PLASMID PROPAGATION IN THE YEAST SACCHAROMYCES CEREVISIAE: FLOW CYTOMETRY STUDIES AND SEGREGATED MODELING

Thesis by
Karl Dane Wittrup

In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

California Institute of Technology
Pasadena, California
1989

(Submitted August 29, 1988)
Acknowledgments

I would like to express my gratitude and friendship to Jay Bailey for showing me a whole new world of interesting problems to work on, and for bringing together and supporting the Bailey research group, which has assumed a life of its own.

Much thanks to Nancy Anderson DaSilva, who introduced me to the ways and means of our mutual acquaintance Saccharomyces cerevisiae, and was always available as a sounding board for the latest lunatic idea or minor triumph.

Thanks to the people who made leisure hours (both of them) livable: John, a true basketball maniac and poolshark; Doug, a great buddy for pumping, pranking, and philosophizing; Kevin, a reluctant fellow monk in the church of Science; and a cast of thousands (you know who you are.)

Thanks to Ree, for providing excuses for occasional diversions and an excuse to be a SoCal tourist. Thanks to Margaret, for providing shelter from the storm when it was sorely needed, and to Aaron and Charlotte, for providing a constant wellspring of hope, joy, and renewal.

Lastly, thanks to my mother would be superfluous. She taught me just about everything I know about being a human being; she is so close to my heart that to thank her would be like thanking myself.
To Mom
Abstract

The baker's yeast *Saccharomyces cerevisiae* is potentially a very useful host for the production of pharmaceutical proteins by recombinant DNA technology. One requirement for efficient overproduction of a foreign protein in yeast is a stable recombinant DNA vector which is maintained at a high number of copies per cell. The rational design of such vectors requires knowledge concerning their propagation in a cell population. The purpose of this work is to develop mathematical and experimental tools for the study of multicopy plasmid propagation, and to apply these tools to the investigation of a particular type of yeast vector: a conditional centromere plasmid.

A method for measuring the distribution of plasmid copy number in yeast populations was developed, using β-galactosidase activity as a marker for plasmid copy number. Enzyme activity is assayed at the single-cell level using a fluorogenic substrate and flow cytometry. The relationship between single-cell fluorescence and enzyme activity is described by a mathematical reaction-diffusion model.

A segregated mathematical modeling framework was established to link measured copy number distributions with probabilistic models of single-cell plasmid replication and partitioning. Simplifications of the general integral-partial differential population balance equations were obtained for a discrete state variable, resulting in a linear system of ordinary differential equations.

Flow cytometry and segregated modeling were applied to the study of a conditional centromere plasmid. This type of plasmid can be amplified to high copy number by unequal partitioning, but the amplified copy number state is unstable in the absence of selection pressure. A segregated model of this plasmid's propagation was shown to be consistent with experimental observations. The conceptual model of plasmid instability suggests changes in the attributes of the host cell
and plasmid construction to improve stability at high copy number. A segregated mathematical model of this type is necessary for the design of bioreactor operating conditions that optimize productivity.
Table of Contents

Acknowledgements i

Abstract v

Chapter 1. Introduction 1
    References 7

Chapter 2. A Mathematical Model of Recombinational Amplification 11
    of the 2μ Plasmid in the Yeast Saccharomyces cerevisiae.

Chapter 3. A Segregated Model of Recombinant 24
    Multicopy Plasmid Propagation

Chapter 4. A Single-Cell Assay of β-Galactosidase Activity 32
    in Saccharomyces cerevisiae

    Abstract 45
    Introduction 46
    Results and Discussion 48
    References 58
    Figures 60

Chapter 6. Propagation of an Amplifiable Recombinant 69
    Plasmid in Saccharomyces cerevisiae: Flow Cytometry
    Studies and Segregated Modeling 70
    Abstract 71
    Introduction 74
    Materials & Methods 76
    Results 88
    Discussion 92
    References 95
    Figures
CHAPTER 1.

INTRODUCTION
The baker's yeast *Saccharomyces cerevisiae* is an attractive host for the expression of heterologous proteins (1-20). Yeast is extensively used as an industrial microorganism for the production of food-related products, and the technology required to grow yeast to high cell densities in large, process-scale equipment is well established. The open literature contains papers describing the use of yeast to produce heterologous proteins at the biotechnology companies Amgen (4-8), Chiron (9-11), Smith Kline & French Laboratories (12), Genentech (13), Hoffman-La Roche (14), INGENE (15), Cetus (16), Biogen (17-18), and Collaborative Research (19-20). *S. cerevisiae* is Generally Recognized As Safe (GRAS) by the FDA and is free of pyrogens and toxins, factors which are significant for the production of pharmaceutical proteins. The genetics and physiology of yeast have been thoroughly studied, and powerful techniques for genetic manipulation of yeast have been developed (21,22). Being a eucaryotic organism, *S. cerevisiae* carries out some of the post-translational protein modifications that higher organisms do, such as secretion signal peptide cleavage and glycosylation (1,4,9,10,15-18,20).

There are relatively few basic types of recombinant vectors for transformation of *S. cerevisiae* (23). Integrating vectors are incorporated into chromosomes by homologous recombination, and are stable but present at low copy number. ARS plasmids contain an Autonomously Replicating Sequence, and are very unstable. Centromere plasmids are often called artificial chromosomes, because their replication and partitioning functions are directed by fragments of chromosomal DNA. Centromere plasmids are very stable, but usually present at a copy number of one. 2µ-based plasmids contain sequences from the native 2µ plasmid, and can be propagated at high copy number. Most 2µ-based recombinant plasmids are lost from the population after 100 generations of growth in non-selective medium, which limits their usefulness for large-scale processes.
Mitotically stable, high copy number vectors are required for efficient expression of heterologous proteins. The ultimate goal of this research is to gain a sufficient understanding of multicopy plasmid propagation in yeast to allow rational improvement in vector design.

Because yeast plasmids are replicated by the chromosomal replication machinery (24-26), it is not possible to amplify plasmid copy number by initiating multiple replication forks at a single origin of replication during a single cell cycle. Two other types of plasmid amplification have been characterized, however: recombinational amplification, and amplification by biased segregation.

It has been proposed(27) and experimentally verified(28-32) that the native 2μ plasmid in yeast can replicate multiple times from a single origin of replication by converting from theta-mode replication to a double rolling circle mode of replication via the site-specific FLP recombination system. The inherent limits on this type of amplification were studied by use of a simplified mathematical model of the process. The results of this study are presented in Chapter 2.

The biased segregation mode of plasmid amplification takes advantage of an inherent tendency that yeast cells possess to partition circular plasmids to the mother cell at cell division(33). This process can partition a large number of plasmids into a small fraction of the population. By selecting for the subpopulation with amplified copy number, the population mean copy number is increased.

The biased segregation process results in a heterogeneous population. Considerable information is present in the form of this heterogeneity which would be lost in measurement of population mean attributes. As an example of the significance of population heterogeneity, consider the example in Figure 1. Four simulated distributions of a cellular state variable are shown. The simulations are reasonable facsimiles of actual distributions of single-cell mass, DNA, pH, and surface
markers. The mean of each of these distributions is 1.0, but it is readily apparent that there are significant differences between the four distributions that indicate fundamental departures from the condition of the "average" cell. Because cell growth and metabolism are cyclic, at any given time only a small fraction of the cells in a population are in the "average" state. In the context of a biotechnological process, the productivity of a cell population may differ from the productivity of the "average" cell in the population, if single-cell productivity depends in a nonlinear fashion on the cellular state variable.

The paradigm adopted in this work is to measure the plasmid copy number distribution and use segregated mathematical modeling to link the measured distributions to single-cell models of multicopy plasmid propagation. In order to accomplish this goal, it was necessary to develop a mathematical framework for description of multicopy plasmid propagation. Simplifications of the population balance equations were obtained for copy number as a discrete state variable, and the effects of plasmid presence on host growth rate were incorporated separately from the copy number transition probabilities. The results of this work are presented in Chapter 3.

In this work, β-galactosidase activity was used as a marker for plasmid copy number. In order to measure the copy number distribution, it was necessary to develop a quantitative assay of β-galactosidase activity at the single-cell level. Previous attempts to measure single-cell activity have met with limited success due to single-cell reaction product leakage, inadequate quantitation, insensitivity, and inhibition of the enzyme by some of the assay reagents. A quantitative flow cytometric β-galactosidase assay was developed, and is described in Chapter 4. A mathematical model characterizing the assay kinetics is presented in Chapter 5.

The analytical and experimental tools for modeling and measuring copy num-
ber distributions were applied to an investigation of conditional centromere plasmid propagation. This type of plasmid can be amplified to high copy numbers by biased segregation, but the amplified copy number state is unstable in the absence of selection pressure. The results of this study are presented in Chapter 6.
Relative number of cells

Figure 1.
REFERENCES


17. Ernst, J.F., "Efficient Secretion and Processing of Heterologous Proteins in
Saccharomyces cerevisiae is Mediated Solely by the Pre-Segment of α-Factor Precursor." DNA 7: 355 (1988).


CHAPTER 2.

A Mathematical Model of Recombinational Amplification of the
2μ Plasmid in the Yeast *Saccharomyces cerevisiae*
A Mathematical Model of Recombinational Amplification of the 2μ Plasmid in the Yeast Saccharomyces cerevisiae

KARL DANE WITTRUP AND JAMES E. BAILEY

Department of Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, U.S.A.

(Received 26 June 1987, and in revised form 23 October 1987)

A mathematical model of 2μ plasmid recombinational amplification in Saccharomyces cerevisiae has been developed, based on mechanisms of 2μ recombination and replication presented in the literature. A probabilistic description reveals the limits inherent in the recombinational mode of plasmid amplification. These limits correspond well with values calculated from reported results. In the model, copy number control is effected by the constitutive expression of a repressor of recombinase expression. Estimation of the model parameters is accomplished via a set of heuristic rules which restrict the feasible parameter space considerably. It is demonstrated that many parameter sets arbitrarily chosen from the feasible parameter space reproduce the observed characteristics of 2μ plasmid amplification: rapid correction of downward copy number deviations, with a lack of strict control of steady-state copy number.

Introduction

The 2μ plasmid is a 6318 bp circular double-stranded DNA plasmid found in most Saccharomyces cerevisiae strains at copy numbers from 50 to 100 (Broach, 1980). Each plasmid generally replicates once during S phase (Zakian et al., 1979), and replication depends on the same set of genes as chromosomal replication (Livingston & Kupfer, 1977). Under certain circumstances, however, the copy number of the 2μ plasmid can be amplified (Sigurdson et al., 1981; Brewer & Fangman, 1980). No significant phenotypic effect due to the presence of 2μ plasmid has been found.

The 2μ plasmid contains four large open reading frames, designated A, B, C, and D. The B and C products act in trans with the cis-acting STB (stability) site (also known as the REP3 site) to promote efficient plasmid partitioning (Cashmore et al., 1986). The B product intercalates into the nuclear lamina (Wu et al., 1987), and appears to bind to the STB site (Veit & Fangman, 1985).

The A product, also called FLP, efficiently catalyzes site-specific recombination at a site present in each of two 599 bp inverted repeats on the plasmid (Andrews et al., 1987). A model has been proposed whereby this recombination can enable multiple replications of a plasmid template from a single initiation of replication (Futcher, 1986). This type of recombinational amplification has been verified experimentally (Volkert & Broach, 1986).
The purpose of this paper is to integrate current knowledge concerning the regulation of 2μ plasmid replication into a mathematical model, and demonstrate that this model can describe 2μ copy number control.

**Amplification Model**

The amplification model proposed by Futcher (1986) consists of the following steps (see Fig. 1):
1. Bidirectional (theta-type) replication is initiated at the origin of replication.
2. One of the recombination sites is replicated, shortly after initiation.
3. A FLP-mediated recombination occurs. This converts the plasmid to a double-rolling-circle mode, with replication forks moving in the same direction around the original template, creating a multimer.
4. Another recombination occurs, reverting the plasmid to theta-mode replication, and terminating amplification.
5. Subsequent recombinations resolve the multimer into monomers.

![Diagram of Amplification Model](image)

**Fig. 1.** Steps of recombinational plasmid amplification.

If it is assumed that the probability of a recombinational event is independent of the replication mode of the plasmid, a simple probabilistic model can be used to predict the mean amplification achieved by this mechanism.

Let

- $O$ = the event that an odd number of intramolecular recombinations occurs on one plasmid during the time required for one plasmid template to be replicated.
- $\bar{O}$ = the event that an even number of recombinations occur.
PLASMID AMPLIFICATION MODEL

The number of replications \(n\) of each template capable of amplification depends on the sequence of events after replication initiation as follows:

\[
\begin{array}{c|c}
\text{n} & \text{Events} \\
1 & \mathcal{O} \\
2 & O \ O \\
3 & O \ \mathcal{O} \ O \\
4 & O \ \mathcal{O} \ \mathcal{O} \ O \\
5 & O \ \mathcal{O} \ \mathcal{O} \ \mathcal{O} \ O \\
\vdots & \vdots \\
\end{array}
\]

If \(f(n)\) = probability that the template is replicated \(n\) times, then

\[
f(n) = \begin{cases} 
1-p & n = 1 \\
(n-2)p^2(1-p)^{n-2} & n \geq 2,
\end{cases}
\]

where

\[p = \text{probability of the event } O.\]

The mean value of \(n\) is

\[
\bar{n} = \sum_{n=1}^{\infty} nf(n)
\]

\[
= 1 - p + \sum_{n=2}^{\infty} n p^2 (1-p)^{n-2}
\]

\[
= \begin{cases} 
2 & 0 < p \leq 1 \\
1 & p = 0.
\end{cases}
\]

This surprising result states that a plasmid capable of recombination will be replicated an average of two times regardless of the rate of recombination. This can be viewed as the consequence of two opposing trends: as \(p\) increases, over-replication becomes more likely; however, although the first event \(O\) initiates amplification, the next event \(O\) terminates amplification. So, as the probability of amplification increases, the mean magnitude of amplification decreases.

Only those plasmids bound by FLP can be amplified. Thus, if \(f_A\) is the fraction of plasmids with the correct FLP binding to enable recombination, the mean number of replications per template is

\[
\bar{n} = f_A \cdot \bar{n} + (1-f_A)
\]

\[
= 1 + f_A.
\]

Therefore, the expected total number of plasmids will increase by a factor of \(2+f_A\) for every chromosomal doubling. In other words, mean plasmid number should at most triple during one cell cycle.

Let

\[
r = 2 + f_A
\]

then

\[
2 \leq r \leq 3.
\]
Values of $r$ have been calculated from several experiments in the literature. The results are shown in Table 1.

**Table 1**

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Post-sporulation replication ratios (Brewer &amp; Fangman, 1980)</td>
<td>3.03</td>
</tr>
<tr>
<td>(2) Regulated expression of FLP (Volkert &amp; Broach, 1986)</td>
<td>2.70; 3.4</td>
</tr>
<tr>
<td>(3) $2\mu$ transmission from cytoductant formation (Sigurdson et al., 1981)</td>
<td>2.34-2.40</td>
</tr>
</tbody>
</table>

In experiment (1) (Brewer & Fangman, 1980), the ratio of post-meiotic replication rates was calculated using double-radiolabel techniques. The ratio of fractional rates of increase of chromosomal DNA to $2\mu$ plasmid DNA was found to be 0.66. Thus,

$$r = 2\left(\frac{1}{0.66}\right)$$

$$= 3.03.$$  

It was also found that only 82% of $2\mu$ DNA is transmitted from the diploid parent to the haploid spores. Perhaps this is responsible for the induction of $2\mu$ amplification.

