

Molecular Design of Expression Systems: Comparison of Different Repressor Control Configurations Using Molecular Mechanism Models

Wilfred Chen and James E. Bailey*

Department of Chemical Engineering, California Institute of Technology, Pasadena, California 91125

Sun Bok Lee

Department of Chemical Engineering, Pohang Institute of Science and Technology, P.O. Box 125 Pohang, 790-330, Korea

Received October 4, 1990/Accepted February 7, 1991

Molecular-level mathematical models have been used to evaluate the effectiveness of eight different configurations of repressor synthesis control on the regulation of cloned gene expression initiated from a promoter-operator sequence. Both single and dual-repressor situations were considered, employing genetically structured models for the *lac* and λP_R promoter-operators in example calculations. Simulation results suggest that the most effective mode of cloned gene expression control is a cross-regulation configuration carried on a multicopy plasmid. This system was able to control cloned gene transcription in the uninduced state over a broad range of plasmid copy number and also provided the highest overall transcription rate in the induced state. The general strategies suggested by these simulations should be applicable for other repressor-operator-promoter systems in diverse hosts.

Key words: mathematical models • cross regulation • repressor synthesis control • gene expression

INTRODUCTION

Maximizing the amount of protein produced from a cloned gene in a recombinant organism requires careful consideration of the tradeoffs involved between cloned-gene expression and host cell growth and biosynthetic activity. High levels of cloned gene expression have been shown to reduce host cell growth rate and, concomitantly, overall protein synthesis activity.^{8,23} This occurs because of competition between plasmid-directed and host cell-directed activity for common pools of precursors, chemical energy and electrons, activator species, repressor molecules, transport apparatus, and enzymes and other catalytic assemblies. Under some cases, high expression levels can even cause cell death. Moreover, it has been observed that induction or derepression of cloned gene expression accentuates this inhibitory effect.²⁴ Therefore, it is very important to regulate the expression of cloned genes in the manufacture of valuable protein using recombinant cells.

In order to maximize product synthesis, cells are typically grown to high densities with minimal cloned gene expression followed by a production phase in which high expression results from induction or derepression of cloned gene transcription.^{17,24} Several different promoter-operator systems, including *lac*, *trp*, and λP_L , have been cloned and utilized to regulate transcription of cloned genes.^{4,25} In each of these systems, the transcription activity of the promoter-operator depends upon interaction with a corresponding specific repressor protein. This interaction may be influenced by adjusting the temperature (ts regulatory mutant) or the composition (inducer or corepressor concentration) of the growth medium in order to switch from low to high promoter activity.

Limitations of many current regulated promoter-operator systems are loss of control of cloned gene expression if the number of cloned gene copies per cell exceeds some threshold or poor expression under fully activated conditions. In order to explore genetic strategies for addressing these limitations, it is useful to consider the different types of repressor expression configurations which can be established using recombinant DNA and modern genetic techniques.

Three different modes of repressor synthesis can be identified: (1) Constitutive synthesis, which proceeds at a constant level independent of the repressor content and the repressor configuration. (2) Autogenous regulation, in which the synthesis of repressor is influenced by the content of repressor in the cell. (3) Cross regulation, in which the synthesis of the repressor is controlled by the presence of a second repressor molecule for a different promoter. With current methods for genetic manipulation, two different types of locations for repressor genes are possible. Repressor can be provided by repressor gene(s) encoded in the chromosome, or the repressor may be formed by the expression of a repressor gene included in an extrachromosomal element. Combining these possibilities, eight different combinations of the

* To whom all correspondence should be addressed.

repressor gene locus and regulation of the repressor gene transcription can be identified. These are illustrated schematically in Figure 1.

In this work, molecular level mathematical models are used to evaluate the effectiveness of each of these different configurations. The preferred system is the one which, for a range of vector copy numbers, retains control of the promoter under the uninduced condition and provides the highest transcription level under the induced state. In general, any promoter-repressor system can be used for these different configurations. However, detailed models for the *lac* promoter-operator system^{12,13} and the λP_R promoter-autorepressor system^{14,15} which have been formulated and tested extensively will be used here to evaluate different modes of

providing and regulating repressor in the host cell. These models allow simulation of various situations in which the host cell contain many copies of the promoter-operator system and multiple copies of the repressor gene. This study provides an illustration of the application of molecular mechanism models to compare alternative molecular designs of recombinant DNA molecules and host genotype. In this context, such models serve as a guide for the genetic design of the host-vector expression system.

MODEL DESCRIPTIONS

As described previously,¹²⁻¹⁵ repressor synthesis can be represented by the following cellular material bal-

