THEORETICAL MODELING AND EXPERIMENTAL INVESTIGATION OF HOST-PLASMID INTERACTIONS IN RECOMBINANT *Escherichia coli*

Thesis by

Steven William Peretti

In Partial Fulfillment of the Requirements

of the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1987

(Submitted August 11, 1986)

Acknowledgements

Many people contribute, both directly and indirectly, to such a large undertaking. I would be remiss not to acknowledge those whose influence, while not in the foreground, is incorporated into the fabric upon which others would leave more noticeable impressions. My parents, Frank and Linda, have given me so much for which I am grateful. In providing a supportive, encouraging and challenging environment, they have engendered in me the self-confidence and stamina to consistently bring forth my best effort, and the intellectual curiosity necessary to fuel a lifetime of discovery. My grandfather, John Peretti, directed my attention to science in his own inimitable style. While I was an undergraduate, Professors Daniel Merriman and Charles Walker showed me, by example, the joy and creativity that is part and parcel of both the pursuit of knowledge and its passage from one to the other.

Those who have had a direct impact on my work at Caltech have likewise aided in my professional development, and to them I owe a great debt. I would therefore like to thank Dr. Jay Bailey for his advice, encouragement, prodding and financial support during my graduate career. Thanks also to Dr. Judy Campbell, who donated strains and plasmids, and who made her expertise and that of her co-workers readily available.

Jamie Lee's collaboration was central to the timely completion of the hybridization procedure and analysis of the results which are at the core of Chapter 4. In addition, his irreverent humor made the completion of this work a pleasure and an adventure. Many thanks, Jamie.

My deepest appreciation goes to my wife, Linda. Her tireless understanding, patience and faith in my abilities made it all possible. Caring for and supporting four children and a graduate student is a Herculean task, and she managed to survive with her humor intact.

Finally, thanks to my children: Chris, Kate, Marc and John, for reminding me daily that there is more to life than doctoral work.

Abstract

Microbial metabolism has been mathematically represented with sufficient mechanistic rigor to allow consideration of the ramifications of introducing recombinant vectors on host cell metabolic activity. The model was first verified using details of transcription and translation sufficiently stringent to test the simulation of RNA polymerase equilibrium distribution on promoter regions and the activation of ribosomes by binding of mRNA.

Plasmids were added to the formulation, and the dependence of productivity and growth rate was simulated. The results show good agreement with those obtained by other researchers. This model structure was then used to simulate potentially beneficial metabolic engineering scenarios in an attempt to identify those processes limiting productivity for recombinant systems. Simulation results indicate that transcription is the metabolic bottleneck that limits product synthesis. In addition, the most efficient strategy for enhancing plasmid product synthesis was shown to be increasing the efficiency of plasmid mRNA translation.

An asynchronous population was simulated in order to study the transient behavior of the cell model. Based on plasmid-free simulation results, the model accurately reflects the predominant metabolic structure for control of macromolecular synthesis. Transient response considerations for plasmid-containing populations suggest that dynamic reactor operation, in the form of induction of plasmid promoter activity, leads to transient accumulation of product which exceeds that attained during the subsequent balanced growth.

Finally, the effect of different plasmid copy numbers and of the concommitant expression of a constitutive plasmid-borne gene were investigated experimentally. Radio-labeling techniques combined with filter hybridizations were used to study the transcription rate from the β-lactamase promoter as well as the level of corresponding mRNA present in the cell. Coupled with product activity

measurements, the stability of the plasmid-derived message, as well as the efficiency of its translation into protein are reduced significantly as copy number increases.

TABLE OF CONTENTS

INTRODUCTION

Genetic methods have been developed whereby DNA coding for molecules not normally produced in a host bacterium can be introduced into the cell and foreign genes expressed. Redirecting the biosynthetic activity of the cell allows the production of relatively large quantities of complex macromolecules that either cannot be manufactured using available chemical techniques or whose production or harvest from tissues is prohibitively expensive. This new technology enables the design of gene expression systems. Both the product gene dosage in the host cell and the cloned gene promoter activity are subject to external manipulation. Consequently, the potential exists for the production of an extremely wide range of molecules, far beyond the set of biologically active compounds presently being synthesized. Unfortunately, such metabolic modification can lead to impaired synthetic capacity or cell death. Maximizing productivity for recombinant systems requires more knowledge about the interactions between the cellular machinery and the carriers of the foreign DNA than is currently available. Integrating the information known about cellular subsystems into a coherent representation of metabolism and using that structure to predict the cell's response to perturbations of exponential growth call for a mathematical formulation of the reaction network.

The objectives of this thesis are twofold. These are to establish both the theoretical and experimental framework necessary to rationally approach the problem of optimizing recombinant system performance. Modeling a system with extremely complex, interwoven reaction mechanisms requires a highly structured model of cell metabolism. Chapter 1 describes the formulation of a plasmid-free, single-cell model, based on work done previously by researchers at Cornell (1), and contains comparisons between model simulations and reported experimental results for a wide variety of cell behavior. The model incorporates sufficient detail regarding the processes of transcription and translation to allow subsequent con-

sideration of the plasmid-containing cell without recourse to empirically derived formulations for host-plasmid interactions.

Chapter 2 presents the recombinant system model. The effects of plasmid copy number, cloned gene promoter strength and the ribosome binding site strength of the resulting mRNA on cell growth and cloned gene product formation are simulated, and results compared to available data for similar systems. In addition, metabolic engineering scenarios are presented and evaluated. Simulation results give some indication of processes that are most important to determining productivity and the leverage that parameters have upon system performance. In modeling the recombinant cell, it was necessary to assume certain parameters and control mechanisms were unchanged from the plasmid-free case, despite the significant perturbation imposed by plasmid addition to the cell.

Chapter 3 details a population simulation using an ensemble of single cells. This formulation is shown to simulate observed growth rates. The model was further tested using transient data obtained from shifts in growth rate via medium manipulation and exhibited response curves qualitatively similar to experimental results. In addition, the transient response of recombinant systems is simulated and the potential for productivity enhancement through dynamic bioreactor operation is explored.

Chapter 4 contains experimental findings regarding the effect of plasmid copy number on several key metabolic variables, including RNA polymerase levels, and the rates of transcription and translation. Results indicate that vector insertion and gene expression significantly effect macromolecular stability. High copy number cases exhibited greatly elevated RNA and protein decay rates relative to rates in plasmid-free populations.

Models for microbial systems can be generally categorized as either structured or unstructured. Unstructured models represent the population by a single

variable. Models have been developed which adequately describe cell growth rates for many conditions, including growth at different substrate concentrations, on various carbon sources and in the presence of specific inhibitors. Developed empirically, these models are adequate for representing undifferentiated behavior such as biomass accumulation rates, but insufficient for simulating the complex behavior exhibited by recombinant systems.

Structured models divide the biomass into constituent components and subsystems. A simple example is the case where biomass is divided into two components, carriers of information (DNA and RNA) and structural/enzymatic molecules (proteins). Likewise, total biomass synthesis is divided into equations describing the formation of these two components. Depending upon the type of behavior one wishes to study and the involvement of this behavior with total cell metabolism, the model formulation required to adequately represent the system of interest may be quite complex. Such is the case when one wishes to describe the details of host-vector interactions.

Recombinant systems consist of a host cell and a vector containing foreign genetic material, as shown in Figure 1. Host cells are typically either *Escherichia coli, Bacillus subtilis,* or *Saccharomyces cerevisiae,* although recent development of recombinant DNA techniques for higher eucaryotic systems opens the door to more complex host systems. *Escherichia coli,* the most widely studied microorganism, was chosen as the host system for this work. More is known about the molecular details of its metabolism than for any other organism, making it suitable for detailed metabolic modeling. The vector can take many forms, the most common being phage, cosmid or plasmid DNA. Again, as a result of the wealth of data available for particular vectors, the plasmid ColE1 was chosen as the model vector, and plasmids with similar replication origins were used in the experimental studies.

Essential to the construction of a vector is the presence of a replication origin, a selection gene and a product gene. The replication origin makes possible the autonomous control of vector concentration within the cell, setting the frequency of replication initiation for the plasmid. The selection gene encodes for a molecule whose presence indicates that the cell contains the vector. Selection genes often code for antibiotic resistance or for enzymes that relieve auxotrophies of the host cell. The product gene is chosen at the discretion of the experimenter, subject only to some limitation on total vector size. Proteins currently capable of being produced by recombinant cells include interferon, human growth hormone and insulin.

Introducing extrachromosomal DNA into cells directly affects processes central to metabolism: DNA replication, RNA synthesis (transcription) and protein synthesis (translation). Implicit in any discussion of plasmid replication or plasmid-based gene expression is the fact that cellular machinery must be diverted from host-directed functions to carry out the plasmid-related operations. Using an understanding of the mechanisms and cellular components involved in replication, transcription and translation as a foundation, it is possible to postulate interrelations between observed macroscopic behavior and molecular interactions. Most prior research done regarding plasmid-cell interactions has involved overall cellular growth rate and viability relative to plasmid amplification. The parameters studied previously include media composition, temperature history and the effects of inhibition of protein production and chromosomal DNA replication.

Media composition has been shown to affect both plasmid replication rate and cellular viability, apart from its effect on cellular growth rate. Growth rate is inversely proportional to the rate of plasmid replication (2,3). This is consistent with observations that the frequency of initiation of plasmid replication is the controlled parameter and is independent of growth rate. Clewell (4) found that, at

identical growth rates, removal of glucose from the media caused an increase in the rate of ColE1 plasmid replication. Glucose removal also increased the fraction of plasmids present complexed with the chromosomal DNA. Plasmid-chromosome complexes are believed to be intermediates in plasmid replication. Another anomalous observation is that plasmid amplification leads to cellular death in rich media, but not in minimal media (5).

The effects of amplification on the cell have been generated using three means to increase plasmid copy number: amino acid starvation, chloramphenicol addition or temperature shifting in conjunction with a mutant plasmid. Amino acid starvation (6) or addition of chloramphenicol to the growth medium (7) will inhibit protein synthesis, which stops chromosomal DNA synthesis but does not stop ColE1 replication. Chloramphenicol addition does transiently stimulate ColE1 replication (7) while inhibiting the formation of plasmid-chromosome complexes. Temperature shifts can cause runaway replication of plasmids with a temperature-sensitive replication mutation (8,9).

The lethal effect of plasmid amplification in rich media referred to previously is reversible to a certain extent. During runaway replication by a temperature-sensitive copy number plasmid, cellular growth and protein synthesis continue at a normal (preshift) rate for 4-5 cell doublings (9). This is followed by one cell doubling at a slower rate; then growth ceases (5). The longer the cell is grown under amplifying conditions before being returned to a temperature where copy number is controlled, the greater the lag time between temperature shift and the attainment of steady-state growth (5). The growth rate upon return to lower temperatures also decreases with increasing incubation time at amplifying temperatures. If the cells are shifted after 4-5 cell doublings at the pre-shift rate have occurred, they are unable to resume growth (5). Obviously, a balance must be struck between increasing gene copies and decreasing synthetic capacity.

Perhaps the most interesting amplification data from a modeling and optimization viewpoint are in Table 1. Uhlin and coworkers (9) were studying β-lactamase expression from a temperature amplifiable plasmid in *E. coli*. As can be expected, the β-lactamase activity increases as amplification progresses. Higher copy number should generate more transcripts, leading to more gene expression. However, the increase in β-lactamase activity after a temperature reduction and the resultant drop in plasmid replication rate suggests that optimal productivity involves more than simply unrestricted amplification.

Much of what has been observed may be caused by competition for the required enzymes between the plasmid and chromosome. This is suggested by the observation that the percentage of DNA present as plasmid DNA after full amplification is constant regardless of the size of the plasmid (9). This could occur by titration, on a mass basis, of some factor required for replication by both plasmid and cell. A possible candidate is RNA polymerase.

Plasmid presence in a cell may have a more direct effect on the chromosomal DNA. Complexes of chromosomal and plasmid DNA have been observed (10) and replicative intermediates of Col E1 are found almost exclusively associated with the chromosome (11,12). Amplification does not alter the ratio of bound to unbound plasmid, so up to 500 plasmids have been found bound to the *E. coli* chromosome (10). This may cause distortion or blockage of replication and transcription sites on the chromosome.

Plasmid amplification might also affect the distribution of genes that get transcribed by increasing the copy number of the plasmid-encoded gene. However, at any growth rate, less than 50% of the total RNA polymerase present in the cells is active (13-15). Most of the non-functioning RNA polymerase is bound to the DNA through non-specific binding by the β′ subunit (13). It may associate non-specifically with DNA, then bind the sigma subunit and remain bound to DNA

after the sigma subunit is released. It has been suggested (15) that the rate-limiting step in transcription is either the release of the product or the liberation of the polymerase from the DNA. If either is the case, much of the inactive polymerase might be unavailable for use elsewhere. If so, transcription of chromosomal genes could be limited as the amount of plasmid DNA available for non-specific binding approaches the amount of chromosomal DNA present.

Control of translation can reduce or amplify the effect of excessive mRNA produced from plasmid-encoded genes. There is no obligate coupling between transcription and translation. The number of times that a specific mRNA functions as a template for protein synthesis is under metabolic control and will increase if the demand for protein increases faster than the increase in supply of the mRNA template (16). Also, regardless of growth rate, an increase in temperature increases the synthesis of RNA polymerase (17).

Treating plasmid amplification and the resulting overexpression of the plasmid-carried gene as starvation of the cell, the ill effects of plasmid presence will be magnified. Plasmids may increase the cell's requirements for specific ions (18) or decrease their host's growth rate (19). Furthermore, starvation, or a reduction in growth rate, stimulates the selective repression of the transcription of ribosomal and transfer RNA genes, limiting the amount of protein synthetic machinery available (20-23). Reduction in growth rate also slows the synthesis rate of the sigma subunit of RNA polymerase, probably lowering the fraction of total polymerase which is active (22).

There are two more aspects of gene expression that deserve mention, promoter strength and ribosomal binding to mRNA. Frequency of transcription is related to the strength of the promoter placed in front of the gene. Maximum transcription rates are attainable using the phage lambda promoter p_L (24-26), or the lac promoter (27,28). Without a strong promoter in place, gene expression will

be at low levels. Efficient translation requires a ribosome binding site on the mRNA. This site includes the start codon and a Shine-Dalgarno [S-D] sequence (29), which varies in length from 3 to 9 bases. S-D sequences are complementary to bases on the 3′ end of 16S ribosomal RNA, which is contained in the 30S ribosomal subunit. It has been shown that the makeup of the ribosome binding site affects the relative extent of mRNA translation.

These genetic factors are directly accounted for in the mathematical model described herein. In addition, cell composition and the macromolecular synthesis rates are explicitly calculated. Experiments involving these parameters give an indication of the nature of the cell's response to the additional burden imposed by the presence of plasmid in the cell.

REFERENCES

1. M.M. Domach, S. Leung, R.E. Cahn, G.G. Cocks and M.L. Shuler, *Biotechnol. Bioeng.*, **26**, 203 (1984).

2. B. Engberg and K. Nordstrom, *J. Bacteriol.*, **123**, 179 (1975).

3. P. Gustafsson and K. Nordstrom, *J. Bacteriol.*, **141**, 106 (1980).

4. D.B. Clewell, *J. Bacteriol.*, **110**, 1135 (1972).

5. B.E. Uhlin and K. Nordstrom, *Mol. Gen. Genet.*, **165**, 167 (1978).

6. M. Bazaral and D.R. Helinski, *Biochemistry*, **9**, 399 (1970).

7. D.B. Clewell, *J. Bacteriol.*, **110**, 667 (1972).

8. P. Gustafasson and K. Nordstrom, *Plasmid*, **1**, 134 (1977).

9. B.E. Uhlin, S. Molin, P. Gustafsson and K. Nordstrom, *Gene*, **6**, 91 (1979).

10. B.C. Kline, J.R. Miller, D.E. Cress, M. Wlodarczyk, J.J. Manis and M.R. Otten, *J. Bacteriol.*, **127**, 881 (1976).

11. J.R. Miller, J.J. Manis. B.C. Kline and A. Bishop, *Plasmid*, **1**, 273 (1977).

12. P. Gustafasson and K. Nordstrom, *J. Bacteriol.*, **123**, 443 (1975).

13. N.S. Shepherd, G. Churchward and H. Bremer, *J. Bacteriol.*, **141**, 1098 (1980).

14. H. Matzura, B.S. Hansen and J. Zeuthen, *J. Mol. Biol.*, **74**, 9 (1973).

15. N. Shimamoto and C.-W. Wu, *Biochemistry*, **19**, 842 (1980).

16. D. P. Nierlich, *J. Mol. Biol.*, **72**, 765 (1972).

17. Y. Nakamura and T. Yura, *J. Mol. Biol.*, **97**, 621 (1975).

18. R.M.M. Klemperer, N.T.A. J. Ismail and M.R. W. Brown, *J. Gen. Microbiol.*, **115**, 325 (1979).

19. J.W. Dale and J.T. Smith, *Ant. von Leeuwen.*, **45**, 103 (1979).

20. R.A. Lazzarini and A.E. Dahlberg, *J. Biol. Chem.*, **246**, 420 (1971).

21. H. Brunschede, T.L. Dove and H. Bremer, *J. Bacteriol.*, **129**, 1020 (1977).

22. Y. Iwakura and A. Ishihama, *Molec. Gen. Genet.*, **142**, 67 (1975).

23. N.S. Shepherd, G. Churchward and H. Bremer, *J. Bacteriol.*, **143**, 1332 (1980).

24. R.N. Rao and S.G. Rogers, *Gene.*, **3**, 247 (1978).

25. J. Hedgepeth, M. Ballivet and H. Eisen, *Molec. Gen. Genet.*, **163**, 197 (1978).

26. H.U. Bernard, E. Remaut, M.V. Hershfield, H.K. Das, D.R. Helinski, C. Yanofsky and N. Franklin, *Gene.*, **5**, 59 (1979).

27. K. Backman and M. Ptashne, *Cell.*, **13**, 65 (1980).

28. L. Guarente, G. Lauer, T.M. Roberts and M. Ptashne, *Cell*, **20**, 543 (1980).

29. J. Shine and L. Dalgarno, *Proc. Nat. Acad. Sci. U.S.A.*, 1342 (1974).

CHAPTER 1

A Mechanistically Detailed Model of Cellular Metabolism for Glucose-Limited Growth of *Escherichia coli* B/r-A

INTRODUCTION

Cellular metabolism is a complex network of interdependent reactions. Its control systems are highly interwoven, contain multitiered feedback loops and regulate catalyst concentration, activity, and selectivity. The complexity is such that, in order to understand fully the effect of perturbations to the system, detailed representation of numerous reaction pathways is necessary. *E. coli*, for example, contains genetic material sufficient to direct synthesis of ca. 3000 proteins and the number of possible reactions is of the same order of magnitude. Reducing a system of this size to a manageable set of material balances that accurately reflects metabolic behavior is a difficult task. It is necessary to identify the important pathways and control systems involved in the processes one wishes to represent. The resultant model will be as detailed as the processes being studies. "Successful" models may be as simple as the Monod equation for biomass increase as a function of limiting substrate concentration or as complex as the single-cell models for *E. coli* growth developed at Cornell (1-3).

Much contemporary biological research involves the elucidation of mechanisms resulting in observed metabolic behavior or clarification of control system structure for very specific pathways. This work is pivotal to understanding observed responses to mutations which may block or modify control pathways or which can alter enzymatic activity and specificity. Of similar importance is determining the overall cellular response to metabolic imbalances and alterations. This requires integration of the models for cellular subsystems into a dynamically interacting formulation of cellular metabolism. Development of such a model depends upon successful identification of critical parameters and relationships and the exclusion of extraneous detail. The complexity of even the simplest bacterium and the nonlinearity of the control mechanisms necessitate the use of computers to solve the system model and to verify proposed subsystem integration.

Structured models resulting from such an effort can be further used as experimental tools to verify subsequently proposed reaction mechanisms and control relationships. Due to the nonlinearity and simultaneous multicomponent nature of the pathway interactions, intuitive, qualitative reasoning is often inadequate to identify and evaluate coherently the possible ramification of metabolic disturbances upon particular reactions or overall cellular activity. Again, structured models provide a means to analyze systematically possible interactions and responses. These models can thereby aid in hypothesis testing and efficient experimental design.

The major objectives of this work are the formulation and verification of a structured, single-cell model for *E. coli* with sufficient reaction rate and control system rigor to allow metabolic subsystem integration and testing with minimal base model alteration. While model accuracy is desired for gross measures of cell behavior such as doubling time and cell yield, the emphasis of this work is on the details of the system response. Adding accurate structural rigor that simulates finer levels of cellular activity provides the framework for the systematic study of the plausibility of different hypothetical kinetic and control schemes. For example, with such a model it is possible to simulate the cell response to mutations affecting components involved in the control relationships and kinetic mechanisms included in the model. Other behavior amenable to such treatment includes dynamic response to glycolytic oscillation, secondary metabolite synthesis, cell septation, cell division, and metabolic distortion due to recombinant vector insertion.

Such mechanistic, structured models require the inclusion of many of the fundamental processes involved in cellular metabolism in correct reaction sequences. When simulating progressively more complicated, detailed behavior, we find it less and less justifiable to identify *a priori* which pathways or control systems will be rate limiting. Also, extensive lumping of metabolic pathways is no longer

appropriate since perturbations of cellular processes or intracellular and extracellular conditions may uncouple otherwise coordinated control systems and responses. For this reason, the model structure should include mechanistically accurate representations for DNA replication, gene transcription, and translation of messenger RNA. In particular, transcription must be based on the distribution of active RNA polymerase on mRNA and sRNA promoters. Simulation of translation should be based on the distribution of ribosomes on different mRNAs. Furthermore, the control system for initiation of chromosomal replication should be dependent on some indicators of overall metabolic activity. Taking these considerations into account, we chose the most recent version (1) of the model developed at Cornell as a basis for further development, because it already included a plausible mechanism for chromosomal replication.

In this chapter a single-cell model is described, which is capable of predicting, at the most fundamental levels, aspects of cellular metabolism crucial to the understanding of bacterial growth regulation and the cell's response to perturbations of intracellular and extracellular conditions. RNA polymerase metabolism is accounted for and its interactions with chromosomal DNA are explicitly calculated. In addition, the activation of free ribosomes has been successfully represented by the binding of messenger RNA prior to the initiation of translation.

MODEL DEVELOPMENT

In order to maximize the predictive power of the model, development proceeded using as few adjustable parameters as possible. Almost all parameters were determined by reference to independent experiments. Control systems that were included were based on known mechanisms demonstrated through either *in vitro* or *in vivo* experiments. Figure 1 illustrates the general structure of the model. All nutrients with the exception of nitrogen and carbon sources are assumed in excess

and readily available as reactants. Waste product and inhibitor accumulation in the cell is presumed minimal and has no effect upon reaction rates. Reactant dependencies of the kinetics of individual model reactions are assumed saturating. The adequacy of respreesenting reactions using saturation kinetics has been well documented for most cellular reactions. Table I presents the mass balance equations, Table II the parameter values. Most of these equations are similar in form if not identical to those found in the paper by Domach et al (1). However, there are substantial alterations and additions regarding protein and RNA synthesis, initiation of translation and distribution of RNA polymerase along DNA, and chromosomal replication initiation. These extensions are next described, documented and evaluated in detail.

TRANSCRIPTION

Initiation

The most extensive addition to the Cornell model (1) is the explicit accounting for RNA polymerase equilibrium binding distribution. The current model formulation accounts for the three general types of polymerase-DNA binding, nonspecific core enzyme binding and holoenzyme binding to promoter and nonpromoter sites. These possible polymerase configurations are indicated in Figure 2. Sigma subunit attachment to core polymerase causes structural alterations in another of the enzyme's subunits. The resulting holoenzyme gains selectivity for transcriptional promoter sites. In addition, non-specific holoenzyme binding decreases vis-a-vis core polymerase binding. Binding of any kind is assumed to block approximately 40 base pairs, making them unavailable for nonspecific binding. RNA polymerase holoenzyme moves along double-stranded DNA through a series of rapid, cyclic bind-release steps until a promoter site is encountered. Attachment at these sites is tighter than non-specific binding and leads to local

unwinding of the DNA strands and initiation of transcription. Initiation rate is limited by the time required by RNA polymerase to find a promoter site via the "hopping" mechanism described (4-6).

Polymerase movement following initiation quickly vacates the promoter site, leaving it available to other polymerase molecules for further binding. Therefore, the equilibrium distribution of polymerase among various DNA sites is calculated assuming that the concentration of available promoter sites (DNAM and DNAR in Equations 12 and 14, Table III) remains constant regardless of the level of polymerase binding to that particular promoter. Core enzyme binding is considered completely nonspecific so that any nucleotide sequence is available as a binding site. The same is assumed for holoenzyme-non-promoter binding.

Additionally, two distinct promoter types are used in the equilibrium calculations, those for stable RNA operons and for an "average" messenger RNA operon. The sRNA operons have the stronger promoter, as has been noted experimentally (7,8). Messenger RNA producing operons (m-operons) are assumed to be evenly distributed along the chromosome, whereas the sRNA operons (s-operons) are individually accounted for, and their number is incremented as the replication fork passes their position on the chromosome.

The functional dependences for polymerase binding are taken from many sources. Binding affinity and its dependence upon mono- and di- valent ions is treated theoretically by Record et al (9). Experimental determination of parameters for ionic dependence as well as for temperature and pH effects has also been reported (10-13). Table III contains the resulting equations for polymerase binding as a function of mono- and di- valent ion concentration and pH. Correlations for temperature dependence are included as are the equilibrium relationships for RNA polymerase.

Furthermore, holoenzyme affinities for m-operon and s-operon promoters are affected by the intracellular concentration of the nucleotide 5′, 3′-guanosine tetraphosphate (PG) and by the rate of protein synthesis. PG is produced in a reaction involving idling ribosomes and decreases the affinity of RNA polymerase for s-operon promoters (14-16), reducing ribosome production. The effect of protein synthesis rate is indirect. Polymerase specificity has been shown to be affected by translation initiation factor (17) and f-met tRNA (18,19), the transfer RNA involved in the initiation of protein synthesis. Free initiation factor (IF2) stimulates s-operon transcription, leading to an increase in cellular protein synthetic capacity. Conversely, excess initiator tRNA signals underutilization of the ribosomes and stimulates m-operon transcription. The coordinated control of RNA polymerase specificity by PG and protein synthetic cofactors is illustrated in Figure 3. These feedback mechanisms allow the cell continually to monitor the capacity and efficiency of its protein synthesis machinery and to adjust the levels of message and ribosome accordingly.

Messenger RNA promoter concentration increases in this model as the replicating fork position advances, with the promoters uniformly distributed along the chromosome. In addition, messenger RNA synthesis is dependent on the distribution of RNA polymerase upon cellular operon promoters and the rate of ribonucleotide addition attainable by RNA polymerase. This is in contrast to the Cornell model, which evaluates mRNA synthesis assuming proportionality to protein synthesis.

Synthesis of stable RNA is likewise determined by the distribution of RNA polymerase upon sRNA promoters. The location of the relevant operons and the progress of chromosomal replication determine the number of promoters found in the cell, thereby affecting RNA synthesis. The operons containing genes for stable RNA (7 rRNA operons represented) are found at 56, 72, 74, 85, 86, 89 and 90

minutes on a 100 minute map of the *E. coli* chromosome. The chromosome undergoes bidirectional replication from an origin located at 84 minutes, so the sRNA operons are found at the following replication fork positions, normalized with respect to the replication origin-terminus distance: 0.02, 0.04, 0.10, 0.12, 0.20, 0.24, and 0.56. Each operon is preceded by two promoters, P1 and P2. These tandem promoters exhibit markedly different strengths and different responses to changes in cellular metabolic activity (7, 8, 54). Specifically, the upstream promoter (P1) has higher transcriptional activity than the downstream promoter (P2) under normal growth conditions. However, P1 is subject to stringent regulation, making its activity sensitive to ppGpp concentration. P2 is free from this repression. This parallel control system allows the dominant promoter to change and the expression of these operons to continue regardless of the particular configuration of cellular variables.

