44]. The first 223 amino acids were expressed in *E. coli* as three-part fusion protein to the intein and the chitin-binding domain (CBD). After purification on chitin beads, the thiol-induced cleavage released the first 223 amino residues with a thioester at the C-terminal alpha carbon. This product was ligated to a 28 amino acids synthesized peptide starting with an N-terminal cysteine. This strategy allows a free choice of the truncated polypeptide expression system. Provided that the synthesis of the second peptide is not limiting, the strategy can be adapted to produce large quantities of the highly toxic protein [45]. Constraints over the split site, dictated by the presence of a cysteine, can be bypassed by mutating a non-essential residue to cysteine.

III. MANIPULATION OF THE COPY NUMBER

Basal expression levels correlate directly with vectors copy numbers. While the stable maintenance of a highly toxic gene in *E. coli* is favored by the use of a low-copy number vector, its quantitative expression necessitates an increase in gene dosage prior to induction. Therefore, a system whereby the plasmid copy number is low but can be significantly increased prior to induction seems to be an attractive solution.

III.1. Thermal Control

At permissive temperatures, runaway-replication vectors can be maintained in *E. coli* at as low as one copy per chromosome. At restrictive temperatures, as a result of loss of replication control, their numbers rise enormously and approach 1000 copies per chromosome [46]. pOU61 is an example of runaway-replication plasmids [47]. RepA, CopA and CopB are responsible for the regulation of plasmid replication. Their expression is under the control of the thermo-sensitive allele of the cI repressor (cI857). Fig. (**4**) shows the thermal response of a system in which we cloned, for illustration, a non-toxic version of phage T4 RegB protein mutant under the control of the T7 promoter in vector pOU61. Expression of the wild type *regB* gene using this system was not possible in the BL21(DE3) strain, but yielded substantial amounts of the highly toxic endoribonuclease by combining the thermal control of the copy number with the phage-mediated delivery of T7 RNAP (CE6, see section I.1). The overall expression yield did not, however, exceed 0.4 mg/L of culture because of the combination of three factors that are detrimental to *E. coli* normal growth: restrictive temperature, CE6 infection and uncontrolled plasmid replication. A system in which the copy number could be raised to a number (~50) more compatible with cell physiology would have been more suitable.
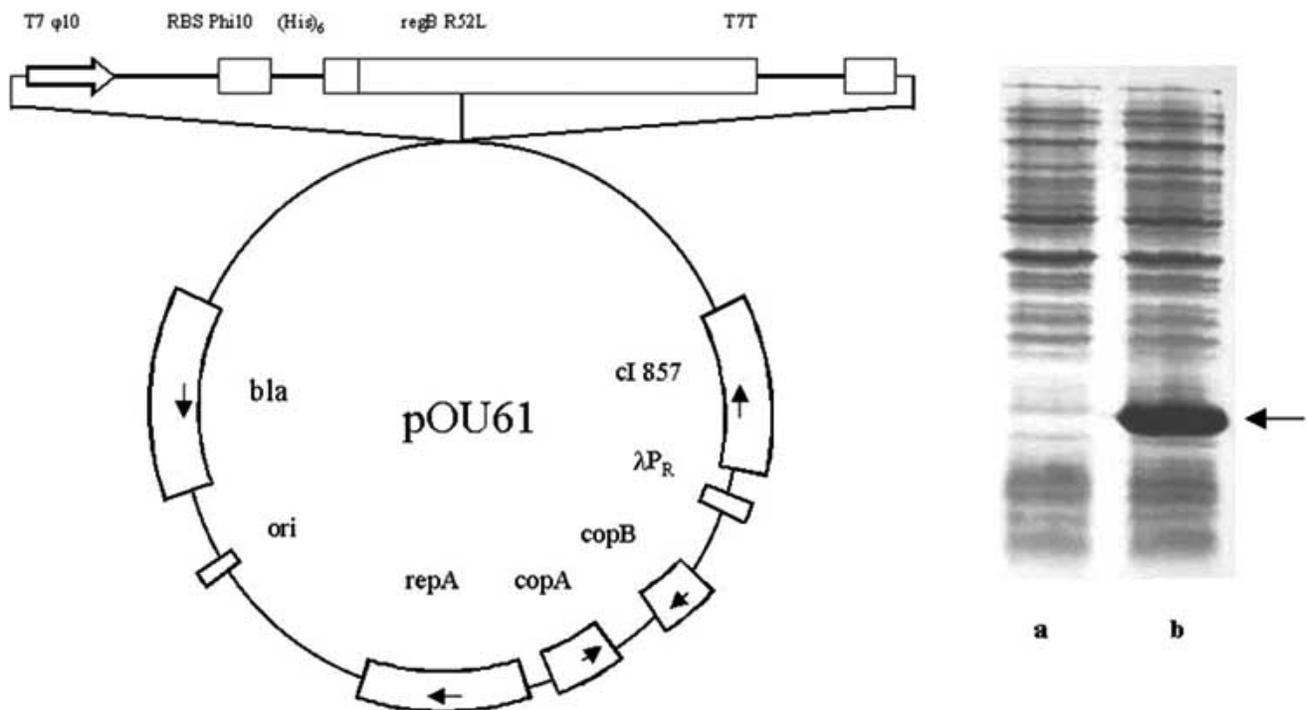
Fig. (4). Production of RegB R52L using the thermo-responsive pOU61 system. Left: architecture of the pOU61 plasmid. *repA*, *copA* and *copB*: replication regulators; bla: ampicillin-resistance determinant; T7φ10, RBSφ10 and T7T: bacteriophage T7 gene 10 promoter, ribosome-binding site and T7 terminator respectively; (His)$_6$: six histidines-tag. Right: SDS-PAGE analysis of the expression system. Lane a: before induction. Lane b: incubation for 30 min at 42°C then shift to 37°C for 2 hours 30 min in the presence of 1 mM IPTG. The arrow indicates the RegB band.