In experiment (2) (Volkert & Broach, 1986), FLP was inserted into the chromosome of *S. cerevisiae* under the control of a galactose-inducible promoter. Induction of FLP expression in a cir$^0$ strain caused the excision of a flp$^{-}$ $2\mu$ circle integrated into another chromosome. This single copy of the $2\mu$ plasmid was then observed to undergo "1.7 doublings per S phase". This gives the value 3.4 in Table 1. It was also noted that the $2\mu$ plasmid is present at "more than 20 copies per cell" after "at least 10 generations". From this information,

$$r = 2(20)^{1/10}$$

$$= 2.70.$$  

In experiment (3), $2\mu$ plasmid was transferred to cir$^0$ strains by cytoductant formation. From experiments with different types of plasmids, it was determined that more than one plasmid was transmitted in less than 2% of the matings. $2\mu$ copy number was measured at "close to the 50 or 100 copies per cell normally found", after "approximately twenty-five divisions". Thus, $r$ is calculated as follows:

$$r = 2(50)^{1/25}$$

$$= 2.34.$$  

$$r = 2(100)^{1/25}$$

$$= 2.40.$$
PLASMID AMPLIFICATION MODEL

All but one of the values in Table 1 are in the range expressed in eqn (4). This is remarkable, considering the diversity of experimental conditions. The most direct measurement of the capacity for amplification was performed by Volkert & Broach (1986). In their system, FLP expression was not regulated by the plasmid, so amplification should have been maximal. The fact that \( r \) was greater than 3 in this case could be due to several factors: experimental uncertainty; stochastic variation of the actual sample mean; or shortcomings of the simplified model. Nevertheless, the overall agreement between experimental results and the prediction of the model is excellent, providing strong support for its acceptance as a simplified description of recombinational amplification.

2µ Copy Number Control Model

A mathematical model of 2µ copy number control has been formulated, based on the following assumptions:

1. Plasmid amplification can be described by the model in the previous section.
2. There are four FLP binding sites, and all four must be bound in order to recombine. Evidence indicates that there are six sites capable of binding FLP on the 2µ plasmid; however, only four are necessary for efficient recombination (Andrews et al., 1987).
3. FLP binding at each site is independent of binding at other sites, and is at equilibrium. It has been suggested that a protein–protein interaction stabilizes the binding of two adjacent FLP molecules (Andrews et al., 1987). This effect will be neglected to maintain the simplicity of the model.
4. A repressor molecule \( R \) is constitutively expressed at a level proportional to plasmid number. The assumption of product-gene dosage proportionality at low gene dosage has been verified with a different expression system in S. cerevisiae (Srienc et al., 1986). This assumption will be less valid at high gene dosages.
5. Repressor \( R \) binds to the FLP promoter region, and this binding is at equilibrium. Nuclease protection experiments have implicated the C gene product as a regulator of FLP expression, since its presence alters chromatin structure at the 5' end of the FLP gene (Veit & Fangman, 1985). 2µ plasmids containing disrupted C and D reading frames, but with an intact FLP gene, are amplified to a copy number of 650 (Veit & Fangman, 1985).
6. FLP protein level is proportional to the number of FLP genes not bound by repressor. On-off control by the repressor is the simplest assumption.
7. Plasmids are partitioned equally between cells at division. The purpose of this model is to study replication and recombination effects. In order to incorporate partitioning effects, it would be necessary to use a segregated modeling approach, as in Wittrup & Bailey (1987).

Repressor Expression and Binding

Constitutive expression of repressor is modeled by the following equation:

\[
R_e = \alpha_R P_e. \tag{5}
\]
where
\[ R_o = \text{total repressor concentration}, \]
\[ P_o = \text{total plasmid concentration}, \]
\[ \alpha_R = \text{constant}. \]

Equilibrium binding is modeled by
\[ R + P \rightleftharpoons RP \]
\[ K_R = \frac{RP}{R \cdot P}, \]  \hspace{1cm} (6)

where
\[ R = \text{free repressor concentration}, \]
\[ P = \text{free plasmid concentration}, \]
\[ RP = \text{plasmid–repressor complex concentration}, \]
\[ K_R = \text{repressor binding constant}. \]

Species balances yield
\[ P_o = P + RP \]  \hspace{1cm} (7)
\[ R_o = R + RP. \]  \hspace{1cm} (8)

Assuming that \( R_o \approx R \), and putting eqns (7) and (8) into eqn (6), the following is obtained:
\[ f_R = \frac{\alpha_R K_R P_o}{1 + \alpha_R K_R P_o}, \]  \hspace{1cm} (9)

where
\[ f_R = \text{fraction of plasmids bound by repressor, } \frac{RP}{P_o}. \]

The assumption that \( R_o \approx R \) is valid if the number of repressor molecules significantly exceeds the number of plasmids. Although this assumption has not been verified experimentally, its use significantly simplifies the analysis.

**FLP Expression and Binding**

The following model for FLP binding is proposed:

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Equilibrium constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P + F \rightleftharpoons PF )</td>
<td>( 4K_F )</td>
</tr>
<tr>
<td>( PF + F \rightleftharpoons PF_2 )</td>
<td>( \frac{1}{2}K_F )</td>
</tr>
<tr>
<td>( PF_2 + F \rightleftharpoons PF_3 )</td>
<td>( \frac{1}{3}K_F )</td>
</tr>
<tr>
<td>( PF_3 + F \rightleftharpoons PF_4 )</td>
<td>( \frac{1}{4}K_F )</td>
</tr>
</tbody>
</table>
where

\[ F = \text{free FLP concentration}, \]
\[ PF_i = \text{concentration of plasmid–FLP complexes with } i \text{ FLP molecules bound}, \]
\[ K_F = \text{equilibrium constant for single site binding}. \]

The numerical factor in each equilibrium constant above denotes the ratio of vacant FLP sites on the reacting complex to the number of occupied FLP sites on the product complex.

Species balances are written as follows:

\[ P_o = P + PF + PF_2 + PF_3 + PF_4 \]  
\[ F_o = F + PF + 2PF_2 + 3PF_3 + 4PF_4, \]  
where

\[ F_o = \text{total FLP concentration}. \]

FLP expression is modeled by the following equation:

\[ F_o = \alpha_F (1-f_R) P_o, \]

where

\[ \alpha_F = \text{ratio of FLP molecules to plasmid molecules, if under constitutive expression}. \]

The problem is nondimensionalized as follows:

\[ \hat{F} = \frac{F}{\alpha_F P_o} \]
\[ \hat{\rho} = \frac{P}{P_o}. \]

These quantities vary between 0 and 1.

Combining equilibrium expressions for the FLP binding reactions with eqns (10)-(12), the following is obtained:

\[ \alpha_F (1-f_R) = \alpha_F \hat{F} + \frac{4\phi + 12\phi^2 + 12\phi^3 + 4\phi^4}{1 + 4\phi + 6\phi^2 + 4\phi^3 + \phi^4} \]  
\[ \hat{\rho} = \frac{1}{1 + 4\phi + 6\phi^2 + 4\phi^3 + \phi^4}, \]

where

\[ \phi = \alpha_F K_F P_o \hat{F}. \]

The amplifiable fraction \( f_A \), defined as \( PF_4/\rho \), is found to be

\[ f_A = \phi^4 \hat{\rho}. \]

Equations (13a–c) express \( f_A \) as a function of \( P_o \), with the parameters \( \alpha_F, K_F \), and \( \alpha_R K_R \). These parameters can be transformed to more intuitively appealing
quantities as follows:

$$f_{R_1} = \frac{\alpha_R K_R}{1 + \alpha_R K_R},$$

(14)

where

$$f_{R_1} = \text{fraction of plasmids in a cell population bound by repressor under constitutive expression when plasmid number is 1 in all cells},$$

and

$$f_{F_1} = \frac{\alpha_F K_F}{1 + \alpha_F K_F},$$

(15)

where

$$f_{F_1} = \text{hypothetical fraction of FLP sites bound by FLP under constitutive expression when plasmid number is 1}.$$

Thus, $f_A$ can be expressed as a function of $P_0$ with the parameters $\alpha_F, f_{F_1}$, and $f_{R_1}$.

**Parameter Estimation and Simulation Results**

It is interesting to consider the model in terms of a system under feedback control. The controlled variable is $P_0$, the manipulated variable is $f_A$, and the controller parameters are $f_{R_1}, f_{F_1}$, and $\alpha_F$. In order to consider the behavior of this model, it is necessary to obtain reasonable estimates of the parameters $f_{R_1}, f_{F_1}$, and $\alpha_F$. Experimental estimates of these quantities are not readily available, so the following heuristic rules are utilized:

1. Amplification is maximal at low copy number, i.e. $f_A$ is close to 1 at low copy number.
2. Amplification is "turned off" at the steady-state average copy number, i.e. $f_A$ is close to 0 at copy numbers from 50 to 100.

![Figure 2: Effect of repressor expression and binding ($f_{R_1}$) on amplification potential ($\alpha_F = 200, f_{F_1} = 0.995$)](image-url)
These admittedly vague assumptions reduce the size of the feasible parameter space surprisingly well.

The behavior of the copy number control for different parameter sets is shown in Figs 2-4. The relationship between $f_A$ and $P_a$ is calculated using eqns (13a-c). It is apparent that the dominant parameter affecting the maximal value of $f_A$ at low copy numbers is $f_{F1}$. Since this seems to be the most significant effect of $f_{F1}$, heuristic rule 1 can be used to estimate $f_{F1}$. The value $f_{F1} = 0.995$ is used in subsequent calculations.

Since $f_{F1}$ most strongly affects the maximum "strength" of the controlling action, it can be considered to be analogous to a gain parameter. The effect of $f_{F1}$ on copy number control is shown in Fig. 5. The magnitude of $f_{F1}$ strongly influences the rate of amplification.
The parameters $\alpha_F$ and $f_{R1}$ seem to affect interactively the copy number at which amplification is "turned off". In this way, these parameters are analogous to a set point for the copy number control system. A parameter region in the $\alpha_F$-$f_{R1}$ plane was calculated wherein $f_a = 0.01$ for copy numbers from 50 to 100. This region is shown in Fig. 6. The transient controlling action of the model for $\alpha_F$-$f_{R1}$ values arbitrarily chosen from this region are strikingly similar, as shown in Fig. 7. Note in particular that a copy number of 50 has been attained after 25 generations for
all of the simulations in Fig. 7; this is in agreement with the observations of Sigurdson et al. (1981).

Plasmid amplification is never completely "turned off" unless $f_A = 0$. Thus, there will always be a slight over-replication of $2\mu$ plasmids, leading to an upward drift in copy number. Two factors would compensate for this drift. First, yeast plasmids tend to preferentially segregate to the mother cell, which grows more slowly after several budding cycles. This segregation bias is large in recombinant plasmids containing the cis-acting STB site in the absence of the B and C products, and is present to a much lesser extent when the B and C products are provided in trans by the endogenous $2\mu$ plasmid (Murray & Szostak, 1983). Second, a slight selective disadvantage is conferred on cells with high copy numbers, because a larger fraction of cellular resources is devoted to plasmid-directed activities (Futcher & Cox, 1983).

**Conclusions**

Based on mechanisms of $2\mu$ plasmid replication and recombination elucidated in the literature, a mathematical model of $2\mu$ copy number control has been formulated. System parameters estimated from heuristic rules yield model behavior which qualitatively agrees with observed copy number amplification. It is not the purpose of this study to obtain quantitative estimates of the model parameters or amplification kinetics. It is instead our intent to demonstrate that the proposed $2\mu$ copy number control system can be expressed in terms of a mathematical model which reproduces the most significant features of $2\mu$ plasmid amplification.

A picture of $2\mu$ plasmid maintenance is emerging which includes two critical components: a partitioning system and an amplification system. The partitioning system acts to overcome a segregation bias toward the mother cell, allowing random distribution of plasmids at cell division. The amplification system rapidly increases the copy number in cells which have received a small number of plasmids through
inefficient partitioning. The amplification system also promotes a trend towards increasing copy number which counteracts the tendency of plasmids to segregate to cells of reduced growth potential. The combination of these two systems allows the 2p plasmid to be very efficiently maintained, even though the plasmid's presence produces a slight selective disadvantage for the host.

In this work, we have shown that a copy number control system utilizing recombinational amplification and a constitutively expressed repressor of recombinase expression can correct downward deviations in copy number without promoting "runaway" amplification. More sophisticated representations of gene expression and protein-DNA binding could doubtless improve the predictive power of the model. However, the bounds placed on amplification by eqn (4) would not be affected by these refinements.

This research was supported by the National Science Foundation and by the Energy Conversion and Utilization Technology (ECUT) program of the U.S. Department of Energy. K. D. Wittrup is the recipient of a National Science Foundation Graduate Fellowship.

REFERENCES

CHAPTER 3.

A Segregated Model of Recombinant Multicopy Plasmid Propagation
A Segregated Model of Recombinant Multicopy Plasmid Propagation

K.D. Wittrup and J.E. Bailey
Department of Chemical Engineering, California Institute of Technology, Pasadena, California 91125
Accepted for publication February 10, 1987

A segregated model of multicopy plasmid propagation has been formulated which incorporates plasmid replication and partition functions, as well as the effect of plasmid presence on host growth rate. Growth of plasmid-free cells in selective medium is explicitly analyzed. The model parameters can be determined from experimentally measurable quantities. Propagation of a recombinant multicopy plasmid in the yeast Saccharomyces cerevisiae is analyzed using this model.

INTRODUCTION

Populations of microorganisms are typically heterogeneous with respect to composition, age, and morphology. The observed properties of a microbial population are the sum of the contributions from individual cells. Any model attempting to utilize mechanistic single-cell kinetics to predict the behavior of a large number of cells must take into account the segregated nature of the population.

Population balance equations provide the mathematical link between individual cellular states and observed population characteristics. Population balance models have been used to describe heterogeneity in cell cycle age, cell length, cell volume, and cellular "maturity." In all the models mentioned, the state variable varies continuously; however, some state variables of biological interest, such as plasmid number, vary discretely. Simplifications of the population balance equation can be accomplished for the case of a discrete state variable. In this paper we present a segregated modeling framework for the description of plasmid number distribution in a cell population. This modeling strategy is applied to the propagation of a recombinant plasmid in a Saccharomyces cerevisiae population.

Model Equations for Plasmid-Containing Cells

In many host-plasmid systems, it is possible for plasmid-free daughter cells to be spawned from plasmid-containing cells. Thus, the total population is composed of two subclasses: plasmid-containing (P⁺) and plasmid-free (P⁻) cells. The models describing these subpopulations differ considerably. The plasmid-containing population model is structured with respect to plasmid number, resulting in a linear system of ordinary differential equations. The plasmid-free population is structured with respect to time since plasmid loss. This population is related to the plasmid-containing population by a boundary condition representing the rate of birth of plasmid-free cells by asymmetric plasmid segregation upon division of P⁺ cells. The parameters of the resulting models can be related to experimentally determined population attributes through model solutions developed below.

A general one dimensional population balance equation is:

\[ 2 \int p(z, z') \mu(z')f(z') dz' = \mu(z)f(z) + \frac{df}{dt} + \frac{\partial f}{\partial z} \]  \hspace{1cm} (1)

where

- \( z \) is the cellular state variable;
- \( f(z) dz \) is the number of cells in state \( z \) to \( z + dz \);
- \( \mu(z) \) is the specific growth rate of cells in state \( z \);
- \( p(t, z') dz \) is the probability that division of a mother cell of state \( z' \) will result in a cell of state \( z \) to \( z + dz \);
- \( \bar{z} \) is the mean time derivative of \( z \).

Plasmid number \( (PN) \) is defined here as the number of plasmids in a cell at birth. Taking \( PN \) as the cellular state variable, \( z = 0 \). Because \( PN \) can only assume integral values, the density function \( f(PN) \) will take the form of a sum of delta functions:

\[ f(PN) = \sum_{n} N_{n} \delta(PN - PN_{n}) \]  \hspace{1cm} (2)

where

- \( PN \) is the plasmid number;
- \( N_{n} \) is the number of cells with \( PN = PN_{n} \);
- \( \delta(x) \) is the Dirac delta function; and
- \( N_{n} \) is the number of plasmid number subclasses.
Combining eqs. (1) and (2),
\[
2 \sum_{j=1}^{N} p(N, PN) \mu_{i} N_{i} = \mu_{i}(PN) \sum_{i=1}^{N} N_{i} \delta(PN - P_{N}) \\
\quad + \sum_{i=1}^{N} \frac{dN_{i}}{dt} \delta(PN - P_{N}) \tag{3}
\]

By integrating eq. (3) with respect to \(PN\), but only over one plasmid number, \(PN_{1}\), the following equation is obtained:
\[
\sum_{i=1}^{N} P_{i} \mu_{i} N_{i} = \mu_{i} N_{i} + \frac{dN_{i}}{dt} \tag{4}
\]

where
- \(\mu_{i}\) is the specific growth rate of cells with \(PN = PN_{i}\) and
- \(P_{i}\) is the mean number of cells with \(PN = PN_{i}\), arising from one division of a cell with \(PN = PN_{i}\).