Elongation

During model development, the results of simulations based on extracellular glucose concentrations, which experimentally result in doubling times near one hour, indicated mRNA levels much higher than those found *in vivo* for the chosen glucose levels. Overproduction of mRNA occurred despite the concurrent calculation of correct RNA polymerase content, fraction of polymerase actively transcribing and the correct pattern of RNA polymerase distribution onto m-operon promoters. Also, the chain elongation rate used in the calculation of mRNA synthesis was the same as that reported in the literature (20). A possible explanation of this result is the attenuation or premature termination of transcription. These phenomena are known to occur for some operons and are accounted for by the terms E_{ts} and E_{tm} in Equations 8a and 10a found in Table I. These terms signify the efficiency of transcription of stable and messenger RNA, respectively. Reduc-

tion of transcription level may occur as a result of abortive transcription initiation or transcription attenuation within mRNA operons. Attenuation and termination result from RNA polymerase that pauses during transcription. Such transcriptional stalling of polymerase for kinetically significant periods of time at certain sites along the chromosome has been observed and , in fact, RNA polymerase often spends more time pausing than transcribing (21). An efficiency near 0.5 for transcription is more likely to give correct cellular RNA concentration than efficiencies near unity.

An E_{tm} value of 0.43 was found to give good agreement for specific growth rates near $\mu = 1.0\ hr^{-1}$. Increasing negative deviation from experimental results for mRNA levels was noted for growth conditions yielding $\mu < 0.6\ hr^{-1}$. As before, the incorrect mRNA level resulted despite attainment of experimentally verified RNA polymerase level, activity and pattern of distribution. These results are reasonable if, instead of remaining constant, the fraction of potential transcription undergoing attenuation decreases as growth rates decline.

RNA polymerase pausing and release from the template are mediated in *E. coli* by at least two proteins, rho and nusA. Each inhibits transcriptional efficiency (22-25). In the absence of any prevailing hypothesis regarding control of transcriptional efficiency, the following regulatory process is proposed: Attenuation can be thought of as the result of a bimolecular reaction between an actively elongating transcript and an effector protein. The level of attenuation is therefore proportional to the ratio of effector protein to target site. In the absence of detailed information regarding the synthesis of rho and nusA protein, presume that these proteins are a constant fraction of the total cell protein. The target sites are stretches of DNA containing specific nucleotide sequences. We assume these sequences to be evenly distributed along the chromosome. A measure of the coverage of the effector proteins onto target sites can then be approximated by the protein/DNA

ratio. As a result, the mass ratio of total protein to chromosomal DNA mass (M1/M3) was used in calculating the attenuation of mRNA transcription (Eq. 10b). While more complex control schemes can easily be envisioned, limited understanding of the mechanisms surrounding the phenomena which comprise attenuation precludes the inclusion of a more complicated control scheme.

The simulations that indicated substantial overproduction of mRNA described above also resulted in slightly elevated levels of stable RNA. As with mRNA, it was first assumed that the attenuation was a constant fraction of the potential transcription, and such treatment was found to be adequate for all subsequent simulation. The value of E_{ts} is much higher than those calculated for E_{tm}. As mentioned above, E_{tm} is near 0.5 (0.4-0.6), whereas an E_{ts} value of 0.98 gives good simulation results. This fraction is well within the margin of error for the experimental determination of the RNA chain elongation rate and may represent nothing more than the error in the estimate. Of greater interest is the obvious disparity of both treatment and magnitude between attenuation for mRNA and sRNA operons, which will be further addressed in the discussion section.

TRANSLATION

Initiation

Ribosome formation depends upon synthesis of rRNA. Once formed, ribosomes are assumed to undergo equilibrium-based binding with mRNA, leading to the onset of protein synthesis. The importance of complementarity between the 3′ terminal sequence of 16S rRNA and ribosome binding sequences on mRNA was recognized by Shine and Dalgarno (26) a decade ago. More recent work supports the conclusion that S-D sequences are not sufficient to indicate a ribosome binding site (26). Analyses of cellular mRNA ribosome binding sites with measures of their relative strengths (27-30) indicate that the secondary structure near the initiator

codon has a major effect upon the relative binding strength. Binding sites flanked by stem and loop structures in the 5′ direction are more efficient translation initiators. However, for those messages that have some stem-loop structure, initiation rates correlate with the binding constant of the Shine-Dalgarno site.

Assuming that most cellular mRNA contain stem-loop structures, which has been documented for many messages (27), we determined an average binding site equilibrium constant using base-pair hydrogen bonding for available S-D sequences. Since base pairing between mRNA and rRNA prior to polypeptide elongation has been shown to be a significant factor in determining the relative distribution of mRNA species bound to active ribosomes (31), this model includes equilibrium calculations of 30S ribosomal binding to different classes of mRNA transcripts. This equilibrium is represented as;

$$RIBO + mRNA <=> RM \qquad \text{(R1)},$$

with RM representing an active ribosome-mRNA complex. The active complex concentration can be calculated using the equation;

$$Keq = [RM]/([RO-RM]*[MO-RM]) \qquad \text{(R2)},$$

where square brackets indicate concentrations, RO and MO are total concentrations of ribosome and mRNA transcripts, and Keq is the equilibrium constant previously described. One simplification is introduced to these calculations. Ribosome complexed to message subsequently conducts translation, freeing the ribosome binding site on that message for interaction with another ribosome. Repetition of this sequence leads to the well-documented (48) situation of simultaneous translation of a message by multiple ribosomes. Therefore, it is reasonable to treat the concentation of ribosome binding sites as the total mRNA concentration, given

$$Keq = [RM]/(MO*[RO-RM]) \qquad \text{(R3)},$$

which can be rearranged to give an expression for RM explicitly.

$$RM = Keq^{*}MO^{*}RO/(1 + Keq^{*}MO]) \qquad \text{(R4)}.$$

This equation is used to calculate the fraction of ribosomes active in translation ($FRACM = RM/RO$) found in Table I, Equation 7a.

Multiple attachment of ribosomes to mRNA occurs as a result of the relative rates of the steps that comprise translation initiation and elongation. Ribosome activation is a multistep sequence, as proposed by Gualerzi and co-workers (31) and illustrated in Figure 4. In the mechanism shown, the limiting step is the irreversible addition of a 50s ribosomal subunit to the $mRNA - tRNA - 30s$ complex, not mRNA binding. Subsequent polypeptide chain elongation occurs very rapidly, freeing the ribosome binding site for subsequent ribosome attachments. This approach correctly calculates ribosome activation under a wide range of growth conditions.

Synthesis and Degradation

Mechanistically, protein synthesis kinetics (Eq. 7a) is now correctly formulated. Cellular translational activity depends upon the availability of mRNA transcripts and of non-translating ribosomes. Once the initiation complex is formed, polypeptide chain elongation is assumed to proceed at a constant rate, modulated only by the availability of precursors and an energy source. These effects follow saturation kinetics, as indicated by the terms containing amino acids (P1) and glucose (A2) in Equation 7a.

Proteolysis has been completely reworked relative to the Cornell model. Pine (33) has reported two kinetically distinct degradative reactions. Immediately following translation, approximately 5% of the newly made protein is degraded. This is probably the result of maturative cleavage and is represented by the term 0.05 DM1S in Equation 7c. Furthermore, work done in Goldberg's lab (34, 35) indicates a complex dependence of proteolysis upon intracellular concentrations of PG,

ATP and amino acid. In studies of the effect of PG upon proteolysis, it has been found that increasing levels stimulate protein degradation. This effect is modulated by amino acid availability. Starvation for amino acids results in higher levels of PG than glucose starvation but gives lower proteolysis rates. Cellular energy level also affects the rate of protein decay. Reducing ATP levels up to 80% has no direct effect upon decay rates, but reduction beyond that point leads to inhibition of proteolysis. The representation for these dependences involves splitting the process into energy dependent and independent processes, with the PG sensitive process independent of cellular energy levels (Eq. 7b).

MACROMOLECULAR PRECURSOR METABOLISM

It was also necessary to revise deoxyribonucleotide synthesis relative to the Cornell model, primarily in recognition of the possible ramifications of extrachromosomal DNA addition to the model. Deoxyribonucleotide synthesis is catalyzed by the enzyme ribonucleotide diphosphate reductase (rdr). This enzyme is involved in the reduction of all ribonucleotides, each with different rate and saturation parameters. Control of rdr synthesis is poorly understood. Experiments indicate a control pattern resulting in a correlation between rdr synthesis and the ratio of intracellular protein to DNA (36, 37). If DNA synthesis is blocked but protein production is unaffected, eventually rdr synthesis will be induced. The induction is not a burst of productivity but rather a steady, elevated level of synthesis requiring continued protein formation for its expression. As indicated in Equation 14, rdr synthesis is dependent upon net protein synthesis and the ratio protein/DNA.

Activity and specificity of rdr are allosterically affected by ATP, dATP and dTTP (38-40). The enzyme is a dimer, containing two sites each for the binding of substrate and effector. Catalytic activity is inhibited by dATP binding at the effector sites. ATP is capable of displacing dATP from these sites, giving enhanced

activity relative to the unbound state. In the absence of effector site binding, dTTP increases catalytic activity and purine specificity. However, the presence of either ATP or dATP in the effector site in conjunction with dTTP binding inhibits activity. Since the dTTP effect is poorly studied, rdr activity is calculated using ATP induction and dATP inhibition, as indicated in the formulation of deoxyribonucleotide synthesis, Equation 5 in Table I. In that equation, ADP replaces ATP to distinguish this phenomenon's dependence upon ATP as a nucleotide rather than as an energy source.

DNA REPLICATION INITIATION

Finally, the equation used to calculate the synthesis of the anti-repressor protein (ARP) responsible for the derepression of replication initiation has been altered. When the rate of DNA synthesis in *E. coli* is insufficient for the growth conditions encountered by the cell, the initiation frequency of DNA synthesis increases. This effect has been observed for DNA synthesis rate reduced via thymine auxotrophy (41), nalidixic acid treatment (42), and dnaB mutation (36). Filpula and Fuchs (36) hypothesize that, as protein synthesis continues in the absence of DNA synthesis, the rate of production of a compound that triggers DNA replication increases. We have successfully incorporated this hypothesis into our model by making ARP synthesis dependent upon the intracellular protein/DNA ratio (Eq. 22). This term replaces one involving intracellular amino acid concentration found in the Cornell formulation.

This is the extent of the major alterations to the model developed at Cornell. Discussion of the rationale behind the formulations for macromolecular precursor metabolism, energy accounting, parameter estimations and cell morphology can be found in descriptions of the base model and its predecessor (1-3). The justification and sources of estimates of new parameters are outlined in Table IV.

COMPARISON OF MODEL SIMULATIONS TO EXPERIMENT

When the performance of simulations are evaluated, multiple levels of verification are often possible. The level of verification possible depends upon the modeling goals as well as on the model structure. Unstructured models can apply only to cumulative aspects of cellular metabolism, such as doubling time or cell yield, and verification of these models is limited to such overall behavior. Structured models may explicitly represent many control schemes and component metabolisms, but again, application is limited to the expressed goals of the modeling effort. The Cornell models provide good examples of this point. The expressed goals of those efforts were to accurately predict general cellular composition (total protein, total RNA), cell morphology, DNA initiation timing, cell growth rate and cell yield for balanced growth under glucose- (1) or nitrogen-limited (56) conditions. Verification of the model is therefore limited to those conditions and to corresponding levels of complexity. Those goals were achieved despite the non-mechanistic formulation of various component synthesis kinetics and control schemes.

Model success in the absence of a mechanistic framework is not surprising, given the nature of the behavior used to verify the model. Variables such as total cell protein or RNA levels, growth rate and cell yield are integrated model responses. That is, they are the result of integration of the model equations over many successive steps. Since the result of such integration gives no indication of the trajectory of the system reponse, correlation of simulation results with experimental data indicates only that the integral of the actual process and the model formulation are the same. It does not validate the details of the formulated kinetics and control schemes. That can be accomplished only by correspondence of transient behavior between simulation and experimental results. DNA initiation timing is a transient activity, and its accurate prediction is a sensitive indicator of

the plausibility of the relevant kinetic mechanism and control scheme embodied in the model. Other cellular behavior of this type includes RNA polymerase activation and distribution, as well as ribosome activation.

Indicators of the suitability of models for cellular behavior are as numerous as the varied cellular responses catalogued. At the total system level, indicators include the specific growth rate dependence upon limiting substrate concentration, the growth yield and morphological features of the cell. In Figure 5a, model predictions for cell volume are compared to experimental data over a range of glucose-limited growth conditions. The model predictions fall between data obtained using a Coulter Counter and that using the more traditional technique of electron microscope observation of cell dimensions following osmium tetroxide fixation. Larger volumes are obtained following fixation, probably due to swelling. The agreement with experimental results is striking, particularly in light of the fact that no model parameters were chosen with reference to volumetric data.

Figure 5b shows a Lineweaver-Burke plot for specific growth-rate dependence on limiting substrate concentration. Model predictions for the Cornell formulation as well as the model herein described are compared to experimental data. The simulation results are consistent with the data presented. There is systematic deviation for each model. This is reasonable in light of the previous argument regarding model verification. No effort was made in either model to account completely for all aspects of cellular energetics. Instead, known energetic usage was accounted for, as was some consideration of non-specific use. Cell energetics as currently formulated is obviously either somewhat incomplete or inaccurate. However, it is not the goal of this work to address such discrepancies.

Figure 5c indicates that the observed growth yields predicted by the cell model are somewhat lower (~10%) than those observed experimentally by Cornell researchers (1-3). These results can be further analyzed using the relationship

developed by Pirt (44):

$$(1/Y_{glu})obs = (1/Y_{glu})max + m_{glu}/u\,. \qquad \text{(A)}$$

The solid line in Figure 5c represents a least-squares regression fit of the model predictions. Combined with Equation A, we get $(Y_{glu})max = 0.50$ and $m_{glu} = 0.25$. Data from aerobically grown cells at 30 C give parameter values in the range 0.36-0.52 gram cell/gram glucose consumed for $(Y_{glu})max$ (45-47) and 0.04-0.06 gram glucose/gram cell hr for m_{glu} (44-46). Data for cells grown at 37 C (ref. 51) indicate m_{glu} values as high as 0.25. The values derived from simulations are comparable to those predicted using the Cornell formulation. It would appear that the modifications and additions to that model which are herein described have little impact upon cellular energetics.

The cell volume, growth rate and yield indicate on a superficial level the appropriateness of the model. A more refined standard would be the cellular composition, specifically the levels of protein, messenger RNA and stable RNA found in the cell at various growth rates.

Figures 6a-c illustrate comparisons between model predictions and actual data for intracellular levels of protein, sRNA and mRNA. The fit for all of these components is quite good. The cellular ribosome fraction is directly proportional to the amount of stable RNA present, so these results indicate the accurate prediction of the protein synthetic capacity of the cell. As previously discussed, correlation for such integrated variables is not sufficient to validate the underlying kinetic and control structure. However, this formulaion utilized experimentally verified kinetic mechanisms and correctly calculates the transient formation of translation initiation complexes. The combination of accurate prediction of mRNA levels, ribosome activation by complexing with available mRNA, and protein levels strongly supports the assertion that this model includes correctly identified and formulated

control and mechanistic detail regarding protein synthesis. The agreement for both RNA types also suggests that the scheme used to calculate the distribution of RNA polymerase may be quantitatively as well as qualitatively correct, though as was described previously, cumulative performance does not prove mechanistic rigor.

The inherent structure of the model allows simulation of very detailed aspects of cellular metabolism, parameters which are often difficult to measure experimentally. In addition, these detailed aspects give further indication of the appropriateness of the mechanisms that have been added to the model formulation. In particular, there are two parameters which are extremely important in evaluating the extension of this model to include other cellular phenomenon. These are the distribution of RNA polymerase and the duration of the C and D periods pertaining to DNA replication.

As described earlier, RNA polymerase selectivity for various promoters changes as a function of the overall growth rate. Specifically, m-operon promoters are favored relative to s-operons as growth rate declines. This expands the capacity of the cell to respond quickly to fluctuations in its environment. Figure 7 shows the fraction of RNA polymerase actively transcribing s-operons as a function of growth rate. The model predictions are in excellent agreement with experimental data. This is a sensitive indicator of the accuracy of the equilibrium calculations and the underlying assumptions regarding the calculations.

Figure 8 illustrates the C and D period estimates of this cell model and of the Cornell model, as well as experimental data reported by Helmstetter and Pierucci (49). There has been little change in the estimates using this model relative to the Cornell model. This aspect of the model performance will be particularily important for extensions of the model, which include extrachromosomal DNA, such as phage infection or plasmid introduction and maintenance.

DISCUSSION

It has been shown that this formulation of a single-cell model for *E. coli* accurately predicts cell performance for a range of experimental conditions, most of which played no part in the parameter estimation process. This model was formulated, however, with the expectation that it would serve as a basis for further investigation of detailed aspects of cellular metabolism. It has already served that function with regard to the formulation for transcriptional efficiency of mRNA operons and the disparity between the treatment of mRNA and sRNA transcription efficiency. As was described previously, E_{ts} can adequately be assumed to be constant over a wide range of intracellular conditions, whereas E_{tm} cannot likewise be treated. The model has accurately estimated both mRNA and sRNA cellular levels using mechanistically correct kinetic formulations, so this discrepancy may be real. However, since average cellular levels for any component are the results of integration over the entire cell cycle, there is no certainty that the mechanism formulated is fully representative for the cellular activity involved. The indicated discrepancy in transcription efficiency suggests cellular regulation differences not previously considered. Both transcription inhibitors nusA and rho bind to RNA transcripts and move toward the transcription complex. RNA secondary structure influences protein binding. If the sRNA operons produce transcripts that are relatively free of attenuation, then sRNA may have significantly different structure from mRNA, particularly with regard to nusA and rho protein binding affinity. An alternative is that, while attenuation of mRNA operons is controlled by inhibitor-bound pausing of the transcription complex, attenuation for the sRNA operons is under the control of DNA secondary structure, which may also cause RNA polymerase to pause at specific sites. Another possibility, mentioned during the discussion of RNA transcript elongation, is that the attenuation efficiency empirically determined for transcription from stable RNA operons is an artifact of

the chain elongation rate chosen for this formulation. This possibility does not alter the conclusion that m-operons and s-operons are significantly different with reference to attenuation of transcription from their respective promoters. Further mechanistic structure would be required to test these alternatives, but even in its present form, the model suggests new avenues for investigation.

Recently, results from experiments that address the question of transcription efficiency in sRNA operons have been reported (50). While the operon studied was found to have rho protein-dependent termination sites, no rho mediated transcription attenuation was exhibited. It is proposed that *E. coli* contains an antitermination system similar to that found for phage lambda. The proposed system would include an antiterminator protein acting at a control site unique to sRNA operons. The latest study of the question of transcription from rRNA promoters, published as this manuscript was being prepared (55), involved monitoring the effect or rRNA transcription of plasmid containing phage promoter regions which form anti-termination complexes during transcription. The significant decrease in rRNA transcription shown for this sytem could be explained by titration of anti-termination factors normally involved in rRNA synthesis by transcription from the phage promoter regions. That results from this model suggest the existence of such a system further supports the contention that the mechanism and control of transcription found in the present formulation allow model exploration of fine-scale regulatory features, suggesting interesting avenues for future experimental studies.

Another function of this model as presently formulated is as a basis to simulate the effects of mutation affecting chromosomal structure or enzymatic activity and specificity. These manipulations may be technically impossible to perform at the present time, extremely difficult to detect and assign accurately, or result in sufficient metabolic distortion to render the resultant cell inactive. Any kinetic

mechanism rigorously formulated provides the opportunity to study the relative importance of various relationships within that mechanism. Rearrangement of the distribution of sRNA operons along the chromosome was chosen as a mutation difficult to accomplish physically, which could be easily studied given the present formulation. This manipulation also presents an opportunity to assess the sensitivity of the model formulation to this aspect of cellular metabolism.

One additional application of structured models is also demonstrated by this example. It has been proposed that, in general, bacterial organization and genetic structure represent an optimized system in terms of evolutionary development. Simulation of this genetic rearrangement may aid in the identification and understanding of the evolutionary choices encountered and the forces leading to particular mechanistic structures presently existent.

Stable RNA operons are found at seven replication fork positions. The distribution is definitely skewed to chromosome positions replicated in the early stage of DNA synthesis. Table V contains the results of simulations with uniform s-operon distribution and a distribution that is skewed inversely to the normal situation. While it would be inappropriate to claim that this model can be used to assess all the possible ramifications of such mutations, simulations using the present model indicate the general trends and highlight the magnitude of the dependence of various processes upon ribosome formation. The primary effect is to decrease ribosome synthesis. Protein synthesis decreases nearly as much, but the effect is damped by a small increase in the fraction of ribosomes that have been activated. The cell also corrects for the changed distribution by increasing the fraction of active RNA polymerase that are complexed to sRNA promoters. Most surprising is the result that the doubling time decreases under these conditions. This can be understood by considering the alteration in cell morphology caused by these changes and the effect upon the D period. The length of the D period is primarily determined by

the cell width at the end of DNA synthesis and the rate of protein synthesis. In these cases, the protein synthesis rates are depressed and the cell widths are significantly smaller when compared to the values for normal chromosomal structure. The amount of crosswall material required for completion of septation is reduced 10%-40%. It is this phenomenon that results in decreased doubling times for these cells. In summary, the placement of sRNA operons on the chromosome relative to the replication origin is very important to the overall performance of the model as formulated, and by implication, to the organism itself.

If the primary result of the skewed distribution that exists naturally is to increase greatly overall protein synthesis capacity while slightly decreasing growth rate, of what evolutionary advantage is the skewed distribution? One hypothetical advantage which is outside of the scope of this model formulation, is an improved capability for cellular adaptation to changing growth conditions. Maintenance of expanded protein synthesis capacity and elevated protein levels allow the cell to adjust rapidly to changes in the type and availability of substrates. One concludes that metabolic diversity is favored relative to small changes in net growth rate in a particular environment. This is related directly to the observation that, in general, *E. coli* does not exist in a static environment but rather must be continually adapting to changes in extracellular conditions.

To further evaluate the present model formulation regarding the synthesis of anti-repressor protein for DNA synthesis initiation, simulations of the response to decreased DNA replication fork velocity were performed. Experimental measurements show that the cell cycle parameter C is affected when the cell undergoes fork slowdown due to exposure to nalidixic acid, deprivation of required deoxyribonucleotides or dnaB mutation. This slowdown was simulated by decreasing the maximum elongation rate for DNA synthesis, u_{M3} in Equation 11. The dependence of the average DNA/protein mass ratio upon the relative duration of the C period is

shown from experimental measurements (circles) and from simulations (triangles) in Figure 9a. Two points should be considered when one is evaluating the simulation performance for this behavior. First, the most important parameter underlying the model response (K_{PD}) was chosen with no consideration for this behavior. Second, the calculated and experimentally observed synthesis rates for protein, RNA and DNA are 3-5 times faster under these conditions than the maximum rates attainable in glucose minimal medium under normal conditions, circumstances well beyond those used as a basis for model develoment. That the simulation results agree with experimental findings as closely as indicated is very encouraging both with regard to the control of replication initiation and the processes underlying protein synthesis.

It has been reported that such perturbations have a minimal effect upon growth rate, indicating that DNA replication velocity is not coupled to the cell growth rate. Figure 9b shows that the model formulation expresses the uncoupled behavior noted experimentally. Taken together with the agreement for replication under slow elongation conditions, this performance supports the claim that not only has integral behavior been correctly represented for unstressed metabolic conditions, but that much of the kinetic and control detail formulated in the present model is correct as well.

The preceding examples illustrate two very important applications of highly structured models. First, the inclusion of correct metabolic mechanisms for segments of the model formulation will tend often to indicate flaws in the treatment of associated or dependent phenomena, possible leading to the design of novel experiments, as resulted from consideration of the treatment of transcriptional efficiency. Second, as in the case of the skewed sRNA promoter distributions, the simulation results allow a fuller and more systematic assessment of the range of variables impacted by a metabolic alternation than would be feasible using intui-

tive, qualitative reasonings.

SUMMARY

A highly structured model for *E. coli* has been presented, which includes mechanistically correct formulations for transcription and translation. Model development was based upon inclusion of individually identified and quantified control systems and reaction mechanisms. The paucity of adjustable parameters used in model development points to the establishment of a model rather than to an extended correlation. The demonstrated capability to predict simultaneously cell morphology, growth rate and yield, cell composition, C and D period duration as well as such details as RNA polymerase distribution and translation initiation to very high degrees of accuracy supports the structure and assumptions upon which the model is based.

Furthermore, as indicated in Figures 5a-c and 8, this model equals the Cornell model in terms of simulation accuracy. This is significant, since there has been major reformulation of critical metabolic processes as well as extensive additions made to the original model. These alterations and additions are necessary in order to expand the range of conditions under which the model can be expected to give reasonably accurate results. An effort was made to account explicitly for kinetic mechanisms and control schemes whenever possible, in recognition of the expectation that perturbation of cell process and controls or conditions within and without the cell may uncouple otherwise coordinated control systems and responses. In particular, transcription and translation were more rigorously formulated to allow subsequent studies of the parasitic effect of vector insertion into a host cell and the expression of any vector-mediated genes. Also, as demonstrated in the discussion section, this formulation will be useful for discrimination between alternative hypothetical kinetic and control mechanisms.