III.2. Arabinose Control

T7-based expression vectors that allow copy number amplification from 1-2 copies per cell to 20-50 copies per cell have been developed (pETcoco-1^{TM}, Novagen). The expression vector is kept at 1-2 copies per cell by using the *oriS* and *repE* elements of the F episome together with the *parABC* partition determinants. 20-50 copies per cell are obtained by using the RK-2 derived *trfA* replicator that acts at *oriV*. Switching from low to high copy numbers is achieved by adding arabinose, which tightly control the expression of the *trfA* gene (see section I.2). Similarly, Bowers designed a host/vector system that allows tight and uniform modulation of the copy number of plasmids with a replication origin (ori) [48]. The regulation is achieved through arabinose-inducible expression of the required Rep protein ( ) whose gene was inserted into the chromosome of *E. coli* under the control of the P$_{BAD}$ promoter. The system was successfully used to express the highly toxic colicin E3 in the absence of its cognate immunity protein [48].

III.3. Use of the "CopyCutter" *E. coli* Strain

The copy number of ColE1-type plasmids is controlled by the *pcnB* (plasmid copy number) gene. This gene, constitutively expressed in *E. coli*, was deleted and replaced by a modified *pcnB* gene under the control of a patented inducible promoter, creating the CopyCutter^{TM} EPI400^{TM} strain (EPICENTRE). Under repressed conditions, this strain decreases the copy number of pUC19 plasmid from ~216 to ~9 copies. When induced, the copy-number is restored to ~200

copies per cell [49]. Similarly, this strain decreases the copy number of the T7-based expression vector pET9 from 33 to 9 copies. When induced, the copy number reaches 19 per cell. This system was successfully used to clone a promoterless 461-bp PCR fragment, containing the phage T4 *regB* gene, into the T7 expression system pET11a. When the ligation mixture was electroporated into a standard *E. coli* strain, 19 clones were obtained. 17 of them contained no inserts and 2 had large deletions. Using the "CopyCutter" cells under repressed conditions, all 29 clones obtained were recombinant [49]. The author, however, did not specify the efficiency of *regB* expression, if there is any, using the obtained recombinant plasmids. Table **2** lists other useful *E. coli* strains that manipulate the copy number.