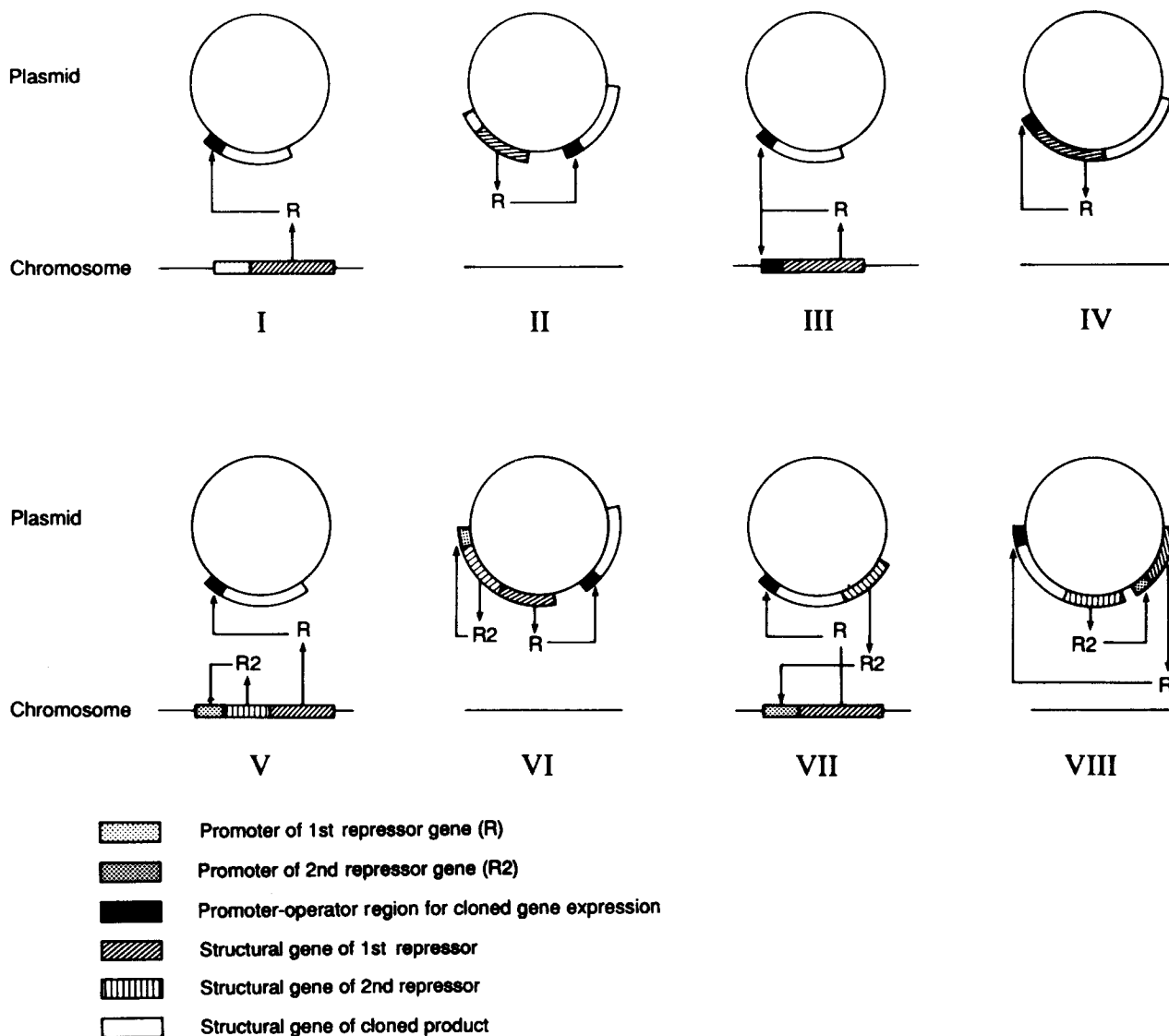


Figure 1. Schematic illustration of the eight different configurations for the regulation of cloned gene transcription considered in this paper. These cases differ with respect to their modes of repressor synthesis and their source of repressor. For cases I, III, V, and VII, the repressor gene is encoded in the chromosome, whereas for cases II, IV, VI, and VIII, the repressor gene is included in the plasmid. For cases I and II, repressor synthesis is constitutive. For cases III and IV, the content of repressor influences its synthesis. Finally, for cases V-VIII, the content of a second repressor influences repressor synthesis. However, for cases V and VI, the second repressor gene is fused to form an operon with the first repressor gene ($[mRNA]_{R_2} = [mRNA]_R$). For the last two cases, the second repressor gene is fused to form an operon with the product gene ($[mRNA]_{R_2} = [mRNA]_P$).

ance equations on the repressor message and protein, respectively:

$$\frac{d[mRNA]_R}{dt} = k_p^o \eta_R [G]_R - k_d [mRNA]_R - \mu [mRNA]_R \quad (1)$$

$$\frac{d[R]_b}{dt} = k_q^o \xi_R [mRNA]_R - k_e [R]_b - \mu [R]_b \quad (2)$$

Similarly, material balances on the cloned gene message and protein can be written as follows:

$$\frac{d[mRNA]_P}{dt} = k_p^o \eta_P [G]_P - k_d [mRNA]_P - \mu [mRNA]_P \quad (3)$$

$$\frac{d[P]}{dt} = k_q^o \xi_P [mRNA]_P - k_e [P] - \mu [P] \quad (4)$$

where the transcription rate constant (k_p^o) and the translation rate constant (k_q^o) are the average rates per template and, as such, incorporate the RNA polymerase elongation rate and the polypeptide chain elongation rate, respectively.¹⁶ Substantial deviations from the cell-average value of the frequency of transcription initiation as well as the regulatory effect on transcription have been incorporated into the transcription efficiency term (η). An analogous definition also applies for the translation efficiency (ξ).

In eqs. (1)–(4), subscripts R and P denote repressor and product protein, respectively. To minimize the number of different parameters in these calculations, it has been assumed that the transcription rate constant (k_p^o), the translation rate constant (k_q^o), the mRNA decay rate constant (k_d), and the protein decay rate constant (k_e) are the same for the repressor and the product. The translation efficiencies for the repressor and the product, ξ_R and ξ_P , respectively, could normally be influenced by the ribosome binding site nucleotide sequence (Shine–Dalgarno sequence), and the distance between the initiation codon (ATG or GTG) and the S–D sequence. These two parameters may be different but are here assumed to be constant unaffected by the plasmid copy number or the design of repressor synthesis regulation. Clearly, these two parameters and other could be adjusted by corresponding changes in nucleotide sequence to the appropriate values for a particular system in which significant differences occur between the repressor and the product kinetic properties.

The major attention in this work is focused primary upon the product gene transcription efficiency η_P . The overall transcription efficiency of a promoter is influenced by the interaction of the promoter–operator region of the DNA with regulatory proteins such as repressor, activator, and RNA polymerase. In this paper, the expression of the product protein will be assumed to be controlled by the *lac* promoter–operator system. This particular promoter–operator is considered as an example here because detailed genetically structured

models for the *lac* operon have been formulated previously.^{12,13} The modeling approach presented here could be readily applied to other promoter–operator systems. The objective is to elucidate central qualitative features of different repressor expression configurations.