NOMENCLATURE

$a1$	stoichiometric coefficient for the anabolic use of ammonia
$A1$	ammonia
$A2$	glucose
ADP	concentration of ATP (g/mL) as an allosteric effector of ribonucleotide reductase activity
$ATPs$	rate of consumption of ATP used as an energy source to drive synthetic reactions
b_i	stoichiometric coefficients for the synthesis of macromolecular precursors from glucose
$CA1$	extracellular concentration of ammonia (mg/L)
$CA2$	extracellular concentration of glucose (mg/L)
CL	length of cylindrical portion of the cell (cm)
d_i	ATP requirement to drive a given reaction (mol ATP/g)
$dATP$	concentration of deoxyadenosine triphosphate (g/mL)
$dNTP$	deoxyribonucleotide triphosphates
e_i	stoichiometric coefficient for interconversion of precursors (g/g)
Ec	RNA polymerase core enzyme
E_{tm}, E_{ts}	attenuation efficiencies for mRNA operons and sRNA operons, respectively
$E1$	ribonucleotide reductase, which catalyzes the conversion of ribonucleotides to deoxyribonucleotides
$E2$, $E3$	enzymes which catalyze the formation of cell envelope and crosswall, respectively

f	proportionality constant relating surface area to non-protein cell envelope mass (cm^2/g)
$FRACI$	fraction of ribosomes which are inactive
$FRACM$	active ribosomal fraction, defined as RM/RO, Eqn. R4 in the text
g_i	stoichiometric constant for conversion of precursor to macromolecule (g/g)
k_i	maximum synthesis rate for precursors
k_T	rate of decomposition
K_{xy}	saturation constant, component y affecting synthesis of component x (g/mL)
K_i	saturation constant for inhibition
K_T	saturation constant for components affecting decomposition
Mi	macromolecules: 1 = protein, 2i = immature sRNA, 2r = mature sRNA, 2m = mRNA, 3 = DNA, 4 = cell envelope, 5 = glycogen
p_c	cytoplasmic density (g/ml)
p_{M4}	cell envelope density (g/mL)
$P1$	amino acids (g)
$P2$	ribonucleotides (g)
$P3$	deoxyribonucleotides (g)
RI	mass of inactive ribosomal RNA, assuming rRNA is 85% of total sRNA
R_x	transport rate into cell of component x
rdr	ribonucleotide reductase
S	cell surface area (cm^2)
$SEPF$	area of septum formed (cm^2)

SL	length of septum formed across cell width (cm)
$sRNA$	"stable" RNA, equals tRNA + rRNA
$TOTM$	total mass of the cell
u_i	maximum synthesis rate for macromolecule i
v_i, v'_i	maximum transport rate (g/h cm^2)
V	cell volume (cm^2)
w_i	stoichiometric constant for use of reducing equivalents for energetics or biosynthesis: ion = ion transport, bio = cell mass reduction, SO4 = sulfur reduction, M = membrane recharge
W	cell width (cm)
X	total pool of reducing power
Z	(moles acetate formed)/(moles glucose dissimilated)

REFERENCES

1. M.M Domach, S. Leung, R.E. Cahn, G.G. Cocks and M.L. Shuler, *Biotechnol. Bioeng.*, **26**, 203 (1984).

2. M.L. Shuler, S. Leung and C.C. Dick, *Ann. N.Y. Acad. Sci.*, **326**, 35 (1979).

3. S.V. Ho and M.L. Shuler, *J. Theor. Biol.*, **68**, 415-435 (1977).

4. G.A.T. Mahon, P. McWilliam, R.L. Gordon and D.J. McConnell, *J. Theor. Biol.*, **87**, 483-515 (1980).

5. G. Churchward, H. Bremer and R. Young, *J. Bacteriol.*, **150**, 572 (1982).

6. B.N. Belintsev, S.K. Zavriev and M.F. Shemyakin, *Nuc. Acid Res.*, **8**, 1391 (1980).

7. J. Hamming, M. Gruber and G. Ab, *Nuc. Acid Res.*, **7**, 1019 (1979).

8. M. Kajitani and A. Ishihama, *Nuc. Acid Res.*, **11**, 3873 (1983).

9. M.T. Record Jr., C.F. Anderson and T.M. Lohman, *Q. Rev. Biophys.*, **11**, 103 (1978).

10. P.L. deHaseth, T.M. Lohman, R.R. Burgess and M.T. Record Jr., *Biochemistry*, **17**, 1612 (1978).

11. H.S. Strauss, R.R. Burgess and M.T. Record Jr., *Biochemistry*, **19**, 3496 (1980).

12. H.S. Strauss, R.R. Burgess and M.T. Record Jr., *Biochemistry*, **19**, 3504 (1980).

13. T.M. Lohman, C.G. Wensley, J. Cina, R.R. Burgess and M.T. Record Jr., *Biochemistry*, **19**, 3516 (1980).

14. Y. Sokawa, J. Sokawa and Y. Kaziro, *Cell*, **5**, 69-74 (1975).

15. A. Travers, *Molec. Gen. Genet.*, **147**, 225-232 (1976).

16. J. Ryals, R. Little and H. Bremer, *J. Bacteriol.*, **151**, 1261 (1982).

17. A.A. Travers, P.G. Debenham and O. Pongs, *Biochemistry*, **19**, 1651 (1980).

18. O. Pongs and N. Ulbrich, *Proc. Nat. Acad. Sci., U.S.A.*, **73**, 3604 (1976).

19. A.A. Travers, P. Buckland and P.G. Debenham, *Biochemistry*, **19**, 1656 (1980).

20. H. Manor, D. Goodman and G. Stent, *J. Mol. Biol.*, **39**, 1 (1969).

21. R.E. Kingston and M.J. Chamberlin, *Cell*, **27**, 523 (1981).

22. D.F. Ward and M. Gottesmann, *Nature (London)*, **292**, 212 (1981).

23. J. Greenblatt, M. McLimont and S. Hanly, *Nature (London)*, **292**, 215 (1981).

24. S. Adhya and M. Gottesmann, *Ann. Rev. Biochem.*, **47** 967 (1978).

25. J. Greenblatt, *Can. J. Biochem. Cell Biol.*, **62**, 79 (1983).

26. J. Shine and L. Dalgarno, *Proc. Nat. Acad. Sci.. U.S.A.*, **71**, 1342 (1974).

27. D. Iserentant and W. Fiers, *Gene*, **9**, 1 (1980).

28. G.F.E. Scherer, M.D. Walkinshaw, S. Arnott and J. Morre', *Nuc. Acid Res.*, **8**, 3895 (1980).

29. M.B. Bahramian, *J. Theor. Biol.*, **84**, 103 (1980).

30. G.D. Stormo, T.D. Schneider and L.M. Gold, *Nuc. Acid Res.*, **10**, 2971 (1980).

31. C. Gualerzi, G. Risuleo and C.L. Pon, *Biochemistry*, **16**, 1684 (1977).

32. J.A. Steitz and K. Jakes, *Proc. Nat. Acad. Sci., U.S.A.*, **72**, 4734 (1975).

33. M.J. Pine, *J. Bacteriol.*, **103**, 207 (1970).

34. K. Olden and A.L. Goldberg, *Biochim. Biophys. Acta*, **542**, 385 (1978).

35. R. Voellmy and A.L. Goldberg, *J. Biol. Chem.*, **255**, 1008 (1980).

36. D. Filpula and J.A. Fuchs, *J. Bacteriol.*, **130**, 107 (1977).

37. D. Filpula and J.A. Fuchs, *J. Bacteriol.*, **135**, 429 (1978).

38. A. Larsson and P. Reichard, *J. Biol. Chem.*, **241**, 2540 (1966).

39. N.C. Brown and P. Reichard, *J. Mol. Biol.*, **46**, 39 (1969).

40. H. Warner, *J. Bacteriol.*, **115**, 18 (1973).

41. R.H. Pritchard and A. Zaritsky, *Nature (London)*, **226**, 126 (1970).

42. H. Schaller, B. Otto, V. Nusslein, J. Hof, R. Herrmann and F. Bonhoeffer, *J. Mol. Biol.*, **63**, 183 (1972).

43. C.L. Woldringh, M.A. DeJong, W. Van den Berg, and L. Koppes, *J. Bacteriol.*, **131**, 270 (1977).

44. S.J. Pirt, **Principles of Microbe and Cell Cultivation** (Wiley, New York, 1975).

45. J.J. Heijmen and J.A. Roels, *Biotechnol. Bioeng.*, **23**, 739 (1981).

46. A.H. Stouthamer, *Symp. Soc. Gen. Microbiol.*, **27**, 285 (1977).

47. W.J. Payne, *Ann. Rev. Microbiol.*, **24**, 17 (1970).

48. P.P. Dennis and H. Bremer, *J. Bacteriol.*, **119**, 270 (1974).

49. C.E. Helmstetter and O. Pierucci, *J. Mol. Biol.*, **102**, 477 (1976).

50. S. Aksoy, C.L. Squires and C. Squires, *J. Bacteriol.*, **159**, 260 (1984).

51. P.J. Senior, *J. Bacteriol.*, **123**, 407 (1975).

52. A.C. St. John and A.L. Goldberg, *J. Biol. Chem.*, **253**, 2705 (1978).

53. Y. Iwakura, K. Ito and A. Ishihama, *Mol. Gen. Genet.*, **133**, 1 (1974).

54. M. Kajitani and A. Ishihama, *J. Mol. Biol.*, **259**, 1951 (1984).

55. R.A. Sharrock, R.L. Gourse and M. Nomura, *J. Bacteriol.*, **163**, 704 (1985).

56. M.L. Shuler and M.M Domach, **Foundations of Biochemical Engineering**, ACS Vol. 207, 93 (1983).

57. E.R. Wagner, J.D. Fabricant and M. Schweiger, *Eur. J. Biochem.*, **120**, 231 (1979).

58. J.D. Manwaring and J.A. Fuchs, *J. Bacteriol.*, **130**, 960 (1977).

Table I. Rate equations for each cellular component. The notation used was chosen to facilitate computer implementation of the equations as presented. Synthesis rates are denoted by D (component name) S, and degradation rates by D (component name) D. Examples of these are DP1S and DP1D for the synthesis and degradation of amino acids (P1). The net time rate of change of the amount of a component is simply D (component name). A and C as subscripts indicate anabolic and catabolic processes, respectively. All other symbols are described in the Nomenclature.

TABLE I

Nutrient Metabolism

Ammonia

$$DA1 = R_{A1} * S - a_1 * (DP1S - DP1D) \tag{1a}$$

$$R_{A1} = v'_{A1} * \left[\frac{CA1}{CA1 + K_{CA1}} - \frac{A1/V}{A1/V + K_{A1}} \right] + v_{A1} * \frac{CA1}{CA1 + K'_{CA1}} * \frac{K_{iA1}}{A1/V + K_{iA1}} *$$

$$\frac{A2/V}{A2/V + K_{A1A2}} \tag{1b}$$

Glucose

$$DA2 = R_{A2} * S - (DA2_A - DA2_C) \tag{2a}$$

$$R_{A2} = v_{A2} * \frac{CA2}{CA2 + K_{CA2}} * \frac{K_{iA2}}{A2/V + K_{iA2}} \tag{2b}$$

$$DA2_A = b_1 * (DP1S - DP1D) + b_2 * (DP2S - DP2D) + b_4 * DP4S + g_5 * DM5 \tag{2c}$$

$$DA2_C = \frac{[DATPs + DX * (P/O)max]}{4 + (12 - 4 * Z) * (P/O)max} * 180 \tag{2d}$$

$$DATP_s = d_{A2} * DA2 + d_{P1} * DP1S + d_{P2} * DP2s + d_{P4} * DP4S + d_{M1} * DM1S$$

$$+ d_{M2} * (DM2iS + DM2mS) + d_{M3} * DM3 + d_{M4} * DM4S$$

$$+ d_{M5} * DM5 + d_{PG} * DPGS + d_V * DV \tag{2e}$$

$$DX = w_1 * R_{A1} * S + (w_{ion} + w_{bio} + w_{so4}) * (\Sigma DA_i + \Sigma DP_i + \Sigma DM_i) + w_M * S$$

$$+ w_{P3} * DP3S \tag{2f}$$

$$Z = u_z * \frac{DM4S}{DM4S + K_{ZM4}} \tag{2g}$$

Precursor Metabolism

Amino Acids

$$DP1S = k_1 * \frac{K_{iP1}}{K_{iP1} + P1/V} * \frac{A1/V}{A1/V + K_{P1A1}} * \frac{A2/V}{A2/V + K_{P1A2}} * V \quad (3a)$$

$$DP1D = k_{TP1} * \frac{K_{TP1A2}}{K_{TP1A2} + A2/V} * P1 \quad (3b)$$

$$DP1 = DP1S - DP1D - g_1 * DM1 - e_2 * (DP2S - DP2D) - e_4 * DP4S \quad (3c)$$

Ribonucleotides

$$DP2S = k_2 * \frac{K_{iP2}}{K_{iP2} + P2/V} * \frac{P1/V}{K_{P2P1} + P1/V} * \frac{A2/V}{K_{P2A2} + A2/V} * V \quad (4a)$$

$$DP2D = k_{TP2} * \frac{K_{TP2A2}}{K_{TP2A2} + A2/V} * P2 \quad (4b)$$

$$DP2 = DP2S - DP2D - g_2 * (DM2i + DM2r + DM2m) - e_3 * DP3S \quad (4c)$$

Deoxyribonucleotides

$$DP3S = k_3 * E1 * \frac{ADP}{ADP + K_{P3ADP}} + \frac{1}{1 + \frac{K_{P3dATP}}{DATP} + \frac{dATP}{K_{iP3dATP}}} \quad (5a)$$

$$DP3D = k_{TP3} * P3 \quad (5b)$$

$$DP3 = DP3S - DP3D - g_3 * DM3 \quad (5c)$$

Cell Envelope Precursors

$$DP4S = k_4 * \frac{K_{iP4}}{K_{iP4} + P4/V} * \frac{P1/V}{K_{P4P1} + P1/V} * \frac{A2/V}{K_{P4A2} + A2/V} * V \quad (6a)$$

$$DP4 = DP4S - g_4 * DM4 \quad (6b)$$

Macromolecule Metabolism

Protein

$$DM1S = u_1 * \frac{A2/V}{K_{M1A2} + A2/V} * \frac{P1/V}{K_{M1P1} + P1/V} * 0.85 * M2r * FRACM \quad (7a)$$

$$DM1D = \left[\frac{P1/V}{P1/V + K_{TM1P1}} * \frac{PG/V}{K_{TM1PG} + PG/V} * \frac{K_{TiM1A2}}{K_{TiM1A2} + A2/V} \right.$$

$$\left. + \frac{A2/V}{K_{TM1A2} + A2/V} \right] * k_{TM1} * M1 \quad (7b)$$

$$DM1 = DM1S - DM1D - 0.05 * DM1S \quad (7c)$$

RNA

$$DM2iS = u_{M2} * E_{ts} * \frac{P2/V}{K_{M2P2} + P2/V} * EDNAR * V \quad (8a)$$

$$DM2iD = k_{TM2} * \frac{K_{TM2i}}{K_{TM2i} + DM1S/V} * M2i \quad (8b)$$

$$DM2i = DM2is - DM2iD - u_{M2r} * M2i \quad (8c)$$

$$DM2r = u_{M2r} * M2i - k_{TM2r} * M2r * \frac{K_{TM2r}}{K_{TM2r} + A2/V} \quad (9)$$

$$DM2mS = E_{tm} * u_{M2} * \frac{P2/V}{K_{M2P2} + P2/V} * EDNAM * V \quad (10a)$$

$$DM2m = DM2mS - k_{TM2m} * M2m \quad (10b)$$

$$E_{tm} = 1.07 * \frac{K_{iAT}}{K_{iAT} + (M1/M3)} \quad (10c)$$

DNA

$$DM3 = u_{M3} * NTOT * \frac{P3/V}{K_{M3P3}+P3/V} * \frac{A2/V}{K_{M3A2}+A2/V} \tag{11}$$

Cell Envelope

$$DM4 = u_{M4} * E23 * \frac{A2/V}{K_{M4A2}+A2/V} * \frac{P4/V}{K_{M4P4}+P4/V}$$

$$- k_{TM4} * M4 * \frac{A2/V}{K_{TM4}+A2/V} \tag{12}$$

Glycogen

$$DM5 = u_{M5} * \frac{A2/V}{K_{M5A2}+A2/V} * V - k_{TM5} * \frac{K_{TM5A2}}{K_{TM5A2}+A2/V} * M5 \tag{13}$$

Enzymes and Cofactors

$$DE1 = n_{E1} * (M1/M3) * NORI * DM1 \tag{14}$$

$$DE2 = n_{E2} * DM1 \tag{15}$$

$$DE3 = n_{E3} * M1 \tag{16}$$

$$E23 = E2 + E3 \tag{17}$$

$$DEc = n_{Ec} * DM1S - (Ec/M1) * DM1D \tag{18}$$

$$DSIG = n_{SIG} * DM1S - (SIG/M1) * DM1D \tag{19}$$

$$DPG = u_{PG} * RI * \frac{K_{iPGP1}}{K_{iPGP1}+P1/V} - k_{TPG} * \frac{A2/V}{K_{TPGA2}+A2/V} * PG \tag{20}$$

$$RI = 0.85 * M2r * FRACI \tag{21}$$

Chromosome Replication Initiatiion Proteins

$$DARPS = n_{ARP} * \frac{(M1/M3)}{K_{iPD} + (M1/M3)} \tag{22}$$

$$RP_{burst} = n_{RP} * NORI * \frac{K_{iRP}}{K_{iRP} + [DM1/M1]_{max} * M1/V - DM1S/V} \tag{23}$$

Cell Morphology

$$TOTM = \Sigma A_i + \Sigma P_i + \Sigma M_i \tag{24}$$

$$V = \frac{(TOTM - M4)}{\rho_C} + \frac{M4}{\rho_{M4}} \tag{25}$$

$$V = \frac{1}{6} * \pi * W_3 + \frac{1}{4} * \pi * W_2 * CL + \frac{1}{2} * \pi * W_@ * SL - \frac{2}{3} * \pi * SL_3 \tag{26}$$

$$S = \pi * W^2 + \pi * W * CL + 2 * \pi * W * SL \tag{27}$$

$$DS = f_2 * DM4 \tag{28}$$

$$SL = SEPF/(2 * \pi * W) \tag{29}$$

$$SEPF_{t+dt} = SEPF_t + (E3/E23) * DS \tag{30}$$

$$CL_{t+dt} = CL_t + (E2/E23) * \frac{DS}{\pi * W} \tag{31}$$

TABLE II

These are the parameter values for the constants shown in Table I.

Ammonium ion

$v_{A1}' = 2.06 * 10^{-7} g/cm^2 h$ $\quad$ $K_{CA1} = 1.08 * 10^{-7} g/mL$

$v_{A1} = 9.5 * 10^{-7} g/cm^2 h$ $\quad$ $K_{A1} = 1.0 * 10^{-5} g/mL$

$K'_{CA1} = 3.6 * 10^{-6} g/mL$ $\quad$ $K_{iA1} = 2.3 * 10^{-3} g/mL$

$$K_{A1A2} = 7.2 * 10^{-5} g/mL$$

Glucose

$v_{A2} = 1.0 * 10^{-5} g/cm^2 h$ $\quad$ $K_{iA2} = 2.0 * 10^{-3} (g/mL)$

$$K_{CA2} = 9.3 * 10^{-6} g/mL$$

Amino acids

$k_1 = 0.39 g/mL\ h$ $\quad$ $K_{iP1} = 5.0 * 10^{-2} g/mL$

$K_{P1A1} = 2.5 * 10^{-5} g/mL$ $\quad$ $K_{P1A2} = 2.5 * 10^{-4} g/mL$

$k_{TP1} = 0.025 h^{-1}$ $\quad$ $K_{TP1A2} = 1.1 * 10^{-5} g/mL$

Ribonucleotides

$k_2 = 0.21 g/mL\ h$ $\quad$ $K_{iP2} = 9.0 * 10^{-3} g/mL$

$K_{P2P1} = 9.9 * 10^{-4} g/mL$ $\quad$ $K_{P2A2} = 2.5 * 10^{-4} g/mL$

$k_{TP2} = 0.03 h^{-1}$ $\quad$ $K_{TP2A2} = 1.1 * 10^{-5} g/mL$

Deoxyribonucleotides

$k_3 = 1.1 * 10^2 g/h\ g\ E_1$ $K_{P3ADP} = 5.2 * 10^{-6} g/mL$

$K_{P3dATP} = 1.2 * 10^{-6} g/mL$ $K_{iP3dATP} = 1.45 * 10^{-5} g/mL$

Cell Envelope Precursors

$k_4 = 0.06 g/mL\ h$ $K_{iP4} = 5.08 * 10^{-3} g/mL$

$K_{P4P1} = 9.9 * 10^{-4} g/mL$ $K_{P4A2} = 1.2 * 10^{-4} g/mL$

Protein

$u_1 = 7.10 g/h\ gRNA$ $K_{M1A2} = 7.2 * 10^{-5} g/mL$

$K_{M1P1} = 2.2 * 10^{-4} g/mL$ $K_{TM1P1} = 1.03 * 10^{-7} g/mL$

$K_{TM1PG} = 3.9 * 10^{-8} g/mL$ $K_{TiM1A2} = 1.404 * 10^{-4} g/mL$

$K_{TM1A2} = 2.73 * 10^{-3} g/mL$ $k_{TM1} = 0.07\ h^{-1}$

RNA

$u_{M2} = 6.78 * 10^7 g/h\ mole$ $K_{iAT} = 11.5\ g\ prot/g\ DNA$

$E_{ts} = 0.96$ $u_{M2r} = 14.0 h^{-1}$

$k_{TM2} = 21.0\ h^{-1}$ $K_{M2P2} = 7.2 * 10^{-5} g/mL$

$k_{TM2r} = 0.07\ h^{-1}$ $K_{TM2i} = 1.8 * 10^{-2} g/mL\ h$

$K_{TM2r} = 1.1 * 10^{-5} g/mL$

DNA

$$u_{M3}=9.8*10^{-5}g/h\ fork$$

$$K_{M3A2}=1.7*10^{-4}g/mL$$

$$K_{M3P3}=4.0*10^{-6}g/mL$$

Cell Envelope

$u_{M4}=67.3g/h\ g\ E_{23}$ $K_{M4A2}=3.0*10^{-5}g/mL$

$K_{M4P4}=5.0*10^{-4}g/mL$ $k_{TM4A2}=0.23\ h^{-1}$

$$K_{TM4A2}=1.8*10^{-4}g/mL$$

Glycogen

$u_{M5}=0.02g/mL\ h$ $K_{M5A2}=0.002\ g/mL$

$K_{TM5}=0.14\ h^{-1}$ $K_{TM5A2}=0.001\ g/mL$

Enzymes and Cofactors

$n_{E1}=1.3*10^{-4}$ $K_{TPGA2}=7.2*10^{-5}g/mL$

$n_{E3}=1.6*10^{-2}h^{-1}$ $n_{E2}=3.1*10^{-3}$

$n_{SIG}=7.0*10^{-4}$ $n_{Ec}=1.15*10^{-2}$

$K_{iPGP1}=1.265*10^{-4}g/mL$ $u_{PG}=9.0*10^{-3}mol/h\ g\ RNA$

$k_{TPG}=125\ h^{-2}$

DNA Initiation Proteins

$$n_{ARP} = 2.60 * 10^{-3} g/gM4 \qquad K_{iPD} = 5.5\ g\ prot/g\ DNA$$

$$n_{RP} = 5.0 * 10^{-17} \qquad [DM1/M1]_{MAX} = 2.1$$

$$K_{iRP} = 0.22 g/ml\ h$$

Cell Morphology

$$\rho_c = 0.258\ g/mL$$

$$f_s = 2.86 * 10^6 cm^2/g$$

$$\rho_{M4} = 0.553\ g/mL$$

Acetate

$$u_Z = 0.9$$

$$K_{ZM4} = 1.0 * 10^{-14} g/mL$$

Oxidation

$$(P/O)max = 1.5$$

Stoichiometric Coefficients

$a_1 = 0.179$ $b_1 = 1.128$ $b_2 = -0.456$ $b_4 = 1.26$

$e_3 = 1.149$ $e_3 = 1.049$ $e_4 = 0.128$ $g_1 = 1.167$

$g_2 = 1.057$ $g_3 = 1.053$ $g_4 = 1.10$ $g_5 = 1.11$

Energy requirement coefficients (mol ATP/g)

$d_{A2} = 0.0056$ $d_{P1} = 0.0025$ $d_{P2} = 0.002$ $d_{P4} = 0.0016$

$d_{M1} = 0.039$ $d_{M2} = 0.0067$ $d_{M3} = 0.0071$ $d_{M4} = 0.0081$

$d_{M5} = 0.0124$ $d_{PG} = 3 mol/mol$ $d_V = 0.00633 mol/mL$

Reductant requirements (mol H/g)

$w_1 = 0.028$ $w_{P3} = 0.0031$ $w_{ion} = 0.002\ mol/g\ cell$

$w_{bio} = 0.0139$ $w_{SO4} = 0.00121$ $w_M = 5.8 * 10^{-4} mol/cm^2 h$

TABLE III: RNA POLYMERASE EQUILIBRIUM

Equations Defining Equilibrium Constants

$$log\ K_o^{Mg++} = 0.5 - 1.71 * log\ [Na^+] \quad (1)$$

$$DMG = log\ (0.5 * (1+\sqrt{1+4 * K_o^{Mg--} * [Mg^{++}]})) \quad (2)$$

$$log\ K^{HD} = -2.5 - 10.8 * log\ [Na^+] - 2 * DMG - 0.4 * (pH - 7.8) \quad (3)$$

$$log\ K^{CD} = -8.5 - 21.2 * log\ [Na^+] - 0.5 * DMG - 0.3 * (pH - 7.8) \quad (4)$$

$$log\ K_o^{Mg--'} = 0.32 + 1.75 * log\ [Na^+] \quad (5)$$

$$DMG' = log\ (0.5 * (1+\sqrt{1+ 4 * K_o^{Mg--} * [Mg^{++}]})) \quad (6)$$

$$DPH = log\ (1+(10^{-7.4})/[H^+]) \quad (7)$$

$$DTEMP = 0.48 + 0.52 * (T/310.16) \quad (8)$$

$$log\ K^{HP} = (-1.3 - 14.7 * log\ [Na^+] - 2 * DMG' - 2 * DPH) * DTEMP \quad (9)$$

$$K^E = 10^{10}$$

Equilibrium Relationships and Associated Mass Balances

$$E_H = K^E * Ec * (SIG_o - E_H\ DNAC - E_H) \quad (10)$$

$$E_H\ DNAC = K^{CD} * Ec * DNAC \quad (11)$$

$$E_H\ DNAM = K_M * K^{HP} * E_H * DNAM \quad (12)$$

$$K_M = 1.4 * \frac{0.43}{0.43 + DM1S/V} \quad (13)$$

$$E_H\ DNAR = K_S * K^{HP} * E_H * DNAR \tag{14}$$

$$K_S = \frac{2.2*10^{-7}}{PG/V+2.2*10^{-7}} * \left(\frac{1.1*10^{-2}}{1.1*10^{-2}+([DM1/M1]_{MAX}*M1/V-DM1S/V)^3}\right) \tag{15}$$

$$E_H\ DNAC = K^{HD} * E_H$$

$$* \frac{(DNAC_o - 40*(Ec\ DNAC_o + E_H\ DNAM + E_H\ DNAR))}{1 + 40 * K^{HD} * E_H} \tag{16}$$

$$Ec = Ec_o - E_H\ DNAM - E_H\ DNAR - E_H\ DNAC - E_H - Ec\ DNAC \tag{17}$$

$$DNAC = DNAC_o - 40 * (E_H\ DNAM + E_H\ DNAR$$

$$+ E_H\ DNAC + Ec\ DNAC) \tag{18}$$

$DNA(C,M,R)$	= molar concentration of non-promoter, mRNA promoter, and sRNA promoter sites on the chromosome
SIG_o	= molar concentration of RNA polymerase sigma subunit
$[Na^+]$	= intracellular sodium ion concentration
$[Mg^{++}]$	= intracellular magnesium concentration
$[H^+]$	= intracellular hydrogen ion concentration
T	= temperature, Kelvin
HD	signifies the non-specific holoenzyme-DNA complex
HP	signifies the holoenzyme-promoter complex
CD	signifies the core polymerase-DNA complex
Ec	is core polymerase
E_H	is holoenzyme
$E_H\ DNAC$	is the non-specific holo-DNA complex
$E_H\ DNAM$	is the holo-mRNA promoter complex
$E_H\ DNAR$	is the holo-sRNA promoter complex
$Ec\ DNAC$	is the core polymerase-DNA complex
K^E	is the equilibrium constant for the core-sigma interaction

TABLE IV

Sources of Parameter Estimates For All New Parameters.

Parameter	Source of Estimate
$K_{P3ADP}, K_{P3dATP}, K_{iP3dATP}$	Allosteric control of ribonucleotide reductase activity (36-39), particularly (39)
K_{TiM1A2}, K_{TM1A2}	Based on finding by St. John and Goldberg (52) that the rate of proteolysis goes through a maximum as the ATP level decreases. Estimate based on data from (51)
K_{TM1PG}, K_{TM1P1}	Effect of ppGpp on proteolysis is modulated by the availability of amino acids (35).
k_{iAT}	Titration of transcription termination proteins by mRNA promoters described in the text
n_{Ec}, n_{SIG}	Fraction of total protein synthesized which is core polymerase and sigma subunit for a doubling time of 1 hour (52)
K_{iPD}	Filpula and Fuchs (36) suggest that the protein/DNA ratio correlates to the synthesis of a trigger for DNA replication initiation. Estimate based on simulation experience.
v_{A2}, K_{iA2}	Values reported in (57) from direct measurements.
$ADP, dATP$	Calculated as a fraction of the total ribonucleotide and deoxyribonucleotide pools, respectively. Taken from Manwaring and Fuchs (58), ADP = 0.466 * [NTP], dATP = 0.364 * [dNTP]

TABLE V

Comparison of simulation results for different distributions of stable RNA operon promoters to results using the correct distribution. Data set A is the uniformly distributed case; set B is the inversely skewed configuration. Results are presented relative to the correctly distributed case.

Cell rates for the correctly distributed cases				
	$1.03\ hr^{-1}$		$0.63\ hr^{-1}$	
Cell Behavior	A	B	A	B
Growth rate	1.09	1.15	1.02	1.17
C period	0.97	0.96	0.99	0.98
D period	0.69	0.34	0.91	0.44
Volume	0.87	0.77	0.94	0.76
Protein	0.85	0.70	0.92	0.70
sRNA	0.84	0.64	0.91	0.66
mRNA	0.94	0.86	0.97	0.86
Active fraction of RNA polymerase	0.98	1.08	0.98	1.17
Cell width	0.89	0.79	0.95	0.76
Cell length	1.21	1.41	0.98	1.44
Active fraction of ribosomes	1.02	1.03	1.02	1.04

FIGURE 1 A schematic representation of the model for *E. coli* growing aerobically in a minimal medium with either glucose or ammonia as the limiting nutrient. Solid lines indicate material flow; dotted lines indicate the flow of information. Extracellular ammonia and glucose enter the cell and participate in the synthesis of macromolecular precursors and macromolecules as well as waste byproducts (W), which normally include CO_2, H20 and acetate. The cell envelope precursors are denoted as CEP; M2i is the "immature" stable RNA; CE is the non-protein cell envelope; PG is ppGpp. There are also enzymes whose metabolism is accounted for but are not included in the material balance. RNAP is RNA polymerase; rdr is ribonucleotide disphosphate reductase, which reduces rNTP to dNTP; E2 and E3 are enzymes that direct crosswall formation and cell envelope synthesis.

FIGURE 2 RNA polymerase exists in many conformations. Binding of the sigma subunit to core polymerase alters binding selectivity of the enzyme. The resultant holoenzyme has much greater affinity for promoter sites and decreased affinity for non-promoter regions relative to core polymerase.

FIGURE 3 Transcription and translation are coupled primarily through the effect of components of the protein synthesis apparatus on RNA polymerase selectivity. An indicator of insuffieient ribosome supply, initiation fractor IF-2 increases the affinity of holoenzyme for stable RNA promoters.

Indicators of ribosome underutilization, ppGpp and f-methionine tRNA (fmet-tRNA), stimulate transcription from mRNA promoters.

FIGURE 4 Ribosome activation begins with the formation of an inactive initiation complex, which is a rapidly equilibrating process. The rate-determining step is the addition of the 50S subunit to the initiation complex, immediately followed by polypeptide chain elongation.