Repeating this integration for all possible plasmid numbers yields a system of ordinary differential equations:
\[
\frac{dN_{i}}{dt} = \sum_{j=1}^{N} (P_{j} - \delta_{ij}) \mu_{i} N_{j}; \quad i = 1, \ldots, N_{s}
\]
\[
N_{i} = N_{i0}; \quad t = 0 \tag{5}
\]

where
- \(\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \)

Formulating the problem in matrix notation:
\[
\frac{dN}{dt} = T \cdot N
\]
\[
N = N_{0}; \quad t = 0 \tag{6}
\]

where
- \(T_{ij} = (P_{j} - \delta_{ij}) \mu_{i}\)

If the matrix \(T\) has distinct eigenvalues, the solution is:
\[
N = \sum_{j=1}^{N} C_{j} \xi_{j} e^{\lambda_{j}t} \tag{7}
\]

where
- \(\xi_{j}\) is the \(j\)th eigenvector of \(T\)
- \(\lambda_{j}\) is the \(j\)th eigenvalue of \(T\)
- \(C_{j}\) is the \(j\)th element of vector \(C_{j}\) and
- \(C = \xi_{j}^{-1} \cdot N_{0}\)

where
- \(\xi\) is the matrix whose column vectors are \(\xi_{j}\).

Restating the solution in more compact notation:
\[
N_{i} = \sum_{j=1}^{N} E_{ij} e^{\lambda_{j}t} \tag{8}
\]

where
- \(E_{ij} = C_{j} \xi_{j}\)

In the limit as \(t \to \infty\), the term with the largest exponential will dominate the sum. The asymptotic solution is:
\[
\lim_{t \to \infty} \left( \frac{N_{i}}{N_{i0}} \right) = E_{i0} \tag{9}
\]

where
- \(\lambda_{0}\) is the eigenvalue with greatest real part

**Selection Mechanisms and Growth of the Plasmid-Free Population**

Most recombinant vectors contain a selection gene, which confers a growth advantage on plasmid-containing cells in selective media. If a host cell cannot synthesize certain amino acids or nucleotides, genes coding for the missing biosynthetic enzymes can be placed on a recombinant vector, and only cells containing these enzymes will be capable of growing on medium lacking the necessary nutrients.

In all known selection methods, the presence of some gene product, either protein or RNA, is necessary for selective advantage. It is the presence of this "complementing product," not the gene itself, which produces the selection. If the selection gene is lost to one of the daughter cells after cell division, the plasmid-free cell may still contain the complementing product. This product will be degraded or diluted with continued growth, but the plasmid-free cell will at first be capable of growth and division.

In this model, it is presumed that the specific growth rate, \(\eta\), of plasmid-free cells may be written as a function of the intracellular complementing product concentration \(CP\), as in the model proposed by Slienc and co-workers.\(^{9}\) In the following analysis, this concentration and subsequently the \(P^{+}\) cell growth rate are related to the time since birth of the plasmid-free cell by division of a plasmid-containing cell. The population balance equation for the \(P^{+}\) population is:
\[
\frac{df}{dt} + \frac{df}{da} = \eta f, \quad f(a,t) = f(t); \quad a = 0 \tag{10}
\]

where
- \(f(a,t) \, da\) is the number of \(P^{+}\) cells with age between \(a\) and \(a + da\) spawned by cells with \(PN = PN_{0}\);
- \(a\) is the genealogical time since plasmid loss; and
- \(\eta\) is the specific growth rate of \(P^{+}\) cells.

This equation may be solved by the method of characteristics to obtain:
\[
f(a,t) = f(0,t) \exp\left( \int_{0}^{a} \eta \, dr \right) \tag{11}
\]

The boundary condition is given as the rate at which \(P^{+}\) cells are generated from divisions of \(P^{-}\) cells:
\[
f(0,t) = \alpha \mu_{i} N_{i}(t) \tag{12}
\]
where \( \alpha \) is the mean number of \( P^- \) cells created per division of cells with \( PN = PN' \).

In order to determine the specific growth rate \( \eta(a) \), the concentration of complementing product must be determined as a function of \( a \). The dilution and degradation of the complementing product are modeled as follows:

\[
\frac{dCP}{da} = -k_e CP - \eta(CP)CP
\]

\[
CP = CP_0, \quad a = 0
\]

where

\( CP \) is the complementing product concentration in \( P^- \) cells;

\( k_e \) is the rate constant for deactivation of \( CP \); and

\( \eta(CP) \) is the growth rate of \( P^- \) cells expressed as a function of \( CP \).

The solution to eq. (13) is:

\[
a = \frac{\int_{CP_0}^{CP} \frac{dX}{X[k_e + \eta(X)]}}{\int_{CP}^{CP'} \frac{dX}{X[k_e + \eta(X)]}}
\]

This equation defines a functional relationship between \( a \) and \( CP \) which will be denoted:

\[
a = \psi(CP) \quad (14b)
\]

It is convenient to transform variables in the number density form from \( a \) to \( CP \):

\[
f(a, t) da = f[\psi(CP), t] \psi'(CP) dCP
\]

so that the frequency function for cellular \( CP \) content in \( P^- \) cells spawned by \( P^+ \) cells with \( PN = PN' \) is given by

\[
f^{CP}(CP, t) = f[\psi(CP), t] \psi'(CP)
\]

Combining eqs. (11), (12), (14), and (15b) yields:

\[
f^{CP}(CP, t) = \frac{\alpha \mu N(t - \psi(CP))}{CP[k_e + \eta(CP)]} \cdot \exp \left( \int_{CP}^{CP'} \frac{\eta(X)}{X[k_e + \eta(X)]} dX \right)
\]

\[
\lim_{t \to \infty} F_+ = \frac{\sum_{i=1}^{N} E_{i,n} \left( 1 + \alpha \mu \int_{0}^{CP} \exp(-\lambda_{ew} \psi'(CP)) \frac{\eta(X)}{CP[k_e + \eta(CP)]} \exp \left( \int_{CP}^{CP'} \frac{\eta(X)}{X[k_e + \eta(X)]} dCP \right) \right)}{\sum_{i=1}^{N} E_{i,n}}
\]

The total number of \( P^- \) cells is found by integrating over all possible values of \( CP \):

\[
N_-(t) = \int_{0}^{\infty} F^{CP}(CP, t) dCP
\]

where

\( N_- \) is the number of \( P^- \) cells.

**EXPERIMENTAL QUANTITIES**

The mean plasmid content of the \( P^- \) population, the overall population specific growth rate, and the fraction of plasmid-containing cells in the population (all readily accessible from experiment), can be related to the parameters in the model as follows:

\[
\frac{\sum_{i=1}^{N} E_{i,n} PN_i}{N_-} = \frac{\bar{\mu}}{N_+ - N_-}
\]

\[
F_+ = \frac{N_-}{N_+ - N_-}
\]

The asymptotic values for these measurable quantities are found by substituting eqs. (9), (16), and (17) into eqs. (18)–(20):

\[
\lim_{t \to \infty} \frac{\sum_{i=1}^{N} E_{i,n} PN_i}{\sum_{i=1}^{N} E_{i,n}} = \frac{\sum_{i=1}^{N} E_{i,n}}{\sum_{i=1}^{N} E_{i,n}}
\]

\[
\lim_{t \to \infty} \bar{\mu} = \lambda_{ew}
\]

**Example: Plasmid pLGSD5 Propagation in \textit{S. cerevisiae}**

Based on flow cytometric \( \beta \)-galactosidase assays of \textit{S. cerevisiae} D603 transformed with pLGSD5, a simplified description of pLGSD5 propagation was proposed. Figure 1 shows the proposed description. The term replication success will be used to denote the case where every plasmid replicates once, and replication failure is defined as the failure of all plasmids to replicate. Segregation success denotes a symmetric division of plasmids between daughter cells, and segregation failure refers to the birth of a plasmid-free cell. For this situation, the plasmid numbers
Plasmid Propagation Model

<table>
<thead>
<tr>
<th>$j$</th>
<th>Replication</th>
<th>Segregation</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_i$</td>
<td>+</td>
<td>+</td>
<td>$pq$</td>
</tr>
<tr>
<td>$P_i$</td>
<td>+</td>
<td>-</td>
<td>$(1-q)$</td>
</tr>
<tr>
<td>$N_i$</td>
<td>-</td>
<td>-</td>
<td>$(1-p)$</td>
</tr>
</tbody>
</table>

Plasmid number = $2^i$

Figure 1. Proposed model of pLGDS5 propagation in S. cerevisiae.

$PN_j = 2^{-j}; i = 1, 2, \ldots, N_m$. For this type of plasmid propagation, the transition probabilities may be written as follows:

$$T_i = \begin{cases} p(2q-1)\mu_j; & j = i \\ p(1-q)\mu_{j-1}; & j = i - 1 \\ 0; & \text{otherwise} \end{cases}$$  \hspace{1cm} (24)

where

$p$ is the probability of replication success and $q$ is the probability of segregation success.

Combining eqs. (24) and (6) and solving yields:

$$N_j = \sum_{i=0}^{j} N_i \left( \frac{1-q}{2q-1} \right)^{j-i} \prod_{i=j}^{j} \left( \frac{\mu_j}{\mu_{i-1}} - 1 \right) \exp[p(2q-1)\mu_j \cdot]$$

which has the corresponding normalized asymptotic form:

$$N_j = 1$$

where

$$\mu_{\max}$$ is the maximum value of $\mu$ with respect to plasmid number $j$.

It should be noted that the functionality of the specific growth rate function $\mu_j$ is critical to the nature of the solution. In the degenerate case where all $\mu_j$ are equal, it can be shown that the asymptotic solution $N_j$ grows without bound as $j \to \infty$. It is the growth inhibition which occurs at high plasmid numbers which bounds the distribution.

The resulting plasmid distribution is a consequence of two opposing trends: segregation failure leading to plasmid number increase, and growth inhibition at high plasmid number which reduces the relative number of high plasmid number cells.

Next, the plasmid-free population is considered. Assuming that the complementing product concentration in plasmid-containing cells is proportional to the gene dosage: $CP = E \cdot PN$, the initial complementing product concentration for $P^+$ cells formed from $P^-$ cells of class $i$ is $E \cdot 2^{-i}; i = 1, 2, \ldots, N_m$. In the model presented here, degradation of the complementing product (the URA3 gene product) is neglected. This assumption provided good qualitative agreement between theory and experiment for ARS1/CEN4 plasmids propagated at one plasmid per newborn cell. These plasmids used the same host and same plasmid selection gene (URA3) as the multicopy D603::pLGDS5 system. The specific growth rate of plasmid-free cells is presumed to be a hyperbolic function of $CP$:

$$\eta = \frac{\mu_{\max} CP}{K_{cp} + CP}$$  \hspace{1cm} (27)

This functional form, which should reasonably approximate the dependence of growth rate on $CP$ for other types of host-vector systems, gave overall model results qualitatively consistent with experimental data for D603::ARS1/CEN4 strains of recombinant $S. cerevisiae$.

The density function $F'(y; t) dy$, which is the number of $P^-$ cells at time $t$ with dimensionless complementing product concentration between $y$ and $y + dy$, is given by

$$F'(y; t) = \sum_{i=0}^{j} (1-pq)\mu_j N_i \phi_i(y)g_i(y)S_i(y; t, i)$$  \hspace{1cm} (28)

where

$$\phi_i(y) = \frac{1}{\mu_{\max}} \left( 1 - \ln(y) + \ln(y_0) - \frac{1}{y_0} \right)$$

$$\phi_i' = \frac{1 + y}{\mu_{\max}}$$

$$S_i(y; t, i) = H(y - \phi_i^{-i}(t)) - H(y - y_0)$$

$H(x) =$ Heaviside step function

$$g_i(y) = y_0/y$$

$$y_0 = 2^i/K_{cp}$$

The asymptotic form of this density is:

$$\lim_{i \to \infty} F'(y; t) = \frac{1-pq}{\mu_{\max} K_{cp}} \left( \frac{1+y}{y} \right)^{\frac{y}{\gamma}} \cdot e^{\gamma y} \sum_{i=0}^{j} \mu_j N_i(t) e^{(1-y) \gamma} \exp \left( \frac{y}{y_0} \right) \cdot \left[ 1 - H(y - y_0) \right]$$  \hspace{1cm} (29)

where

$$\gamma = \frac{p(2q-1)\mu_{\max}}{\mu_{\max}}$$

Substituting eq. (29) into eq. (17), the asymptotic solution for the number of plasmid free cells is:

$$N_i \to \left( \frac{1-pq}{\gamma \mu_{\max}} \right) \sum_{i=0}^{j} \mu_j N_i(t) \left[ 1 + \xi_i^{-i} e^{(1-y) \gamma} \cdot \left[ 1 - H(y - y_0) \right] \right] \cdot \left[ 1 - P(2\xi_i [2 - 2y]) \right]$$  \hspace{1cm} (30)
where

$$\xi_i = \frac{\gamma}{\nu_0}$$

$P(\chi^2|\nu) = \chi^2$ cumulative distribution function with $\nu$ degrees of freedom and $\Gamma(x) =$ gamma function.

**COMPUTATIONAL RESULTS**

The growth model used for the D603:pLGSD5 system was developed by Srie and co-workers. The presence of plasmids provides an "activation" through the production of complementing product. The plasmid also causes an "inhibition" of growth rate due to plasmid maintenance and other plasmid-directed activities. These counterbalancing trends are modeled as follows:

$$\mu = A(PN)I(PN)$$

$$A(PN) = \frac{\mu_\mu PN}{K_{PN} + PN}$$

$$I(PN) = \frac{K_i}{K_i + PN}$$  \hspace{1cm} (31)

where

- $\mu$ is the specific growth rate;
- $PN$ is the plasmid number;
- $\mu_\mu, K_{PN},$ and $K_i$ are model parameters.

The following values for these parameters have been fit to experimental data for the *S. cerevisiae* strain D603 transformed with a plasmid containing the *URA3* gene as a selection marker:

- $\mu_\mu = 0.668 \text{ h}^{-1}$;
- $K_{PN} = 0.828$ plasmids/cell; and
- $K_i = 13.88$ plasmids/cell.

The possibility of replication failure was included in the proposed model because it was consistent with the observed plasmid number distributions obtained with flow cytometry. However, since *S. cerevisiae* plasmids replicate using the chromosomal replication machinery, replication failure as defined is a very unlikely event. Therefore, the parameter $p$ is fixed at 1.0 in all subsequent calculations.

The asymptotic values for the mean plasmid number, growth rate, and plasmid-containing fraction were determined as a function of the segregation probability, $q$. The results are shown in Figures 2–4. Calculated asymptotic plasmid number distributions are shown in Figures 5 and 6. The experimentally determined mean plasmid number for D603:pLGSD5 is 10; the mean growth rate is $0.36 \text{ h}^{-1}$; and the plasmid-containing fraction is 0.9. For $q = 0.9$, the model yields $PN = 9.0$; $F_c = 0.54$; and $\mu = 0.32 \text{ h}^{-1}$.

The large discrepancy between model and experiment for $F_c$ could be due to inability of the growth model to correctly describe cell behavior at high plasmid number. It

---

**Figure 2.** Simulated mean copy number for *S. cerevisiae* D603:pLGSD5 as a function of the probability, $q$, of segregation success.

**Figure 3.** Effect of segregation success probability, $q$, on the calculated specific growth rate for *S. cerevisiae* D603:pLGSD5.

**Figure 4.** Plasmid-containing fraction for *S. cerevisiae* D603:pLGSD5 evaluated for different values of segregation success, $q$. 

---

308  BIOTECHNOLOGY AND BIOENGINEERING, VOL. 31, MARCH 1988
may also be an indication that the proposed description of plasmid propagation is oversimplified. If it were common for segregation "near failures" to occur, with one daughter cell containing many plasmids and the other daughter cell containing only one, the plasmid number distribution would be closely approximated by the present model, but the plasmid-containing fraction would differ drastically.

**DISCUSSION**

A general segregated model for plasmid propagation in recombinant populations has been developed. Although the original motivation for this model was the representation of yeast populations, its assumptions and results should also apply to other multicopy host–vector populations. A distributed model such as this can be used in conjunction with flow cytometry experiments to elucidate single cell plasmid replication and segregation behavior. This model can also be combined with reactor performance equations to obtain a more detailed representation of a bioreactor.