The transcription efficiency for the complete wild-type *lac* operon is determined by the interaction of the *lac* promoter–operator regions with three regulatory molecules^{3,18}: *lac* repressor protein, cyclic AMP receptor protein (CRP) complexed with cyclic AMP (cAMP), and RNA polymerase (RNP) activated by the sigma factor. However, for the L8UV5 *lac* promoter–operator and the *lac* portable promoter which have been widely used for the regulation of cloned gene expression, the CRP–cAMP complex does not affect the overall transcription efficiency.² Such a situation is assumed here. Since the major objective of this study is to determine the influence of different types of repressor synthesis on cloned gene expression, it is assumed that the interaction between RNA polymerase and the *lac* promoter is unaffected by any of the different repressor synthesis design discussed here. Furthermore, according to previous results,¹³ it has been estimated that RNP binding activity varies only slightly with copy number up to 100. Thus, effects of copy number on RNP binding are neglected in these calculations.

First, product transcription efficiency (η_P) will be considered. Assuming that only the *lac* repressor protein is involved in the regulation of cloned gene transcription, η_P can be expressed as

$$\eta_P \approx \Phi_P = (1 - \Psi_P), \quad (5)$$

where Ψ_P is the binding probability of the repressor molecules to the operator sequence. As shown previously,¹² the binding probability of the repressor to the operator is given by the equation

$$\Psi_P = \frac{1}{2} \left[\left(1 + \frac{[R]_b}{[O]_b} + \frac{1}{\alpha [O]_b} \right) - \left\{ \left(1 + \frac{[R]_b}{[O]_b} + \frac{1}{\alpha [O]_b} \right)^2 - 4[R]_b/[O]_b \right\}^{1/2} \right] \quad (6)$$

where

$$\alpha = \frac{K_{A1} + K_{C1} K_{D1} [I]}{1 + K_{B1} [D]_b + K_{C1} [I] + K_{C1} K_{E1} [D]_b [I]} \quad (7)$$

$[R]_b$, $[O]_b$, $[D]_b$, and $[I]$ are the total repressor concentration, the operator concentration, the concentration of nonspecific binding sites for the repressor, and the inducer concentration, respectively.

The transcription efficiency for repressor synthesis (η_R) is dependent upon the cases considered (see Fig. 1). When the repressor transcription is constitutive as in cases I and II, η_R is not influenced by the repressor level and thus is assumed to be constant. Parameter η_R for a constitutive promoter deviates from unity depending on the extent to which its promoter strength (frequency of initiation) deviates from the cell-average

value. For autogenous repressor synthesis, the transcription of both the repressor and the product are controlled by the *lac* promoter-operator; therefore, η_R is equal to η_P in cases III and IV.

For cases V to VIII, the transcription of the *lac* repressor is controlled by a second controllable promoter. In this study, the λP_R promoter will be considered in particular, again because of a good prior modeling framework for this particular promoter.^{14,15} Again, it should be noted that this approach to genetic design is not restricted to this particular choice of a second regulated promoter. Transcription of the λP_R promoter is influenced by interactions between RNA polymerase and the *cro* repressor. By assuming that the concentration of the operator is much lower than that of the repressor ($[CRO] \approx [CRO]_b$), the transcription efficiency for the λP_R promoter is derived as follows¹⁴:

$$\eta_\lambda = \frac{1 + K_3[CRO]_b}{1 + a[CRO]_b + b[CRO]_b^2 + c[CRO]_b^3} \quad (8)$$

where

$$a = K_1 + K_2 + K_3, \quad (9)$$

$$b = K_1K_2 + K_2K_3 + K_3K_1, \quad (10)$$

$$c = K_1K_2K_3. \quad (11)$$

where K_1 , K_2 , and K_3 represent the binding affinity of the *cro* repressor at the three different operator sites, respectively, and $[CRO]_b$ represents the total *cro* repressor concentration.

However, in this study the concentration of the operator will frequently exceed the concentration of the repressor especially for those cases where the repressor gene is included in the plasmid. Therefore, the actual free repressor concentration must be considered in determining the transcription efficiency. As shown in detail in the Appendix, the free repressor concentration, $[CRO]$, can be determined as:

$$[CRO]_b = [CRO] + \frac{K_1[P_R O_R]_b [CRO]}{1 + K_1[CRO]} + \frac{K_2[P_R O_R]_b [CRO]}{1 + K_2[CRO]} + \frac{K_3[P_R O_R]_b [CRO]}{1 + K_3[CRO]}. \quad (12)$$

where $[P_R O_R]_b$ is the total concentration of the λP_R operator. Since this is a fourth order polynomial equation

for $[CRO]$, four solutions will exist. However, in all cases, there is only one physically feasible solution (other solutions either complex or negative). As expected, the concentration of free repressor is very close to that of total repressor when $[CRO]_b$ is much greater than $[P_R O_R]_b$. On the other hand, $[CRO]$ deviates greatly from $[CRO]_b$ at the other end. With $[CRO]$ determined, the transcription efficiency can be obtained by replacing $[CRO]_b$ by $[CRO]$ in eq. (8).

The gene concentration $[G]$ is given by

$$[G] = \frac{G}{V_c N_A}, \quad (13)$$

where, for the chromosome,¹²

$$G^b = \exp \left[\int_{t-C-D}^{t-fC} \mu(\tau) d\tau \right], \quad (14)$$

and for a multicopy vector

$$G^p = G^b \bar{N}. \quad (15)$$

Parameter \bar{N} in equation (15) is the average vector copy number.

Depending upon which of the cases shown in Figure 1 is considered, the definitions and equations used for calculations must be adjusted as summarized in Table I.

MODEL PARAMETERS

Most of the parameters appearing in the model equations can be determined from information in the literature. A summary of the parameter values used for these simulations are listed in Table II. The transcription rate constant and the translation rate constant have been estimated from the relationship between the growth rate and these rate constants established previously.¹⁶ The doubling time of *E. coli* is assumed to be one hour. The parameters characterizing binding affinities for the *lac* repressor (K_{A1} to K_{E1}) and the *cro* repressor (K_1 to K_3) to their respective operator sites are assigned values corresponding to the wild-type system. Concentrations of the bacterial chromosome ($[G]^b$), the operator ($[O]_b$), and the nonspecific binding sites ($[D]_b$) have been calculated by using eqs. (13) and (14) with $C = 40$ min, $D = 20$ min, and $f = 0.5$.¹²

In the presence of cloning vectors, the total operator concentration ($[O]_b$), and the total nonspecific binding

Table I. A summary of model parameters and equations used for calculations for the different repressor expression designs as indicated in Figure 1.