FIGURE 5 In each figure, the solid lines indicate simulation results based on the present model, whicle the dashed lines denote results from the Cornell formulation. (A) Average cell volumes for *E. coli* B/r-A relative to the growth rate for an asynchronous culture. Coulter counter results from Domach et al. (Ref. 1) are denoted by (0), while osmium tetroxide-fixed cell data reported by Woldringh et al. (Ref. 43) are denoted by (Δ). Model results indicate the geometric mean of dividing and newborn single-cell volumes. (B) Growth rate dependence on glucose concentration for glucose-limited conditions. Data are taken from Ref. 1 for chemostat cultiviation of *E. coli* B/r-A at 37 C. (C) Observed growth yield under glucose limitation for chemostat-grown *E. coli* B/r-A, Ref. 1

FIGURE 6 The data presented are for *E. coli* B/r grown in minimal medium with different carbon sources utilized to achieve different growth rates, as reported by Dennis and Bremer, Ref. 48. All results are expressed as mass of macromolecule per genome. Shown are the growth rate effect on relative protein content (A), stable RNA content, which is proportional to ribosome content (B), and mRNA content (C). The lines indicate simulation results, not curves fitted to the data. It is important to note that the model formulation was developed using data for a doubling time of one

hour.

FIGURE 7 Fraction of active RNA polymerase transcribing from sRNA promoters. Data shown (o) are from Ref. 48; the line was obtained from model simulations.

FIGURE 8 C and D periods for *E. coli* B/r-A as a function of cellular growth rate. Circles represent data reported by Helmstetter and Pierucci (49). Filled symbols are estimates of the D period, open symbols the C period. The solid lines are results from this formulation, while the dashed and dotted lines are from Ref. 1. The C period represents the time between initiation and completion of chromosomal replication. The D period extends from the termination of DNA synthesis to the completion of cell division. These parameters are sensitive indicators of the suitability fo the model formulation regarding the timeing of DNA replication initiation.

FIGURE 9 In these figures, the circles and line represent experimental data (41), while the triangles denote simulation results.; (A) The response of the DNA/protein mass ratio to elongation of the DNA replication interval C (reduced replication fork velocity); (B) Dependence of doubling time upon DNA synthesis rate evaluated from the model. Experiments also show that growth rate is not coupled significantly to replication fork velocity.

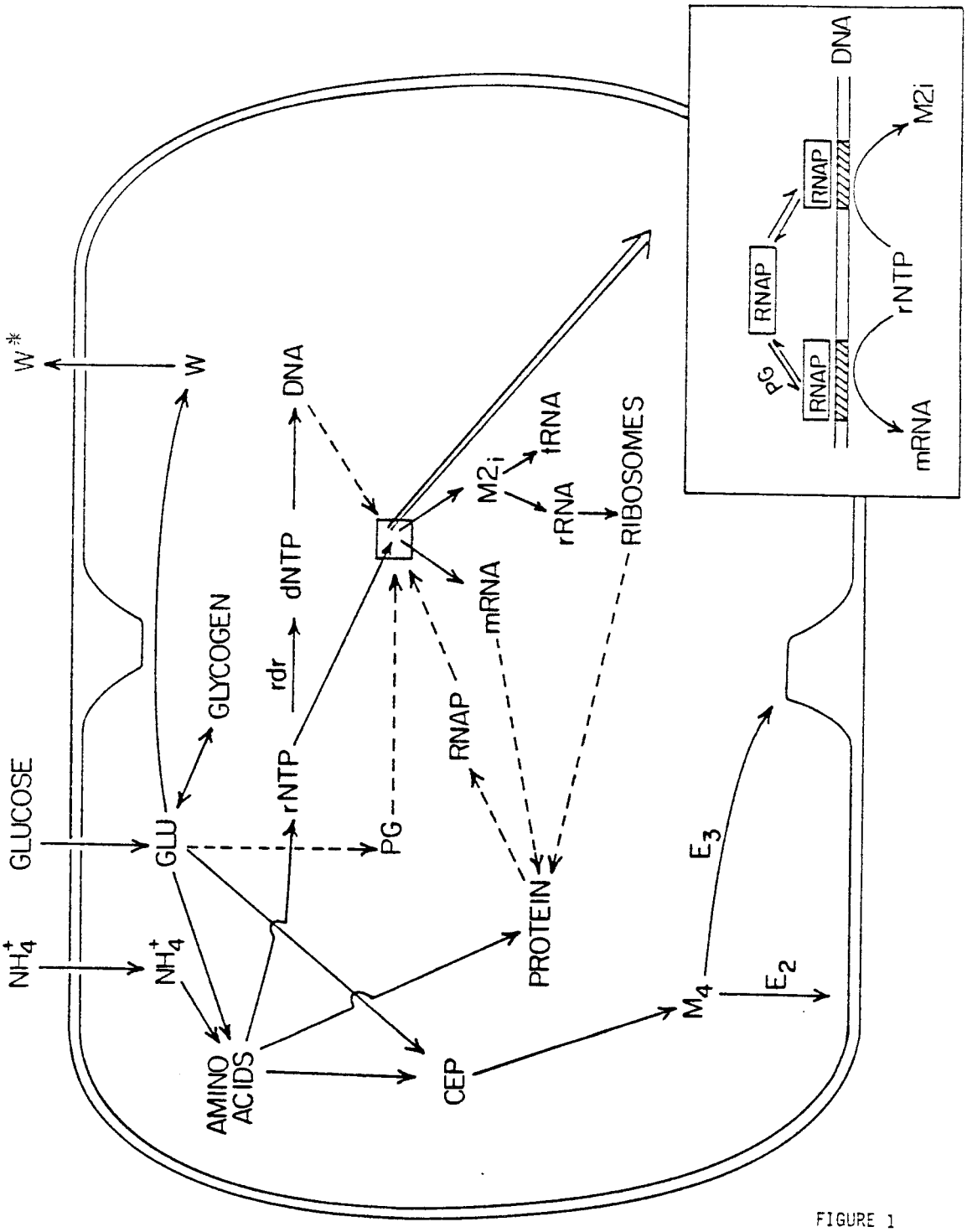

W*
W
DNA
M2i
tRNA
rRNA
RIBOSOMES
GLYCOGEN
rdr
dNTP
rNTP
PG
RNAP
mRNA
GLUCOSE
GLU
NH4+
NH4+
AMINO ACIDS
CEP
PROTEIN
E3
M4
E2
RNAP
RNAP
RNAP
DNA
M2i
rNTP
mRNA

FIGURE 1

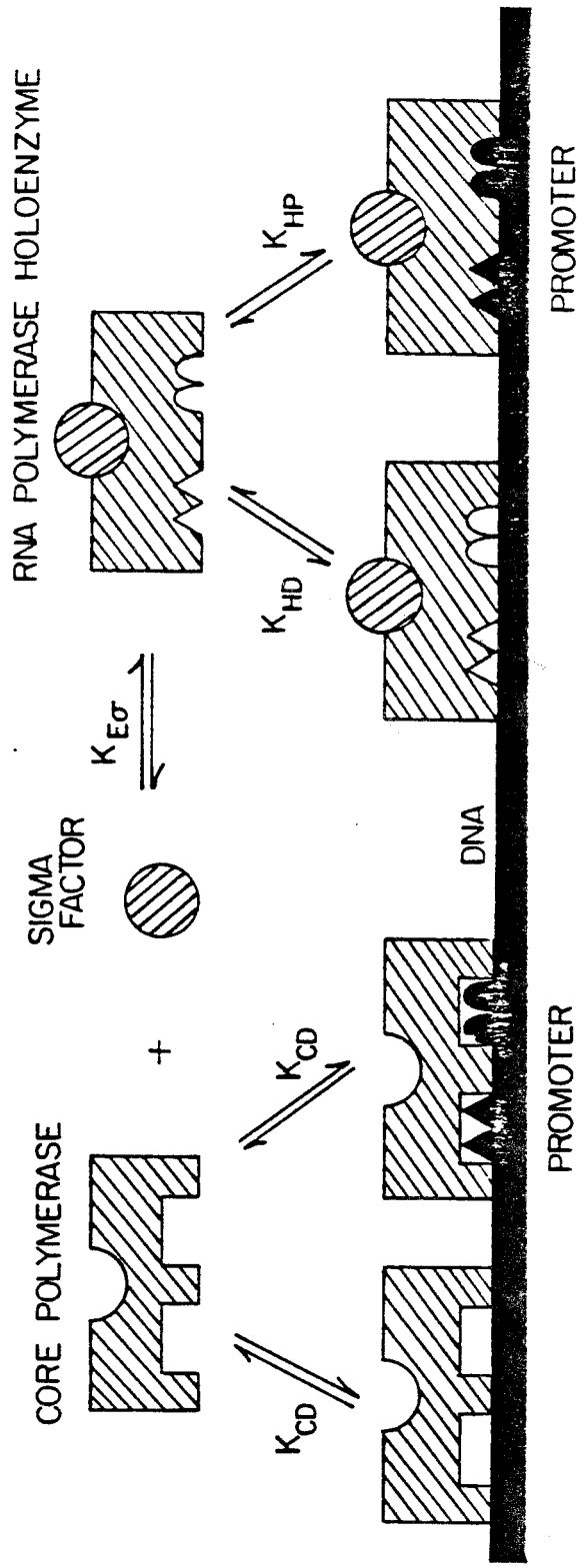

RNA POLYMERASE HOLOENZYME
CORE POLYMERASE
+
SIGMA FACTOR
$K_{E\sigma}$
K_{CD}
K_{CD}
K_{HD}
K_{HP}
DNA
PROMOTER
PROMOTER

FIGURE 2

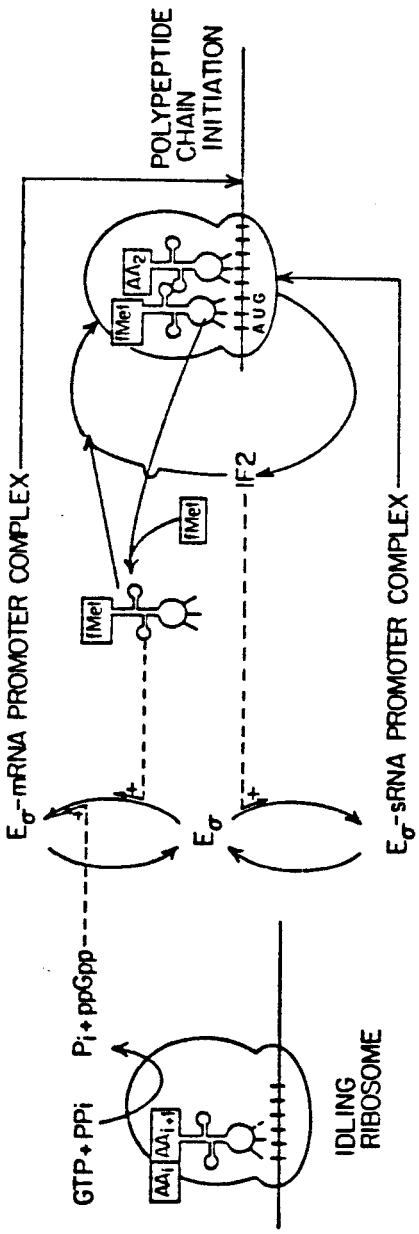

POLYPEPTIDE CHAIN INITIATION
AUG
IF2
fMet
fMet
fMet
E_σ-mRNA PROMOTER COMPLEX
E_σ
E_σ-sRNA PROMOTER COMPLEX
GTP + PPi
P_i + ppGpp
IDLING RIBOSOME

FIGURE 3

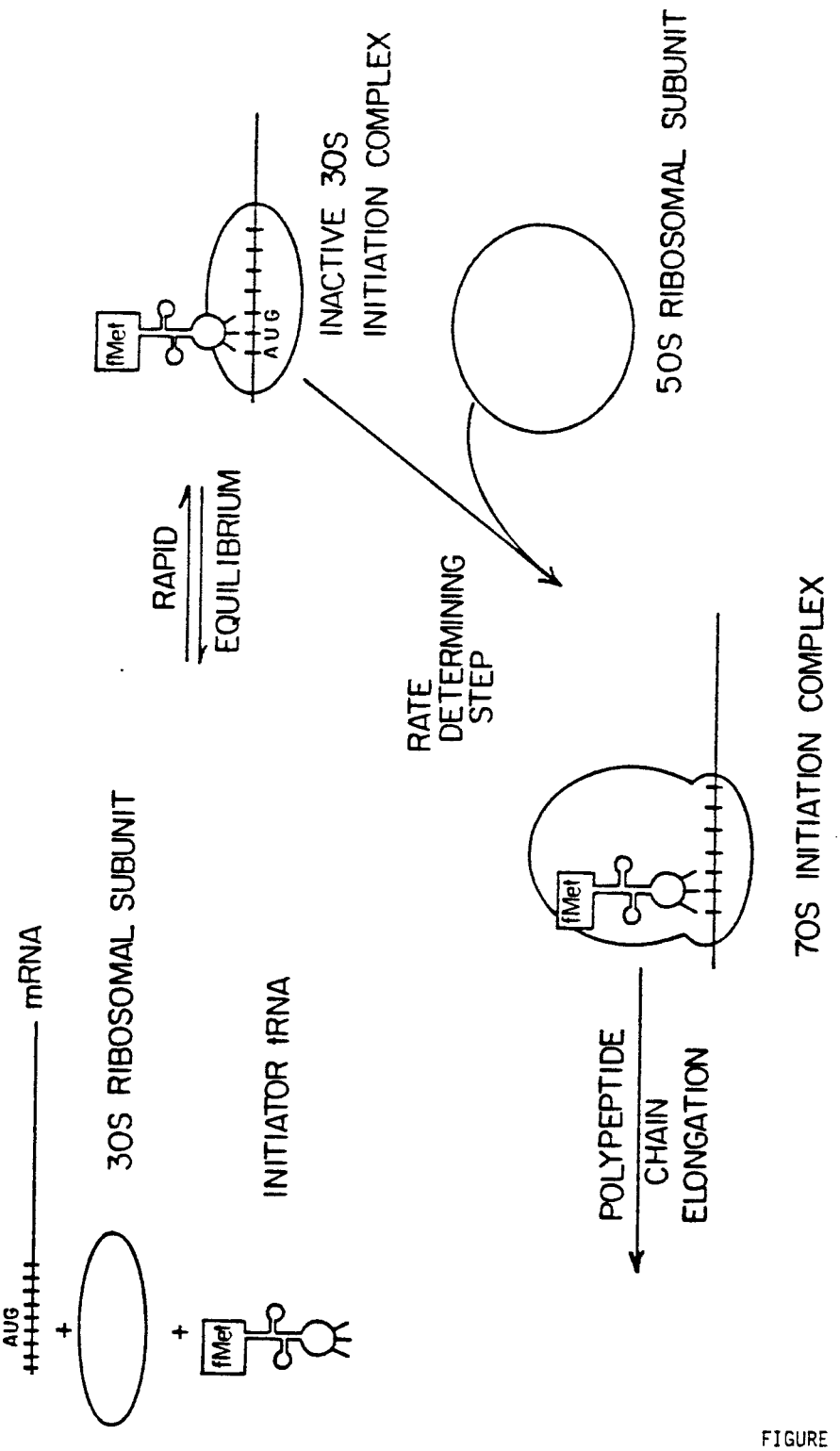

RIBOSOME ACTIVATION
AUG
mRNA
30S RIBOSOMAL SUBUNIT
fMet
INITIATOR tRNA
RAPID
EQUILIBRIUM
fMet
AUG
INACTIVE 30S
INITIATION COMPLEX
RATE
DETERMINING
STEP
50S RIBOSOMAL SUBUNIT
fMet
70S INITIATION COMPLEX
(ACTIVE)
POLYPEPTIDE
CHAIN
ELONGATION

FIGURE 4

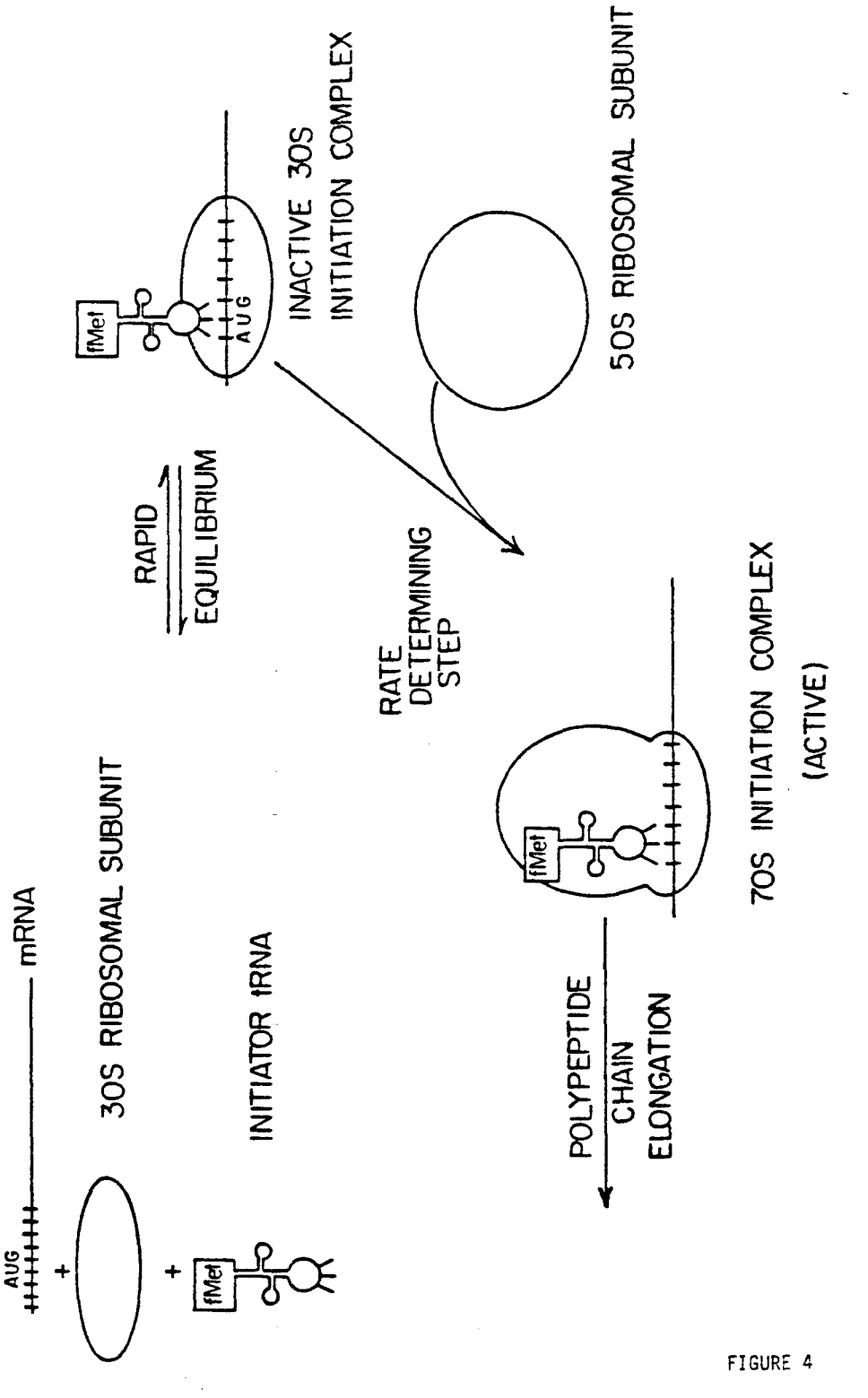

RIBOSOME ACTIVATION
AUG
mRNA
30S RIBOSOMAL SUBUNIT
INITIATOR tRNA
fMet
RAPID
EQUILIBRIUM
INACTIVE 30S
INITIATION COMPLEX
RATE
DETERMINING
STEP
50S RIBOSOMAL SUBUNIT
70S INITIATION COMPLEX
(ACTIVE)
POLYPEPTIDE
CHAIN
ELONGATION

FIGURE 4

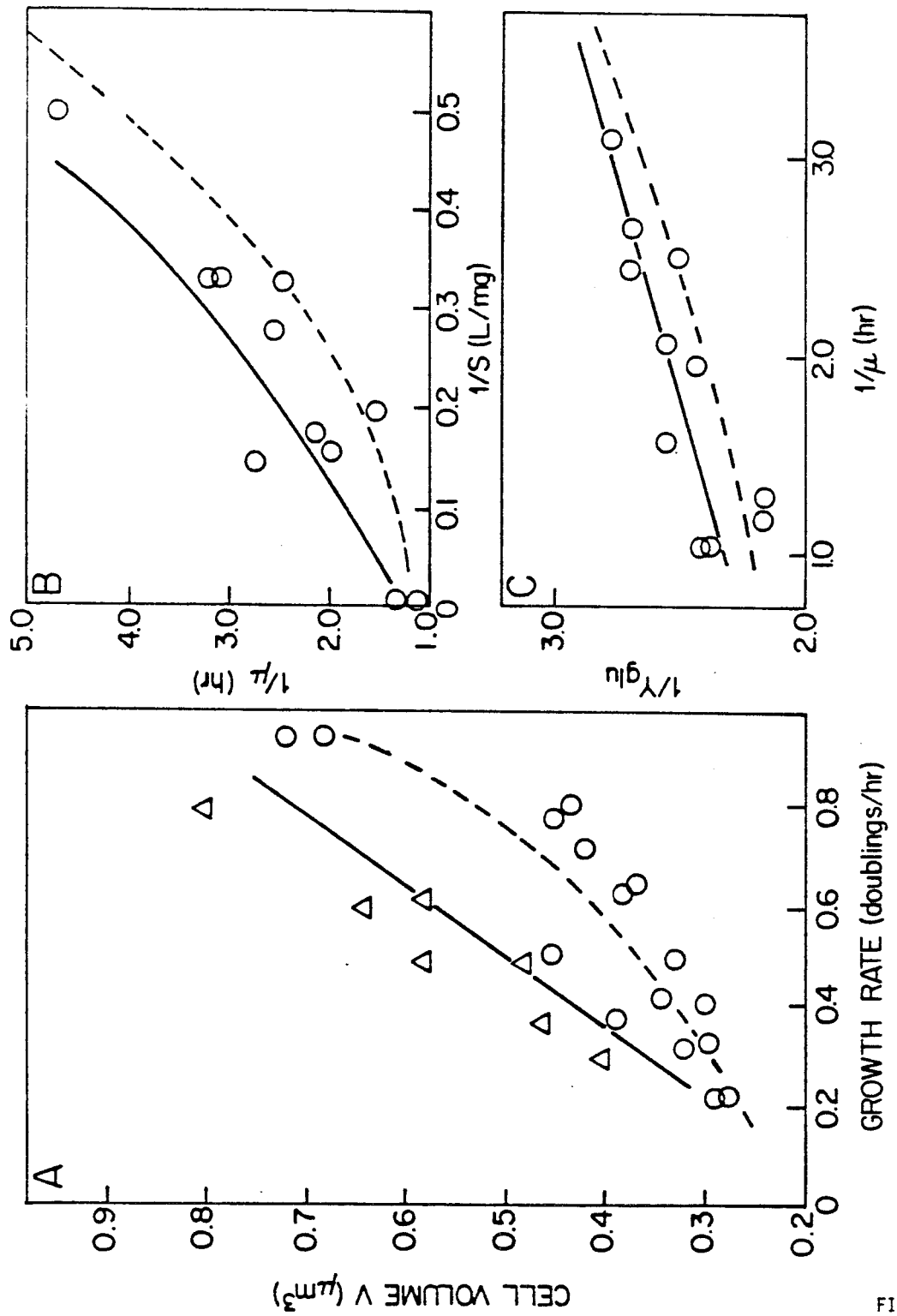

A
B
C
CELL VOLUME V (μm^3)
GROWTH RATE (doublings/hr)
1/μ (hr)
1/S (L/mg)
$1/Y_{glu}$
1/μ (hr)

FIGURE 5

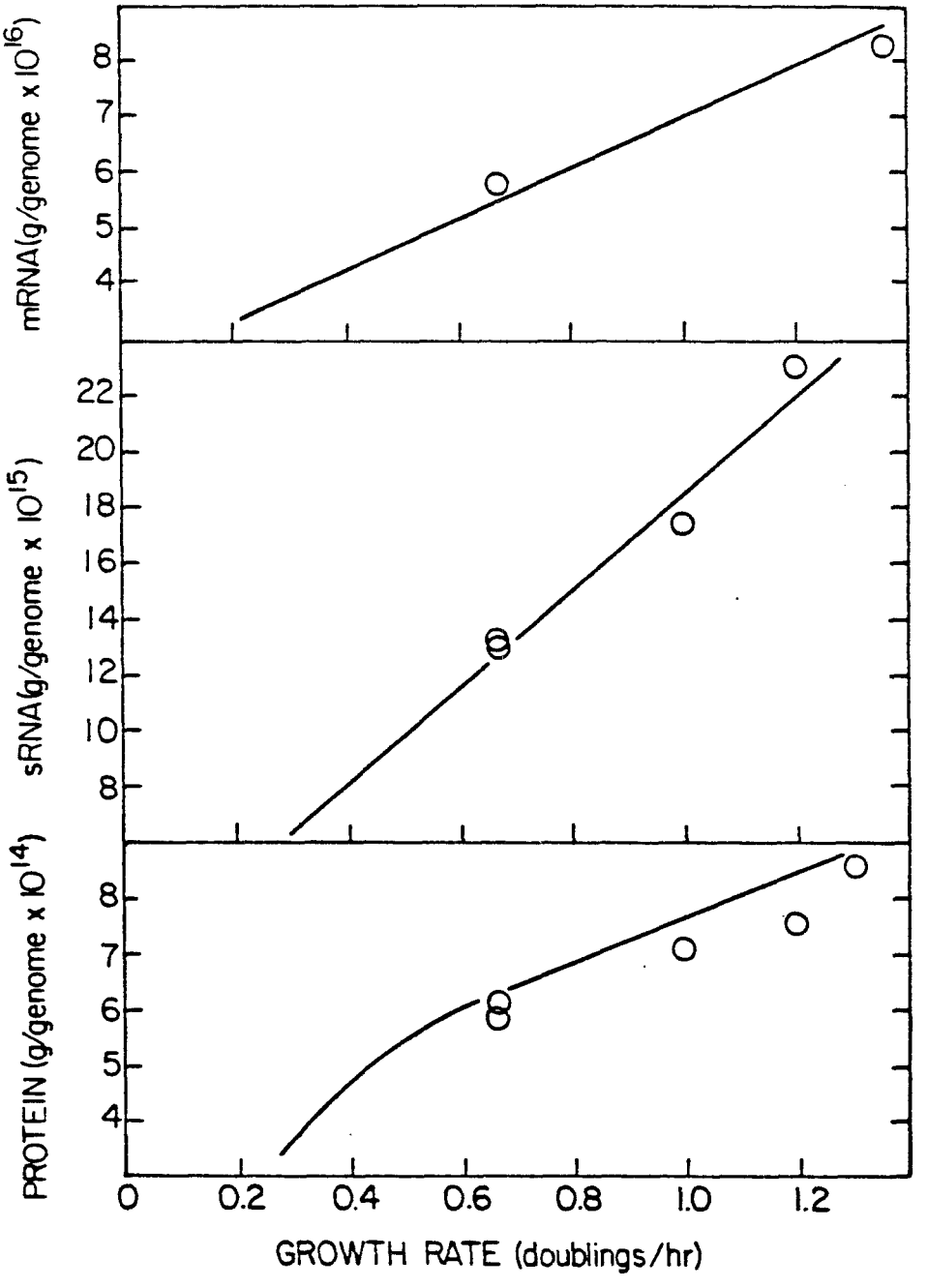

mRNA(g/genome x 10^16)
8
7
6
5
4
sRNA(g/genome x 10^15)
22
20
18
16
14
12
10
8
PROTEIN (g/genome x 10^14)
8
7
6
5
4
0
0.2
0.4
0.6
0.8
1.0
1.2
GROWTH RATE (doublings/hr)