IV. ADDITION OF STABILIZING SEQUENCES

The main problems associated with the expression of highly toxic genes is plasmid instability during the growth phase. Because of their physiological advantages, plasmid-free cells quickly overtake the culture. This phenomenon dramatically decreases the expression yield because only a very small fraction of the surviving cells retained its ability to express the target gene. To circumvent this problem, expression vehicles carrying genes that cause cell death upon plasmid loss are required. Two *E. coli* loci are of particular interest: *parB* and *ccd*. The *parB* locus (*hok/sok*) cell death program involves a 52 amino acids membrane damaging protein (Hok) whose mRNA is very stable but translationaly inactive. *sok* is a highly instable antisense RNA that binds the *hok* mRNA leader region. In plasmid-free cells, the pool

of *sok* decays rapidly. This leads to the processing of the 3' end of *hok* and the production of an active transcript [50]. The *ccd* locus (*ccdA/ccdB*) involves a proteolytically stable CcdB protein that inhibits DNA gyrase [51]. CcdA binds to CcdB and blocks its action. In plasmid-free cells, the pool of CcdA decays rapidly because of its short half-life. This leads to cell death. Expression of the highly toxic phospholipase D cloned in a standard pET vector illustrates this method. Without stabilization loci, the percentage of plasmid-bearing cells was as low as 10% [52]. When the *parB* or the *ccd* loci were added to the plasmid, this percentage reached nearly 100 % just before induction.

V. EMPIRICAL SELECTION OF *E. coli* STRAINS EXPRESSING HIGHLY TOXIC GENES

In the previous sections, we described the major strategies that should be tested in order to express, in *E. coli*, highly toxic products. The rationale design of these strategies is mainly based on the very well known genetics of the bacterium. The important number of parameters that should be taken in account illustrates the complexity of the biological process leading to a quantitative expression of highly toxic genes. In some cases, this whole complexity can be encompassed by empirical selection of those *E. coli* strains that tolerate the expression of the highly toxic product. This takes advantages of the silent mutations that accumulate over generations of growth and that could be selected for the advantage they confer to *E. coli* in the presence of a detrimental recombinant gene. Hence, a strain as standard and as stable as BL21(DE3) should be carefully analyzed after transformation for phenotypic differences on agar plates, for differences in the basal expression levels of T7 RNAP using standard reporter genes, for differences in the kinetic of accumulation of the recombinant product. HIV-2 protease could not be maintained under the control of the T7 promoter in the BL21(DE3)pLysS supplied by Novagen [53]. However, a BL21(DE3)pLysS variant empirically selected for its ability to express HIV-1 protease, was shown to tolerate HIV-2 protease. The BL21(DE3)pLysS variant was shown to have a slower rate of mRNA accumulation when compared to the commercially available strain [53]. A complete protocol for selection of useful variants of BL21(DE3) is described by Miroux B. and Walker J.E. [54]. The authors described the selection of two derivatives strains called C41(DE3) and C43(DE3) that tolerated the expression of some highly toxic genes. In the case of the C41(DE3) strain, not only basal recombinant mRNA levels were five times as low as in BL21(DE3), they also accumulated twice as slow after induction. A third variant called BL21(DE3)NH was recently isolated during attempts to express unstable genes originating from *Mycobacterium tuberculosis* [55]. This variant was confirmed to be *E. coli* by standard biochemical and cultural tests even though it formed small opaque colonies and was shown, using scanning electron microscopy, to have extremely long bacillary shapes [55]. The mechanism by which this variant tolerates the expression of highly toxic genes is unknown.

VI. SPECIAL EXPRESSION PROTOCOLS

Table **3** summarizes special expression protocols that could be used to improve the expression of highly toxic

genes using the standard pET/BL21(DE3) expression system.