Case	η_R	η_P	$[G]_R$	$[G]_P$	$[O]_b$	$[D]_b$
I	constant	Φ_P	$[G]^b$	$[G]^b \bar{N}$	$[O]_b^b \bar{N}$	$[D]_b^b (1 + c_1 \bar{N})$
II	constant	Φ_P	$[G]^b \bar{N}$	$[G]^b \bar{N}$	$[O]_b^b \bar{N}$	$[D]_b^b (1 + c_1 \bar{N})$
III	Φ_P	Φ_P	$[G]^b$	$[G]^b \bar{N}$	$[O]_b^b (1 + \bar{N})$	$[D]_b^b (1 + c_1 \bar{N})$
IV	Φ_P	Φ_P	$[G]^b \bar{N}$	$[G]^b \bar{N}$	$[O]_b^b \bar{N}$	$[D]_b^b (1 + c_1 \bar{N})$
V	η_λ	Φ_P	$[G]^b$	$[G]^b \bar{N}$	$[O]_b^b \bar{N}$	$[D]_b^b (1 + c_1 \bar{N})$
VI	η_λ	Φ_P	$[G]^b \bar{N}$	$[G]^b \bar{N}$	$[O]_b^b \bar{N}$	$[D]_b^b (1 + c_1 \bar{N})$
VII	η_λ	Φ_P	$[G]^b$	$[G]^b \bar{N}$	$[O]_b^b \bar{N}$	$[D]_b^b (1 + c_1 \bar{N})$
VIII	η_λ	Φ_P	$[G]^b \bar{N}$	$[G]^b \bar{N}$	$[O]_b^b \bar{N}$	$[D]_b^b (1 + c_1 \bar{N})$

Table II. Summary of model parameter values.

Parameter	Value	Reference
k_p^o	5.0M mRNA/M DNA min	16
k_q^o	10.0M protein/M mRNA min	16
k_d	0.5 min ⁻¹	16
k_e	0.01 min ⁻¹	16
K_1	$1.25 \times 10^8 M^{-1}$	14
K_2	$1.25 \times 10^8 M^{-1}$	14
K_3	$1.0 \times 10^9 M^{-1}$	14
K_{A1}	$2 \times 10^{12} M^{-1}$	12
K_{B1}	$1 \times 10^3 M^{-1}$	12
K_{C1}	$1 \times 10^7 M^{-1}$	12
K_{D1}	$2 \times 10^9 M^{-1}$	12
K_{E1}	$1.5 \times 10^4 M^{-1}$	12
$[G]_b^p$	$4 \times 10^{-9} M$	12
$[O]_b^p$	$4 \times 10^{-9} M$	12
$[D]_b^p$	$4 \times 10^{-2} M$	12
c_1	0.001	12

site concentration ($[D]_b$) are given as:

$$[O]_b = [O]_b^p + [O]_b^o = [O]_b^p(1 + \bar{N}) \quad (15a)$$

$$[D]_b = [D]_b^p + [D]_b^o, \quad (16)$$

For the case in which the promoter–operator site is not present in the chromosome,

$$[O]_b = [O]_b^p \bar{N} \quad (17)$$

For nonspecific binding, it follows by definition that the probability of protein binding is independent of nucleotide sequence. Accordingly, the number of nonspecific binding sites increases with the amount of DNA. Therefore, $[D]_b$ can be represented as

$$[D]_b = [D]_b^p(1 + c_1 \bar{N}) \quad (18)$$

where c_1 is the ratio of nonspecific binding site contents of the plasmid to those of the chromosome. The size of the cloning vector, which influences the parameter c_1 , has been chosen to be 0.1% of the *E. coli* chromosome. This corresponds, for example, to a vector such as pBR322.¹³

SIMULATION RESULTS

This work is concerned only with the steady-state solutions of the model described above. However, because these equations are highly nonlinear and not amenable to analytical solution, it is convenient to determine the steady-state conditions numerically by integrating the system of differential equations using a variable step-size Runge–Kutta algorithm with predictor–corrector until time-invariant values are approached asymptotically by all intracellular concentrations. This method was used to simulate a number of different scenarios as summarized next. Simulation results are numerically stable independent of initial conditions.

Transcription Efficiency under Uninduced Condition

With the current advance in genetic manipulation techniques, it is possible to modify the translation efficiency

by altering the ribosome binding site nucleotide sequence. For example, a ribosome binding site nucleotide sequence (TTAAAATTAAGGAGG, barred portion represents the Shine–Dalgarno sequence) with very high translation efficiency was synthesized and applied successfully to the production of cloned gene products up to 10–20% of the total intracellular protein.⁷ In Figures 2 and 3, the cloned operator activity Φ_P , which is directly related to the cloned gene transcription efficiency [eq. (5)], is plotted as a function of the vector copy number \bar{N} for two different translation efficiencies ($\xi_P = \xi_R$). Trends for all of the different repressor synthesis and regulation designs illustrated in Figure 1 are similar for both translation efficiencies.

For case I, in which repressor is provided by constitutive expression of a chromosomal gene, regulation of the cloned promoter–operator is lost as the vector copy number increases. For case II, in which repressor is obtained from constitutive expression of a repressor gene included in the expression plasmid, the transcription efficiency decreases as a function of the vector copy number.

Different relationships between plasmid copy number and the transcription efficiency is obtained for the auto-genous regulation cases (III and IV). When the repressor modulates its own promoter–operator in a chromosomal gene as in case III, there is modest loss of control with increase in vector content of the cell. However, η_P increases only gradually over the whole range of copy number. For case IV, in which the repressor gene with its autoregulated promoter–operator is included in the plasmid, the transcription efficiency is essentially independent of the vector copy number.

For cases V to VIII, expression of the *lac* repressor is controlled by cross regulation with the λP_R promoter.