FIGURE 6

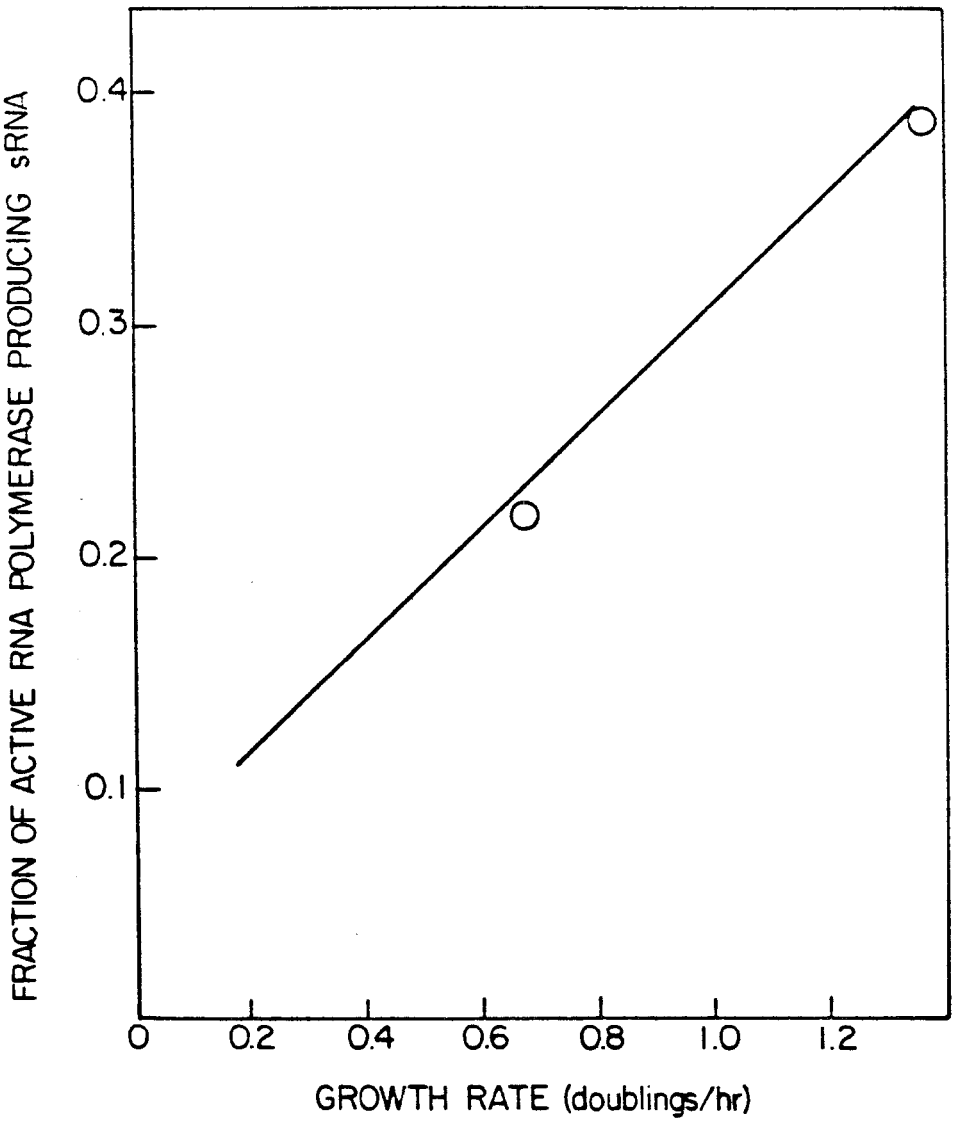

FRACTION OF ACTIVE RNA POLYMERASE PRODUCING sRNA
0.4
0.3
0.2
0.1
0
0.2
0.4
0.6
0.8
1.0
1.2
GROWTH RATE (doublings/hr)

FIGURE 7

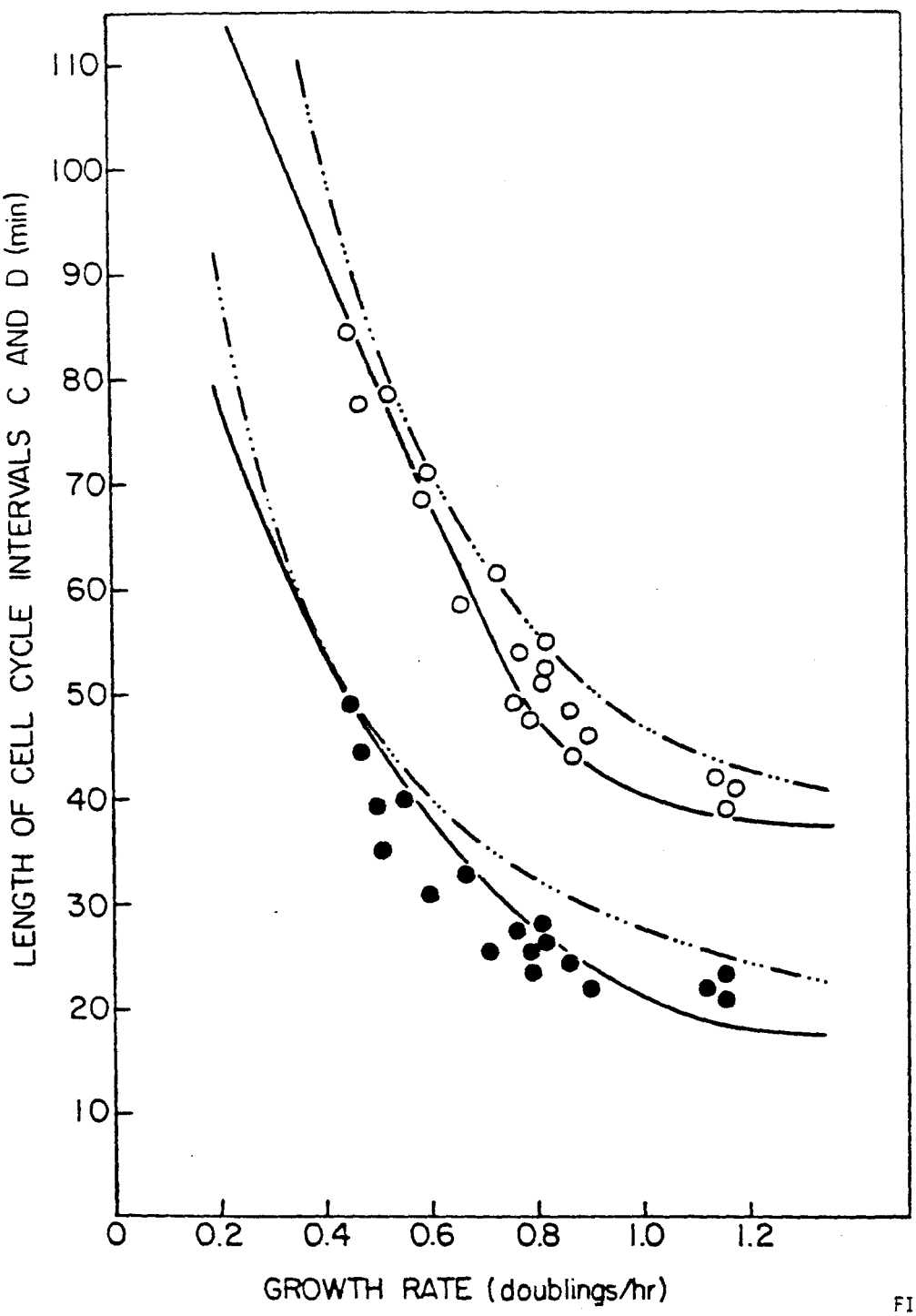

LENGTH OF CELL CYCLE INTERVALS C AND D (min)
110
100
90
80
70
60
50
40
30
20
10
0
0.2
0.4
0.6
0.8
1.0
1.2
GROWTH RATE (doublings/hr)

FIGURE 8

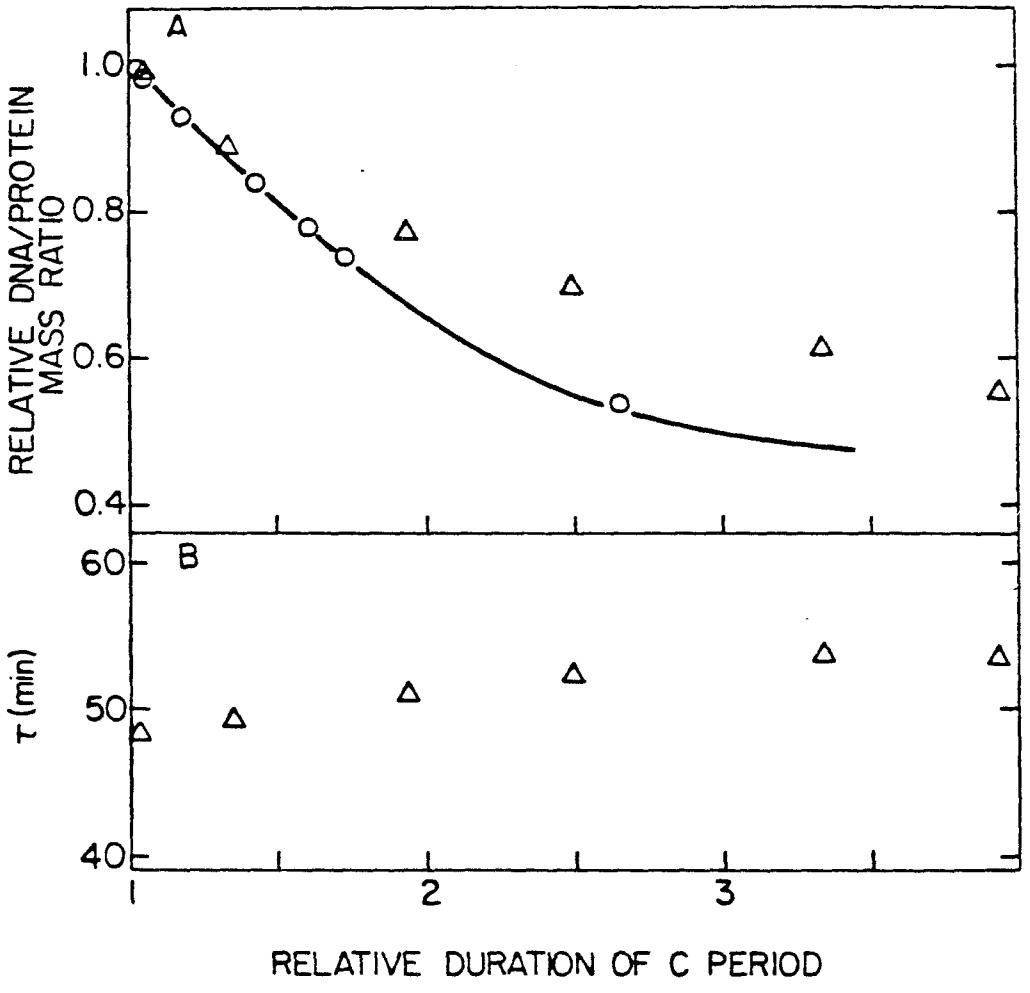

A
B
1.0
0.8
0.6
0.4
RELATIVE DNA/PROTEIN MASS RATIO
60
50
40
τ (min)
1
2
3
RELATIVE DURATION OF C PERIOD

FIGURE 9

CHAPTER 2

Simulations of Host-Plasmid Interactions in *Escherichia coli*: Copy Number, Promoter Strength and Ribosome Binding Site Strength Effects on Metabolic Activity and Plasmid Gene Expression

INTRODUCTION

Recombinant cell metabolism is a network of interdependent reactions, complicated further by the inclusion of the macromolecular products derived from the genetically engineered DNA. Introducing foreign DNA into cells alters the genome of the host organisms. This situation is in many ways analogous to that which results when mutations occur within chromosomal DNA. In each instance, the cell is presented with a metabolic stress to which it must adapt utilizing the pathways and regulatory responses that constitute cellular metabolism. The adaptive response will depend, qualitatively and quantitatively, upon the pathways affected by the stress, the magnitude of the perturbation, and the regulatory structure controlling the affected reactions.

As with mutation, the range of perturbations to cell metabolism potentially arising from recombinant vector insertion is enormous. There is nearly an infinite variety of DNA that can be introduced into bacterial cells. Vectors can be constructed using both procaryotic and eucaryotic genetic material. There are also three general vector types used to carry cloned genes: plasmid, cosmid, and phage.

Host cell metabolism, in the case of the simple bacterium *Escherichia coli*, may include up to several thousand different reactions. Highly structured models of cellular metabolism are required in order to evaluate the ramifications of vector insertion. At the same time, the modeling must necessarily involve substantial lumping of components and reactions, since neither the kinetic parameters nor the reaction sequence itself is known for many cellular functions. Sufficient information exists, however, for the processes involved in RNA and protein synthesis. A single cell model which includes detailed formulations for transcription and translation has been developed (1). The objectives of this work are to adapt the existing model to encompass recombinant systems and to use the resulting formulation both to identify the key host-vector interactions and to evaluate alternative stra-

tegies for improving cloned gene productivity in these systems.

We have developed a very general model structure into which particular vector systems may be easily placed. In order to evaluate the suitability of the formulation, it was necessary to choose a specific vector system. Due to the relative abundance of experimental data surrounding plasmid-based recombinant systems, plasmid DNA was chosen as the model vector. The alterations made to the metabolic model are not specific to plasmid systems, but are valid for any vector system. The effect of considering plasmid as the model vector is primarily related to the number and type of genes carried on the vector.

In developing a recombinant system model, it is necessary to use a formulation containing detailed kinetics and control structures for the metabolic pathways that significantly interact with vector-derived molecules. Figure 1 shows the metabolic pathways involved in host-plasmid competition. Those processes are DNA replication, transcription and translation. Plasmid replication, once initiated, requires much the same enzymatic machinery as chromosomal replication. It competes directly with the chromosome for the deoxyribonucleotides, enzymes and cofactors involved in DNA replication. Also, competition for replication initiation factors may occur, depending on the plasmid origin of replication and its function.

Plasmid DNA (DNAp) will also compete with chromosomal DNA (DNAc) for DNA binding proteins, including RNA polymerase (RNAP). Central to transcription, this enzyme binds both specifically and non-specifically with DNA. Non-specific binding leads to the redistribution of RNAP between DNAc and DNAp on the basis of the number of base pairs of each DNA type present. That is to say, part of the RNAP distribution depends upon the relative masses of the two DNA species. Specific polymerase binding affects the distribution of RNAP upon various operon promoters, dependent upon the relative strength and number of promoters found on each type of DNA. As with replication, competition occurs for the full

range of enzymes and precursors involved in transcription.

The same type of competition occurs for translation. Plasmid derived mRNA (mRNAp) and chromosomal mRNA (mRNAc) are the molecules which direct protein synthesis in conjunction with ribosomes. The composition of the protein made in the cell will be determined by the relative abundance of the different mRNA species, as well as by the affinities of mRNAc and mRNAp for free ribosomes. Consequently, plasmid DNA competes with the chromosome for control of the available protein synthetic machinery of the cell.

Considering the variety of competitive interactions which come into play for host-vector systems and the centrality of these processes for normal cell metabolism, it is necessary to address the influence of the vector on cell activity, on vector replication and on the expression of the genes it carries. Plasmids normally contain no genetic information for enzymes involved in any of the competitive interactions shown in Figure 1. DNAp uses host-mediated transcription and translation apparatus to regulate plasmid replication initiation and to synthesize the desired cloned gene product(s) (CGP). The nature of this product determines whether the vector is completely parasitic or is in some way symbiotic relative to the host cell. The product may be a molecule one wishes to harvest and purify, valuable for its pharmacological or enzymatic activity. The CGP may also be an enzyme meant to enhance the activity of the host by opening novel metabolic pathways, or by enlarging the capacity of existent pathways. The major thrust of effort in biotechnology thus far has been with CGP intended for harvest and purification. The focus of this work will also be on the case of CGP with no positive contribution to cellular activity.

Assessing the productivity of recombinant systems and correctly simulating the response of the host cell to the attendant stresses involves evaluating the kinetics for reactions constituting most of the key metabolic pathways in cellular meta-

bolism. Detailed representation of the kinetics and control for individual steps in transcription and translation as well as for the synthesis of energy sources and precursors is necessary. While the primary goal of this effort is to understand the impact of the host-vector interaction upon the kinetics of CGP synthesis, this in turn depends upon the network structure of the metabolic system. Integration of all the key reactions and major processes is necessary, since the kinetics of any process is ultimately determined by the redistribution of metabolic activity among all pathways, resulting from the introduction of a metabolic stress.

MODEL STRUCTURE

The mathematical formulation used in this work details many of the reactions central to DNA replication, transcription and translation for glucose-limited, balanced growth of the bacterium *Escherichia coli*. Control interactions and reaction mechanisms have been formulated using as much detail as possible for the processes of transcription and translation. Major features of the host cell model are summarized next; further details are provided in the first chapter.

DNA replication frequency is controlled by a repressor/antirepressor system analogous to the inhibitor dilution model of Pritchard, Barth & Collins (2). Repressor protein and antirepressor protein synthesis are directly related to chromosomal protein synthesis, and their relative amounts determine the timing of initiation. Repressor protein is synthesized in a burst at the time of replication initiation; antirepressor synthesis continues throughout the cell cycle. The progress of the replication forks and their locations along the chromosome are monitored. In this way the availability and distribution of promoter and non-promoter sites are always known, and are used to calculate the RNA polymerase equilibrium distribution.

RNA polymerase is accounted for in both the core enzyme and holoenzyme

configurations. Nonspecific binding of core polymerase to DNA and binding of holoenzyme to promoter and non-promoter sites are also calculated. Different promoter strengths, as well as promoters with different regulatory features, have been included. Synthesis of a particular RNA species requires the formation of the appropriate RNAP-promoter complex. Ribosome synthesis is limited by the availability of rRNA, which in turn depends upon the promoter activity and gene dosage of rRNA operons. The operons are located at specific chromosomal positions. In conjunction with an explicit accounting of the replicating fork positions along the chromosome, this scheme allows the supply of protein synthetic machinery calculated by the model to closely parallel the supply observed experimentally, as was reported in Ref. 1.

Protein synthesis results from the activation of an idle ribosome by formation of an mRNA-ribosome complex. Ribosomes recognize and interact with different mRNA species primarily through the ribosome binding site located on the mRNA. Protein synthesis is then distributed among various proteins as a result of the relative abundance of the relevant mRNA and the strength of the ribosome binding sites found on the particular mRNA.

The model formulation described previously (1) was specifically developed to enable easy extension to recombinant systems. There are three features that make this model formulation of *E. coli* cellular metabolism well suited for extension to recombinant systems. The first is calculation of the timing of chromosomal replication based solely upon factors directly related to chromosomally derived protein synthesis within the cell, rather than on the concentrations of some macromolecular precursors found inside the cell. There is no need, unless demonstrated experimentally, to postulate a specific effect of vector presence upon the timing of chromosomal replication. This would be difficult to formulate justifiably at this time, insofar as control of chromosome replication initiation is still poorly understood.

Second, protein synthesis is driven by the interaction of mRNA species with ribosomes. Competition between chromosomal- and vector-derived mRNA, with attendant effects on the synthesis of both chromosomal and vector-mediated protein (PROc and PROp respectively), can thus be easily simulated and studied. No empirical correlations between cellular vector number (Np) and protein synthetic capacity are used in this model.

Third, and at the core of the model's suitability, is the detailed calculation of the distribution of RNA polymerase upon various promoter regions found along DNA. RNA synthesis from vector genes and cellular genes is determined by such calculations. In this way, cellular plasmid content and cloned gene expression affect at every level the metabolism of the cell. The magnitude and manner of the effects can be studied, since they are not fixed *a priori* by the use of empirical correlations. Rather, one chooses relative promoter and ribosome binding site strengths for the vector genes and mRNAp, and the model simulates the alterations in host cell metabolism and CGP synthesis which result.

Using this framework, one is able to explore the full range of possibilities with regard to cloned gene promoter strength, promoter regulatory structure, mRNAp ribosome binding site strength, vector copy number and CGP stability, in order to understand the dynamic response of the host cell to such perturbations. Once the details of cellular response to such stresses are uncovered, one can develop suitable cultiviation schemes or metabolic engineering approaches to expand the capacity of those pathways found to be limiting, with the ultimate objective being the optimization of CGP productivity. In particular, since this framework allows the alteration of every aspect of the vector used, from the replication control mechanism to the CGP promoter strength (PS) to the mRNAp ribosome binding site strength (RBS), the model can be used to test and to optimize recombinant vector design, taking into account concomitant, induced alterations in recombinant cell

kinetic properties.

Vector Implementation

Implementation of recombinant vector metabolism into the basic model formulation is for the most part straightforward. Functional dependence of synthesis and degradation reactions upon relevant precursors and cofactors for all macromolecules (DNA, RNA, protein) is assumed to be the same regardless of the DNA type from which the macromolecule was derived. For example, PROp synthesis has the same functional dependence upon amino acids and ATP level as does PROc synthesis. Degradation rates are treated in a similar fashion. Table I contains the kinetic expressions for the synthesis and degradation of chromosomal- and plasmid-derived macromolecules.

As indicated in the equations of Table I, the synthesis terms for mRNA and protein are dependent upon the number of RNAP-promoter and ribosome-mRNA complexes formed for the particular compound. The calculation of RNAP distribution is treated the same way for the recombinant system as for the host cell case. Vector addition to the cell increases the number of non-specific binding sites available for polymerase binding. The plasmid promoter sites are treated as additional promoter sites with specified affinity for RNA polymerase binding. Table II shows the equilibrium reactions involving RNA polymerase. As is the case for chromosomal promoters, transcription subsequent to polymerase binding at plasmid promoter sites makes the promoter site available for further binding interactions. As a result, the promoter site concentration is considered constant during the equilibrium calculation.

The assumption of invariant concentration during equilibrium calculations is also made for mRNAc and mRNAp with regard to translation initiation on ribosomes. Translation is presumed to expose the ribosome binding site on mRNA to

repeated mRNA-ribosome interactions. The fraction of total ribosomes involved in translation of mRNAc and mRNAp are, respectively, FRACM and FRACP in Table I, Equations 1a and 2a. These fractions are calculated using the following equations:

$$FRACM = Keq_c * mRNAc_o / (1+Keq_c * mRNAc_o + Keq_p * mRNA_{P_o}) \quad (R1)$$

$$FRACP = Keq_p * mRNAp_o / (1+Keq_c * mRNAc_o + Keq_p * mRNAp_o) \quad (R2)$$

In these equations, Keq_i are the equilibrium ribosome binding constants and $mRNAi_o$ are the concentrations of mRNA species at the time of the calculation.

Terms have been added to the material balances for macromolecular precursors and energy compounds reflecting the consumption of these molecules during the synthesis of DNAp, RNAp, and PRPp, where RNAp includes mRNAp and other untranslated RNA derived from the plasmid. Table III lists the amended equations, with the plasmid-related terms in boldface.

Two changes were made to the formulation unrelated to those mentioned above. The first pertains to the reduction of PROc concentration within plasmid-containing cells relative to levels in plasmid-free cells. Developed using vector-free metabolic data, the kinetic expressions for the rates of synthesis of amino acids, ribonucleotides, cell wall precursors, glycogen, the enzyme involved in crosswall synthesis, and the antirepressor protein involved in chromosomal initiation depend upon the concentrations of various cellular compounds in conjunction with total cell volume or the ratio of protein to DNA. This is adequate for the plasmid-free case, since normal cell regulation is in force, so that cell volume and various precursor levels are coupled to PROc levels. Vector insertion upsets this coordination. Normal pathway balance is disrupted, leading to the uncoupling of chromosomal protein levels and cell volume. At a given cell volume this leads to a reduction in

the concentration of PROc for the recombinant system relative to the plasmid-free case, since plasmid proteins and RNA contribute to the total volume. The account for this dilution of metabolically active protein, a multiplier was added to the kinetic expressions for the compounds listed previously. The multiplier is equal to the fractional decrease of PROc concentration relative to the average level for the plasmid-free cell. The correction factor used,

$$[PROc]\ recombinant\ /\ [PROc]\ vector-free,$$

was determined by referring to the simulation results for the vector-free case. The resulting expression is:

$$PROc\ \ recombinant\ /\ 0.68\ *\ total\ \ cell\ \ mass.$$

The second change made was the addition of a term to the expression for the rate of synthesis of deoxyribonucleotides. It has been postulated that the enzyme responsible for the conversion of ribonucleotides to deoxyribonucleotides, ribonucleotide diphosphate reductase (rdr), is somehow closely associated with the chromosome, and that the regulation of its level and activity is such that the cellular supply of dNTP is always sufficient for DNA synthesis (3), which is the case for plasmid-free conditions. Data presented by Seo and Bailey (4) suggest that the same holds for the plasmid case. Measuring the time needed to replicate the chromosome, they found that the plasmid-bearing cell synthesized chromosomal DNA faster than the plasmid-free cell, despite a decrease in cellular growth rate. This result further suggests that competition between DNAp and DNAc for replication enzymes and DNA precursors is not a significant component of the host-plasmid interaction. Avoiding dNTP limitation for the model cell with plasmids necessitated augmenting "normal" dNTP synthesis levels by a factor relative to the vector/chromosome DNA mass ratio. The resulting term, [1+2*DNAp/DNAc],

was determined by simulation experience. These are the only changes from the plasmid-free model, and they reduce to unity for a plasmid-free cell.

One final consideration underlying implementation of recombinant system modeling is the control of vector replication initiation. In the case of ColE1 plasmids, initiation is primarily controlled by the interaction of two RNA species, and is further modulated by a very small plasmid-derived protein (5-7). Any particular replication scheme may readily be accommodated in the single-cell formalism, given the flexibility inherent in this model. What is lacking for most vector replication schemes is sufficient information regarding the kinetic and equilibrium constants for the constituent interactions. This model formulation could be used to investigate the suitability of hypothesized vector replication initiation mechanisms. Since the goal of this work was to evaluate the recombinant cell model, not particular replication mechanisms, the frequency of initiation was an input to the system. In this way, we were able to simulate the effect that maintenance of a particular copy number has upon host cell activity. Furthermore, we have modeled the vector in this work after a ColE1-type plasmid. To approximate the metabolic load on the cell caused by plasmid presence, each plasmid was modeled to contain two promoters with strength and regulatory attributes identical to those of an "average" chromosomal promoter. The RNAp species resulting are not translated, simply degraded with half-lives similar to mRNAc. These represent approximately the RNA species involved in plasmid replication initiation-frequency control described previously.

VECTOR EFFECTS ON BACTERIAL GROWTH

Simulation and Experimental Results Comparison

Testing the suitability of the revised formulation requires a comparison of simulation results with available experimental findings. The most widely reported

effect of host-plasmid interactions is growth rate depression, although this has not been reported for all cases. Several experimental studies have indicated that growth under unrestricted conditions, using complex medium or supplemented minimal medium with excess glucose, exhibits no slowdown for plasmid-carrying cells (P^+) relative to plasmid-free cells (P^-) when both are grown in pure culture (8-10). This is not surprising in light of the findings of Klemperer et al. (11) that P^+ strains have greater nutritional requirements than P^- strains, particularily for Mg2+, K+, Fe2+ and $PO3^{3-}$. For growth conditions in which one of these nutrients (11,12) or the carbon source (10,12) was growth limiting, plasmid-free cells grew more quickly than those containing plasmid.

Populations behave differently in mixed cultures than in pure cultivation. P^+ strains show significant growth disadvantages relative to P^- strains for all media types when grown in mixed culture (8-10). For the recombinant systems studied, the plasmids were found to be inherited stably, so the observed growth rate depression for P^+ cells in mixed culture was not an artifact caused by a significant rate of plasmid loss by the P^+ population. A careful study of recombinant cell growth and productivity reported by Seo and Bailey (13) using ColE1-type plasmids and work with R plasmids done by Engberg, Hjalmarsson and Nordstrom (14) showed that plasmid number dependent growth-rate depression occurs for cultivation in complex, unrestricted medium or minimal medium. One concludes, then, that host-plasmid interactions often result in a decrease in cell activity necessary for growth.

CGP synthesis in host-plasmid systems is most easily manipulated experimentally by controlling the number of plasmids and thereby the number of cloned genes found in the cell. Of particular importance to those concerned with optimal CGP productivity are the effects of plasmid size and plasmid number upon growth rate. It has been found that increasing either plasmid size or number leads to

greater degrees of growth-rate depression (13-17).

Table IV summarizes experimental results obtained using ColE1-type plasmids (13) and simulation output showing the effect of plasmid number on growth rate. The qualitative agreement is striking. The model predicts the relatively sharp initial drop in growth rate as plasmids are added to the cell, followed by a gradual decrease as Np increases. The accentuated decrease in specific growth rate indicated by the model for high plasmid number can be likened to the experimental observations (14,17) that R plasmid amplification leads to increasing decline in cell growth rate, culminating in loss of cell viability for extremely high plasmid number (2000-3000 plasmid/cell). Overall, the model accurately simulates the qualitative aspects of plasmid number effect upon host growth rate.