CONCLUSION

We reviewed in this paper special strategies and genetic tools useful for the expression of highly toxic genes. Although we were probably not exhaustive, the set of solutions summarized in this review highlight two important points: the need to characterize the level of toxicity of problematic genes not only qualitatively but also quantitatively (using standard plasmid stability tests for example) and the need to tune the tightness of the inducible expression system to match the basal expression level that *E. coli* tolerates. Although the desirable tightness can be achieved by combining many rationale strategies based on the well-known genetics of *E. coli*, empirical selection of *E. coli* strains can be rewarding and time-saving.

For those unfortunate cases, where an extensive investigation of *E. coli* expression systems did not circumvent the toxicity problem, there are several alternative strategies, *in vitro* and *in vivo*, that should be tested. The *in vitro* strategy is based on cell-free translation systems. The Rapid Translation System (RTS) uses enriched *E. coli* lysate to couple *in vitro* transcription and translation reactions. This system has been reviewed in a previous issue of this journal [61]. Large protein amounts, compatible with structural analysis, can be obtained by using a continuous flow of an input buffer (containing amino acids, adenosine triphosphate and guanosine triphosphate) and by removing continuously the newly synthesized polypeptide [62]. Commercially available kits (Roche Diagnostics, Invitrogen) have been used to successfully express 32 different rickettsial proteins (many of these proteins are known to be highly toxic to *E. coli*) [63]. Alternative *in vivo* strategies include the expression of highly toxic genes in yeast, in mammalian cells and in insect cells. In these strategies, the abolition of the toxic problem is directly related to the differences between the physiology of a prokaryotic cell and the physiology of a eukaryotic cell. Yeast *Saccharomyces cerevisiae*, the most sophisticated eukaryotic model for recombinant DNA, expressed many highly toxic virulence proteins from bacterium *Yersinia Pestis* [64]. The methylotrophic yeast *Pichia pastoris* combines the advantages of a eukaryotic cell physiology with an excellent expression yield (see reference [65]). Expression in mammalian cells is an additional strategy for the expression of toxic genes. Chinese Hamster ovary (CHO) and non-secreting myeloma cells are very popular in recombinant protein expression [66]. The use of mammalian cells, however, suffers from several limitations including cultivation complexity, expensiveness and low expression yields. In contrast, insect cells transformed by baculovirus vectors are easier to manipulate, more resistant and more productive [67]. Interestingly, an ecdysone-inducible system has been recently developed in insect cells [68]. This system tightly regulates the transcription of the recombinant gene in insect cells using an ecdysone agonist as inducer. The highly toxic diphteria toxin (DT) has been expressed using this system [68].

ACKNOWLEDGMENTS

REFERENCES

[1]   Sorensen, H.P. and Mortensen, K.K. (**2005**) *J. Biotechnol.*, *115*, 113-28.

[2]   Jonasson, P., Liljeqvist, S., Nygren, P.A. and Stahl, S. (**2002**) *Biotechnol. Appl. Biochem.*, *35*, 91-105.

[3]   Baneyx, F. (**1999**) *Curr. Opin. Biotechnol.*, *10*, 411-21.

[4]   Hannig, G. and Makrides, S.C. (**1998**) *Trends Biotechnol.*, *16*, 54-60.

[5]   Makrides, S.C. (**1996**) *Microbiol. Rev., 60*, 512-38.

[6]   Studier, F.W. and Moffatt, B.A. (**1986**) *J. Mol. Biol.*, *189*, 113-30.

[7]   Studier, F.W. (**1991**) *J. Mol. Biol., 219*, 37-44.

[8]   Unnithan, S., Green, L., Morrissey, L., Binkley, J., Singer, B., Karam, J. and Gold, L. (**1990**) *Nucleic Acids Res., 18*, 7083-92.

[9]   Sanson, B., Hu, R.M., Troitskaya, E., Mathy, N. and Uzan, M. (**2000**) *J. Mol. Biol., 297*, 1063-74.

[10]   Komai, T., Ishikawa, Y., Yagi, R., Suzuki-Sunagawa, H., Nishigaki, T. and Handa, H. (**1997**) *Appl. Microbiol. Biotechnol.*, *47*, 241-5.

[11]   Guzman, L.M., Belin, D., Carson, M.J. and Beckwith, J. (**1995**) *J. Bacteriol., 177*, 4121-30.