An analytical solution to eq. (6) is in principle always obtainable, but for some systems the eigenvalue problem may be untenable, and a numerical solution must be pursued. This model structure facilitates comparison of different plasmid propagation mechanisms, since all replication and segregation rates are embodied in the $P_r$ parameters. In a general programming space constructed to solve the eigenvalue problem numerically, different propagation mechanisms can be represented merely by changing the subroutine which assigns values to the $P_r$ parameters.

A major advantage of this modeling strategy is the inclusion of plasmid effects on host growth rate. The resulting model structure allows a clean separation of the effects of plasmid replication and partitioning from the host–plasmid metabolic interactions. The usefulness of the plasmid-free population model depends heavily on the accuracy of the growth model, but the parameters of the growth model may not be easily determined experimentally. Methods for extracting this information from experimental flow cytometric data on plasmid-content distributions are being pursued.

This research was supported by the National Science Foundation and by the Energy Conversion and Utilization Technology (ECUT) Program of the U.S. Department of Energy. K. D. Wittrup is the recipient of a National Science Foundation Graduate Fellowship.

** NOMENCLATURE**

- $a$: genealogical time since plasmid loss
- $CP$: complementing product concentration
- $exp$: constant defined in eq. (8)
- $f(a,i)$: genealogical age frequency function (eq. (10))
- $f^{CP}(i,1)$: complementing product frequency function (eq. (15))
- $F^{CP}(i,1)$: complementing product frequency function (eq. (16))
- $F_0$: plasmid-containing fraction
- $H(x)$: Heaviside step function
- $k_t$: rate constant for deactivation of $CP$
- $K$: growth model parameter (eq. (27))
- $K_P$: growth model parameter (eq. (31))
- $K_m$: growth model parameter (eq. (31))
- $N^+$: number of $P^+$ cells
- $N^-$: number of $P^-$ cells
- $N^0$: number of cells with $PN = PN^0$
- $N^+$: number of $PN$ subclones
- $p$: probability of replication success
- $P^+$: plasmid-containing
- $P^-$: plasmid-free
- $P_n$: transition probability (eq. (4))
- $PN$: plasmid number
- $PN^0$: plasmid number in subclass $i$
- $PN^+$: mean plasmid number
- $q$: probability of segregation success
- $T$: transition probability matrix (eq. (6))
- $y$: dimensionless $CP$ (eq. (28))

**Greek characters**

- $\alpha$: plasmid loss rate (eq. (12))
- $\beta$: Dirac delta function
- $\eta$: specific growth rate of $P^+$ cells
- $\lambda$: $j$th eigenvalue of $T$
- $\lambda_m$: eigenvalue of $T$ with greatest real part
- $\mu$: specific growth rate of cells with $PN = PN^0$
- $\mu$: mean specific growth rate
\( \mu_a \) growth model parameter [eq. (31)]

\( \xi^i \) \( i \)th eigenvector of \( T \)

\( \theta \) function defined in eq. (14)

References

11. F. Sannic, unpublished.
CHAPTER 4.

A Single-Cell Assay of β-Galactosidase

Activity in *Saccharomyces cerevisiae*
A Single-Cell Assay of β-Galactosidase Activity in Saccharomyces cerevisiae

Karl Dane Wittup and James E. Bailey
Department of Chemical Engineering, California Institute of Technology, Pasadena, California 91125
Received for publication October 20, 1987; accepted February 23, 1988

A novel assay of single-cell exogenous β-galactosidase activity in Saccharomyces cerevisiae has been developed. Intracellular fluorescence due to the hydrolysis of resorufin-β-D-galactopyranoside attains a steady state between production of resorufin and its subsequent leakage from the cell. The cells are permeabilized with Triton X-100, and the assay is performed at 0°C. These conditions were chosen to minimize intercellular fluorescence communication. Free resorufin in the extracellular space is bound by bovine serum albumin to prevent its uptake by cells.

Two regimes of fluorescence accumulation are observed, one limited by the rate of diffusion of substrate into the cell, and one limited by the rate of enzymatic cleavage of the substrate. A quantitative correlation between fluorescence and β-galactosidase activity is obtained under optimized assay conditions.

Key terms: Fluorogenic substrate, resorufin, β-galactosidase, fluorescence accumulation

Fluorogenic substrates are compounds that change their fluorescence characteristics when acted upon by an enzyme—typically, a fluorescent moiety is liberated by the enzymatic cleavage of a non-fluorescent substrate. Fluorogenic substrates have been developed for a variety of enzymes (5,7,8). These substrates can be used to assay enzyme activities at the single-cell level with flow cytometry. One advantage of studying intact cells is the opportunity to observe activity in the enzyme's native environment; assays of cellular extracts display reduced activity.

A major difficulty with many fluorogenic substrates is the leakage of fluorescent product from one cell and subsequent product diffusion into other cells. This intercellular communication can prevent the accurate determination of population heterogeneity. One way to circumvent this problem is through the use of a trapping agent to localize the product as an intracellular precipitate. The most commonly used trapping reagents are diazo salts, which couple with naphthol products to form precipitates. Trapping agents may inhibit enzyme activity (12,13), however, and appropriate coupling compounds are not available for some of the most commonly used fluorophores. Also, some trapped precipitates are either nonfluorescent, or the emission is shifted into the infrared (7,8).

The dependence of fluorescence accumulation on substrate concentration (20-23) and enzyme activity (14) has often been found to differ significantly from Michaelis-Menten kinetics. This discrepancy has been attributed to the presence of multiple non-specific enzymes (20,21), diffusion of fluorescent product away from the cell (6,22), product inhibition of enzyme activity (6), and limitations on the diffusion of substrate into the cell (3-6,20,23). Several investigators have studied the effect of membrane-active detergents and antibiotics on the accumulation of fluorescein in cells incubated with fluorescein diazoate, which is hydrolyzed by non-specific esterases (15-18). These studies demonstrate that membrane transport processes can have a dominant effect on fluorescence accumulation.

We have developed a flow cytometric assay for β-galactosidase activity in Saccharomyces cerevisiae using the fluorogenic substrate resorufin-β-D-galactopyranoside (RG). After permeabilization with Triton X-100, a steady state is rapidly established between production of fluorescent resorufin and its leakage out of the cell. Free resorufin in the extracellular fluid is bound by bovine serum albumin to minimize its diffusion into other cells. The correlation between steady-state fluorescence and β-galactosidase activity is biphasic, displaying a region of reaction limitation and a region of transport limita-

1This research was supported by the National Science Foundation and by the Energy Conversion and Utilization Technology (ECUT) program of the U.S. Department of Energy. K.D. Wittup is the recipient of a National Science Foundation Graduate Fellowship.

2Address reprint requests to Dr. J.E. Bailey, Department of Chemical Engineering, California Institute of Technology, Pasadena, CA 91125
tion. This technique allows a quantitative determination of enzyme activity in single yeast cells.

MATERIALS AND METHODS

Strains and Cultivation

Saccharomyces cerevisiae diploid strain D603 (2) contains the reg mutation, which inhibits glucose catabolite repression. The reg mutation allows the induction of galactose-inducible transcription in the presence of glucose. Cells were grown at 30°C in 50 mL of SD minimal medium buffered with sodium citrate (0.05 M, pH 4.5) and supplemented with adenine sulfate (20 mg/L), L-histidine (20 mg/L), L-methionine (20 mg/L), and L-lysine (30 mg/L). Plasmid pLGSD5 (9) contains the URA3 gene for selection of plasmid-containing cells, and a portion of the Escherichia coli lacZ gene for β-galactosidase under the control of the galactose-inducible GAL10-CYC1 hybrid promoter. Expression of an active β-galactosidase fragment is induced by adding galactose to a final concentration of 2% w/v. S. cerevisiae does not possess endogenous β-galactosidase activity, so all β-galactosidase activity is expressed from the inducible lacZ gene. D603 has one copy of the lacZ gene chromosomally integrated, and D603(2) has two integrated copies.

Spectrophotometric β-Galactosidase Assay

The spectrophotometric assay used is a variation of the protocol developed by Celniker et al. (2). Cells are centrifuged out of the growth medium, washed twice with phosphate buffered saline (PBS: 0.01 M phosphate, pH 7.6, 0.85% w/v NaCl, 20 mM MgCl₂), permeabilized in 50% v/v isopropanol/PBS for 10 min at 0°C, then washed twice with PBS. The assay mixture contains 2.7 mL of phosphate buffer (0.1 M, pH 7.6), 0.01 M KC₁, 0.001 M MgSO₄, 0.05 M β-mercaptoethanol pre-warmed to 37°C, 0.3 mL of 0-nitrophenyl-β-D-galactopyranoside (10 mg/mL in H₂O), Sigma, St. Louis, MO), and 0.3 mL of permeabilized cell suspension. Absorbance at 420 nm is measured using the kinetics program of the Shimadzu UV-260 (Shimadzu, Kyoto, Japan) Spectrophotometer. The cuvet is allowed to equilibrate thermally with the 37°C thermostat controlled block for 3 min, then the rate of increase of the absorbance at 420 nm is monitored for 3 min. Units of β-galactosidase activity are standardized by comparison to frozen aliquots of a solution of E. coli β-galactosidase (Sigma). One unit of β-galactosidase will hydrolyze 1.0 μmole of 0-nitrophenyl-β-D-galactoside to 0-nitrophenol per minute at pH 7.3 and 37°C. The number density of cells in the permeabilized cell suspension is determined using a Coulter Counter 2M with a Channelizer 256 (Coulter Electronics, Hialeah, FL).

Flow Cytometry and Spectrofluorometry

A Cytofluorograf 50H flow cytometer (Ortho Instruments, Westwood, MA) was used for single-cell fluorescence measurements, using a krypton ion laser (INNOVA 90K, Coherent, Palo Alto, CA) producing 50 mW of light-stabilized power at 568 nm. Fluorescence signals were collected after passing through a 590 nm high-pass colored glass filter (65.1345, Rolyent Optics, Covina, CA). The residence time in the feed tubing was minimized to prevent warming of the sample. The feed tubing was wrapped with cotton and aluminum foil. A Shimadzu RF540 spectrofluorometer was used for the measurement of fluorescence in sonicated cell homogenates.

Single-Cell β-Galactosidase Assay

The base assay mixture contained the following: 40 μL Triton X-100 (2% v/v in H₂O); 60 μL resorufin-β-D-galactopyranoside (Molecular Probes, Eugene, OR) (2 mg/mL in DMSO); 0.55 mL PBS at 0°C; 0.4 mL bovine serum albumin (Sigma) (2% w/v in PBS at 0°C); and 50 μL of cells in PBS (washed twice with PBS after spinning out of medium). Final cell density is approximately 10⁴ cells/mL. The components are added in order to 1.5 mL Eppendorf centrifuge tube, shaken well, and placed in an ethanol-ice bath in the sample chamber of the flow cytometer. Flow is initiated immediately, and fluorescence distributions are collected after 5 min. At the beginning of the experiment, a sample is run with double RG and Triton X-100 concentrations for 15 min, in an attempt to saturate any substrate or product binding in the feed tubing.

RESULTS

Assay Development

RG appears orange in solution, while free resorufin is bright red. The excitation and emission spectra of resorufin, the product of enzymatic hydrolysis of resorufin-β-D-galactopyranoside, are shown in Figure 1. RG has an absorbance maximum at 470 nm, and excitation at this wavelength produces fluorescence emission that is about 0.1% the intensity of an equal concentration of resorufin. The fluorescence of resorufin is significantly quenched at concentrations above 20 μM, as shown in Figure 2. When Triton X-100-permeabilized β-galactosidase-containing cells were added to an RG solution at room temperature, the solution turned red within minutes, and fluorescence from individual cells was masked by the surrounding resorufin when the sample was run on the flow cytometer. In order to reduce the reaction rate, the reaction mixture was kept on ice. Under these conditions, the solution remains orange, but cellular fluorescence signals are detectable on the flow cytometer. The final fluorescence represents a steady state between resorufin production and its leakage from the cell. Cells spun out of substrate solution and immediately resuspended in PBS lose all measurable fluorescence before the flow of sample can be started. The temperature of the sample is kept as low as possible by taking the following steps: the sample Eppendorf tube is immersed in an ice-ethanol bath; the length of sample feed tubing is minimized; the feed tubing is insulated by wrapping with cotton and aluminum foil; and the volume of the feed tube inlet is minimized.
Fig. 1. Excitation (---) and emission (---) spectra of resorufin. Resorufin-D-galactopyranoside (RG) is essentially nonfluorescent.

Fig. 2. Fluorescence-concentration relationship for resorufin. Different symbols represent independent experiments.
Fig. 3. Dependence of mean single-cell fluorescence on Triton X-100 concentration. Triton X-100 permeabilizes the cells, allowing the fluorogenic substrate RG to enter. The results of three independent experiments are shown.

Fig. 4. Binding of resorufin by bovine serum albumin (BSA). The fluorescence intensity of bound resorufin is less than that of free resorufin. The solid curve represents a least squares fit to the model in equation 1, with $K = 0.069 \mu M^{-1}$. 
It is somewhat surprising that the reaction proceeds at a measurable rate at temperatures near 0°C. This may be attributed to the high efficiency of RG hydrolysis by β-galactosidase. The Michaelis-Menten kinetic constant $k_{on}$ for RG hydrolysis by β-galactosidase is 730 s$^{-1}$ at 22°C (11), as opposed to a $k_{cat}$ of 10 s$^{-1}$ for fluorescein di-β-D-galactopyranoside at 22°C (10), and a $k_{cat}$ of 340 s$^{-1}$ for hydrolysis of fluorescein diacetate by porcine liver esterase (10).

When the flow of sample through the inlet tubing is initiated, the distribution of single-cell fluorescence displays an interesting transient behavior for 1–5 min. The fluorescence histogram exhibits three peaks; the middle peak is the only one that persists past the initial transient, and it is stable in intensity for at least 20 min. The stable middle peak occurs at a reproducible fluorescence intensity, and this steady-state fluorescence histogram is the distribution examined in all of the subsequent analysis. The measurement of a steady-state fluorescence distribution is a fundamentally different approach from the measurement of initial reaction rates. This approach circumvents the difficulty of obtaining a linear initial reaction rate from kinetic data with significant curvature and also provides a stable fluorescence distribution that can be measured without significant changes during data collection. The unusual initial transients are being explored further, and appear to be due to thermal transients in the feed tubing. Regardless of the cause of this unusual dynamic behavior, the final steady state is stable, reproducible, and correlates with enzyme activity in a quantitative fashion.

**RG Hydrolysis Kinetics and Permeabilization**

Using sonicated yeast cell suspensions under the conditions used for the flow cytometry assay, the Michaelis-Menten binding constant $K_m$ for the hydrolysis of RG was found to be 93 μM, and isopropyl β-D-thiogalactopyranoside (IPTG) inhibited the reaction as a competitive substrate with $K_I = 24$ μM.

Several techniques of permeabilization were evaluated, using ethanol, isopropanol, and Triton X-100. Resorufin leaked rapidly out of cells fixed with ethanol and isopropanol. The effect of Triton X-100 concentration on mean single-cell fluorescence at steady state is shown in Figure 3. At Triton X-100 concentrations below 0.01%, very little RG enters the cell, while at concentrations above 0.1%, leakage of resorufin from the cell becomes significant. Subsequent assays were performed at a Triton X-100 concentration of 0.073%.

**Binding of Resorufin by Bovine Serum Albumin**

While studying the de-ethylolation of ethoxyresorufin in isolated rat hepatocytes, Burke and Orrenius (11) discovered that bovine serum albumin (BSA) binds extracellular resorufin, preventing it from re-entering the cell and undergoing sulphate conjugation. For this reason, the ability of BSA to scavenge free resorufin and pre-
Fig. 6 Fluorescence distribution of a mixture of β-gal-containing (D6032) and β-gal free (D603) cells. In the sample represented by the top histogram, no BSA was added, and β-gal-free cells became fluorescent by taking up free resorufin in the buffer. In the bottom histogram, the nonfluorescent population is more clearly separated from the fluorescent population by the addition of BSA. The log scale is approximately 36 channels per decade.
vent "crosstalk" between fluorescent and nonfluorescent yeast cells was studied.

Various concentrations of BSA were added to a 12 μM solution of resorufin under the conditions of the flow cytometry assay. The reduction of fluorescence upon BSA addition was measured in a spectrofluorometer, and is shown in Figure 4. The data were described by the following model:

\[
BSA + R = BSA-R
\]

\[
K = \frac{[BSA-R]}{[BSA][R]}
\]

Fluorescence Intensity = α[R] + β[BSA-R],

where [BSA] = bovine serum albumin concentration, [R] = resorufin concentration, [BSA-R] = concentration of resorufin bound by BSA, \(α\) = molar fluorescence intensity of free resorufin, and \(β\) = molar fluorescence intensity of resorufin bound by BSA.