One additional aspect of cellular metabolism affected by vector presence is DNA metabolism, in particular, the cell cycle parameters C and D. These parameters represent the time necessary to replicate the chromosome and the time period between completion of DNA replication and cell division, respectively. Seo and Bailey (4) have found that both time periods decrease for cell growth faster than one generation per hour. Figures 2a and b illustrate the calculated and experimental C and D period dependence upon the number of plasmid per cell. The model results indicate a decrease in C period, although the dependence of C period on plasmid content obtained from simulation is somewhat different from that experimentally observed for pMB1 plasmids in *E. coli* HB101. The calculated dependence can be shown to be affected by the values chosen for the cloned gene promoter strength. Two simulations are presented; one with PS=1 and RBS=1, the other with PS increased to 2. The simulated dependence of C period on plasmid number does not quite match experimental results, but the general trend of C period decrease with increasing plasmid content is predicted.

D period behavior is, however, quite different from available experimental

data. The model indicates a decrease in D interval relative to a plasmid-free cell at high plasmid number. Contrary to experimental data for pMB1-type plasmids, simulations show an increase in D period at low plasmid content. This discrepancy illustrates the central problem encountered when attempting to model large, complex systems. Care must be taken to include as much information as possible for important model processes if one wishes to apply the formulation to circumstances significantly different from those used in model construction. This model was developed using vector-free behavior for different growth rates. (For a full explanation of model formulation, see (1).) The C and D periods calculated under those conditions agree closely with experimental values. Only a general dilution of enzyme concentration was considered in treating the effect of vector presence upon the enzyme involved in crosswall formation and upon the triggering of cell division. The particulars of cell division control and crosswall formation will have a major impact on D period length. Some researchers (14,17) have noted specific effects of plasmid presence on cell division behavior, as extreme as inhibiting crosswall formation for multiples of the cell doubling time. Unfortunately, there is insufficient information available at present regarding the mechanism and timing of cell division to account for such specific interactions in the model. By the use of a model formulated at the present level of detail, however, such discrepancies between model and experiment can be used to suggest undiscovered mechanisms, in this case, the influence of plasmid presence on cell division control. Application of such highly structured models for biological discovery has been emphasized by Domach and Shuler (18), and this theme was evident in the preceding single-cell model development (1) in which a new antitermination process was indicated.

Data detailing the quantitative aspects of host-plasmid interactions are very limited, particularly with respect to CGP productivity. One systematic study of this question was reported by Seo and Bailey (13). Their results indicate that the

specific productivity of recombinant systems, expressed as CGP synthesized per gene copy, decreases as the plasmid number in the cell increases. Table V summarizes that data as well as simulation results using this model. Qualitative agreement between theoretical and experimental results is evident.

The data available for *E. coli* plasmid systems support the trends indicated by model simulations. It is important to note that the behavior used to test the model is primarily integrated behavior. That is, the behavior represent time-averaged quantities, integrated over the course of a cell doubling. We do not claim to have included all possible host-plasmid interactions in this formulation. However, using a model for cellular metabolism which was evaluated using detailed cellular parameters, we have some expectation that trends indicated by the model would have some correspondence with reality, at least qualitatively.

Since a molecular description of cell metabolism is used to calculate recombinant cell growth, the model can simulate details of the system response that are difficult to assay experimentally. Understanding the metabolic rearrangements which underly observable cellular responses to the stress applied by plasmid presence may facilitate the optimal design and operation of these systems. Detailed simulation results are presented in the spirit of suggesting possible cell responses and providing predictions for testing in future experiments.

Table VI lists details of transcription and translation resulting from simulations of cell growth with amplified plasmid number. Results indicate that the computer cell expands the utilization of both transcription and translation pathways by enlarging the fraction of RNA polymerase and ribosomes that are active as the load presented by plasmid metabolism increases. Acting counter to the expansion of macromolecular synthetic capacity is the more vigorous competition for available RNAP and ribosomes represented by elevated plasmid numbers. This directs synthesis away from chromosomally derived molecules, diminishing the capacity of

the computer cell to sustain plasmid-free growth rates. Under plasmid-free conditions, experiments indicate that decreasing limiting nutrient concentration will cause the growth rate to drop for glucose-limited growth. This triggers a redistribution of RNA polymerase among the chromosomal promoters, leading to an increase in the fraction actively transcribing mRNA. As a result, ribosome pool size is diminished as sRNA synthesis decreases, reflecting a reduced need for protein synthetic capacity.

Vector presence triggers the same regulatory mechanism. As the simulations indicate, increasing the number of plasmid gene copies in the cell leads to a decrease in the fraction of polymerase transcribing sRNA or mRNAc. The decline in sRNA transcription is the more striking and more significant, exhibiting reduction of 50% or more as plasmid number increases. Additionally, the alteration of mRNA population engendered by the increase of mRNAp leads to decreased translation of mRNAc. The expansion of the system capacity by activating greater fractions of RNA polymerase and ribosomes ultimately proves insufficient to counter the decrease in PROc synthesis.

An interesting result, unverified experimentally because of the difficulty involved in measurement, is the change of rate-limiting process as amplification increases. Referring to data for transcriptional and translational efficiency in Table VI, we note that first ribosome availability, then RNA polymerase is the more significant bottleneck as copy number increases. As the plasmid number increases, transcription efficiency decreases less slowly than translation efficiency initially but then continues to decline while translation efficiency goes through a local maximum before continuing to decrease. Further simulation results for transcription, translation and CGP productivity are presented in greater detail in the Appendix.

PREDICTIVE SIMULATIONS

Comparison of model results with experimental data is useful to establish the accuracy of the formulation. The goal of modeling is, however, predicting behavior for conditions different from those used to develop the model and to use model structure to learn about some underlying system responses not amenable to direct measurement. The model can then be used as a research tool, with simulation results indicating interesting parameters to study experimentally. The simulations and discussions that follow exemplify such model utility.

Promoter and Ribosome Binding Site Strength Variation

To further investigate the relative contributions of transcription and translation to the host-vector interaction, a series of simulations were run, varying the CGP promoter strength (PS) or the mRNAp ribosome binding site strength (RBS). Table VIIa shows growth parameters for the case PS=1, RBS=1; PS=5, RBS=1; PS=1, RBS=5, where unity denotes the base case considered above (average *E. coli* promoter and ribosome binding site strengths). The plasmid replication frequency was assumed equal for each case. These results indicate that the most significant effect of vector presence is the disruption of RNA polymerase equilibrium, not ribosomal redistribution. Both cases have the same redistribution of ribosomes onto mRNAp. The difference in growth rates for the two cases seems to be due predominantly to the higher number of ribosomes maintained in the RBS=5 case, resulting from the lower redistribution of RNA polymerase onto plasmid promoters.

Furthermore, these results suggest that the more effective means of increasing recombinant system productivity is by maximizing the strength of the mRNAp ribosome binding site. This boosts CGP level by the same factor as increasing the promoter strength of the cloned gene without the attendant depression of cellular

activity caused by the decrease in chromosomal transcription, which would occur with increased promoter strength. This conclusion is indicated by the productivity data of Table VIIb. Most pertinent are the values for CGP synthesis rates and the relative success of competition for RNAP and ribosomes for the two cases where PS or RBS has been changed from the base value of one. Enhancing the equilibrium constant of the CGP promoter by a factor of 5 increases the efficiency of polymerase-promoter formation by only a factor of 3 relative to the case of PS=1. RBS=5, resulting in a ten percent increase in product synthesis rate from the base case (PS=1, RBS=1). The same increase in the ribosome binding constant results in a greater than sixfold increase in ribosome utilization by the plasmid-derived mRNA relative to the case with raised promoter strength, leading to a rate of product synthesis fifty percent higher than that predicted for the base case.

Metabolic Engineering Approaches

Metabolic engineering is the term used to describe the use of vector-borne genes to alter the function or capacity of existent metabolic pathways. Use of this approach in conjunction with the system being discussed would allow the simultaneous application of metabolic perturbations and augmentation of the most seriously affected aspect of cellular activity. Two sets of simulations representing metabolic engineering applications further support the contention that enhanced vector-specified affinity for ribosomes is a better strategy for improving productivity than enhanced promoter strength. The plasmids considered contained two cloned genes. These encode for a protein to be harvested and either RNA polymerase core enzyme (rpo operon) or rRNA (rrn operon). In this fashion, the capacity of the processes central to the host-vector interaction and recombinant system productivity are expanded. The rpo and rrn genes were considered to be expressed using "average" promoter strength and transcription patterns, identical to those

for mRNAc operons. The resulting mRNAp had a ribosome binding site strength of one relative to mRNAc RBS. Table VIII shows the growth characteristics for cases of enhanced PS or RBS for the harvested CGP. In all cases evaluated, growth rate increased as did productivity, indicating an expansion of cellular synthetic capacity.

The difference in magnitude between adding rpo or rrn operons to the plasmid system is, however, more dramatic than one might expect. The large difference in effect is indicated by the productivity data shown in Table VIII, particularly the fraction of total protein which is CGP for each case. Only twenty percent of the total protein is plasmid-derived for the base case (PS=1, RBS=1). One can see that the addition of the rpo operon to the plasmid expands the entire system, not simply a subsystem, as is the outcome for augmentation of translation via rrn gene addition. Adding the rpo gene opens up the whole range of metabolic functions more directly than would expanding the ribosome population. Surprisingly, the ribosome population for this case is also more than twice that calculated after addition of the rRNA operon to the plasmid.

DISCUSSION

A detailed single-cell model for *E. coli* growth has been extended to include recombinant DNA in the cell. The model formulation is quite general, allowing simple implementation of hypothetical vector replication mechanisms. In addition, the effect of vector system parameters on recombinant system growth and productivity may be easily investigated. In this work the effect of cloned gene promoter strength and mRNAp ribosome binding site strength were evaluated for ColE1-type plasmid systems. One implication is, given the regulatory structure of the host cell, competing for transcription and translation activity is about twice the perturbation upon cell activity that would result if the only interaction for the

host-plasmid case were competition for ribosomes. The other implication is that expanding transcription capacity is more efficient than enlarging translation capacity for the purpose of increasing overall metabolic activity, and thereby cloned gene protein productivity.

Models of this complexity allow one to evaluate cellular activity at a molecular level to explore the underlying causes of observable system behavior. This advantage can be a major liability if one does not recognize the limitations of modeling in general. The fact that this model can correctly simulate, on a qualitative level, recombinant system behavior does not validate every aspect of the system response detailed by the model. Every model of cellular behavior is necessarily an approximation of reality. Until more information is available for the detailed aspects of recombinant systems, these modeling results are speculative at best. This does not, however, invalidate the inclusion of this detail in discussion of such systems. The major function of this model is as a research tool that is useful in the evaluation of hypothesis within an internally self-consistent, quantitative framework. The work presented in this paper is to be understood in that light. The results suggesting particular cloned gene regulatory changes or metabolic engineering approaches, which give increases in productivity, are to be used as indicators of interesting parameters to manipulate and study, not as rules to follow for recombinant system construction and cultivation. As these parameters are studied, the modeling will become more detailed and more able to be applied accurately to novel, unstudied systems.

NOMENCLATURE

$A2$	glucose
$CA2$	extracellular concentration of glucose (mg/L)
CGP	product of the cloned gene carried on the plasmid
$DATPs$	rate of consumption of ATP used as an energy source to drive synthetic reactions
d_i	ATP requirement to drive a given reaction (mol ATP/g)
$DNAc,p$	chromosomal and plasmid DNA, respectively
Dxx	rate of accumulation of molecule xx per hour
$DxxS,D$	the rates of synthesis and degradation, respectively, for molecule xx
$EDNAMc,p$	equilibrium concentration of RNA polymerase bound to chromosomal and plasmid mRNA promoters, respectively
e_i	stoichiometric coefficient for interconversion of precursors (g/g)
E_{tm}	transcription efficiency of mRNA operons
$FRACM$	fraction of active ribosomes translating mRNAc
$FRACP$	fraction of total ribosomes translating mRNAp
g_i	stoichiometric constant for conversion of precursor to macromolecule (g/g)
k_T	rate of decomposition
K_{xy}	saturation constant, component y affecting synthesis of component x (g/mL)
K_i	saturation constant for inhibition

K_T	saturation constant for components affecting decomposition
Mi	macromolecules; 1 = protein, 2i = immature sRNA, 2r = mature sRNA, 2m = mRNAc, 2p = mRNAp, 3c = DNAc, 3p = DNAp, 4 = cell envelope, 5 = glycogen
$mRNAc,p$	chromosomal- and plasmid-derived messenger RNA, respectively
Np	number of plasmids per cell
$NPLAS$	number of plasmids available for replication
$NTOT$	number of chromosome replication origins available for initiation
Pi	macromolecular precursors; 1 = amino acids, 2 = ribonucleotides, 3 = deoxyribonucleotides
P^+	plasmid-containing cells
P^-	plasmid-free cells
$PROc,p$	chromosome- and plasmid-derived protein
PS	strength of a promoter relative to an "average" mRNA promoter
$RNAP$	RNA polymerase
rpo	operon name denoting RNA polymerase
rrn	operon name denoting ribosomal RNA
u_i	maximum synthesis rate for macromolecule i
V	cell volume (cm^2)

REFERENCES

1. S.W. Peretti and J.E. Bailey, *Biotechnol. Bioeng.*, in press (1986).
2. R.H. Pritchard, P.T. Barth and T. Collins, *Symp. Soc. Gen. Microbiol.*, **19**, 263 (1969).
3. D. Filpula and J.A. Fuchs, *J. Bacteriol.*, **135**, 429 (1978).
4. J-H. Seo and J.E. Bailey, submitted 1986.
5. J. Tomizawa, *Cell*, **38**, 861 (1984).
6. J. Tomizawa and T. Som, *Cell*, **38**, 871 (1984).
7. G. Cesarini, M.A. Meusing and B. Polisky, *PNAS*, **79**, 6313 (1982).
8. A.J. Alldrick and J.T. Smith, *Antoine van Leuwen.*, **49**, 133 (1983).
9. J.W. Dale and J.T. Smith, *Antoine van Leuwen.*, **45**, 103 (1979).
10. J. Adams, T. Kinney, S. Thompson, L. Rubin, and R.E. Helling, *Genetics*, **91**, 627 (1979).
11. R.M.M. Klemperer, N.T.A.J. Ismail and M.R.W. Brown, *J. Gen. Microbiol.*, **115**, 325 (1979).
12. D. Godwin and J. H. Slater, *J. Gen. Microbiol.*, **111**, 201 (1979).
13. J-H. Seo and J.E. Bailey, *Biotechnol. Bioeng.*, **27**, 1668 (1985).
14. B. Engberg, K. Hjalmarsson and K. Nordstrom, *J. Bacteriol.*, **124**, 633 (1975).
15. M. Weinberger and C.E. Helmstetter, *J. Bacteriol.*, **137**, 1151 (1979).
16. P.T. Barth, H. Richards and N. Datta, *J. Bacteriol.*, **135**, 760 (1978).
17. B.E. Uhlin and K. Nordstrom, *Mol. Gen. Genet.*, **165**, 167 (1978).
18. M.M. Domach and M.L. Shuler, *J. Theor. Biol.*, **106**, 577 (1984).

TABLE I

Mass balance equations for chromosome- and plasmid-derived macromolecules

Chromosomal Protein

$$DM1S = u1 * \frac{A2/V}{K_{M1A2} + A2/v} * \frac{P1/V}{K_{M1P1} + P1/V} * 0.85 * M2r * FRACM \quad (1a)$$

$$DM1D = \left[\frac{P1/V}{P1/V + K_{TM1P1}} * \frac{PG/V}{K_{[TM1PG} + PG/V} * \frac{K_{TiM1A2}}{K_{TiM1A2} + A2/V} \right.$$

$$\left. \left[+ \frac{A2/V}{K_{TM1A2} + A2/V} \right] * kTM1 * M1 \right. \quad (1b)$$

$$DM1 = DM1S - DM1D - 0.05 * DM1S \quad (1c)$$

Plasmid Protein

$$DCGPS = u_1 * \frac{A2/V}{K_{M1A2} + A2/V} * \frac{P1/V}{K_{M1P1} + P1/V} * 0.85 * M2r * FRACP \quad (2a)$$

$$DCGPD = \left[\frac{P1/V}{P1/V + K_{TM1P1}} * \frac{PG/V}{K_{TM1PG} + PG/V} * \frac{K_{TiM1A2}}{K_{TiM1A2} + A2/V} \right.$$

$$\left. \left[+ \frac{A2/V}{K_{TM1A2} + A2/V} \right] * k_{M1} * CGP \right. \quad (2b)$$

$$DCGP = DCGPS - DCGPD - 0.05 * DCGPS \quad (2c)$$

RNA Transcription Attenuation and Termination

$$E_{tm} = 1.07 * \frac{K_{iAT}}{K_{iAT} + (M1/M3)} \quad (3)$$

mRNAc

$$DM2m = E_{tm} * u_{M2} * \frac{P2/V}{K_{M2P2} + P2/V} * EDNAMc * V - k_{TM2m} * M2m \quad (4)$$

RNAp

$$DM2P = E_{tm} * uM2 * \frac{P2/V}{K_{M2P2} + P2/V} * EDNAMp * V - k_{TM2m} * M2p \quad (5)$$

DNA

$$DM3c = u_{M3} * NTOT * \frac{P3/V}{K_{M3P3} + P3/V} * \frac{A2/V}{K_{M3A2} + Aw/V} \quad (6)$$

$$DM3P = u_{M3} * NPLAS * \frac{P3/V}{K_{M3P3} + P3/V} * \frac{A2/V}{K_{M3A2} + A2/V} \quad (7)$$

TABLE II

Equilibrium reactions involving RNA polymerase

Conformation	Equilibrium Constants
$E_C + o <=> E_H$	K_E
Non-specific Binding	
$E_C - DNA_C <=> E_C\ DNA_C$	K_{CD}
$E_C - DNA_p <=> E_C\ DNA_p$	K_{CD}
$E_H - DNA_C <=> E_H\ DNA_C$	K_{HD}
$E_H - DNA_p <=> E_H\ DNA_p$	K_{HD}
Promoter binding	
$E_H - DNA_{CM} <=> E_H\ DNA_{CM}$	K_{HP1a}
$E_H - DNA_{PM} <=> E_H\ DNA_{PM}$	K_{HP1b}
$E_H - DNA_{CS} <=> E_H\ DNA_{CS}$	K_{HP2}
$E_H - DNA_{PN} <=> E_H\ DNA_{PN}$	K_{HP1c}

TABLE II continued

$E_H - RNA\ \ polymerase\ \ holoenzyme$

$DNA_p - plasmid\ \ non-promoter\ \ sites\ =\ total\ \ sites - DNA_{PM} - DNA_{PN}$

$E_C - RNAP\ \ core\ \ enzyme$

$DNA_{PM} - plasmid\ \ mRNA\ \ promoter$

$\sigma - RNAP\ \ sigma\ \ subunit$

$DNA_{PN} - other\ \ plasmid\ \ promoters$

$DNA_C - chromosomal\ \ non-promoter\ =\ total\ \ sites\ - DNA_{CM} - DNA_{CS}$

$DNA_{CM}\ -\ chromosomal\ \ mRNA\ \ promoter$

$DNA_{CS} - sRNA\ \ promoters$

K_I - equilibrium constants, where I =

a) E denotes core~sigma binding
b) CD denotes core enzyme~DNA binding
c) HD denotes holoenzyme binding to non-specific DNA sites
d) HP1 denotes holoenzyme binding to mRNA promoters
e) HP2 denotes holoenzyme binding to sRNA promoters

The equilibrium "constants" are actually dependent on intracellular concentrations of mono- and di-valent ions, H^+, guanosine tetraphosphate, various molecules involved in translation and the temperature (dependence detailed in Chapter 1)

TABLE III

Impact of plasmid molecules and their synthesis on energetic and precursor metabolism

ATP Consumption

$$DATPs = d_{A2} * DA2 + d_{P1} * DP1S + d_{P2} * DP2S + d_{P4} * DP4S + d_{M1} * DM1S + d_{M1} * \mathbf{DCGPS} + d_{M2} * (DM2iS + DM2mS + \mathbf{DM2pS}) + d_{M3} * (DM3c + \mathbf{DM3p}) + d_{M4} * DM4S + d_{M5} * DM5 + d_{PG} * DPGS + d_V * DV \quad (1)$$

Amino Acids

$$DP1 = DP1S - DP1D - g_1 * (DM1 + \mathbf{DCGP}) - e_2 * (DP2S - DP2D) - e_4 * DP4S \quad (2)$$

Ribonucleotides

$$DP2 = DP2S - DP2D - g_2 * (DM2i + DM2r + DM2m + \mathbf{DM2p}) - e_3 * DP3S \quad (3)$$

Deoxyribonucleotides

$$DP3 = DP3S - DP3D - g_3 * (DM3C + \mathbf{DM3p}) \quad (4)$$

TABLE IV

Comparison of Experimental and Simulation Results for the Growth Rate as a Function of the Relative Number of Plasmids Per Cell

Np	Relative Growth Rates			
	Experimental			Calculated
0		1		1
1	0.93[a]	0.92[b]	0.97[c]	0.94
2	0.88	0.91	0.94	0.90
5	0.88	0.87	0.88	0.89
10	0.78	0.82	0.84	0.87
20	--	--	--	0.66
34	0.68	0.77	0.82	--

Experimental values are taken from Ref. 13 for *E. coli* HB101 with pMB1 derived plasmids, grown at 37 C in:

(a) M9 minimal medium, enriched with leucine, proline and thiamine
(b) Luria broth enriched with leucine, proline and thiamine
(c) M9 minimal medium enriched with thiamine and 0.4% casamina acids

All medium contained 0.2% glucose.

TABLE V

Comparison of experimental and simulation results for the cloned gene product level, variation with the number of plasmids per cell

Relative number of plasmids per cell	Cloned gene product, plasmid specific synthesis	
	Experimental	Calculated
1	1	1
2	2.1	1.9
5	4.2	3.6
10	5.4	4.1
20	---	5.2
34	7.0	---

(a) Luria broth enriched with leucine, proline, thiamine and 0.2% glucose, as described in Ref. 13.

TABLE VI

Plasmid content effects on macromolecular synthesis parameters and cloned gene productivity

Several columns show relative values, normalized by the value at the smallest plasmid content

TRANSCRIPTION

Np	RNAP	Active RNAP (%)	RNAP~s[1]	RNAP~m	RNAP~p	Transcription[2] Efficiency
			(% of Active RNA Polymerase)			
0	1	35	29	71	0	-
24	1	44	24	72	4.1	1
59	0.97	49	24	67	9.7	0.89
112	0.84	42	22	60	18	0.86
166	0.77	46	19	56	25	0.53
244	0.54	48	17	46	36	0.39
510	0.35	50	11	33	56	0.19

(1) The notation RNAP~s,m,p denotes the percentage of active RNAP transcribing stable RNA, chromosomal mRNA or plasmid-derived mRNA operons, respectively.

(2) The relative ratio of CGP mRNA synthesized/plasmid copy per unit time

TABLE VI continued

TRANSLATION

Np	mRNAc	mRNAp	Ribosomes (1000)	RIBO~m[3] (% of Total RIBO)	RIBO~p	Tl Eff.	CGP[4] Synthesis Rate
0	--	--	8.8	78	--	--	--
24	1	1	8.6	76	5.1	1	1
59	1.2	2.3	8.3	73	9.2	0.89	2.06
112	1.3	4.2	7.2	67	17	0.76	3.16
166	1.3	3.9	6.8	64	22	0.96	3.87
244	0.80	4.2	4.3	56	28	0.92	3.96
510	0.61	5.7	2.6	46	36	0.67	3.68

(3) Ribo~m,p denotes the percentage of total ribosomes translating chromosomal mRNA or plasmid mRNA, respectively

(4) The relative ratio of CGP synthesized/plasmid mRNA per unit time

TABLE VII

Alteration of promoter strength or ribosome binding site strength

A. Effect on growth-related parameters

Case	Specific Growth Rate (1/hr)	Np	RNAP	RNAP~p	Ribosomes (1000)	RIBO~p % Of Total Ribosomes
PS = 1 RBS = 1	0.59	112	1	1	7.2	17
PS = 5 RBS = 1	0.34	150	0.32	2	2.1	49
PS = 1 RBS = 5	0.42	127	0.41	0.5	2.9	49

B. Effect on productivity

Case	RNAP~p Per Plasmid	RIBO~p Per Plasmid	CGP Per Cell	CGP Synthesis Rate Per Unit Time
PS = 1 RBS = 1	1	1	1	1
PS = 5 RBS = 1	1.5	0.55	1.9	1.1
PS = 1 RBS = 5	0.46	3.5	2.1	1.5

TABLE VIII

Enhanced RNA polymerase or ribosome activity through addition of the specified gene to the plasmid

Case	Plus[1]	Specific Growth Rate (1/hr)	CGP Per Cell	% Protein As CGP	RNAP	Ribosomes (1000)	CGP Synthesis Rate Per Unit Time
PS = 5 RBS = 1	0	0.34	6.4	59	1	2.1	1.0
	rrn	0.46	8.1	54	1.6	3.9	1.7
	rpo	0.47	10	44	5.9	8.6	2.2
PS = 1 RBS = 5	0	0.42	7.3	56	1.3	2.9	1.4
	rrn	0.47	8.8	53	1.8	4.6	1.9
	rpo	0.44	18	49	7.3	12	3.5

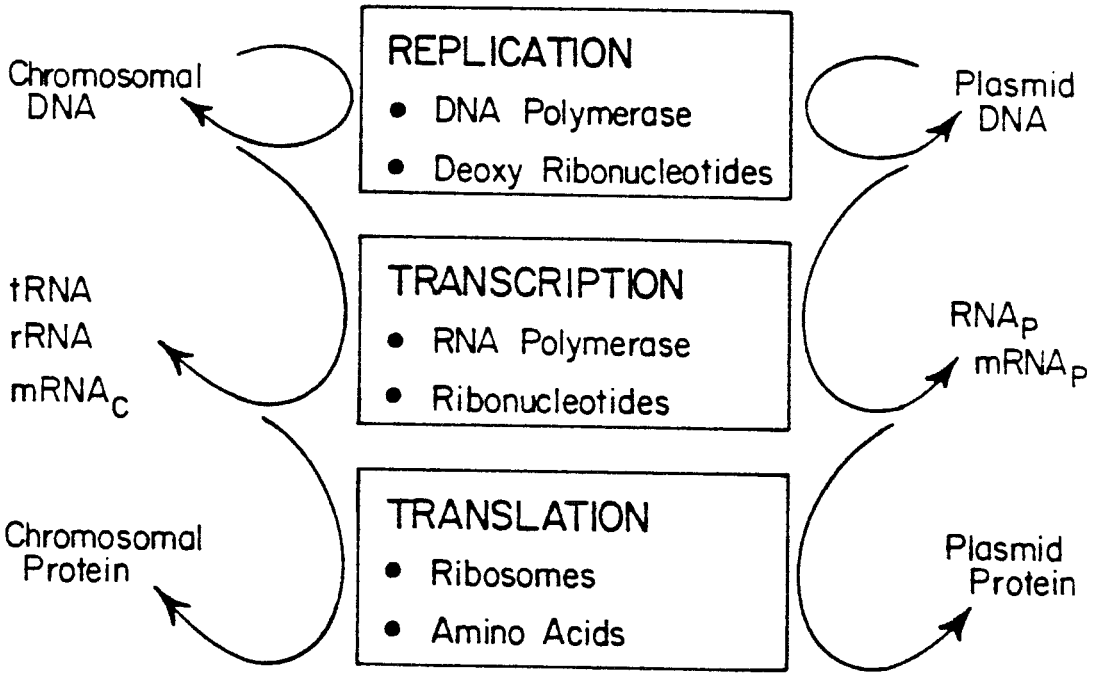
1) Genes added to the plasmid; rpo and rrn denote rRNA and RNA polymerase genes, respectively

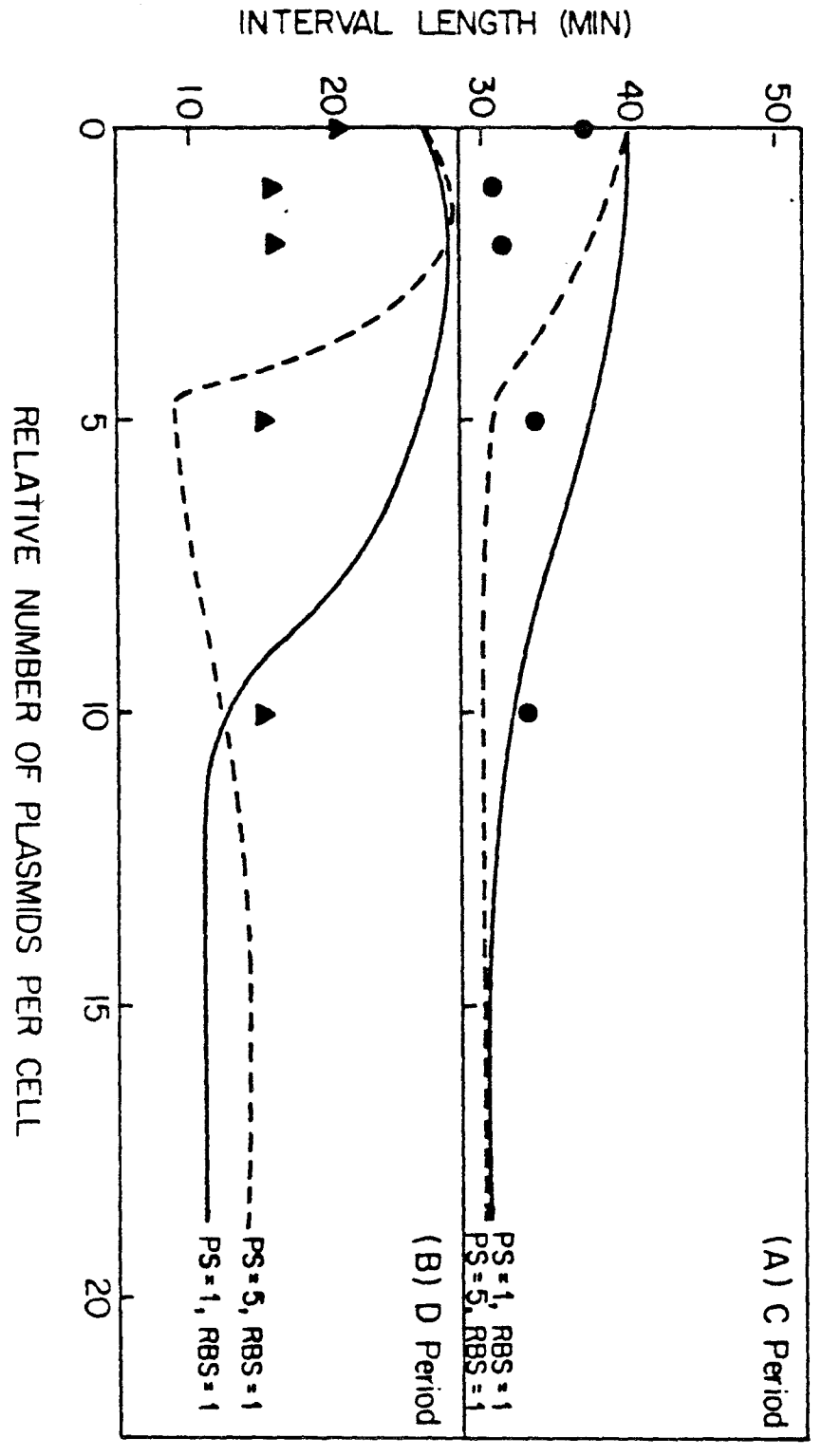
Note: Values for RNAP content per cell and CGP synthesis rate are normalized with respect to the values for the case PS=5, RBS=1, with no additional plasmid-borne genes.