[12]   Boomershine, W.P., Raj, M.L., Gopalan, V. and Foster, M.P. (**2003**) *Protein Expr. Purif., 28*, 246-51.

[13]   Hall, T.A. and Brown, J.W. (**2002**) *RNA, 8*, 296-306.

[14]   O'Connor, C.D. and Timmis, K.N. (**1987**) *J. Bacteriol., 169*, 4457-62.

[15]   Moore, S. (**1981**) in *The Enzymes* (Boyer, P.D., Ed.). pp. 281-296. Academic Press, New York.

[16]   Worrall, A.F. and Connolly, B.A. (**1990**) *J. Biol. Chem., 265*, 21889-95.

[17]   Podhajska, A.J., Hasan, N. and Szybalski, W. (**1985**) *Gene, 40*, 163-8.

[18]   Hasan, N. and Szybalski, W. (**1987**) *Gene, 56*, 145-51.

[19]   Wulfing, C. and Pluckthun, A. (**1993**) *Gene, 136*, 199-203.

[20]   Rose, R.E. (**1988**) *Nucleic Acids Res., 16*, 355.

[21]   Rosenberg, A.H., Lade, B.N., Chui, D.S., Lin, S.W., Dunn, J.J. and Studier, F.W. (**1987**) *Gene, 56*, 125-35.

[22]   Beckwith, J.R. and Zipser, D. (**1970**) in *The Lactose Operon*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

[23]   Muller-Hill, B. (**1975**) *Prog. Biophys. Mol. Biol., 30*, 227-52.

[24]   Reznikoff, W.S. (**1992**) *Mol. Microbiol., 6*, 2419-22.

[25]   Warren, J.W., Walker, J.R., Roth, J.R. and Altman, E. (**2000**) *Plasmid, 44*, 138-51.

[26]   Hillen, W. and Berens, C. (**1994**) *Annu. Rev. Microbiol., 48*, 345-69.

[27]   Lutz, R. and Bujard, H. (**1997**) *Nucleic Acids Res., 25*, 1203-10.

[28]   Bouet, J.Y., Krisch, H.M. and Louarn, J.M. (**1998**) *J. Bacteriol. 180*, 5227-30.

[29]   Manco, G., Adinolfi, E., Pisani, F.M., Ottolina, G., Carrea, G. and Rossi, M. (**1998**) *Biochem J., 332*, 203-12.

[30]   Brown, W.C. and Campbell, J.L. (**1993**) *Gene, 127*, 99-103.

[31]   Saida, F., Uzan, M., Lallemand J.Y. and Bontems, F. (**2003**) *Biotechnol. Prog., 19*, 727-33.

[32]   Saida, F., Odaert, B., Uzan, M. and Bontems, F. (**2004**) *Protein Expr. Purif., 34*, 158-65.

[33]   Brosius, J. (**1984**) *Gene, 27*, 161-72.

[34]   Johansson, J., Mandin, P., Renzoni, A., Chiaruttini, C., Springer, M. and Cossart, P. (**2002**) *Cell, 110*, 551-61.

[35]   Varshney, U. and RajBhandary, U.L. (**1990**) *Proc. Natl. Acad. Sci. USA, 87*, 1586-90.

[36]   Seong, B.L., Lee, C.P. and RajBhandary, U.L. (**1989**) *J. Biol. Chem., 264*, 6504-8.

[37]   Schulman, L.H. and Pelka, H. (**1985**) *Biochemistry, 24*, 7309-14.

[38]   Bouet, J.Y., Campo, N.J., Krisch, H.M. and Louarn, J.M. (**1996**) *Mol. Microbiol., 20*, 519-28.

[39]   Kimura, M., Miki, T., Hiraga, S., Nagata, T. and Yura, T. (**1979**) *J. Bacteriol., 140*, 825-34.

[40]   Asoh, S., Nishimaki, K., Nanbu-Wakao, R. and Ohta, S. (**1998**) *J. Biol. Chem., 273*, 11384–11391.