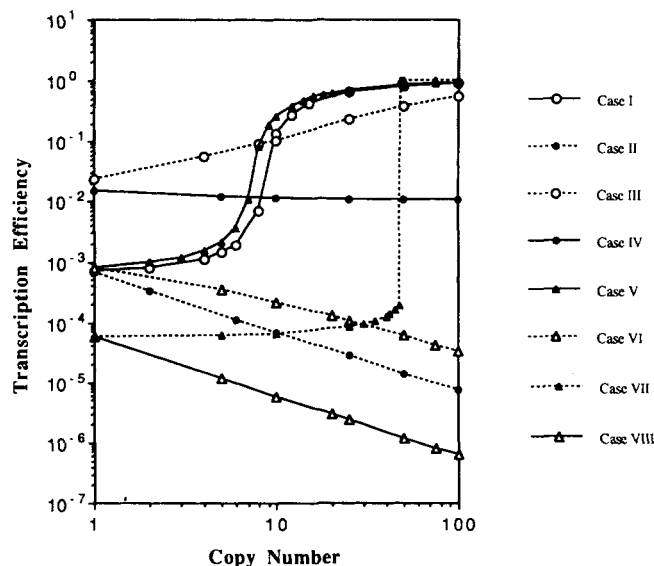


Figure 2. Model calculation of normalized uninduced transcription efficiency as a function of vector copy number for all the different cases with $\xi = 0.02$. For cases I and II, $\eta_R = 0.1$, in order to match the repressor concentration synthesized from the *lacI* promoter as reported.¹²

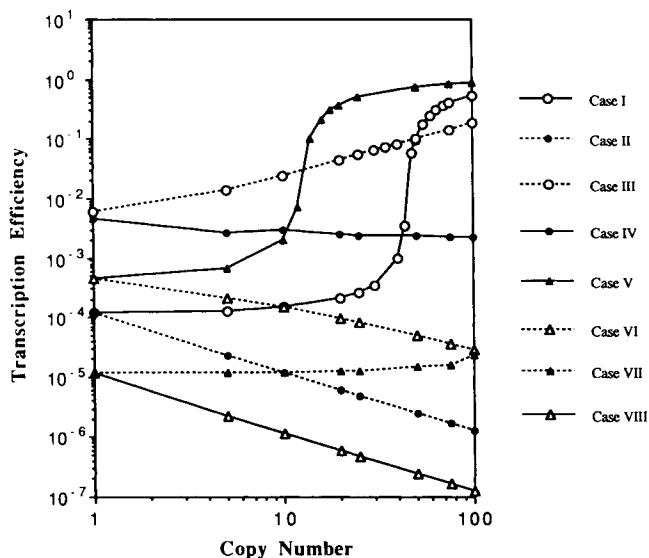


Figure 3. Model calculation of normalized uninduced transcription efficiency as a function of vector copy number for all the different cases with $\xi = 0.1$. This is to illustrate that the same trend is observed for the transcription efficiency with different ξ values.

However, for cases V and VI, the autorepressor (CRO) gene for the λP_R promoter is fused to form an operon with the *lac* repressor gene. In the final two cases, the autorepressor gene is fused to form an operon with the product gene. For case V, the transcription efficiency remains under control of the repressor up to a critical copy number. Beyond that critical copy number, the repressor gradually loses control of transcription. Similar to the behavior of the previous case, case VII also retains control of transcription up to a critical copy number. However, in this case the transcription efficiency changes rapidly from one extreme to another (i.e., changing from complete control to no control of transcription). Local stability analysis reveals that the low transcription efficiency state remains stable until the critical copy number where one of its eigenvalues changes sign (data not shown). Conversely, the opposite is true for the high transcription efficiency state. Finally, for cases VI and VIII, the system is able to control product transcription over the whole range of copy number with η_P decreasing with increasing copy number.

Although Φ_P is independent of ξ_P for the first six cases, it should be noted that cases VII and VIII could indeed be dependent on ξ_P . Simulation results reveal that trends are similar for all ξ_P/ξ_R ratios less than 1. Only when this ratio increases to around 10 will trends be reversed. That is, instead of controlling expression for the product gene, the expression of the repressor gene is being controlled. This situation can always be avoided, however, by appropriate genetic manipulation of, for example, the ribosome binding sites on product and/or repressor message.

Transcription Efficiency under Induced Condition

As mentioned in the Introduction, the production of cloned gene products usually involves two stages. A

growth phase in which cloned gene transcription is uninduced or repressed followed by a production phase in which cloned gene expression is induced or derepressed. Therefore, not only are we interested in understanding the transcription efficiency of the cloned gene under uninduced conditions, but it is also important to know how the transcription efficiency behaves under induced conditions for all cases considered. Inducer (IPTG) concentration for the *lac* promoter employed in all calculations was $10^{-3}M$.

In Figure 4, the overall transcription rate ($\bar{N}\eta_P$) is plotted as a function of copy number for all cases. Increasing copy number usually increases transcription because of increased cloned gene dosage. This trend holds for all cases considered except case II in which the overall transcription rate tends to increase at low copy number, reaches a maximum, and then decreases. This implies that at high copy number, the repressor level is too high for the presence of inducer to completely derepress cloned gene expression. A similar phenomenon is observed for case IV, where the overall transcription rate is always lower than for the other cases. In this case, the addition of inducer not only induces expression of the product, but also induces expression of the repressor. Therefore, high repressor concentration counteracts the derepression effect of the inducer and keeps the overall transcription rate low.