By a nonlinear least squares curve fit, the model parameters were determined to be \(K = 0.069 \text{mM}^{-1}\), \(α/β = 0.156\).

The relevance of this type of equilibrium binding in solution to the actual reduction of background fluorescence in the flow cytometer was investigated. A mixture of β-galactosidase-containing and β-galactosidase-free cells was stained with RG at various concentrations of BSA. The modal fluorescence intensity of the enzyme-free cells as a function of BSA concentration is shown in Figure 5. The solid curve represents the binding model of Equation 1 with \(K = 0.069 \text{mM}^{-1}\). BSA successfully inhibits resorufin that has leaked out of β-galactosidase-containing cells from diffusing into enzyme-free cells. The scavenging capability of BSA allows enzyme-free cells to be clearly distinguished from enzyme-containing cells, as shown in Figures 6 and 7. A BSA concentration of 110 μM was used in all subsequent analysis.

**Fluorescence - Substrate Concentration Relationship**

The dependence of steady-state single-cell fluorescence on RG concentration was studied with two cell types: D603-pLGSD5, which contains the lacZ gene on a multicopy plasmid present at ten copies per cell, and D6032i, which has two copies of the lacZ gene chromosomally integrated. The β-galactosidase activity per cell in D603-pLGSD5 after 4 hr of induction is approximately 5 times the level in D6032i.

As shown in Figure 8, single-cell fluorescence in D603-pLGSD5 is proportional to RG concentration up to a concentration of about 1,000 μM. At higher RG concentrations, fluorescence remains essentially constant. The fluorescence-substrate concentration relationship for D6032i, shown in Figure 9, is similar, except that maximum fluorescence is reached at an RG concentration of about 300 μM.

This substrate concentration dependence is apparently not of the Michaelis-Menten type, because the fluorescence reaches a saturation value at an RG concentration that depends on the enzyme activity. This
behavior can be explained by the existence of two regimes of fluorescence accumulation: at low RG concentrations, fluorescence is limited by the rate of substrate diffusion into the cell, and at high RG concentrations, fluorescence is limited by the rate of enzymatic cleavage of the substrate. This idea is expanded in the Discussion.

Fluorescence-Activity Correlation

Because the lacZ gene cloned into S. cerevisiae strains used here is inducible, a continuous range of single-cell β-galactosidase activity is attainable, merely by varying the length of the induction time. The median single-cell fluorescence of cells stained with RG was measured for cells with various β-galactosidase activities and compared to the mean specific activity measured spectrophotometrically. The results are shown in Figure 10. Three independent experiments were performed, and the fluorescence measured by flow cytometry was calibrated by comparison to the scatter signal from nonfluorescent beads.

For β-galactosidase activities less than 40 × 10⁹ units/cell, single-cell fluorescence is approximately proportional to enzyme activity. At higher activities, fluorescence remains essentially constant, because the diffusion of substrate into the cell is limiting the rate of resorufin production.

For cells with such high activity that substrate transport is limiting, the effective β-galactosidase activity can be reduced by the addition of a known amount of the competitive inhibitor IPTG. This allows the measurement of a fluorescence signal proportional to the effective enzyme activity, from which the true activity can be determined. The effect of IPTG addition on single-cell fluorescence is shown in Figure 11.

DISCUSSION

Two critical problems with single-cell fluorogenic substrate assays have been addressed in the development of this assay: prevention of inter-cellular fluorescence crosstalk and determination of a quantitative correlation between fluorescence and enzyme activity.

Several factors are important in minimizing resorufin leakage and uptake by other cells. By running the assay at 0°C, the reaction rate is slowed considerably, producing less resorufin to leak out of the cell; membrane permeability is also decreased. Use of an optimal concentration of Triton X-100 allows RG to diffuse into the cell, without permitting excessive resorufin leakage. Most importantly, BSA in the assay buffer binds resorufin, preventing it from diffusing back into the cells. To a much lesser extent, BSA also binds RG (1), so BSA is not added in great excess.

Substrate concentration dependences similar to that shown in Figure 8 have been reported previously, and various explanations have been proposed (3-6,20-23). The saturation of fluorescence intensity cannot be attributed to fluorescence quenching due to high resorufin content, because the saturation occurs at a lower substrate concentration and fluorescence intensity in cells with lower enzyme activity (Fig. 9). The saturation also cannot be explained by Michaelis-Menten-type enzyme-binding saturation: the apparent $K_m$ varies as a function of specific activity, and saturation occurs at a concentration greater than 93 μM, the value of $K_m$ measured in vitro.

The explanation for the presence of two regimes of fluorescence accumulation is facilitated by the availability of a continuous range of specific β-galactosidase ac-
Fig. 9. Dependence of single-cell fluorescence on substrate concentration for D6032. Fluorescence intensities were normalized by the maximum value in this figure. The absolute value of the fluorescence intensity is less than that in Figure 7.

Fig. 10. Correlation of median single-cell fluorescence with mean specific activity. Three symbols represent three independent experiments. Single-cell fluorescence was calibrated against the right-angle scatter signal of nonfluorescent beads for comparison between experiments, then normalized to arbitrary units. Mean specific activity was measured spectrophotometrically.
activity. The fluorescence-activity correlation shown in Figure 10 provides the key to understanding the fluorescence saturation in Figures 8 and 9. Fluorescence is proportional to enzyme activity at low activity, but as the rate of reaction increases, it approaches the maximum rate at which substrate can diffuse into the cell. Subsequent increases in enzyme activity fail to increase fluorescence accumulation, which is limited by the rate of RG diffusion into the cell. Addition of a competitive substrate reduces the effective enzyme activity, shifting the fluorescence accumulation back to a reaction-limited regime. A mathematical model describing the two regimes of fluorescence accumulation has been formulated and describes the observed phenomena well (manuscript in preparation).

The rationale for developing a quantitative single-cell enzyme assay in _S. cerevisiae_ has been to provide a measurement of plasmid copy number distribution. Because gene expression is proportional to gene dosage at low copy number (2), a measurement of the distribution of enzyme activity can be deconvoluted to obtain the distribution of single-cell gene dosage in the population. Knowledge concerning the distribution of plasmid content can provide detailed insight into plasmid replication and segregation.

Another technique for measuring single-cell β-galactosidase activity in _S. cerevisiae_ has previously been presented (19); using the fluorogenic substrate α-naphthyl-β-D-galactopyranoside with the trapping agent hexazonium pararosaniline. Spectrofluorometry of whole cells and sonicated cell suspensions stained by this protocol has revealed that the precipitate formed in this reaction is not fluorescent (data not shown). The differentiation between enzyme-containing and enzyme-free cells that was observed can be attributed to enhanced orthogonal light scatter from intracellular precipitates.

An alternative substrate for this type of assay is Naphthol AS-BI-β-D-galactopyranoside (NASBIG). When Naphthol AS-BI, the product of NASBIG hydrolysis, is trapped with various diazo salts, the resulting precipitate fluoresces in the infrared region (λ<sub>em</sub> = 890 nm). This possibility was not explored further because of problems with enzyme inhibition by diazo salts.

Because resorufin is excited by the yellow-green line of a krypton ion laser, an argon ion laser can be used at 488 nm to excite a second fluorophore. If the second fluorophore's emission spectrum overlaps the excitation spectrum of resorufin, resorufin emission may be enhanced by nonradiative energy transfer or secondary excitation. For this reason, the laser beams should be spatially separated to ensure that resorufin excitation is independent of the concentration of the second fluorophore. Resorufin may contribute to the quenching of fluorescence emission from the second fluorophore when the cell passes through the 488 nm beam, however. Because resorufin fluorescence accumulation is sensitive to the cell's membrane permeability, only procedures that do not require fixation can be carried out in conjunction with this β-galactosidase assay. Possible compatible labeling protocols include use of fluorescein-conjugated antibodies or lectins for assay of cell surface markers.

The general strategy used in the development of this assay should prove valuable in application to other fluo-
Progenetic substrates and organisms. The comparatively small size and tough cell wall of S. cerevisiae would seem to make it a more difficult system for the application of flow cytometry than mammalian cells. Therefore, the methods presented here should be easily applicable to mammalian cells. β-galactosidase gene fusions are often employed in the study of gene expression, so this assay could be used to study gene regulation at a single-cell level. Substrates for many intracellular enzymes are now available; the present approach to protocol development may allow methylumbelliferone to be more widely used as a fluorophore.

LITERATURE CITED

CHAPTER 5.

Mathematical Modeling of a Single-Cell Enzyme Assay
ABSTRACT

A quantitative assay of β-galactosidase activity in single cells of *Saccharomyces cerevisiae* has been developed using a fluorogenic substrate and flow cytomtery. The relationship between single-cell fluorescence and enzyme activity is described by a mathematical model which accounts for substrate and product diffusion. It has been demonstrated that diffusion limitation rather than enzyme activity can determine the level of single-cell fluorescence under certain assay conditions. The mathematical model was used to modify the assay to allow measurement of enzyme kinetics rather than diffusion rates.
Introduction

The availability of a broad range of fluorescent probes of cellular structure and function, in conjunction with flow cytometry, allows the measurement of macromolecular composition, ionic content, and surface marker levels at the single-cell level (for review, see 1,2). The measurement of enzyme activity in single cells by flow cytometry, termed flow cytoenzymology, is generally accomplished by the use of a fluorogenic substrate. A fluorogenic substrate is a nonfluorescent compound which can undergo a reaction catalyzed by the enzyme of interest to form a fluorescent product. Fluorogenic substrates for a variety of enzymes have been developed(3,4).

Measurement of single cell enzyme activity has been used to study native enzymes (3-8) and to measure expression levels of "reporter enzyme" gene fusions (9-11). By fusing the gene encoding a well-characterized "reporter enzyme" to the end of a gene of interest, the control of gene expression can be determined by assaying for the reporter enzyme activity, rather than attempting to measure the level of the natural gene product.

Two major problems have hampered the widespread application of flow cytoenzymology: fluorescent product leakage from cells and subsequent diffusion into other cells; and the lack of a quantitative model of the relationships among single-cell enzyme activity, bulk substrate concentration, and single-cell fluorescence(3,6,12-18). An assay for β-galactosidase activity in the yeast Saccharomyces cerevisiae has been devised which circumvents these difficulties(19). In this communication, we describe a mathematical model of the assay kinetics and show how the model was used to guide the development of the assay.

Fluorogenic Substrate Enzyme Assays

A schematic diagram of the processes involved in a single-cell enzyme assay
using a fluorogenic substrate are shown in Figure 1. The enzyme activity is determined by measuring the concentration of the fluorescent product accumulated inside the cell. The substrates used are generally nonpolar and diffuse readily across the cell membrane, while the products are polar and diffuse more slowly out of the cell.

Activity is typically determined in one of two ways: measurement of total product accumulated up to a time point when the reaction is stopped; or measurement of the rate of change of the product concentration with time. The approach taken with the assay described here is to allow the product concentration to reach a steady state between product formation by reaction and loss by diffusion. The use of a steady state fluorescence assay eliminates the need to precisely control the duration of the procedure, and obviates the need to approximate a linear slope from initial rate data which may possess significant curvature or experimental noise. Because the cellular product level is constant, the distribution of single-cell fluorescence does not significantly change during the period required to collect data for 100,000 cells (5 minutes).

Assay Conditions

The conditions of the assay are described in detail elsewhere (19), but will be discussed briefly here. Resorufin-β-D-galactopyranoside (RG, Molecular Probes, Portland, Oregon) is the fluorogenic substrate, which is cleaved by β-galactosidase to form resorufin. Fluorescein-di-β-D-galactopyranoside (FDG) was considered as an alternative substrate, but was not used for several reasons: the rate of enzymatic cleavage of FDG is 73 times slower than that of RG (20,21), and the time required for the assay would be greatly increased; cleavage of FDG results in two products, fluorescein and fluorescein monogalactopyranoside (FMG), and further FMG hydrolysis to fluorescein complicates the analysis of kinetic data (21); and
unlike fluorescein, the excitation and emission maxima of resorufin ($\lambda_{ex} = 575$ nm, $\lambda_{em} = 585$ nm) are well removed from the range where cellular autofluorescence can limit sensitivity.

In this study, $\beta$-galactosidase activity is expressed in $S. cerevisiae$ diploid strain D603 from a GAL10-$lacZ$ fusion, placing expression under the control of a galactose-inducible promoter(11,22). Because $S. cerevisiae$ does not possess a native $\beta$-galactosidase, the level of single-cell activity can be varied by changing the time of induction or the $\beta$-galactosidase gene dosage. The rationale for the development of this assay was to measure the distribution of gene dosage in a yeast population by measuring the distribution of the gene product, $\beta$-galactosidase.

The cells are permeabilized by adding the nonionic detergent Triton X-100 at a concentration of 0.073%. The assay mixture is chilled on ice to reduce resorufin leakage. Bovine serum albumin (BSA) is added to the assay mixture because it binds to extracellular resorufin and prevents it from diffusing back into the cells. Fluorescence from single cells is measured with a flow cytometer (Ortho Cytofluorograph 50H) using the yellow-green 568 nm line of a krypton ion laser (Coherent Innova 90K, Palo Alto, CA) to excite resorufin fluorescence. The level of single-cell fluorescence reaches a steady state after 5 minutes.

Population mean $\beta$-galactosidase activity was measured spectrophotometrically using a chromogenic substrate assay as described previously (19). One unit of $\beta$-galactosidase activity is defined as the amount that will hydrolyze 1.0 $\mu$ mole of O-nitrophenyl-$\beta$-D-galactoside to O-nitrophenol per minute at pH 7.3 and 37°C.

Model Equations

In order to obtain a quantitative correlation between resorufin fluorescence and $\beta$-galactosidase activity, a single-cell model of the diffusion and reaction processes involved in the assay was written. The equations describing product accum-
mulation are:

\[
\frac{dS}{dt} = k_i (S_o - S) - \frac{v_m S}{K_m + S} \quad (1a)
\]

\[
\frac{dP}{dt} = \frac{v_m S}{K_m + S} - k_e P \quad (1b)
\]

\[S = 0, \quad P = 0, \quad \text{at} \quad t = 0\]

where

\[S \equiv \text{internal substrate concentration}\]

\[P \equiv \text{internal product concentration}\]

\[S_o \equiv \text{external substrate concentration}\]

\[v_m, K_m \equiv \text{Michaelis–Menten kinetic parameters}\]

\[k_i, k_e \equiv \text{transport coefficients}\]

The assumptions present in this model are:

\textit{a) External substrate concentration is constant.} This assumption has been validated by performing the assay in a spectrophotometer, and observing that absorbance at the RG maximum of 470 nm remains approximately constant for the duration of the assay.

\textit{b) External product concentration is much less than the internal product concentration.} The concentration of BSA used in the assay is in great excess of the concentration of resorufin released from the cells, so at equilibrium 90\% of the extracellular resorufin is bound by BSA. In addition, if the resorufin concentration in the extracellular space significantly affected single-cell fluorescence accumulation, then the results of the assay would depend on the number density of cells.
In practice, single-cell fluorescence was independent of cell number density at the maximum cell concentrations used.

c) Enzymatic substrate cleavage can be described by Michaelis-Menten kinetics. This assumption has been verified for the cleavage of RG by β-galactosidase-containing cell lysates; however, it is not possible to independently measure the intracellular enzyme kinetics without explicitly accounting for substrate transport.

d) Cell-to-cell variation in surface area and volume are ignored. This assumption has been adopted to simplify the analysis. The cell size distribution will broaden the fluorescence distribution of a population of homogeneous enzyme content. Empirical evidence suggests that this broadening is not significant, since the fluorescence distribution of a monodisperse population with one copy of the β-galactosidase gene per cell is a sharp peak (23).