FIGURE CAPTIONS

FIGURE 1 A schematic representation of the competition between chromosomal - and plasmid-directed macromolecule systems for the cell's macromolecular synthetic machiner.

FIGURE 2 Cell cycle parameter dependence on relative plasmid number. Circles represent data taken from Seo and Bailey (4), lines are simulation results. PS is the promoter strength of the cloned gene and RBS is the strength of the ribosome binding site on the cloned gene's mRNA. A value of unity for these parameters represents the strength of "average" chromosomal mRNA promoters and ribosome binding sites.


REPLICATION
• DNA Polymerase
• Deoxy Ribonucleotides
TRANSCRIPTION
• RNA Polymerase
• Ribonucleotides
TRANSLATION
• Ribosomes
• Amino Acids
Chromosomal DNA
Plasmid DNA
tRNA
rRNA
mRNA$_C$
RNA$_P$
mRNA$_P$
Chromosomal Protein
Plasmid Protein


INTERVAL LENGTH (MIN)
0
10
20
30
40
50
(A) C Period
(B) D Period
PS=1, RBS=1
PS=5, RBS=1
PS=1, RBS=1
PS=5, RBS=1
0
5
10
15
20
RELATIVE NUMBER OF PLASMIDS PER CELL

APPENDIX. DETAILED SIMULATION RESULTS

Model complexity allows the simulation of details of cellular metabolism which are difficult or impossible to measure with current analytical methods. In this formulation, transcription has been decomposed into two reactions; the formation of an active RNA polymerase holoenzyme-promoter complex and subsequent RNA chain elongation. Likewise, translation is modeled as consisting of two steps: equilibrium binding of mRNA to inactive ribosomes and a chain elongation step that encompasses both the activation of the mRNA-ribosome complex and polypeptide chain elongation.

The productivity of recombinant systems depends upon the capability of the cells to perform transcription and translation. Table A1 indicates the effect that Np has upon parameters that have an impact on productivity, including promoter efficiency, ribosome binding efficiency and chain elongation rates. The promoter efficiency is a measure of the number of holoenzyme-promoter complexes formed per promoter site. The ribosome binding efficiency quantifies the binding of ribosomes to mRNAp on a per plasmid message basis.

The number of promoter complexes formed per plasmid copy decreases as plasmid number increases. The increase in the fraction of polymerase bound to plasmid promoters does not keep pace with the increase in plasmid promoters due to the accompanying increase in non-specific binding sites for holoenzyme on the plasmid DNA. To somewhat offset the decline in efficiency of competing for polymerase, mRNA transcription attenuation is decreased. This accounts for the initial increase in effective transcription elongation rate. For the chromosomal operons, decreasing cell activity not only shifts RNA polymerase distribution to favor mRNA operons, but also relieves attenuation and premature termination of transcription of these operons. Transcription of the cloned gene was treated in the same way as described previously for chromosomal genes (1).

Utilization of available ribosomes is reduced for the plasmid system. Increasing mRNAc level damps the effect of increasing mRNAp level on ribosome redistribution. While overall translation efficiency is shown to decrease, simulations indicate that the chain elongation rate of translation increases with plasmid number. This counterintuitive result is caused by the equally surprising trend calculated by the model that the intracellular concentration of glucose-like and energetic compounds increases dramatically as plasmid number rises. One assumption made during model revision is that the composition of the cell membrane is not altered by plasmid presence. The activity of the glucose transport system embedded in the membrane is likewise presumed unaffected. As metabolic stress results in decreasing macromolecular synthesis capacity, lower levels of glucose and energy compounds are consumed, leading to an increase in the concentration of those compounds. Chain elongation for translation requires activated amino-acyl tRNA species. The activation involves an energetic compound, hence the increase in chain elongation rate. The simulation results for this aspect of translation are plausible, but no experimental data related to this behavior are available at this time.

The strength of plasmid promoters and ribosome binding sites have a significant impact upon the productivity of the recombinant system. Table A2 shows the effect on many aspects of cell metabolism resulting from changing the promoter strength or ribosome binding site strength. It is important to note for the RNA polymerase distribution that the plasmid fraction shown in only that bound to the promoter of the CGP. There are additionally two promoters for untranslated RNA species, each with a strength of one, which accounts for the balance of the active RNA polymerase.

Metabolic engineering can also lead to increased levels of productivity. Table A3 shows results from simulation of the addition of a gene for either rRNA or

RNA polymerase to the plasmid. In conjunction with data presented in Table VIII, these results indicate that increased productivity rates are correlated with elevated RNAP and ribosome levels, offsetting decreases in fractional distribution of RNAP on plasmid promoters and ribosomes on plasmid-derived mRNA.

TABLE A1

Plasmid number effect upon macromolecular

binding and chain elongation rates

(normalized by values for the 24 plasmids per cell case)

Plasmids per cell	Promoter Efficiency	Transcription Elongation Rate	Ribosome Binding Efficiency	Translation Elongation Rate
24	1	1	1	1
59	1.03	0.86	0.77	1.18
112	0.76	1.13	0.66	1.13
166	0.70	0.76	0.87	1.14
244	0.51	0.76	0.66	1.42
512	0.28	0.67	0.43	1.50

Notes: Promoter efficiency is the number of transcription complexes formed per plasmids

Ribosome binding efficiency is the number of translation complexes formed per plasmid mRNA

TABLE A2

Cellular response to alteration of binding constants
for RNA polymerase-promoter and mRNA-ribosome interactions

TRANSCRIPTION

Case	Np	RNAP	Active RNAP (%)	RNAP~s (% of Active RNAP)	RNAP~c	RNAP~p	Promoter Eff.	Elong. Rate	Tx Eff.
PS = 1 RBS = 1	112	1	42	22	60	6.1	1	1	1
PS = 5 RBS = 1	150	0.32	45	12	40	35	0.69	0.97	0.67
PS = 1 RBS = 5	127	0.41	38	17	58	8.3	0.46	0.47	0.22

TRANSLATION

Case	mRNAc	mRNAp	Ribosomes (1000)	Ribo~m (% of Total Ribosomes)	Ribo~p	Ribosome Binding Eff.	Elong. Rate	Tl Eff.
PS = 1 RBS = 1	1	1	7.2	67	17	1	1	1
PS = 5 RBS = 1	0.36	1.52	2.1	34	49	0.55	1.31	0.72
PS = 1 RBS = 5	0.44	0.34	2.9	37	49	3.38	1.31	4.59

TABLE A3

Details of metabolic adaptation to the addition of genes for RNA polymerase or rRNA to the plasmid

TRANSCRIPTION

Case	Plus	Np	RNAP	Active RNAP (%)	RNAP~s (% of Active RNAP)	RNAP~c (% of Active RNAP)	Promoter Eff.	Elong. Rate	Tx Eff.
PS = 5 RBS = 1	0	150	1	45	12	40	35	1	1
	rrn	113	1.57	40	13	47	29	1.53	1.36
	rpo	115	5.88	45	15	60	18	3.93	0.98
PS = 1 RBS = 5	0	127	1.27	38	17	58	8.3	0.67	0.48
	rrn	113	1.78	33	17	62	7.0	0.75	0.64
	rpo	118	7.25	42	16	72	4.0	2.10	0.63

TRANSLATION

Case	Plus	mRNAc	mRNAp	Ribo (1000)	Ribo~m (% Total Ribo)	Ribo~p (% Total Ribo)	Ribosome Binding Eff.	Elong. Rate	Tl. Eff.
PS = 5 RBS = 1	0	1	1	2.1	34	49	1	1	1
	rrn	1.3	1.2	3.9	39	45	1.48	0.97	1.44
	rpo	3.2	2.1	8.6	47	38	1.46	0.70	1.03
PS = 1 RBS = 5	0	1.2	0.23	2.9	37	49	6.11	1	6.12
	rrn	1.4	0.26	4.6	40	45	7.65	0.92	7.04
	rpo	5.0	0.83	12	45	44	6.09	0.69	4.22

CHAPTER 3

Transient Response Simulations of Recombinant Microbial Populations

Abstract

An asynchronous bacterial population has been approximated using a finite number of "computer" cells, each based on a complex single-cell model for *E. coli*. This formulation correctly simulates the transient responses of macromolecular synthesis rates (protein, rRNA, mRNA) and of the total mass synthesis rate to sudden increases and decreases in the concentration of limiting energy source in the growth medium.

Simulations of the responses of recombinant populations to plasmid amplification or plasmid promoter induction also resulted in behavior similar to that determined experimentally. The calculated responses for recombinant populations subjected to constant promoter induction or cyclic induction/non-induction lead to the conclusion that inducible systems give greater productivity than those with fixed promoter strength. Furthermore, this formulation is suitable to act as a basis for exploring other aspects of recombinant population dynamics.

INTRODUCTION

Plausible reaction sequences and kinetic expressions based on steady-state experimental data may often be formulated. However, there can be several different models, all of which exhibit the same quality of fit to the data. Additional information is then necessary to discriminate between potential mechanisms. The dynamic response of the system to an applied perturbation will often exhibit behavior which more distinctly reflects the reaction steps underlying the observed overall kinetics (1,2). This method has been widely used in the field of heterogeneous catalysis and has proven effective in model discrimination for reactions whose steady-state behavior can be approximated by several different reaction mechanisms (3,4). Also, extensions of steady-state models to describe dynamic reaction system behavior often fail because important dynamic features are unidentifiable from steady-state experiments. Transient responses of intermediates absent from the steady-state formulation may profoundly influence dynamic features of the reactor.

Transient response analysis is an important tool for modeling biological systems as well. The enormous complexity of the reaction network and control interactions requires analyzing system dynamics to verify proposed kinetic mechanisms. The frequent application of batch processing in microbial systems underscores the need for accurate descriptions of dynamic characteristics of bioreactors. Additionally, the application to biological systems of methodology developed for understanding heterogeneous catalysis is useful and the parallels will be pointed out throughout the presentation.

Simulating the dynamic behavior of a microbial population presupposes a model for an asynchronous culture. In this paper, a highly structured single-cell model (5) is used as the basis for simulating an asynchronous bacterial culture. The dynamic response of the culture to perturbations of limiting substrate is calcu-

lated and compared with experimental results. Simulations of recombinant population transient behavior are presented and analyzed in the context of productivity optimization through process design and dynamic reactor control.

Structured models representing metabolism for a single cell have been developed and used to describe the steady growth of *E. coli* populations (5,6). These formulations are extremely complex and give the appearance of mimicking cell behavior in very fine detail. The use of data from systems at steady state to validate proposed metabolic mechanisms, however, is incomplete. The experimental determination of kinetic parameters or macromolecular levels, subsequently used to construct the model, involves measuring population-averaged quantities. While individually these values represent the best estimate of the model parameters, the simultaneous attainment of the average value for a large number of variables may not be representative of any single cell. The premise that the vector of average values leads to the best representation of the population is most stringently tested by analysis of simulation results for transient behavior. Simulating transient behavior, however, necessitates the use of a population of computer cells. If a single cell were used to simulate transient behavior, there is immediately the question of cell cycle effects. The perturbation driving the transient response would have to be introduced at a certain point during the cell cycle. Since the relative synthesis rate and activity of various molecules vary during the cell cycle, the timing of the perturbation would have some effect on the system response. Only by first successfully simulating an asynchronous population can the response to subsequent perturbations be reasonably compared to the experimentally observed phenomenon.

Domach and Shuler (7,8) have developed a formulation that adequately simulates the variation of cell doubling time and fission mass exhibited by asynchronous microbial populations. Random variation of the activity of the enzyme responsible

for crosswall formation was the only alteration made to their single-cell model (7). The enzyme activity had a mean value of one and was distributed between 0.3 and 1.7. This resulted in a Gaussian distribution of cellular D periods. It was found that this mechanism produced the experimentally observed doubling time versus division size coefficient of variation ratio and the negative doubling time correlation between offspring and parent.

PROGRAM DESCRIPTION

This formulation, adopted from Domach and Shuler (7,8) approximates a population with a finite number of computer cells, each representing a subpopulation. The subpopulations are subject to cell cycle variability through D period imprecision. Two hundred fifty computer cells are used to generate the population. To initiate the program, the single cell model is run without cell cycle variability, and all component values and integration parameters are recorded at fifty evenly spaced time points throughout the cell's doubling. These fifty configurations represent subpopulations in stages of the cell cycle from birth to division. The age distribution, derived by Powell (9)

$$P(t/T)=(ln2)2^{1-t/T}$$

represents the relative probability that a cell in an exponential-phase population will be of age t/T, where T is the generation time. The fifty cell configurations represent discrete cell ages from zero to one. The cell configuration at each of these ages is multiplied by the probability for that cell age to create the initial composition for an exponential-phase population. This procedure is repeated five times, and the results are stored in the initialization array. Each cell configuration is then turned "on" sequentially for one minute of integration time. After all cells have gone through one minute of growth, the composition of the system is tabulated and stored. The tabulation involves multiplying each cell by the number of

cells it represents in the population and summing over all cells for each cell constituent. When this is completed, the process is repeated. Equal cell fission is assumed and results in the doubling of the number of cells represented by that set of state variables. The fission process has been found to be very precise (10), so the effect of this assumption on performance is minor.

The single-cell model used here to simulate a representative cell in the population has been previously developed, and a full discussion of the formulation may be found elsewhere (5). The treatment of transcription is particularly detailed, including different forms of RNA polymerase and the explicit calculation of the equilibrium distribution of the enzyme upon various promoter and non-promoter regions along the chromosome. Protein synthesis is also well formulated, dependent on the formation of an active mRNA-ribosome complex prior to initiation of translation. These processes will be highlighted as we consider the transient response of microbial populations to externally applied perturbations.

Simulation of Growth Transients

The programs are structured to simulate batch growth, though a chemostat could also be modeled. The first test of the formulation was to determine whether exponential phase growth could be simulated and to compare the results with those arising from using a single cell to approximate exponential phase growth. Figure 1 shows the results for the two substrate concentrations considered. Total cell mass increases exponentially as expected, and the doubling times calculated from the slopes agree with experimental values for the different limiting substrate concentrations.

A more stringent test of the suitability of the population simulation and of the mechanistic structure of the single-cell model was then undertaken. Shift of some operating parameter will cause changes in the kinetics of macromolecular

synthesis and accumulation, which will be more indicative of the underlying mechanisms. The parameter chosen to introduce a perturbation to drive a transient response was the concentration of the limiting substrate, glucose. There have been experimental reports published regarding shifts up and down of growth media and the attendant transient effects upon synthetic activity in the population. However, the experimental work done was at growth rates that are in some cases faster than the model can simulate. The single-cell model was developed for glucose-limited conditions, and as such cannot achieve growth rates beyond $\mu = 0.9\ hr^{-1}$. This is a particular liability for the shift-up simulations. Experimental data from Brunschede, Dove and Bremer (11) are for a shift from $\mu = 0.5$ to $\mu = 1.5$ per hr. The analogous simulation was for a shift in glucose concentration from 3.2×10^{-6} g/mL to 2×10^{-3} g/mL, corresponding to a change in growth rate from μ=0.25 hr^{-1} to $\mu = 0.77\ hr^{-1}$.

Cellular control systems have been observed to exhibit qualitatively different behavior for rapidly growing cells relative to slower growing cultures (11,12). For example, the ratio of RNA to DNA is directly proportional to growth rate for $\mu > 0.5\ hr^{-1}$. The ratio exhibits increasingly positive deviation as growth rates decline. Similar results are found for the RNA to protein ratio, with the break point in the range $0.7 < \mu < 0.8\ hr^{-1}$. The experiments were, therefore, performed between growth rates largely within the proportionality range, whereas the simulations must use rates where deviant results are manifest. To compensate for this restriction, the data presented have been normalized with respect to total cell mass. Nevertheless, the experimental and theoretical curves are not expected to be coincident.

If the kinetic formulation of the model has a structure similar to that found in the cell, the shape of the simulated response curve should mirror, qualitatively, the shape of the actual response. Kobayashi (13) has shown that the nature of the

transient response curves of products resulting from reactant concentration step changes is often indicative of the reaction mechanism for heterogeneous catalytic systems. Biomacromolecular synthesis reactions have the same structure as certain systems encountered in non-biological catalysis. An example of such is the Eley-Rideal type mechanism used to describe bimolecular reactions catalyzed by solid surfaces.

$$X1 + S \quad > X1 - S$$

$$X1 - S + X2 \quad > Y - S$$

$$Y - S \quad < \quad > Y + S$$

This mechanism may represent, for example, the process of translation. The first step is the formation of the active mRNA-ribosome complex (X1-S). Step two represents polypeptide chain elongation, while step three is the release of the completed polypeptide from the ribosome. Kobayashi found that different parameter sets gave qualitatively similar response modes for this mechanism. Other mechanisms will give distinctively different responses. While the response curves arising from biological systems will be possibly more complex than those from multistep heterogeneous catalysis, there is still the expectation that transient studies will provide more sensitive indications of underlying mechanisms than do balanced growth experiments.

Figure 2 shows the effect upon mass and protein levels of shifting the culture to a condition sustaining faster growth from one that results in longer doubling times. Experimentally, this was accomplished by diluting the cells, grown in succinate minimal medium, into glucose medium supplemented with 20 mg/L of each amino acid. This was simulated by increasing the extracellular glucose concentration for the computer cell. The shapes of the experimental and theoretical curves agree fairly well for both protein and mass accumulation.

Figure 3 shows total RNA synthesis rates following the shift-up and the qualitative agreement between theory and experiment for this aspect of metabolism is quite important. Transcription is a very complex process and is at the center of cellular metabolism. During a shift up in growth rate, transcription rates reflect many adjustments made by the cell. RNA polymerase level increases as the protein synthesis level rises. The polymerase equilibrium shifts from mRNA synthesis to production of stable RNA species. Higher sRNA transcription leads to more ribosomes, further increasing protein synthesis rates. Increased initiation of chromosomal replication provides more gene copies available for transcription. Qualitative simulation of this response indicates that the major mechanisms and control of transcription have been realistically formulated.

As a further test of the model formulation, responses to a shift down in energy source were simulated. In this case, the experimental and theoretical approaches to effecting the shift down are virtually identical. Whereas shift up was accomplished experimentally by changing the energy source, shift down resulted from adding alpha-methyl glucoside (AMG) to a final concentration of 1% to minimal medium containing 0.1% glucose. AMG acts as a competitive inhibitor of glucose transport. It is non-metabolized but utilizes the glucose transport system to enter the cell, reducing the effective concentration of glucose in the medium. While no specific growth rate values were reported for the experiments (14,15), it was mentioned that the AMG approximately halved the growth rate. Simulating a reduction of glucose concentration is straightforward.

Experiments were done (14,15) with two strains, isogenic except for a rel A mutation leading to the rel^- (relaxed) phenotype. The rel locus controls the synthesis of a novel nucleotide, guanosine tetraphosphate. ppGpp is an important effector in the regulation of stable RNA production, and of the relative distribution of RNA polymerase among different promoter types. Rel^+ (stringent) strains are

characterized by a rapid increase in ppGpp level upon shift down and a concomitant drop in all RNA polymerase activity. Relaxed strains, instead, show continued synthesis of various RNA types upon the onset of starvation.

Figures 4a,b,c present the RNA synthesis rate response curves. In these simulation calculations, the shift was from $\mu = 0.77$ to $\mu = 0.25$ hr^{-1}. As evidenced by the shapes of the curves, the model is representative of a relaxed population. This is somewhat surprising, since ppGpp is explicitly accounted for in this model and its effect on RNA polymerase promoter specificity has been formulated mathematically.

Guanosine tetraphosphate metabolism may be incompletely represented, as indicated by the data shown in Figure 5. The calculated ppGpp levels show characteristics of both stringent and relaxed strains. It appears that transient degradation is inadequately represented in the mathematical formulation. Previous descriptions of the model performance did not show comparisons between calculated and measured ppGpp levels in cells growing at different specific growth rates. The agreement found between experimental and theoretical data for the fraction of active RNA polymerase transcribing rRNA operons (5) may have been due to other control mechanisms included in the model.

As described above, *rel*$^-$ strains continue to synthesize RNA upon energy source depletion, even to the point of transiently increasing mRNA synthesis immediately following shift down. Two observations can be made regarding the agreement between theoretical and experimental results. The congruent qualitative shapes for the response curves suggest that the kinetic structure giving rise to each is similar. There is, however, a noticeable lag evident in the model curves. Calculated RNA synthesis takes approximately 20 minutes to stabilize following the shift, whereas the measured rates are relatively stable within 10-12 minutes of the change. This may reflect differences between the experimental conditions and

the simulation.

In addition, the rate constants for the mechanism will affect the time constant of the response. Deviations between observed and calculated response times may indicate that some model parameters require modification.

Directly dependent upon the levels of messenger and stable RNA is the protein synthesis rate. As shown in Figure 6, *de novo* protein synthesis drops to the final post-shift level within one minute. There is no apparent difference between rel^- and rel^- strains for this behavior, and the agreement of the simulated trajectory is excellent. This is surprising, given the complexity of the steps underlying the response. It has been determined (14,15) that for the relaxed strain, the stability of mRNA and the rate of initiation of translation are the same as for the stringent strain. These experiments also showed that the rel^- strain reduces its protein synthesis rate in tandem with the rel^+ strain while carrying a higher amount of mRNA and the same number of ribosomes per cell. Both strains showed reduced polypeptide chain elongation rate immediately after the shift. In the stringent strain, the normal chain elongation rate was restored within 2 minutes, whereas the relaxed strain maintained the lower rate for approximately 20 minutes. Normal elongation rates were not attained until 50-60 minutes after the shift. By this time, RNA synthesis rates in the two strains were equal (15). It is encouraging that the model formulation is able to encompass such complicated behavior within the existing reaction framework.

PRACTICAL APPLICATIONS OF BIOREACTOR DYNAMICS

To this point, the dynamic response of the system has been used as a means of evaluating the model formulation. The level of agreement between simulation and experiment for the diversity of conditions studied and the complexities of the resulting response suggest that this model, combined with the population simulation routine, is applicable to the simulation of bacterial culture dynamics.

At the present time, most commercial biological reactors are operated in batch mode. Batch fermentations are inherently transient. Throughout population growth, the medium will be changing as nutrients are converted to waste products, cell mass and metabolites. In addition, many commercially important products are synthesized during very specific periods of the cell cycle, or under a narrow range of growth conditions. For example, the storage compound polyhydroxybutyrate is produced immediately following nitrogen source depletion in the growth medium (16).

Recombinant cell cultivation systems, which are becoming increasingly prominent, are also important examples of transient processes. A significant portion of recent research with these systems is focused on understanding the processes surrounding vector gene expression. This information can be exploited to increase the overall rate of product formation.

Increasing the abundance of mRNA molecules derived from a particular gene will elevate the rate of synthesis of the specified protein. This can be accomplished by increasing the gene dosage or the rate of transcription from the gene promoter. Use of high copy number vectors (17,18) and strong promoters (19,20) are direct applications of these approaches. Unfortunately, it has been determined theoretically (21) and experimentally (22-26) that continuous high level expression driven by strong promoters or high copy number is detrimental to the cell. Growth rates decline, relative to uninduced or unamplified rates, ultimately leading to loss of productivity.

As a result, controlling gene transcription and plasmid replication are important to optimizing product formation. Induction and amplification have been studied in batch systems in the following manner: Cultures grown without induction or amplification were induced or amplified and growth continued under the new conditions. The behavior of the cultures was compared to systems with fixed high-

level transcription or elevated copy number. Experimental results suggest that inducible transcription systems (27,28) are more productive than non-inducible counterparts. Likewise, amplifiable vectors exhibit enhanced productivity (29,30) for short term amplification conditions.

Simulation of Induction and Amplification

Simulation of the response of a recombinant bacterial culture to the induction of the promoter of a plasmid-borne gene will test the treatment of plasmid-related metabolism and the host-plasmid interactions that result. A detailed single-cell model including plasmid metabolism has been developed and described previously (21). Briefly, plasmid DNA is assumed to compete with the chromosome for RNA polymerase. The resulting mRNA species compete for ribosomes. Further, the affinity of plasmid-derived mRNA for ribosomes is taken as equal to that for the chromosomal mRNA-ribosome interaction. The strength of the plasmid gene promoter is initially equal to that of an "average" chromosomal operon. Induction is modeled by increasing the binding constant of RNA polymerase for the cloned promoter on the plasmid.

The simulations were run for cells growing in glucose minimal medium with a glucose concentration of 2×10^{-3} g/mL. This results in a specific growth rate of 0.77 hr^{-1}. The plasmid replication frequency was one that resulted in an average copy number of 50 plasmids per chromosome. The results from simulations are presented in Figure 7, along with induction kinetics that were obtained experimentally for a recombinant *E. coli* system (27). An inducer was added to the medium and the fraction of total protein present as plasmid gene product was determined. The shapes of the curves are quite similar although the model significantly underestimates the final product level. This will be affected by the product stability and the efficiency at which the plasmid messenger RNA is translated into protein.

Of importance is the rapid initial increase in plasmid product fraction, indicating that the highest relative rate of product formation immediately follows induction. As indicated in Figure 8, induction causes a transient overshoot in the product formation rate per unit cell mass. It is this period encompassing elevated product formation rate which potentially could be exploited by forced transient operation of the bioreactor.

Plasmid amplification was also simulated, and the results are plotted in Figure 9. At t=0, the replication rate was increased in a stepwise fashion from 50 initiations per hour to 500 initiations per hour. In the model, the plasmid replication frequency is an input. The time between initiation events is fixed. An increase in replication frequency is, therefore, effected through a decrease in the interval between initiation events. In the experiments by Uhlin and Nordstrom (18), the temperature was raised, resulting in more frequent replication initiation due to changes in temperature-sensitive regulatory species. This resulted in a one hundred fold increase in plasmid copy numbers from 2-3 to 200-300 copies per chromosome. The simulation results are in close agreement with experimental findings. The absence of an immediate, rapid rise, as found for increased promoter strength, indicates that amplification is not an effective variable for investigating transient overproduction of plasmid-derived products. Promoter strength induction is a suitable means for causing transient productivity enhancement from plasmid genes. Manipulating plasmid gene promoter strength was used in subsequent investigations of the dynamics of recombinant cell metabolism.