For cases VII and VIII, in which repressor is controlled by cross regulation, the overall transcription rate is the highest. The addition of inducer increases expression not only of the product but also of the cro repres-

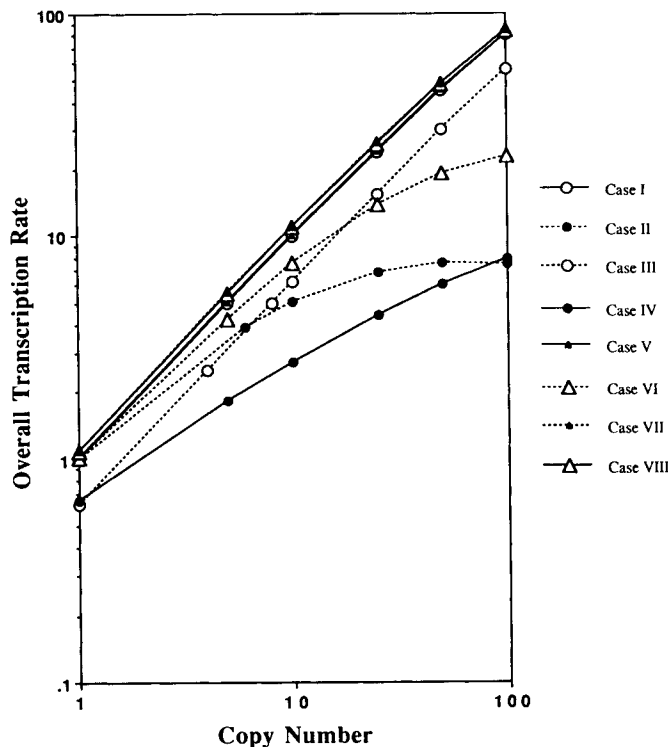


Figure 4. Model calculation of the normalized overall transcription rate as a function of vector copy number for all the different cases with $\xi = 0.02$. Inducer concentration $[I] = 1 \times 10^{-3}M$ for all cases.

sor. This represses expression of the *lac* repressor and further increases product expression. Cases I and V have slightly lower transcription rates and are followed by case VI. For all these configurations, the repression level is essentially independent of the inducer concentration. Therefore, the lower the repressor level, the higher the transcription rate (compare cases V and VI). An intermediate transcription rate is observed for case III. As in case IV, addition of inducer also induces repressor expression. Since the repressor gene is only present in a single copy in the chromosome, repressor level is not high enough to completely counteract the derepression effect of the inducer.

DISCUSSION

Several promoter-operator sequences such as those cloned from the *lac*, *trp*, and λP_L operons are widely used for regulation of cloned gene expression. The corresponding repressors are the *lacI*, *trpR*, and *cl* proteins, respectively. Of these, only *lacI* expression is constitutive. Transcription of the genes for *trpR* and *cl* are regulated by an autorepression system.^{9,21} Previous reports based upon experimental studies of the efficiency of cloned *lac* promoter-operator sequences and *trp* promoter-operators in maintaining control of cloned gene expression vary significantly. Using common cloning vectors with copy numbers in the range 20–30, a cloned *lac* promoter-operator provides transcription at nearly the induced level without any inducer present in the medium in a wild-type *E. coli* host.^{1,6,22} Use of host strains containing various *lacI* mutations giving overproduction of the *lac* repressor have been employed to provide retention of cloned product gene transcription repression in these system at modest copy number.²⁰ In the framework developed here, this corresponds to the host-vector system designated case I. This system has the disadvantage that control of cloned gene transcription will be eventually lost at some copy number. Furthermore, the level of repressor is constant independent of whether inducer is present or not. One solution for providing adequate repressor concentrations in order to maintain control of cloned gene transcription is inclusion of the repressor gene in the cloning vector as in case II. Although this system retains control of transcription, the overexpression of repressor leads to a low transcription rate after the addition of inducer. Therefore, both systems have deficiencies as a regulatory system for cloned gene expression.

By contrast, previous experimental studies show that cloned *trp* promoter-operators provide effective control of cloned gene transcription for multicopy vectors in the autorepression configuration shown in case III.^{10,11} Here, when the *trpR* gene product is titrated as the number of vectors in the cell increases, the autorepression system provides increased transcription of the chromosomal repressor gene, adapting the host-vector system to a variety of vector-host relationships.

The simulation results obtained here agree qualitatively with these previous experimental reports, although the data are limited to date. Also, such comparison must be qualified by the comment that *trp* promoter-operator transcription regulation includes some features that are different from those found in the *lac* promoter-operator system on which the present model and calculations are based. However, it is reasonable to expect that the trends shown here apply equally well to the *trp* system.

Although for case IV the system is able to retain control of transcription over the entire range of copy number (η_P independent of copy number), its overall transcription rate is relatively low in the induced state. This is not a desired feature, since a central goal is to maximize product protein production after induction of cloned gene expression.

Other alternatives considered here for controlling cloned gene transcription involve cross regulation of repressor expression. As of today, no experimental data are available to verify our simulation results. However, bearing in mind the qualitative success of the results from the first four cases as compared to experiment, it is reasonable to expect that the actual system will follow the trends calculated for the cross regulation system. Although many different types of promoter-repressor can be used for cross regulation, the λP_R promoter-autorepressor system was chosen for these example calculations since its kinetic model is well tested. The simulation results of cases V and VI are not significantly different from those of cases I and II; however, they are included in order to complete consideration of all different possibilities. Moreover, such results may not be easy to anticipate without model simulation.

From our simulation results, the best expression control systems are cases VII and VIII. The case VIII configuration is not only able to retain control of cloned product gene transcription in the uninduced state, it also provides the highest overall transcription rate in the induced state. Case VII also has attractive control characteristics, with the drawback of abrupt loss of regulation at a critical copy number. Additional simulations (results not shown) have shown that increasing the number of repressor gene copies in the chromosome will shift the critical copy number to higher values. In particular, by including two tandem copies of the repressor gene on the chromosome in case VII, this system is able to control cloned product gene transcription up to a copy number of 120 (data not shown), while retaining the high transcription rate in the induced state. An advantage for this configuration is a substantial reduction in the amount of repressor molecules synthesized compared to case VIII.