The model equations were nondimensionalized as follows:

\[
\frac{d\hat{S}}{d\tau} = \frac{1}{Da_i} (1 - \hat{S}) - \frac{\hat{S}}{\beta + \hat{S}} \tag{2a}
\]

\[
\frac{d\hat{P}}{d\tau} = \frac{\hat{S}}{\beta + \hat{S}} - \frac{1}{Da_e} \hat{P} \tag{2b}
\]

\[\hat{S} = 0, \quad \hat{P} = 0, \quad \text{at} \quad \tau = 0\]
\[ \hat{S} \equiv \text{Dimensionless substrate concentration} \left( \frac{S}{S_0} \right) \]
\[ \hat{P} \equiv \text{Dimensionless product concentration} \left( \frac{P}{S_0} \right) \]
\[ \tau \equiv \text{Dimensionless time} \left( \frac{t v_m}{S_0} \right) \]
\[ Da_i \equiv \text{Influx Damköhler number} \left( \frac{v_m}{k_i S_0} \right) \]
\[ Da_e \equiv \text{Efflux Damköhler number} \left( \frac{v_m}{k_e S_0} \right) \]
\[ \beta \equiv \text{Saturation parameter} \left( \frac{K_m}{S_0} \right) \]

Since single-cell resorufin fluorescence is allowed to come to a steady state, the relevant solution to Equation 2 is the steady state solution, obtained by setting the derivatives equal to zero. Equation 3 relates the product concentration \( \hat{P} \) to the dimensionless parameters \( \beta, Da_i, \) and \( Da_e. \)

\[ \hat{S} = \sqrt{\beta + \left( \frac{1 - \beta - Da_i}{2} \right)^2 + \frac{1 - \beta - Da_i}{2}} \quad (3a) \]

\[ \hat{P} = Da_e \frac{\hat{S}}{\beta + \hat{S}} \quad (3b) \]

The parameter \( Da_i \) represents the ratio of the characteristic maximum reaction rate to the characteristic substrate diffusion rate. The saturation parameter \( \beta \) represents the degree of saturation of enzyme binding by the substrate. Both \( Da_i \) and \( \beta \) determine whether product accumulation is reaction-limited or diffusion-limited. It is necessary to perform the assay under reaction-limited conditions in order to obtain a strong correlation between enzyme activity and product concentration.

The most useful relationship between fluorescent product concentration \( (P) \) and enzyme activity \( (v_m) \) is proportionality. If \( \beta << \hat{S}, \) then
\[ \dot{P} \approx Da_c \quad (4a) \]

or

\[ P \approx \frac{v_m}{k_e} \quad (4b) \]

This situation corresponds to saturation of the enzyme-substrate binding step by an excess of substrate. For this reason, it is desirable to use the maximum feasible bulk substrate concentration.

The dependence of product concentration on enzyme activity was calculated from Equation 3, and is shown in Figure 2. The values of \( k_i, k_e \), and \( S_0 \) were fixed at 1.0 so that \( Da_i = 1 \) when \( v_m = 1 \). When \( \beta \) is small, there is a sharp transition from the reaction-limited regime \( (Da_i < 1) \) to the diffusion-limited regime \( (Da_i > 1) \). As \( \beta \) increases, the saturation of enzyme binding by the substrate decreases. Because \( v_m \) alone does not specify the reaction rate, product accumulation can be reaction-limited even if \( Da_i \) is greater than 1, provided that \( \beta \) is large.

For a fixed enzyme activity \( v_m \) the fluorescence accumulation will be reaction-limited if \( Da_i \) is small. The assay conditions were manipulated in such a way as to minimize \( Da_i \). The manipulated conditions include: pH, temperature, permeabilizing detergent concentration, and substrate concentration.

Changing the pH of the assay buffer will change the intracellular pH only if the cell membrane is proton-permeable. Since \( \beta \)-galactosidase activity is optimal at a pH of 7.3(24), one could vary the enzyme activity \( v_m \) without changing the quantity of enzyme present in the cell or affecting substrate and product diffusion by varying intracellular pH. However, it was found that varying the pH of the assay buffer from 6 to 8 did not significantly affect fluorescence accumulation. This would indicate either that the cell membrane is not permeabilized to protons
by 0.073% Triton X-100, or that the dependence of enzyme activity on intracellular pH is weak.

The temperature of the assay affects three parameters: $v_m$, $k_i$, and $k_e$. Decreasing the temperature from 37°C to 5°C lowers $v_m$ for β-galactosidase by a factor of 10(9,24). Over this same range, membrane fluidity drops drastically. For example, the flux of fluorescein across cell membranes is reduced by a factor of 200 across this same temperature range(9,25). It was necessary to perform the assay at 0°C in order to minimize resorufin leakage. However, this is at cross-purposes with the goal of reducing $D_{a_i}$, since $k_i$ is reduced proportionately more than $v_m$.

The transport coefficient $k_i$ can be varied independently of $v_m$ by changing permeabilization conditions. The nonionic detergent Triton X-100 was added to permeabilize the cells at a concentration optimized to balance the effects of improved substrate permeability and increased resorufin leakage(19).

Increasing the substrate concentration $S_o$ decreases $D_{a_i}$ and $\beta$. For this reason, the substrate was added at the highest concentration attainable without significant precipitation, which was observed as narrow forward angle light scatter noise with the flow cytometer. This concentration is 290 µM.

**Experimental results**

In developing this single-cell enzyme assay, the original intent was to add substrate at a saturating concentration (i.e., until fluorescence accumulation was unaffected by further increases in substrate concentration). It was found, however, that the saturating substrate concentration depends on the mean enzyme activity of the population (Figures 3 and 4). Furthermore, single cell fluorescence is approximately proportional to the substrate concentration up to the saturating level.

The observed dependence of product accumulation on substrate concentra-
tion could not be due to Michaelis-Menten type saturation kinetics, because the apparent $K_m$ varies with the enzyme activity $v_m$. The saturation cannot be due to "concentration quenching" of resorufin fluorescence, or nonspecific RG hydrolysis outside of the cells, because the maximum fluorescence is less at lower $v_m$.

The observed substrate dependence can be explained in terms of the single-cell model of Equations 1-3. At substrate concentrations below the saturating level, fluorescence accumulation is diffusion-limited. As $S_o$ increases, $D_{a_i}$ decreases, and at the saturating level, fluorescence accumulation is reaction-limited. If this explanation is correct, then $D_{a_i}$ will be approximately 1 at the point where product accumulation becomes saturated; therefore, the substrate concentration at this point will be proportional to the mean enzyme activity. The mean activity of the cells in Figure 4 is 3.5 times that of the cells in Figure 3, and the saturating substrate concentration in Figure 4 is approximately 3-5 times that in Figure 3. This is in agreement with the prediction of the diffusion-reaction model.

The flow cytometric $\beta$-galactosidase assay was performed on a number of samples with a range of mean enzyme activity, and the results are shown in Figure 5. It was found that median single-cell resorufin fluorescence is proportional to mean $\beta$-galactosidase activity at levels up to $4.0 \times 10^{-8}$ U/cell. Further increases in activity do not yield a further increase in fluorescence, because the assay is diffusion-limited at activities higher than $4.0 \times 10^{-8}$ U/cell.

Because it is of interest to measure single-cell activities higher than $4.0 \times 10^{-8}$ units/cell, isopropyl-$\beta$-D-thiogalactopyranoside (IPTG) was added as a competitive inhibitor of $\beta$-galactosidase activity, to shift the accumulation of resorufin to a reaction-limited regime. The effect of IPTG on resorufin accumulation is shown in Figure 6. Since IPTG is a competitive inhibitor, its apparent effect is to increase $K_m$, as opposed to decreasing $v_m$. Increasing $K_m$ corresponds to increasing $\beta$.
The effect of increasing $\beta$ on the relationship between product concentration $P$ and enzyme activity $v_m$ was calculated from Equation 3, and is shown in Figure 7. Although adding an inhibitor lowers the absolute level of fluorescence at all activities, it allows the fluorescence signal at activities greater than $4.0 \times 10^{-8}$ U/cell to be differentiated from signals of cells with lower activity.

Discussion

One of the major stumbling blocks to more widespread application of flow cytoenzymology has been the lack of quantitative correlations between single-cell activity and the signal being measured. In this work, we have developed a mathematical model which describes the accumulation of a fluorescent product as the result of enzymatic cleavage of a fluorogenic substrate. This model predicts that under certain assay conditions, the level of fluorescent product is determined by the rate of substrate diffusion into the cell. The model can also be used to suggest changes in the assay conditions to shift product accumulation to a reaction-limited regime.

Previous investigators have described mathematical models of histochemical staining reactions. Cornelisse and coworkers calculated the steady-state concentration gradient of a diffusible enzyme reaction product in an idealized spherical cell with a spherical enzymatically active particle in its center(17). The steady state equations were numerically integrated, and the model was applied to the generation of phosphate ions by acid phosphatase-containing rat liver lysosomes. Cornelisse and van Duijn also examined the precipitation kinetics of lead phosphate, the trapped product of a histochemical assay of phosphatase activity(18). O'Sullivan modeled the distribution of an enzyme reaction product in planar tissue sections under various conditions of substrate diffusivity and enzyme distribution(26). The cases considered were: rapid substrate penetration, resulting
in Michaelis-Menten kinetics; slow substrate diffusion into a region of uniformly distributed enzyme, which resulted in modified Michaelis-Menten kinetics; and slow penetration into a region of non-uniform enzyme distribution, which was an intractable problem.

In contrast to the models mentioned above, in the present work diffusion limitations inside the cell were ignored, eliminating the difficulty of cell geometry considerations. The resulting simple model should have general applicability to flow cytoenzymological assays.

An assay of β-galactosidase activity in single viable mammalian cells has recently been developed using fluorescein-di-β-D-galactopyranoside (FDG) as the substrate (9). FDG is rapidly loaded into the cells under hypotonic conditions at 37°C. After FDG loading, the cells are placed at 4°C to allow fluorescein to accumulate; the low temperature eliminates leakage problems. With this assay, single-cell fluorescence is proportional to enzyme activity over a range of activities greater than two orders of magnitude. Rapid substrate loading and negligible product leakage eliminates the necessity to account for diffusion processes, as is required for the assay described here. However, the hypotonic method used to permeabilize mammalian cells does not permeabilize the tough cell wall of yeast cells; preliminary experiments with FDG in distilled water revealed no substrate penetration into the yeast cells used to develop the RG assay (data not shown).

Gene fusions with reporter enzymes can be used to study gene expression in cells ranging from bacteria to mammalian cells and plant cells. In order to obtain quantitative data from these systems, the kinetics of the enzyme assay must be well characterized. In particular, artifacts due to diffusion limitation must be clearly differentiated from the kinetics of the enzymatic reaction. Mathematical models of the type described here can be used to guide the development of single-cell enzyme
assays, and to quantitatively correlate the measured signal with single-cell enzyme activity.
References


Figures

**Figure 1.** Schematic diagram of a fluorogenic substrate single-cell enzyme assay. A nonfluorescent substrate diffuses into the cell and is enzymatically cleaved to form a fluorescent product. The fluorescent product is often more polar than the substrate, and diffuses less readily across the cell membrane.

**Figure 2.** Relationship between the steady state fluorescent product concentration $P$ and single-cell enzyme activity $v_m$ as a function of the substrate binding saturation factor $\beta$. The curves were calculated using Equation 3, with $k_i, k_e$, and $S_o$ fixed at 1.0.

**Figure 3.** Steady state median single-cell resorufin fluorescence measured by flow cytometry vs. substrate (RG) concentration. Cells of the strain D603-2i, which has two chromosomally integrated copies of the *lacZ* gene per cell, were assayed after 4 hours of induction of $\beta$-galactosidase expression. The mean enzyme activity is approximately $4.0 \times 10^{-8}$ U/cell. The units of fluorescence intensity are normalized by the maximum value in this figure, and do not correspond to the units in Figure 4.

**Figure 4.** Steady state median single-cell resorufin fluorescence measured by flow cytometry vs. substrate (RG) concentration. Cells of the strain D603:pLGSD5, which possess an average of ten copies of the *lacZ* gene per cell on a multicopy plasmid, were assayed after 4 hours of induction of $\beta$-galactosidase expression. The mean enzyme activity is approximately $1.4 \times 10^{-7}$ U/cell. The units of fluorescence intensity are normalized by the maximum value in this figure, and do not correspond to the units in Figure 3.

**Figure 5.** Median single-cell resorufin fluorescence measured by flow cytometry vs. mean $\beta$-galactosidase activity measured spectrophotometrically. The three different symbols represent three independent experiments. Each experiment was
performed with cells of *S. cerevisiae* strains with 1, 2, and 10 copies of the *lacZ* gene (D603-i, D603-2i, and D603:pLGSD5, respectively), with β-galactosidase expression induced from 2-4 hours. Fluorescence measurements were calibrated with the right-angle light scatter signal from 2 μm polystyrene spheres. The solid line represents a least squares fit of Equation 3 to the data, with the following parameters: \( \beta = 0.0216, k_iS_o = 4.17 \times 10^{-8} \) U/cell, maximum fluorescence intensity = 35.9 arbitrary units.

**Figure 6.** Effect of the competitive β-galactosidase inhibitor IPTG on steady state resorufin fluorescence measured by flow cytometry. The effect of a competitive inhibitor on the model parameters is to increase \( K_m \), which increases \( \beta \).

**Figure 7.** Use of a competitive inhibitor to shift product accumulation to a reaction-limited regime. These curves were calculated from Equation 3. Addition of a competitive inhibitor effectively increases \( K_m \), which increases \( \beta \). When \( \beta = 0.1 \) (curve on the left), the ratio \( P(v_m = 2)/P(v_m = 1) \) is 1.25. When \( \beta = 3.0 \), the ratio \( P(v_m = 2)/P(v_m = 1) \) is 1.69. Although the absolute fluorescence intensity is decreased (note the different vertical scales for the two curves), it is possible to distinguish product concentrations at higher activities when \( \beta = 3.0 \) than when \( \beta = 0.1 \).
Figure 1.
Figure 2.
β-GALACTOSIDASE ACTIVITY (UNITS/CELL * 10^-8)

MEDIAN SINGLE-CELL FLUORESCENCE

Figure 4.
Figure 5.
CHAPTER 6

Propagation of an Amplifiable Recombinant
Plasmid in *Saccharomyces cerevisiae:
Flow Cytometry Studies and Segregated Modeling
ABSTRACT

Efficient expression of a foreign protein product by the yeast *Saccharomyces cerevisiae* requires a stable recombinant vector present at a high number of copies per cell. A conditional centromere yeast plasmid was constructed which can be amplified to high copy number by a process of unequal partitioning at cell division, followed by selection for increased copy number. However, in the absence of selection pressure for plasmid amplification, copy number rapidly drops from 25 plasmids/cell to 6 plasmids/cell in less than ten generations of growth. Copy number subsequently decreases from 6 plasmids/cell to 2 plasmids/cell over a span of 50 generations. A combination of flow cytometric measurement of copy number distributions and segregated mathematical modeling were applied to test the predictions of a conceptual model of conditional centromere plasmid propagation. Measured distributions of plasmid content displayed a significant subpopulation of cells with a copy number of 4-6, even in a population whose mean copy number was 13.5. This type of copy number distribution was reproduced by a mathematical model which assumes that a maximum of 4-6 centromere plasmids per cell can be stably partitioned at cell division. The model also reproduces the observed biphasic instability. The agreement between simulation and experimental results provides support for the proposed model, and demonstrates the utility of the flow cytometry/segmented modeling approach for the study of multicopy recombinant vector propagation.
Introduction

The baker’s yeast *Saccharomyces cerevisiae* presents several advantages for the production of pharmaceutical proteins by recombinant DNA technology(1-7). Yeast is an established industrial microorganism, and can be grown to very high cell densities in large, process-scale fermentors. *S. cerevisiae* is one of only a few microorganisms that are Generally Recognized As Safe (GRAS) by the FDA; yeast possesses no known toxins, and is edible in bread. The capability for genetic manipulation of *S. cerevisiae* is highly developed, and the fundamental biology of this organism is thoroughly studied. To a limited extent, *S. cerevisiae* can mimic the post-translational protein modifications performed by higher eucaryotes, such as glycosylation and proteolytic processing associated with secretion(1,7,8-14).

One limitation of *S. cerevisiae* as a host for heterologous protein expression is the relative lack of mitotically stable high copy number vectors available for transformation(15,16). The capability to increase gene expression by increasing promoter strength is limited by the maximum rate of mRNA transcription from a single promoter. Amplification of the number of copies of the gene provides an alternate route to enhancing gene expression. Integrative vectors and centromere plasmids are stable, but maintained at low copy number. Plasmids based on the native 2µ plasmid are present at higher copy numbers, but are less stable.