[41]   Donnelly, M.I., Stevens, P.W., Stols, L., Su, S.X., Tollaksen, S., Giometti, C. and Joachimiak, A. (**2001**) *Protein Expr. Purif., 22*, 422-9.

[42]   Lee, J.H., Minn, I., Park, C.B. and Kim, S.C. (**1998**) *Protein Expr. Purif., 12*, 53-60.

[43]   Tatsuda, D., Arimura, H., Tokunaga, H., Ishibashi, M., Arakawa, T. and Tokunaga, M. (**2001**) *Protein Expr. Purif., 21*, 87-91.

[44]   Evans, T.C., Benner, J. and Xu, M.Q. (**1998**) *Protein Sci., 7*, 2256-64.

[45]   Amitai, G. and Pietrokovski, S. (**1999**) *Nat. Biotechnol., 17*, 854-5.

[46]   Nordstrom, K. and Uhlin, B. E. (**1992**) *Biotechnology, 10*, 661-666.

[47]   Larsen, J.E., Gerdes, K., Light, J. and Molin, S. (**1984**) *Gene, 28*, 45-54.

[48]   Bowers, L.M., Lapoint, K., Anthony, L., Pluciennik, A. and Filutowicz, M. (**2004**) *Gene, 340*, 11-8.

[49]   Haskins, D. (**2004**) *EPICENTRE Forum, 11*, 6.

[50]   Gerdes, K., Jacobsen, J.S. and Franch, T. (**1997**) *Genet. Eng. (N Y), 19*, 49-61.

[51]   Van Melderen, L. (**2002**) *Int. J. Med. Microbiol., 291*, 537-44.

[52]   Mishima, N., Mizumoto, K., Iwasaki, Y., Nakano, H. and Yamane, T. (**1997**) *Biotechnol. Prog., 13*, 864-8.

[53]   Chen, E. (**1994**) *Methods Enzymol., 241*, 29-46.

[54]   Miroux, B. and Walker, J.E. (**1996**) *J. Mol. Biol., 260*, 289-98.

[55]   Poletto, S.S., da Fonseca, I.O., de Carvalho, L.P., Basso, L.A., and Santos, D.S. (**2004**) *Protein Expr. Purif., 34*, 118-25.

[56]   Suter-Crazzolara, C. and Unsicker, K. (**1995**) *Biotechniques, 19*, 202-4.

[57]   Zhang, C.C., Glenn, K.A., Kuntz, M.A. and Shapiro, D.J. (**2000**) *J. Steroid Biochem. Mol. Biol., 74*, 169-78.

[58]   Yike, I., Zhang, Y., Ye, J. and Dearborn, D.G. (**1996**) *Protein Expr. Purif., 7*, 45-50.

[59]   Bocanegra, J.A., Bejarano, L.A. and Valdivia, M.M. (**1997**) *Biotechniques, 22*, 798-802.

[60]   Pan, S.H. and Malcolm, B.A. (**2000**) *Biotechniques, 29*, 1234-8.

[61]   Betton, J.M. (**2003**) *Curr. Protein Pept. Sci., 4*, 73-80.

[62]   Neerathilingam, M., Greene, L.H., Colebrooke, S.A., Campbell, I.D. and Staunton, D. (**2005**) *J. Biomol. NMR, 31*, 11-9.

[63]   Renesto, P. and Raoult, D. (**2003**) *Ann. N. Y. Acad. Sci., 990*, 642-52.

[64]   Lesser, C.F. and Miller, S.I. (**2001**) *EMBO J., 20*, 1840-9.

[65]   Daly, R. and Hearn, M.T. (**2005**) *J. Mol. Recognit., 2*, 119-38.

[66]   Barnes, L.M., Bentley, C.M. and Dickson, A.J. (**2003**) *Biotechnol. Bioeng., 81*, 631-9.

[67]   Ikonomou, L., Schneider, Y.J. and Agathos, S.N. (**2003**) *Appl. Microbiol. Biotechnol., 62*, 1-20.

[68]   Dai, X., Willis, L.G., Palli, S.R. and Theilmann, D.A. (**2005**) *Protein Expr. Purif., 42*, 236-45.