Although it is true that the cross-regulation configuration shows the highest efficiency in this study, we cannot exclude the possibility that a single-repressor system might be more effective than the dual-repressor system with different sets of parameter values (for ex-

ample, changes in translation efficiency or transcription rate constant by appropriate genetic manipulations). Since the transcription efficiency depends on the level of repressor content, such changes will undoubtedly alter the transcription efficiency by changing the amount of repressor. For example, if the overall transcription rate (OTR) values for different configurations are normalized to those at $\bar{N} = 1$, then the OTR for case III becomes as high as that for case VII or VIII. However, considering the reasonable case in which significant interactions between transcription and translation do not occur, the trend for the induction ratio [i.e. $\eta_P(\text{induced})/\eta_P(\text{uninduced})$] as a function of copy number is independent of the values of parameters such as k_p^o , k_q^o , ξ_P , and ξ_R . Figure 5 shows a plot of the induction ratio versus copy number for all different cases. Case VIII shows the highest induction ratio amongst all cases with induction ratio increasing versus copy number. On the other hand, for Case III, the induction ratio approaches one as copy number increases. Therefore, allowing high level induction of transcription for such a system would result in a loss of transcription control under uninduced conditions.

This study illustrates how a kinetic model of recombinant systems formulated at the molecular level can be used to explore the qualitative and eventually some quantitative features provided by different genetic designs. Given the extremely powerful capabilities available now for precise design of the promoter-operator region at the nucleotide sequence level, it is important to utilize such quantitative and systematic modeling relationships to assess complex interactions in recombinant systems and to compare the effects of many alternative molecular designs which are in principle available. Current capabilities for molecular genetic manipulation provide so many options that empirical trial and error

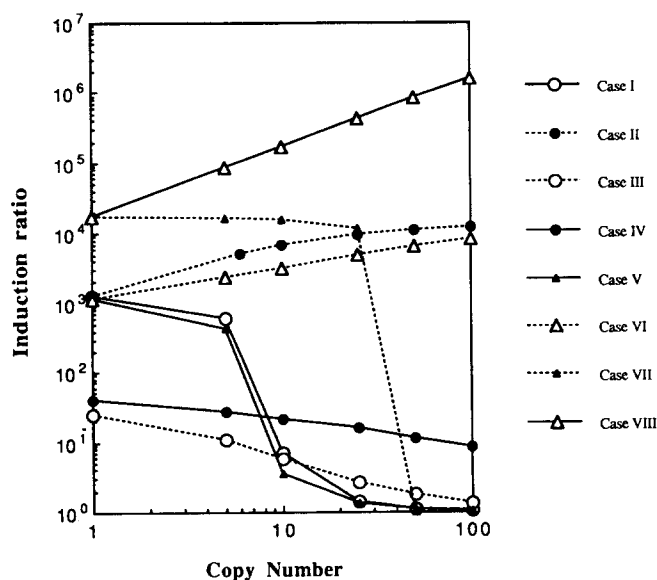


Figure 5. Model calculation of induction ratio for all the different cases with $\xi = 0.02$ and parameter values listed in Table II.

approaches are inadequate for optimization. Instead, quantitative, systematic models are necessary for rational and efficient expression system design.

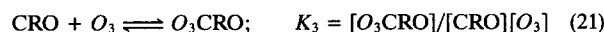
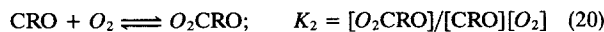
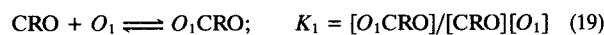
The main objective of this work is to illustrate the major qualitative features of the alternative arrangements for repressor synthesis as summarized in Figure 1. However, in order to optimize the design of the product gene promoter-operator and the production and regulation of the associated repressor, one must take into account different demands on the host cell metabolic and biosynthetic capacity posed by the different repressor expression design and the effects of these demands on product gene expression level. These considerations must be merged with the economic criteria for the process and the available process environmental manipulations in order to evaluate and to optimize the host genotype and the molecular design of the plasmid. In this context, combining our current models with the kinetic models for product formation of recombinant culture such as those developed by Lee, Seressiotis, and Bailey¹⁸ and Betenbaugh and Dhurjati⁵ should be useful.

The success of the present study in simulating trends for alternative regulation configurations for repressor design is quite encouraging and suggests that this framework might be applicable for other promoters. However, these trends are not necessarily universal, and full analysis of any particular system (either experimental or theoretical with appropriate model parameters) is recommended.

This work was supported by the Catalysis and Biocatalysis Program of the Advanced Industrial Concepts Division of the U.S. Department of Energy. One of the authors (SBL) was supported in part by the Korea Science and Engineering Foundation (KOSEF).

APPENDIX

Assuming that nonspecific binding of the cro repressor is not important, the cro repressor can bind to each of the three operator sites on the λP_R promoter-operator region ($P_R O_R$):



Total balance on cro repressor and each of the three operators:

$$[\text{CRO}]_b = [\text{CRO}] + [O_1\text{CRO}] + [O_2\text{CRO}] + [O_3\text{CRO}] \quad (22)$$

$$[O_1]_b = [O_1] + [O_1\text{CRO}] \quad (23)$$

$$[O_2]_b = [O_2] + [O_2\text{CRO}] \quad (24)$$

$$[O_3]_b = [O_3] + [O_3\text{CRO}] \quad (25)$$

Combining eqs. (23)–(25) and eqs. (19)–(21) yields:

$$[O_i\text{CRO}] = \frac{K_i [O_i]_b [\text{CRO}]}{1 + K_i [\text{CRO}]} \quad (26)$$

The index i refers to operator sites 1, 2, or 3. Furthermore, the total concentration of each of the three operator sites is the same;

therefore,

$$[O_1]_b = [O_2]_b = [O_3]_b = [P_R O_R]_b \quad (27)$$

Substituting eqs. (26) and (27) into eq. (22) yields eq. (12).