Circular plasmids in *S. cerevisiae* exhibit a bias to segregate to the mother cell at cell division in the absence of DNA sequences which promote stable plasmid partitioning(23). The result of biased plasmid propagation, as shown in Figure 1, is that a small fraction of the population contains a large number of plasmids, while the majority of cells are plasmid-free. Plasmids which contain an ARS (Autonomously Replicating Sequence) to direct plasmid replication but lack partitioning sequences display this type of plasmid propagation(23).
Two types of partitioning sequences are known to function in yeast plasmids: the 2μ plasmid STB site, and chromosomal centromeres. The STB site interacts with proteins expressed from the native 2μ plasmid to stably partition multicopy plasmids (24-26). Centromeres are the DNA sequences responsible for partitioning of chromosomes at mitosis. Placing a centromere on an ARS plasmid eliminates segregation bias, but plasmid copy number is reduced to one (27,28). From the standpoint of overexpressing a foreign protein in yeast, it is desirable to stably maintain a plasmid at high copy number. Although a centromere imparts stability, it also limits copy number to one.

It has been found that initiating transcription through a centromere interferes with its partitioning function (29,30). By placing an inducible promoter upstream of a centromere, one can control centromere function by varying the level of transcription. Such an inducible promoter-centromere construction is called a “conditional centromere.” Several different types of conditional centromere have been constructed (20,29,30).

When a conditional centromere is placed on a circular plasmid and transcription through the centromere is repressed, the plasmid is stably propagated at a copy number of one. When transcription through the centromere is induced, the plasmid is subject to biased segregation as shown in Figure 1, and after several generations a subpopulation with high copy number will be formed. Assuming that each plasmid replicates once per cell cycle (31-33), the highest single-cell copy number attainable by this strategy is $2^N$, where $N$ is the number of generations of biased segregation. If it is possible to select the subpopulation of cells with high copy number, this provides a method for amplifying plasmid copy number in yeast.

Single-cell multicopy plasmid content is typically heterogeneous, and consid-
erable information concerning plasmid propagation is present in the form of this heterogeneity. A combined experimental/modeling approach which explicitly accounts for the segregation of a population into individual cells is necessary in order to rigorously relate observations of mean population behavior to events at the cellular level. The information flow can also be reversed, from knowledge of single-cell mechanisms to prediction of mean population characteristics. Since the productivity of a population is the sum over the single-cell productivities, segregated mathematical models can link fundamental biological knowledge to bioreactor design and control.

In this paper we describe the construction of a conditional centromere plasmid which can be amplified to high copy number by utilizing the property of biased plasmid segregation in yeast. The instability of the amplified copy number state is studied using a flow cytometric enzyme assay to measure the distribution of single-cell copy number, and a segregated mathematical model to provide the link between a simplified conceptual model and the experimental observations.
MATERIALS & METHODS

Strains and Media

Saccharomyces cerevisiae diploid strain D603(17) (α/α, cir+, ade2 – 101, ura3 – 53, his3, lys2 – 801, met, reg1 – 501) possesses the reg mutation, which inhibits glucose catabolite repression. The reg mutation allows the induction of galactose-inducible transcription in the presence of glucose. β-galactosidase expression from the GAL10-lacZ fusion is induced by adding galactose to a final concentration of 2% w/v. D603-i has one copy of the lacZ gene chromosomally integrated, and D603-2i has two integrated copies.

For growth selection of plasmid-containing cells, ura− SD minimal medium(18) was used, buffered with sodium citrate (0.05 M, pH 4.5) and supplemented with adenine sulfate (20mg/L), L-histidine (20-mg/L), L-methionine (20 mg/L), and L-lysine (30 mg/L). For selection of amplified copy number, YPD-G100 medium was used (YPD complex medium (18) supplemented with G418 (100 μg/mL)(Gibco)).

Plasmid construction

Plasmid pAP251 was constructed from portions of plasmids pLGSD5(19) and pAP736(20). The 3.2 kb EcoRI fragment containing pBR322 sequences and the 4.6 kb EcoRI-HindIII fragment containing URA3 and the GAL10-CYC1-lacZ fusion from pLGSD5 were ligated together with the 4.2 kb HindIII fragment containing CUP1p-CEN3-CUP1T, PGKp-G418R, and ARS1, and the 0.8 kb EcoRI-HindIII fragment containing TRP1 from pAP736. The map of the resulting 12.8 kb plasmid is shown in Figure 2.

β-galactosidase assays

Spectrophotometric β-galactosidase assays were performed as previously described(21). Briefly, the absorbance at 420 nm of O-nitrophenol liberated from
O-nitrophenyl-\(\beta\)-D-galactoside (ONPG, Sigma) was measured with a Shimadzu UV-260 spectrophotometer (Shimadzu, Kyoto, Japan) after addition of a known number of isopropanol-permeabilized cells containing \(\beta\)-galactosidase. The number density of cells was determined using a Coulter Counter ZM with a Channelyzer 256 (Coulter Electronics, Hialeah, FL).

The single-cell \(\beta\)-galactosidase assay is described in detail elsewhere (21). The nonfluorescent fluorogenic substrate resorufin-\(\beta\)-D-galactopyranoside (Molecular Probes) diffuses into cells and is cleaved by \(\beta\)-galactosidase to form fluorescent resorufin. Single-cell resorufin fluorescence is measured with a flow cytometer (Ortho Cytofluorograf 50H). The single-cell fluorescence signal is correlated with single-cell enzyme activity (22).
RESULTS

Plasmid pAP251 and Conditional Centromere Control

The plasmid pAP251 was constructed as described in Materials and Methods to study the propagation of conditional centromere plasmids. A schematic plasmid map is shown in Figure 2. The significance of each portion of the plasmid is outlined here, and is explained in greater detail below. The partitioning function of the centromere CEN3 can be partially repressed by inducing transcription from the yeast CUP1 promoter, which is induced by Cu$^{2+}$ ions. The URA3 gene provides selection for the growth of D603 cells containing the plasmid in minimal SD medium lacking uracil. The G418$^R$ gene product is an enzyme which inactivates the antibiotic G418. G418$^R$ expression is directed by the constitutive yeast PGK promoter. Since increasing levels of the G418$^R$ gene product confer increased resistance to G418, increases in copy number result in increased G418 resistance. $\beta$-galactosidase expression from the GAL10-lacZ gene fusion is induced by galactose. $\beta$-galactosidase activity is used as an indicator of plasmid copy number.

The native CUP1 gene in yeast codes for a small cysteine-rich protein which chelates free copper ions in the cytoplasm, conferring resistance to the toxic effects of copper (34,35). Transcription from the CUP1 promoter can be induced by adding CuSO$_4$ to the growth medium(36,37). Previous studies have shown that CUP1 transcription is maximally induced at CuSO$_4$ concentrations high enough to cause an inhibitory effect on growth rate(37).

An experiment was performed to determine the effect of CuSO$_4$ on centromere function in pAP251. Equal aliquots of a suspension of D603:pAP251 cells in early exponential growth were divided into test tubes with differing total concentrations of CuSO$_4$. The cells were incubated at 30°C for 24 hours, after which the total number of cells were counted to determine copper toxicity. The plasmid-containing
fraction of selected samples was determined.

Because destabilization of centromere function leads to a biased segregation event, the probability of plasmid loss per generation is increased by loss of centromere function. The probability of plasmid loss per cell division is calculated using the following equation:

\[
p = \frac{\ln \frac{F^+(0)}{F^+(N)}}{N \ln 2}
\]

where

\( p \equiv \text{probability of plasmid loss per cell division} \)

\( F^+(i) \equiv \text{plasmid - containing fraction after } i \text{ generations of growth} \)

It is assumed in the derivation of this equation that the growth rate of cells with and without plasmid are the same. This assumption is valid for growth of D603 cells with and without single-copy pAP251 in YPD medium. Deviations from this equation would occur at high copy number, which inhibits cell growth in plasmid-containing cells. This expression is adequate for the present comparison, however.

The effect of CuSO\(_4\) concentration on cell growth and \( p \) are shown in Figure 3. An increase in plasmid instability due to repression of centromere function coincides with the onset of copper toxicity effects, as would be expected if centromere partitioning function is destabilized by CUP1 transcription. It should be noted that the level of centromere destabilization observed with this system is somewhat lower than the effects observed with other conditional centromere constructs(29). This may be a result of lower levels of transcription from the native \( CUP1 \) promoter, relative to other promoters which have been used for this purpose.

Selection of clones with amplified pAP251 copy number
The plasmid pAP251 has two selective markers. The *URA3* gene allows for selection of cells containing the plasmid, regardless of copy number. The presence of one *URA3* gene on a plasmid is sufficient for cells to grow in minimal SD medium lacking uracil. The other selective marker on pAP251 is PGK-G418\(^R\), which provides selection for the growth of cells with high copy number. The G418\(^R\) gene product is an enzyme which inactivates the aminoglycoside antibiotic G418, so increasing the level of this enzyme provides increased resistance to G418(38-39). Since expression of the G418\(^R\) gene product increases with gene dosage, cells with amplified pAP251 copy number are resistant to higher levels of G418. A G418 concentration of 100 \(\mu\)g/mL in YPD plates (YPD-G100) was used to select clones with amplified pAP251 copy number, since D603 cells lacking the plasmid exhibit no growth in this medium.

At the conclusion of the experiment shown in Figure 3, the population mean \(\beta\)-galactosidase activity was found to have increased negligibly in those samples which had exhibited centromere destabilization. This was expected, since the average number of plasmids per cell over the entire population had not increased, only the distribution of the plasmids in the population. The subpopulation with high copy number arises due to biased segregation, not over-replication of the plasmid. To select cells with amplified copy number, the samples exposed to CuSO\(_4\) were plated on YPD-G100 plates, and single colonies were selected for further examination. To determine the effect of pAP251 amplification on G418 resistance, a colony was picked from a ura\(^-\) SD plate (cells with a copy number of one), and a colony was picked from a YPD-G100 plate (amplified copy number). Each entire colony was placed in 1 mL of ura\(^-\) SD medium for 3 hours at 30°C. 0.08 mL of this inoculum was added to each of several test tubes containing 3 mL of YPD and varying concentrations of G418. The tubes were incubated at 30°C.
for 24 hours with shaking, after which the number of cells in each test tube was
determined. The results are shown in Figure 4. The cells with amplified pAP251
copy number are resistant to G418 levels approximately ten times higher than cells
with a single copy of pAP251.

At the conclusion of the growth experiment just described, 0.2 mL of each
sample was diluted to a total of 3 mL YPD at the same G418 concentration. After
5 hours of growth, β-galactosidase expression was induced for 3 hours. Population
mean β-galactosidase activity was determined by a spectrophotometric assay, and
the results are shown in Figure 5. The maximum β-galactosidase activity of the
cells with amplified copy number is approximately nine times higher than that
of single-copy cells. Taken together, the data in Figures 4 and 5 provide a gross
estimate of 10 plasmids/cell for the cells with amplified copy number. Note that
in the absence of G418 mean copy number fell to approximately 6 over the course
of the experiment.

β-galactosidase as a reporter of copy number

The level of expression of β-galactosidase from the GAL10-lacZ gene fusion
on pAP251 can be used to estimate copy number. β-galactosidase may also be
considered a model biotechnological product, in addition to its role as a reporter
molecule. Viewed in this light, increases in β-galactosidase expression provide
direct evidence of the benefit of amplified copy number for the expression of a
foreign protein.

It has been shown that β-galactosidase activity is proportional to the number
of lacZ genes in the S. cerevisiae strain D603 at copy numbers up to 10 (see Figure
6). An attempt was made to verify proportionality between copy number and β-
galactosidase activity in amplified clones of D603:pAP251 by performing Southern
blot hybridizations to directly measure copy number. Although quantitation of
copy number was prevented by high background levels, it was apparent that more plasmid DNA was present in cells with amplified β-galactosidase activity than in cells with one copy of the plasmid (data not shown). Mean β-galactosidase activity equivalent to 45 lacZ genes has been measured in one sample, suggesting that the capacity to express β-galactosidase is not completely saturated by less than 45 lacZ genes per cell.

**Plasmid stability at amplified copy number**

The stability of the amplified copy number state was studied. Amplified clones from YPD-G100 plates were grown for several generations in ura− SD medium lacking G418. The cultures were maintained in continuous exponential growth by serial dilutions. The population mean copy number was periodically estimated by inducing β-galactosidase expression in aliquots from each culture. The results are shown in Figure 7.

After approximately 10 generations in the absence of G418, there is a broad distribution of mean copy number in the samples, ranging from 8 to 24 plasmids/cell. In the next 10 generations, copy number falls rapidly to 4 to 6 copies per cell, a level which drops only slightly over the next 20 generations. The plasmid-containing fraction was measured by replica-plating after 10 generations and 44 generations, and was approximately constant at 0.95.

The slow transient decrease in copy number was examined in more detail, and the results are shown in Figure 8. Although copy number drops from 24 to 6 in 10 generations in Figure 7, 50 generations of growth pass before copy number drops from 6 to 2 in Figure 8.

**Copy number distribution by flow cytometry**

Measurement of mean copy number and plasmid-containing fraction provide little insight into the cause of the two phases of instability of the amplified copy
number state. In order to further examine this instability, the distribution of plasmid copy number in the population was measured.

A quantitative flow cytometric assay of β-galactosidase activity in single cells of *S. cerevisiae* has been developed, and is described elsewhere(21). Briefly, a non-fluorescent substrate analog diffuses into the cells and is cleaved by β-galactosidase to form a fluorescent product. Single-cell fluorescence is measured by flow cytometry, and this fluorescence is related to single cell enzyme activity by a mathematical model accounting for substrate and product diffusion(22).

The copy number distributions of populations with amplified pAP251 copy number were measured, and are shown in Figure 9. A striking feature of the distributions is that although one population’s mean copy number (13.2) is twice that of the other (6.5), the mode or peak of both of the distributions is at the same position, a copy number of about 4. The greater mean copy number of the more amplified population is due to a larger proportion of cells in the high copy number tail of the distribution. Cells with a copy number of one possess a tight distribution of copy number.

**Model of conditional centromere plasmid propagation**

Based on the experimental results presented above and published observations concerning the propagation of plasmids in *S. cerevisiae*, a model of conditional centromere plasmid propagation has been formulated. The assumptions of the model are as follows:

1). *There is a maximum number* $N_c$ *of centromere-containing plasmids that can be equally partitioned at cell division.* This is a simplification of the general assumption that some component of the mitotic apparatus is titrated by increasing the number of centromeres in the cell. The molecular machinery for dividing chromosomes between nuclei at cell division is unlikely to be present in great
excess of the requirements for chromosomal partitioning.

2). All plasmids in excess of \( N_c \) are subject to biased segregation. If the cell’s partitioning apparatus is saturated by excess centromeres, all remaining plasmids would appear to lack a partitioning locus, i.e., to segregate in a biased fashion.

3). Single-cell specific growth rate decreases monotonically with increasing copy number. In this context, inhibition of growth rate is due to a toxic effect of excess cloned centromeres. Futcher and Carbon (40) have observed an elongated morphology and reduced growth rate in cells forced to maintain in excess of five centromere-containing plasmids. In growth experiments with the strain D603:pAP251, population mean specific growth rate is found to decrease with an increase in population mean copy number (Figure 10). At mean copy numbers in excess of 20, a significant fraction of D603:pAP251 cells exhibited the elongated morphologies described by Futcher and Carbon.

4). When copy number is less than \( N_c \), the dominant mode of plasmid loss is replication failure. This assumption is based on the observations of Hieter and coworkers (41) and Koshland and coworkers (42), who found that approximately 80% of plasmid loss events in single-copy centromere-containing cells are due to failure of the plasmid to replicate, rather than aberrant segregation. It is assumed that each plasmid is subject to the same probability of replication failure when copy number is below \( N_c \).

This model of conditional centromere instability states, in effect, that the rapid drop in copy number from amplified levels is due to segregational instability, and that the subsequent slow decrease is due to replication failure. The particularly simplified form in which these assumptions are stated is necessary in order to formulate a tractable mathematical model of plasmid propagation. Mathematical modeling provides the link between assumptions 1-4 and the experimentally
measured copy number distributions.