NOMENCLATURE

C	time required for chromosome replication
c_1	number of nonspecific binding sites per plasmid/number of nonspecific binding sites per chromosome
[CRO]	cro repressor concentration
D	time between completion of a round of replication and subsequent cell division
[D]	concentration of nonspecific binding sites for repressor
G	number of DNA molecules
f	a fractional distance of gene from the origin
[G]	DNA concentration
[I]	inducer concentration
k_d	mRNA decay rate constant
k_e	protein decay rate constant
k_p^o	overall transcription rate constant
k_q^o	overall translation rate constant
K_1, K_2, K_3	binding affinities for cro repressor
$K_{A1}, K_{B1}, K_{C1}, K_{D1}, K_{E1}$	binding affinities (see Ref. 12)
[mRNA]	mRNA concentration
\bar{N}	average plasmid copy number (number of plasmid molecules/number of chromosomes)
N_A	Avogadro's number
[O]	operator concentration
[P]	cloned gene product concentration
[$P_R O_R$]	λP_R operator concentration
[R]	repressor concentration
[R_2]	second repressor concentration
t	time
V_c	cell volume
<i>Greek letters</i>	
η	overall transcription efficiency
ξ	overall translation efficiency
μ	specific growth rate
Ψ_P	probability of repressor binding to its specific binding site
Φ_P	fraction of operators without bound repressor
<i>Subscripts</i>	
P	cloned gene product
R	repressor
o	total concentration
<i>Superscripts</i>	
b	bacterial chromosome
p	plasmid

References

- Backman, K., Ptashne, M., Gilbert, W. 1976. Construction of plasmids carrying the cl gene of bacteriophage λ . Proc. Natl. Acad. Sci. USA. **75**: 4174-4178.
- Beckwith, J., Grodzicker, T., Arditti, P. R. 1972. Evidence for two sites in the *lac* promoter region. J. Mol. Biol. **69**: 155-160.
- Beckwith, J. R., Ziper, D. (ed.) 1970. The Lactose Operon. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Bernard, H. V., Helinski, D. R. 1980. Bacterial plasmid cloning vehicles. p. 133-168 In: J. K. Setlow and A. Hollander (ed.), Genetic Engineering: Principles and Methods, vol 2. Plenum, New York.
- Betenbaugh, M. J., Dhurjati, P. 1990. A comparison of mathematical model predictions to experimental measurements for growth and recombinant protein production in induced cultures of *Escherichia coli*. Biotechnol. Bioeng. **36**: 124-134.
- Clewell, D. B., Helinski, D. R. 1972. Effect of growth conditions on the formation of the relaxation complex of supercoiled ColE1 deoxyribonucleic acid and protein in *Escherichia coli*. J. Bacteriol. **110**: 1135-1146.
- Crowel, R., Seamans, C., Lomedico, R., McAndrew, S. 1985. Versatile expression vectors for high-level synthesis of cloned gene products in *Escherichia coli*. Gene. **38**: 31-38.
- Guarante, L., Yocum, R., Gilford, F. 1982. GAL10-CYC1 hybrid yeast promoter identifies the GAL4 regulatory region as an upstream site. Proc. Natl. Acad. Sci. USA. **79**: 7410-7414.
- Gunsalus, R. P., Yanofsky, C. 1980. Nucleotide sequence and expression of *Escherichia coli trpR*, the structural gene for the trp aporepressor. Proc. Natl. Acad. Sci. USA. **77**: 7117-7121.
- Hallewell, R. A., Emtage, S. 1980. Plasmid vectors containing the tryptophan operon promoter suitable for efficient regulation expression of foreign genes. Gene. **9**: 27-47.
- Hershfield, V., Boyer, H. W., Yanofsky, C., Lovett, M. A., Helinski, D. R. 1974. Plasmid ColE1 as a molecular vehicle for cloning and amplification of DNA. Proc. Natl. Acad. Sci. USA. **71**: 3455-3459.
- Lee, S. B., Bailey, J. E. 1984. Genetically structured models for *lac* promoter-operator function in the *Escherichia coli* chromosome and in multicopy plasmids: *lac* operator function. Biotechnol. Bioeng. **26**: 1372-1382.
- Lee, S. B., Bailey, J. E. 1984. Genetically structured models for *lac* promoter-operator function in the chromosome and in multicopy plasmids: *lac* promoter function. Biotechnol. Bioeng. **26**: 1383-1389.
- Lee, S. B., Bailey, J. E. 1984. A mathematical model for λdv plasmid replication: analysis of wild-type plasmid. Plasmid. **11**: 151-165.
- Lee, S. B., Bailey, J. E. 1984. A mathematical model for λdv plasmid replication: analysis of copy number mutants. Plasmid. **11**: 166-177.
- Lee, S. B., Bailey, J. E. 1984. Analysis of growth rate effects on productivity of recombinant *Escherichia coli* populations using molecular mechanism models. Biotechnol. Bioeng. **26**: 66-73.
- Lee, S. B., Ryu, D. D. Y., Siegel, R., Park, S. H. 1988. Performance of recombinant fermentation and evaluation of gene expression efficiency for gene product in a two-stage continuous culture system. Biotechnol. Bioeng. **31**: 805-820.
- Lee, S. B., Seressiotis, A., Bailey, J. E. 1985. A kinetic model for product formation in unstable recombinant populations. Biotechnol. Bioeng. **27**: 1699-1709.
- Miller, J. H., Reznikoff, W. S. (ed.) 1978. The Operon. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- O'Farrell, P. H., Polisky, B., Gelfand, D. H. 1978. Regulated expression by readthrough translation from a plasmid-encoded β -galactosidase. J. Bacteriol. **134**: 645-654.
- Ptashne, M. 1971. Repressor and its action. p. 221-238 In: A. D. Hershey (ed.), The Bacteriophage Lambda. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Sadler, J. R., Tecklenburg, M., Bete, J. L. 1980. Plasmids containing many tandem copies of a synthetic lactose operator. Gene. **8**: 279-300.
- Seo, J. H., Bailey, J. E. 1985. Effect of recombinant plasmid content on growth properties and cloned gene product formation in *Escherichia coli*. Biotechnol. Bioeng. **27**: 1668-1674.
- Siegel, R., Ryu, D. D. Y. 1985. Kinetic study of instability of recombinant plasmid pPLc23trpA1 in *E. coli* using two-stage continuous culture system. Biotechnol. Bioeng. **27**: 28-33.
- Thompson, R. 1982. Plasmid and phage M13 cloning vectors. p. 1-52 In: R. Williamson (ed.), Genetic Engineering 3. Academic, London.