Model Implementation

A mathematical framework for modeling multicopy plasmid propagation has been described previously (43). By a simplification for the case of a discrete state variable (plasmid number), the population balance equations are reduced to a linear system of ordinary differential equations:

\[
\frac{dN}{dt} = \mathbf{T} \cdot \mathbf{N}
\]

\[
\mathbf{N} = \mathbf{N}_0 \quad ; \quad t = 0
\]

where

\[PN \equiv \text{plasmid copy number}\]
\[N_i \equiv \text{number of cells with } PN = PN_i\]
\[T_{i,j} \equiv (P_{i,j} - \delta_{ij})\mu_j\]
\[P_{i,j} \equiv \text{mean number of cells with } PN = PN_i \text{ arising from one division of a cell with } PN = PN_j\]
\[\mu_j \equiv \text{single - cell specific growth rate of cells with } PN = PN_j\]

\[
\delta_{ij} = \begin{cases} 
1, & i=j \\
0, & i \neq j
\end{cases}
\]

From the assumptions of the model of conditional centromere plasmid propagation, the elements of the transition probability matrix \( \mathbf{T} \) were assigned as follows:

\[P_{i,j} = P^{BIN}(i; j, 1 - P_R) + \delta_{ij}
\]

\[3a)\]
for \(1 \leq j \leq N_c\), \(1 \leq i \leq j\).

\[
P_{N_c+i,j} = P^M(N_c+i,j) + P^M(2j-N_c-i,j)
\]

for \(N_c < j \leq N_{\text{max}}\), \(1 \leq i \leq 2j - 2N_c\),

where

\[
P^M(k,l) = P^{BIN}(2l-N_c-k;2j-2N_c,1-P_S)
\]

\[
P^{BIN}(k;N,p) = \frac{N!}{k!(N-k)!} p^k(1-p)^{N-k}
\]

\(P_{i,j}\) \(\equiv\) element of \(T\) matrix defined in Equation 2

\(P_R\) \(\equiv\) probability that a single plasmid fails to replicate

\(P_S\) \(\equiv\) probability that a single plasmid remains with the

mother cell after cell division

\(N_c\) \(\equiv\) maximum number of centromere – containing plasmids

that can be stably partitioned

When \(PN\) is greater than \(N_c\), all of the plasmids are assumed to replicate, and \(N_c\) plasmids are distributed to the daughter cell. The remaining plasmids are binomially distributed, with a probability \(P_S\) of remaining with the mother cell. When \(PN\) is less than or equal to \(N_c\), all plasmids are considered to possess the probability \(P_R\) that they will fail to replicate. The mother cell receives \(PN\) plasmids, and the daughter cell receives all plasmids that were successfully replicated.
The growth model is a simple empirical form with parameters chosen to approximate the data in Figure 10:

\[ \mu_i = 0.33 \left(1 - \frac{i}{N_{max}}\right) \]

where

\[ \alpha = 0.5 \]

\[ N_{max} = 50 \]

It is incorrect to attempt to directly fit equation 4 to the data in Figure 10, because equation 4 expresses single-cell growth rate as a function of single-cell copy number, and the data in figure 10 represent population averages. Both the copy number distribution and the transition matrix \( T \) would be needed to relate a single-cell growth model to the population mean growth kinetics. However, the qualitative behavior of the propagation model is insensitive to the exact values of the parameters in equation 4.

**Feasible Parameter Space**

Estimates of the model parameters \( N_c, P_S, \) and \( P_R \) are obtainable from the literature and the experimental results presented here. In Figures 7 and 8, it is shown that copy number rapidly falls to a value of 4 to 6 in the absence of selection pressure for plasmid amplification. In Figure 9, it is shown that a substantial subpopulation exists with a copy number of 4-6, even when population mean copy number is 13.2. These observations together suggest that a reasonable estimate of the number of plasmids which can be stably partitioned is \( N_c = 4 - 6 \).

Murray and Szostak(23) present data showing the segregation bias for plasmids lacking a partitioning site. Their data includes the probability of formation of a plasmid-free daughter cell by a biased segregation event, and the mean copy number. From these two values, the single-plasmid segregation bias probability
$P_S$ can be calculated using the following equation:

$$P_S = P_M^{PN}$$

where

$P_M \equiv$ probability of formation of a plasmid – free daughter cell

$PN \equiv$ mean plasmid copy number

From data for several different plasmids, the parameter $P_S$ varies from 0.9 - 0.95.

Using colony-color assays to determine the events responsible for single-copy centromere plasmid loss (41,42,44), it has been estimated that approximately 80% of the single-copy plasmid loss events are due to replication failure. The rate of plasmid loss varies from 0.005 to 0.02. In the present simplified model, the parameter $P_S$ is assumed to vary from 0.005 to 0.04.

**Simulation Results**

The purpose of the mathematical model is to demonstrate the consistency of the conceptual propagation model with the experimental results. In this context, curve fitting the value of a particular parameter is less interesting than the qualitative behavior of the mathematical simulation within a feasible parameter space.

Equation 2 with the $T$ matrix determined by Equation 3 was numerically solved with the IMSL subroutine DVERK, which uses a fifth order Runge Kutta algorithm(45). In all of the calculations shown below, the growth model with the parameter values in Equation 4 was used. In all of the simulations, growth of a plasmid-free subpopulation was ignored, since the plasmid-free fraction was below 5% during growth in ura$^-$ SD minimal medium.

The effect of replication failure probability $P_R$ on the phase of slow copy number decrease is shown in Figure 11. The initial condition for the simulation
is one cell with copy number 6. The slope of the curves approximates that of the data for values of \( P_R \) between 0.02 and 0.04.

The effect of segregation bias probability \( P_S \) on the phase of rapid copy number decrease is shown in Figure 12. The initial condition at 10 generations is a single cell with copy number 25. This allows Figure 12 to be compared with the experimental data shown in Figure 7. The same basic features are present at values of \( P_S \) from 0.8 to 0.99: a drop from 25 plasmids/cell to 6 plasmids/cell within 10 generations, followed by a slow decline in copy number. Increasing the probability of biased segregation increases the rate of decline in mean copy number.

Transient shifts in the copy number distribution were simulated, using the experimentally measured distribution at mean copy number 13.2 from Figure 9 as an initial condition. The transient simulation was stopped when population mean copy number reached 6.5, so that the result could be compared to the experimentally measured distribution shown in Figure 9. To test the sensitivity of the distribution to the model parameters, simulations were performed with all six possible combinations of the parameter sets \( N_c = (5, 6) \) and \( P_S = (0.8, 0.9, 0.99) \). The results are shown in Figure 13. Gaussian measurement noise with a coefficient of variance of 0.25 was superimposed on the calculated distributions. All of the simulated distributions resemble the measured distribution in Figure 9, in that the tail of the distribution shrinks while the mode at copy number 4-6 rises.

The qualitative agreement between the model calculations and the experimental results is good for all feasible parameter values that were tested. Given the simplicity of the conceptual propagation model and the insensitivity of the model results to parameter changes within the feasible parameter space, this agreement provides a convincing argument for the validity of the basic assumptions of the propagation model.
DISCUSSION

A conditional centromere plasmid was constructed to study the stability of this type of amplifiable multicopy plasmid in yeast populations. The inherent tendency of *S. cerevisiae* to segregate circular plasmids to the mother cell in the absence of a partitioning site was exploited to raise the copy number in a subpopulation of cells. This subpopulation was selected by its enhanced resistance to the antibiotic G418, caused by increased G418$^R$ gene dosage. The stability of the amplified copy number state was characterized by measurements of population mean copy number in the absence of G418, and by measurement of the copy number distribution with flow cytometry. Based on these observations, a simplified conceptual model of conditional centromere plasmid propagation was formulated. The predictions of this model were tested in a segregated mathematical modeling framework, and the simulation results were consistent with the assumptions of the conceptual model for a range of parameter values.

The proposed model for instability of amplified conditional centromere plasmids postulates that the cell possesses insufficient partitioning apparatus to stably partition a large number of centromeres. The resulting biased segregation of excess plasmids results in one subpopulation with even further amplified copy number, which strongly inhibits cell growth, and another subpopulation with copy number equal to the maximum number of plasmids which can be stably partitioned. This type of propagation leads to a rapid drop in population mean copy number in the absence of selection for plasmid amplification.

Another possible explanation for conditional centromere plasmid instability is that many of the plasmids fail to replicate at amplified copy number. Several observations argue against this hypothesis. First, it has been shown that other plasmids in yeast replicate on average once per plasmid per cell cycle. This has
been demonstrated for ARS plasmids(32) and the 2μ plasmid(33), which is maintained at copy numbers from 50 to 100. In order for replication failure to explain the observed instability, the cell would need to be unable to faithfully replicate in excess of 6 plasmids per cell. This is unlikely, in light of the observations just mentioned.

*S. cerevisiae* has 17 chromosomes, with 17 centromeres per cell(46). It possesses approximately 300 origins of replication(47). It seems likely that the molecular machinery required for DNA replication will be present in great excess of the apparatus required for mitotic partitioning. Excess centromere plasmids would therefore titrate out any trans-acting partitioning factors before they exhausted the available replication potential of the cell.

It is possible that the observed propagation is due to structural instability of centromere plasmids at high copy number. This is unlikely, since the plasmids remaining at low copy number do not exhibit gross rearrangements which would eliminate the expression of the URA3, G418\textsuperscript{R}, or lacZ genes. Southern blot hybridization did not reveal a spread of plasmid structures, which might result from extensive plasmid rearrangement or degradation. In general, yeast plasmids are structurally stable; they are packaged with histones in a fashion identical to chromosomal chromatin(48-50).

It is of interest to consider how the information obtained in this study might lead to improved design of *S. cerevisiae* expression systems. Amplified conditional centromere copy number might be stabilized in cells of higher ploidy. Yeast can be stably propagated with one (haploid), two (diploid), three (triploid), four (tetraploid), or many (polyploid) copies of each chromosome. It is possible that cells maintaining a greater load of chromosomal DNA will possess a greater excess capacity for centromere partitioning. The strain used in this study was diploid,
because diploid cells are larger than haploid cells, and it was easier to perform the
flow cytometric assay with larger cells.

Carbon and coworkers have reported the isolation of two mutants which dis-
play the phenotype of greater tolerance for excess cloned centromeres(51). These
mutant strains might be directly useful for conditional centromere plasmid prop-
agation.

The results of this work suggest an alternate type of amplifiable plasmid: a
conditional STB 2μ-based plasmid. By applying the principle of transcription-
induced repression to STB site function, segregation bias could be utilized to
amplify copy number. It is already known that 50-100 STB-site-containing native
2μ plasmids can be stably partitioned by the cell. By overexpressing the trans-
acting REP1 and REP2 2μ gene products from a chromosomally integrated locus,
it might be possible to stabilize in excess of this number of recombinant plasmids.

Cell populations are usually heterogeneous with respect to such cellular state
variables as mass, age, and biochemical composition. Considerable information
is present in the form of this heterogeneity, and this information is lost when
measurements are made of population mean quantities. Measurements of the dis-
tribution of a cellular state variable can be linked to models of single-cell kinetics
by segregated population balance mathematical models. These models are often
of daunting complexity, and not amenable to analytical solution. By simplifying
the conceptual model's assumptions and minimizing the number of parameters, it
was possible in this work to demonstrate consistency of the experimental results
with a proposed model of plasmid propagation. The mathematical model repro-
duced several distinctive features of the experimental results: a rapid drop in copy
number from 25 to 6 in less than 10 generations; a subsequent slow drop from
a copy number of 6 to 2 in the next 50 generations; and a shift in copy number
distribution occurring by growth of a subpopulation at copy number 4-6 at the expense of cells in the tail of the distribution.

Segregated mathematical models of multicopy plasmid propagation can be used to optimize bioreactor operating strategies with respect to plasmid stability and copy number. The effect of manipulating environmental conditions (e.g., selection pressure, plasmid amplification induction) on the copy number distribution, and hence, population productivity, can be predicted with this type of model. The experimental and modeling approach adopted in this study might also be successfully applied to studies of other host-vector systems, such as replicative-runaway bacterial plasmids, and DHFR-linked gene amplification in mammalian cells.

Acknowledgement

I would like to thank Barry Ratzkin and Avantika Patel at Amgen, Inc. for constructing the plasmid pAP251.
REFERENCES


FIGURES

Figure 1. Circular plasmids lacking a partitioning element display a segregation bias in *S. cerevisiae*. At the far left of the figure, one cell contains one plasmid. This plasmid is replicated as a daughter bud forms. When the bud separates from the mother cell, the new plasmid has a high probability of remaining with the mother cell, rather than being partitioned to the daughter cell. If this process is repeated, a small subpopulation with many plasmids is formed, while most of the cells are plasmid-free.

Figure 2. Schematic map of the plasmid pAP251. The individual elements of the plasmid are explained in detail in the text. The plasmid was constructed from pLGSD5 and pAP736 as described in Materials and Methods.

Figure 3. Relationship between copper toxicity and centromere destabilization induced by CUP1 transcription. The solid curve on the left shows the normalized number of cells present in a culture grown for 24 hours in the presence of the copper concentration shown on the abscissa. The dashed line connects points representing the probability of plasmid loss per cell division. This probability is a measure of the repression of centromere partitioning function, and demonstrates that centromere destabilization reflects the expected pattern of induction of CUP1 transcription.

Figure 4. G418 sensitivity of a single-copy pAP251 clone and an amplified pAP251 clone. The normalized number of cells present after 24 hours of exposure to varying G418 concentration is shown. Amplified copy number provides increased resistance to G418 by increasing the number of copies of the G418\(^R\) gene.

Figure 5. Effect of G418 on the level of *lacZ* gene expression from single-copy and amplified copy number cells. The line symbols are the same as those in Figure
4. The ordinate is calibrated in terms of lacZ gene equivalents. This calibration is obtained by comparison with β-galactosidase activity expressed in control strains with 1 or 2 copies of the GAL10-lacZ gene fusion chromosomally integrated. This figure, with Figure 4, provide a rough estimate of 10 for the copy number of the amplified state.

**Figure 6.** Proportionality of β-galactosidase expression to lacZ gene dosage. β-galactosidase expression is induced from the GAL10-lacZ gene fusion by addition of galactose. Population mean β-galactosidase activity was measured spectrophotometrically. The strains with 1 and 2 lacZ copies are D603-i and D603-2i, respectively. The strain with copy number 10 is D603:pLGSD5 (pLGSD5 is a multicopy plasmid).

**Figure 7.** Stability of the amplified copy number state in the absence of G418. At time zero, colonies were taken from YPD-G100 plates and used to inoculate cultures in ura- SD minimal medium. The cultures were continually maintained in exponential growth by serial dilutions. Eight separate cultures were simultaneously performed. β-galactosidase activity was induced and measured in aliquots of the culture at the times indicated. The number of generations of growth was determined by optical density measurements, and the β-galactosidase activity was calibrated in terms of lacZ gene equivalents.

**Figure 8.** Long-term plasmid stability. Three parallel cultures of a single clone were performed as described for Figure 7. The error bars represent one standard deviation.

**Figure 9.** Flow cytometrically measured copy number distributions. The distribution of single-cell β-galactosidase activity was measured, and expressed in terms of lacZ gene equivalents. All three distributions correspond to D603:pAP251 populations, with population mean copy numbers of 1, 6.5, and 13.2 plasmids/cell
estimated by spectrophotometric β-galactosidase assays.

**Figure 10.** Effect of amplified copy number on specific growth rate. Population mean copy number was estimated from β-galactosidase activity assays. The dots represent independent experiments, and the solid line represents the empirical growth model in Equation 4.

**Figure 11.** Effect of replication failure probability $P_R$ on long-term plasmid instability. The data points and error bars are from Figure 8, and the solid lines represent numerical solutions to Equations 2 and 3, with $N_c = 6$ and $P_R$ shown on the figure.

**Figure 12.** Effect of segregation bias probability $P_S$ on short-term plasmid instability. The three curves represent numerical solutions of Equations 2 and 3, with $P_S = 0.8, 0.9, \text{ and } 0.99$, from top to bottom. The other model parameters are: $P_R = 0.02, N_c = 6$.

**Figure 13.** Transient shift in copy number distribution. Equations 2 and 3 were numerically integrated with the initial condition shown in the figure as a dotted line. The equations were integrated to the time point where copy number is 6.5, for comparison with Figure 9. All six possible combinations of the two sets $N_c = (5,6) \text{ and } P_S = (0.8,0.9,0.99)$ were used for the simulations. $P_R$ was fixed at 0.02.
Figure 2.
Figure 3.

PROBABILITY OF PLASMID LOSS PER GENERATION

CuSO₄ (mM)

NUMBER OF CELLS (NORMALIZED)
Figure 4.
Figure 6.
Figure 9.
Figure 12.