Simulation and theoretical studies starting in the late 1960's suggested that intentional transient operation of chemical reactors could result in improved performance (31-34). Since that time, researchers have found experimental systems which exhibit enhanced productivity or selectivity as a result of forced periodic operation (35,36).

To investigate the possibility of productivity enhancement for biological systems as a result of applied perturbations, simulations of transient behavior in recombinant systems were performed involving the induction of plasmid gene promoter strength. Inducible operons have been the focus of intense scrutiny and many have been well characterized (37-40). Two major classes of inducible promoters are those that require a chemical effector and those whose activity is temperature-sensitive. In the absence of a good model detailing the temperature-dependence of bacterial metabolism, this simulation will necessarily be limited to those systems where induction is mediated by a chemical inducer or repressor.

The response of the model population to a stepwise increase in promoter strength is indicated in Figure 10. The glucose concentration in the medium used for the simulations was 2×10^{-3} g/mL and the population had an average plasmid copy number of 50 prior to promoter strength induction. For the two cases presented, the product formation rate continuously declines from a high initial rate. Eventually the slopes of product accumulation for the induced populations drop below that for the uninduced population. This point is better indicated by the maximum in Figure 11 that displays the induced population product level following induction normalized by the uninduced population product level at the corresponding time. The normalization factor is, therefore, ever increasing as the uninduced population grows. The maximum in relative product level indicates the point during cultivation when the increase in product accumulation due to induction is balanced by the growth penalty incurred by the resulting metabolic perturbation. After this point, the growth-rate decline is not offset by concomitant increases in cell specific productivity for the induced population. As a result, the rate of product accumulation in the induced system is less than that in the unperturbed culture.

These simulations were done for dilute cultures, such that cell growth has no

effect on substrate concentration in the growth medium. It is possible that cultures in different physiological states (e.g., at different specific growth rates) prior to induction might respond differently to these perturbations. There are other possible scenarios when the timing of induction relative to substrate depletion is considered. These were not calculated explicitly, due to the time and cost involved in computation for this model, but some evaluation is possible. The faster growing culture will deplete limiting nutrients more rapidly, leading to growth-rate reduction. This may have the effect of shifting the location and magnitude of the relative maximum, abolishing the maximum altogether or causing the appearance of other local maxima. In any case, it is unlikely that one would encounter the same relative rate of increase that leads to the maximum for constant substrate concentration. The existence of a maximum for each of these is due to inherent metabolic regulatory mechanisms and has implications for the design and operation of bioreactors.

The maxima in Figure 11 suggest both operational strategies and process designs to maximize system productivity. One implication is that steady-state chemostat cultivation is inappropriate for optimal product formation. Instead, batch cultivation or a multi-stage continuous process are more productive processing schemes. The timing of induction relative to harvesting of the culture is very important for attainment of optimal product levels, as the relatively sharp maxima in Figure 11 indicate. The time spent under induction conditions should be close to that resulting in maximum relative product to take best advantage of the population transient. This can best be achieved in either a batch reactor or plug flow device. Difficulties inherent in maintaining adequate metabolite levels throughout a plug flow reactor lead to the conclusion that batch operation is likely optimal for exploiting systems with inducible promoters directing plasmid mRNA synthesis.

One additional attempt to exploit the response time differences for gene

expression and overall cell synthesis involved simulation of periodic operation of the bioreactor. This entailed the systematic perturbation of the culture, with the goal of attaining productivity beyond that possible for cultivation under fixed conditions. The promoter strength was varied periodically, first increased by a factor of ten for some time period, then returned to the uninduced value. The total period between induction events was 40 minutes. Several operating modes were simulated. These were: a) up 20 minutes, down 20 minutes; b) up 30 minutes, down 10 minutes; c) up 35 minutes, down 5 minutes. Figure 12 presents the results of these manipulations (points; see legend) and compares them to the case where promoter strength is shifted by a factor of ten and maintained at that level (solid curve).

The trends that result are interesting, but do not show major benefits due to cycling promoter strength. There are times when the product level in a periodically controlled culture exceeds that of the steadily induced population. It is also possible that changing from oscillatory to constant, elevated promoter strength at some point would provide even higher product levels. However, benefits achieved by cyclic operation are not large and likely would not justify the complication and expense of implementing this type of bioreactor control.

Summary

An asynchronous population of *E. coli* has been successfully approximated by a finite ensemble of single cells. The single-cell model used is highly detailed, allowing investigation of the transient response of key intracellular processes in a population. Excellent qualitative agreement between experimental results and simulations indicates that for the processes of transcription and translation the single-cell formulation is sufficiently robust to represent the predominant metabolic structure of the cell.

This model was then used to investigate the dynamic behavior of recombinant populations. Again, the details of the transient response determined by population simulation exhibited good agreement with experimental results. The major conclusion from this effort is that transient induction of plasmid promoter strength is more productive than steady cultivation in the induced state.

Unfortunately, there is scant detailed information yet available for recombinant system dynamics. Consequently, rather than claim that this formulation is strictly accurate and capable of predicting transient recombinant system behavior for a wide variety of conditions, we propose that the formulation represents a useful approach to exploring such population responses. In addition, for plasmid-bearing *E. coli* strains, this model affords the opportunity to evaluate in a systematic manner different metabolic mechanisms involved in cellular control or host-vector interactions. Further experimental work is essential for a more detailed evaluation of the population model.

References

1. H. Kobayashi and M. Kobayashi, Catal. Rev.-Sci. Eng., **10**, 139 (1974).

2. C. O. Bennett, Catal. Rev.-Sci. Eng., **13**, 121 (1976).

3. M. Kobayashi, A.C.S. Symposium Series, **196**, 213 (1982).

4. M. R. Prairie and J. E. Bailey, Chem. Eng. Sci., **41**, 937 (1986).

5. S. W. Peretti and J. E. Bailey, Biotechnol and Bioeng., in press (1986).

6. M. M. Domach, S. Leung, R. E. Cahn, G. G. Cocks and M. L. Shuler, Biotechnol. Bioeng., **26**, 203 (1984).

7. M. M. Domach and M. L. Shuler, J. Theor. Biol., **106**, 577 (1984).

8. M. M. Domach and M. L. Shuler, Biotechnol. Bioeng., **26**, 877 (1984).

9. E. O. Powell, J. Gen. Microbiol., **15**, 492 (1956).

10. C. N. Newman and H. E. Kubitshek, J. Mol. Biol., **121**, 270 (1978).

11. H. Brunschede, T. L. Dove and H. Bremer, J. Bacteriol., **129**, 1020 (1977).

12. P. P. Dennis and H. Bremer, J. Bacteriol., **119**, 270 (1974).

13. M. Kobayashi, Chem. Eng. Sci., **37**, 383 (1982).

14. K. Johnsen, S. Molin, O. Karlstrom and O. Maaloe, J. Bacteriol., **131**, 18 (1977).

15. S. Molin, K. von Meyenberg, O. Maaloe, M. T. Hansen and M. L. Pato, J. Bacteriol., **131**, 7 (1977).

16. Unpublished results, this laboratory.

17. K. Nordstrom, L. C. Ingraham and A. Lundback, J. Bacteriol., **110**, 562 (1972).

18. B. E. Uhlin and K. Nordstrom, Molec. Gen. Genet., **165**, 167 (1978).

19. P. E. Stevis and N. W. Y. Ho, Enzyme Microb. Technol., **7**, 592 (1985).

20. K. Backman and M. Ptashne, Cell, **13**, 65 (1978).

21. S. W. Peretti and J. E. Bailey, Biotechnol. Bioeng., in press (1986).

22. J. W. Little, Molec. Gen. Genet., **162**, 51 (1979).

23. H. Shimatake and M. Rosenberg, Nature, **292**, 128 (1981).

24. N. P. Fiil, D. Bendiak, J. Collins, and J. D. Friesen, Molec. Gen. Genet., **173**, 39 (1979).

25. J. W. Dale and J. T. Smith, Ant. van Leewen., **45**, 103 (1979).

26. P. Zund and G. Lebek, Plasmid, **3**, 65 (1980).

27. E. Amann, J. Brosius and M. Ptashne, Gene, **25**, 167 (1983).

28. M. A. Osburne, R. J. Craig and D. M. Rothstein, J. Bacteriol., **263**, 1101, (1985).

29. E. Remaut, H. Tsao and W. Fiers, Gene, **22**, 103 (1983).

30. J. M. Nicaud, N. Mackman and I. B. Holland, J. Biotechnol., **3**, 175 (1985).

31. J. M. Douglas and D. T. Rippin, Chem. Eng. Sci., **21**, 305 (1966).

32. F. J. M. Horn and R. C. Lin, I. and E. C. Process Des. Develop., **6**, 21 (1967).

33. J. E. Bailey, Chem. Eng. Commun., **1**, 111 (1973).

34. J. E. Bailey, "Periodic Phenomena," Chemical Reactor Theory: A Review, Prentice-Hall Inc., N. J., 758 (1977).

35. M. R. Prairie and J. E. Bailey, Chem. Eng. Sci., in press (1986).

36. Y. Barshad, X. X. Zhau, E. Gulari, J. Cat., **94**, 128 (1985).

37. A. Kepes, Prog. Biophys. Mol. Biol., **19**, 201 (1969).

38. I. P. Crawford and G. V. Stauffer, Ann. Rev. Biochem., **49**, 163 (1980).

39. H. A. DeBoer, L. J. Constock and M. Vassar, Proc. Nat. Acad. Sci., USA, **80**, 21 (1983).

40. J. Hedgepeth, M. Ballivet and H. Eisen, Molec. Gen. Genet., **163**, 197 (1978).

Figure Captions

Figure 1. Simulated increase in mass for asynchronous populations growing in media with with different limiting glucose concentrations. The lower line is for growth in medium with a glucose concentration of of 3.27 mg/L while the solid line is for a glucose concentration of 1000 mg/L.

Figure 2. The response of cell mass and the protein/mass ratio to a shift up in energy source. Data for cell mass (triangles) and protein per cell mass (diamonds) are from Brunshede, Dove and Bremer (11), solid lines are simulation results. All results are normalized to values at the time of upshift (t=0).

Figure 3. Total RNA synthesis rate response to a shift up in energy source. Data (Δ) taken from Ref. (11); dotted line added to highlight the trajectory of the response. Solid line represents simulation results. All results are normalized by the rate at the time of shift (t=0).

Figure 4. The response of the synthesis rates for mRNA (a), rRNA (b) and total RNA (c) to a shift down in limiting energy source. Data are taken from Ref. (15) for two strains, one designated rel^+ (diamonds), the other rel^- (Δ). The dashed and dotted lines are drawn to highlight trends, while the solid line represents simulation results. Normalization was done using rates at the time of shift down (t=0).

Figure 5. Guanosine tetraphosphate levels following a shift down in the limiting energy source for rel^+ (diamonds) and rel^- (Δ) strains (Ref. 14). The solid line was calculated using the model.

Figure 6. Protein synthesis rates decline almost instantaneously in both rel^+ (diamonds) and rel^- (Δ) strains (Ref. 14). The solid line is from the simulation. Normalization is with synthesis rates at t=0.

Figure 7. Change in fraction of total protein which is plasmid-derived product following promoter induction. The experimental data (Δ) are taken from Reference 27. The dashed line represents an induced promoter strength (RNA polymerase equilibrium binding constant, normalized by the equilibrium constant for an average *E. coli* promoter) of five, while the solid line is for a tenfold induction. The results have been normalized to the protein fraction established prior to induction.

Figure 8. Rate of product formation following promoter induction. Triangles represent experimental data, the solid line represents simulation results.

Figure 9. Relative product level during plasmid amplification. The experimental data (Δ) come from Uhlin and Nordstrom (18).

The solid line comes from a simulation. There is a slight concavity in the simulation data immediately following initiation of amplification, though not nearly as pronounced as that exhibited experimentally.

Figure 10. Calculated levels of plasmid product found in different populations. The induced cultures were identical to the uninduced (dashed line) population until t=0, when induction occurred. The dotted line represents fivefold induction of the cloned gene promoter, the solid line tenfold.

Figure 11. The solid and dashed lines represent population simulations of data for the cloned-gene product levels following five and tenfold induction, respectively, of the cloned-gene's promoter. These values are normalized by the product level in a control culture which is uninduced.

Figure 12. Model calculations of product mass relative to that of the uninduced culture. The solid line represents the tenfold steadily induced system. The other points are from simulations of populations with plasmid promoters which are periodically switched between tenfold induction and no induction. The diamonds represent the case of repeatedly having 20 minutes of promoter induced (up), followed by 20 minutes of uninduced (down). Triangles represent 30 minutes up, 10 minutes down, while octagons represent the case of 35 minutes up and 5 minutes down.

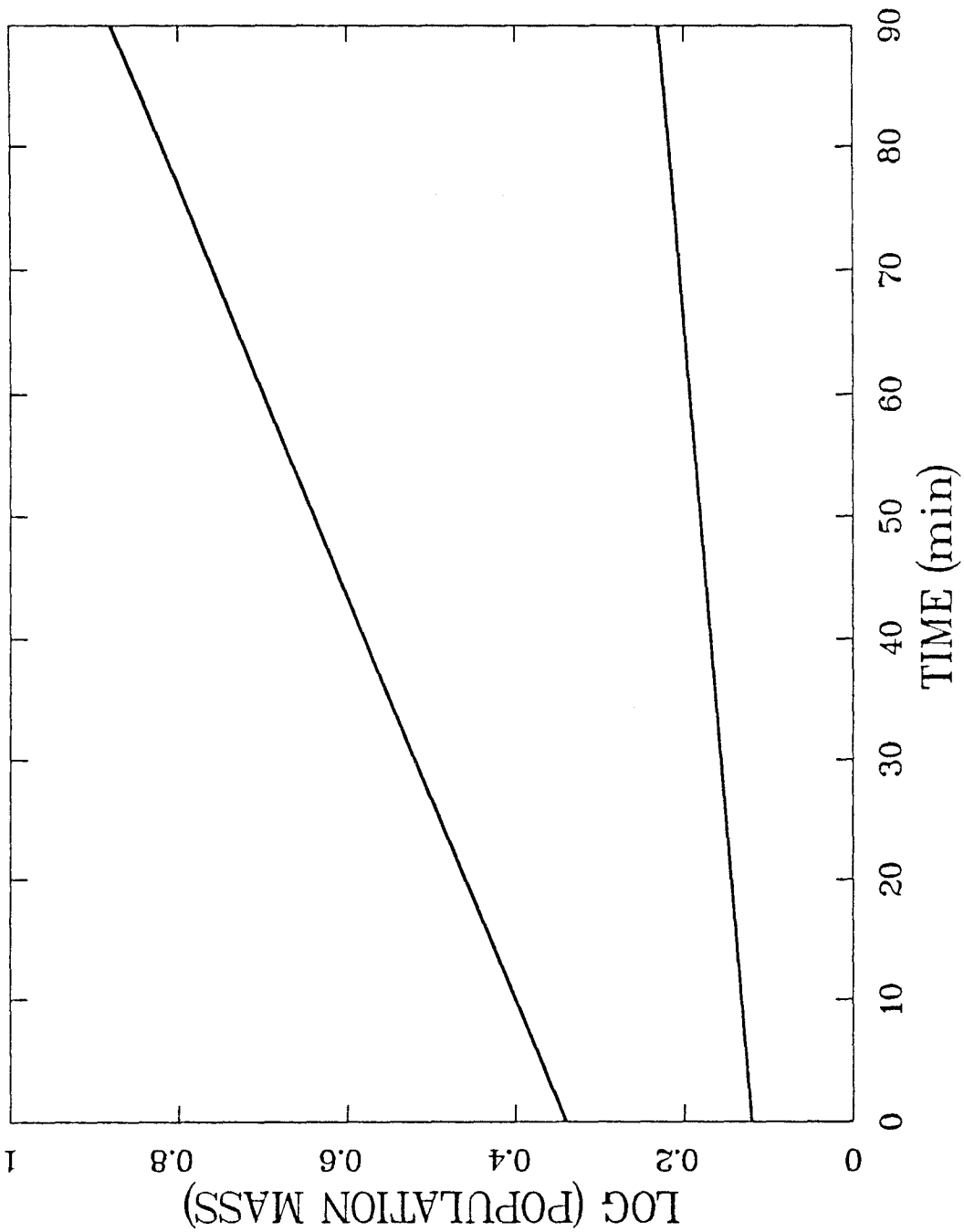

LOG (POPULATION MASS)
0
0.2
0.4
0.6
0.8
1
TIME (min)
0
10
20
30
40
50
60
70
80
90

FIGURE 1

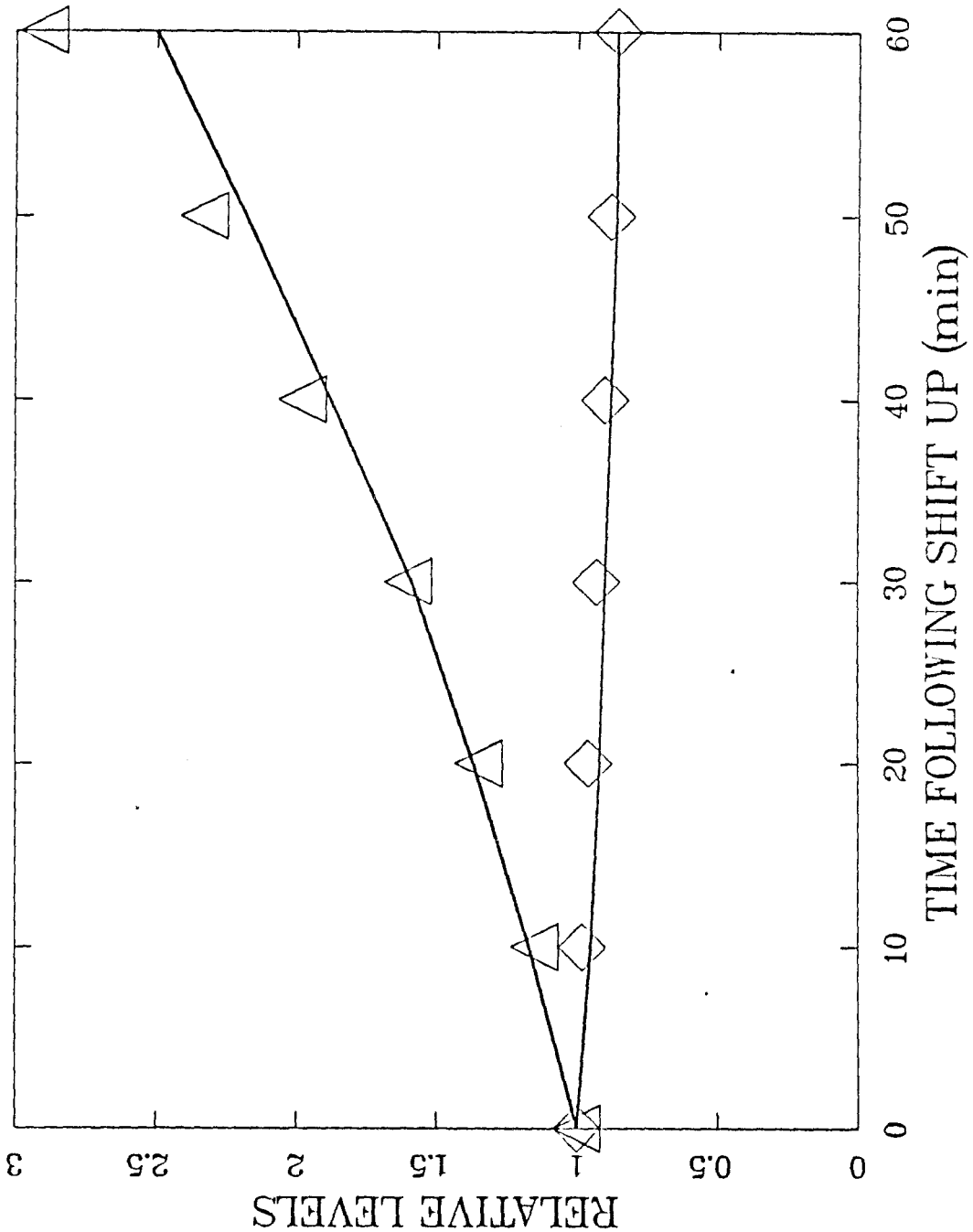

RELATIVE LEVELS
0
0.5
1
1.5
2
2.5
3
TIME FOLLOWING SHIFT UP (min)
0
10
20
30
40
50
60

FIGURE 2

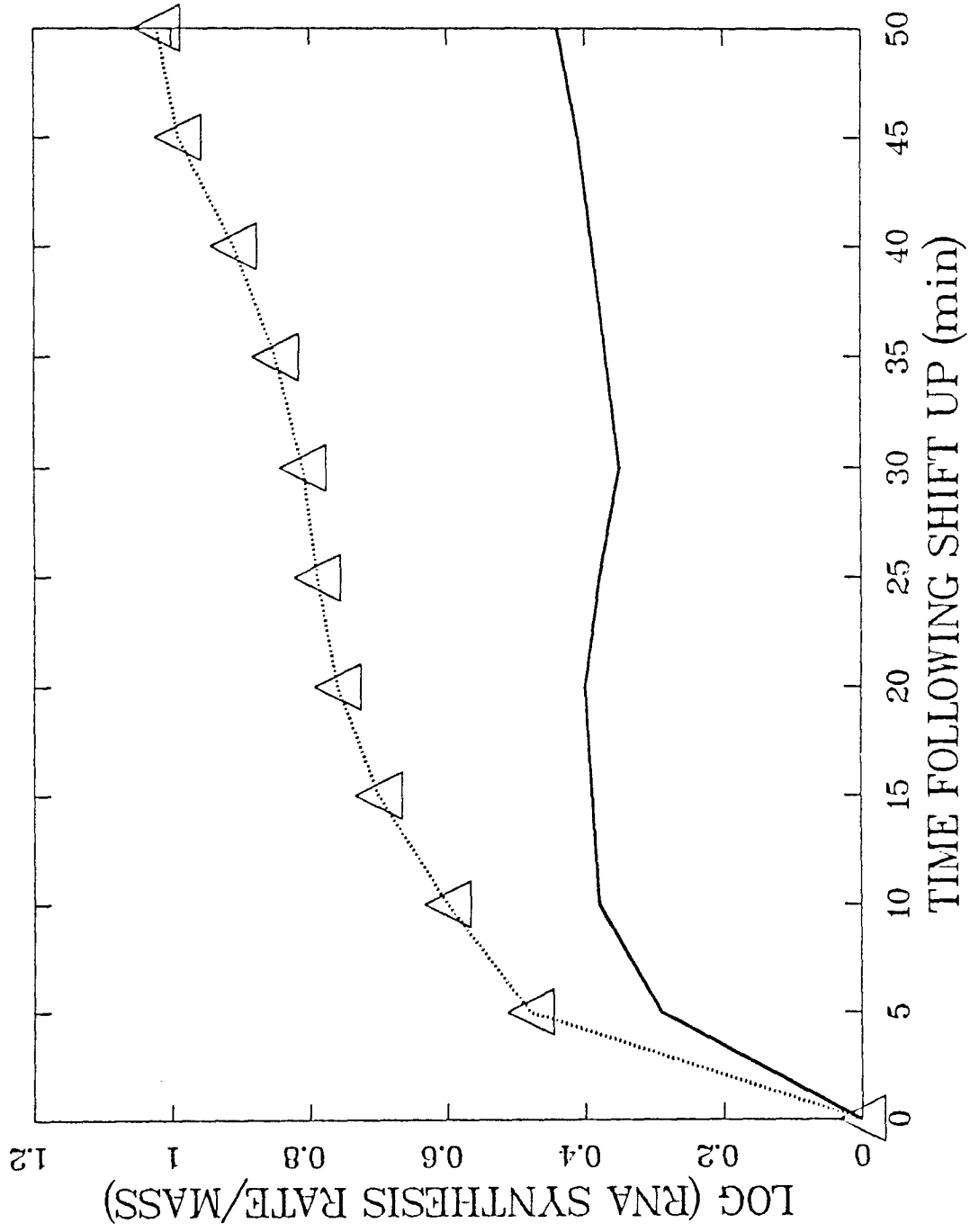

LOG (RNA SYNTHESIS RATE/MASS)
0
0.2
0.4
0.6
0.8
1
1.2
TIME FOLLOWING SHIFT UP (min)
0
5
10
15
20
25
30
35
40
45
50

FIGURE 3

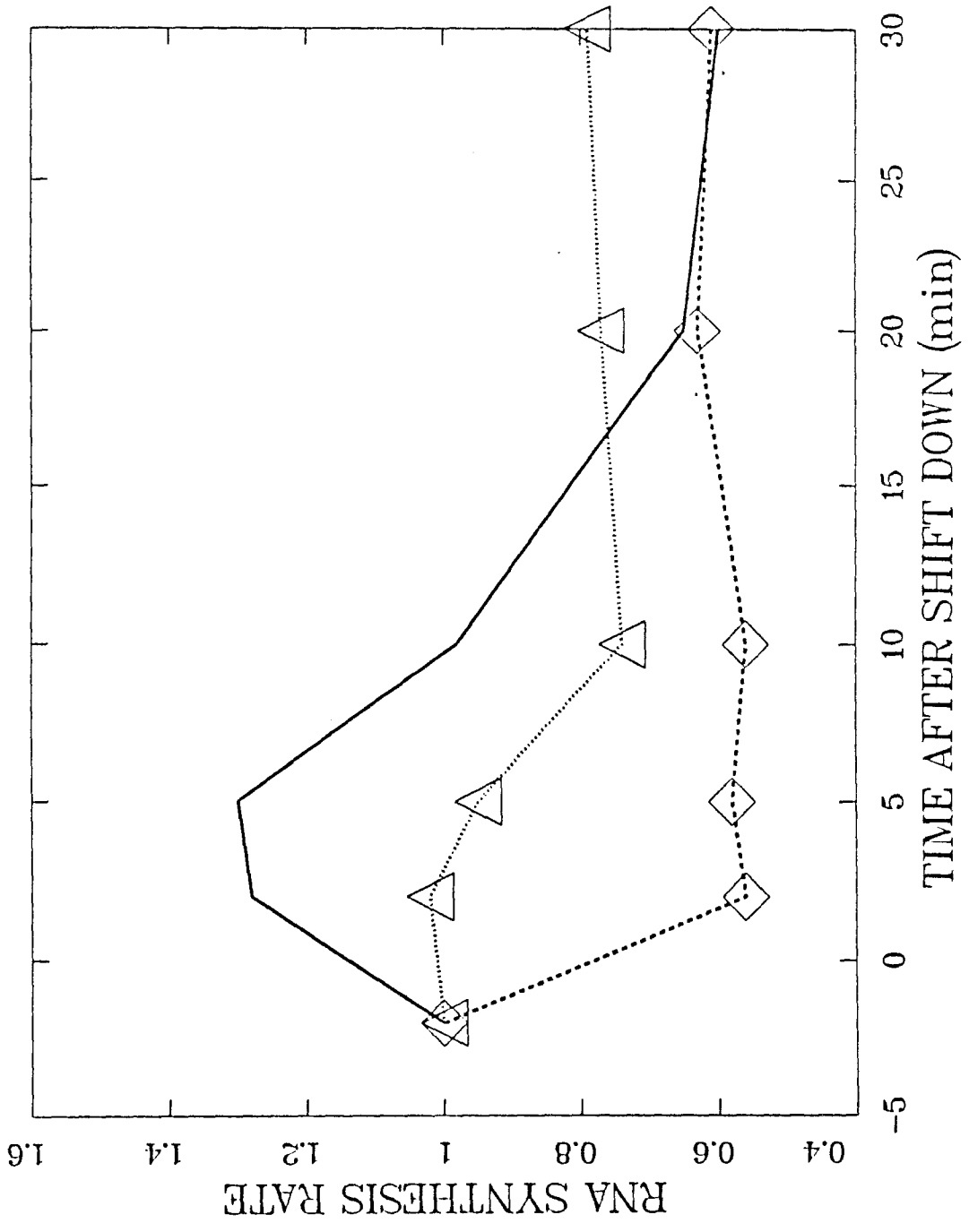

RNA SYNTHESIS RATE
0.4
0.6
0.8
1
1.2
1.4
1.6
TIME AFTER SHIFT DOWN (min)
-5
0
5
10
15
20
25
30

FIGURE 4a

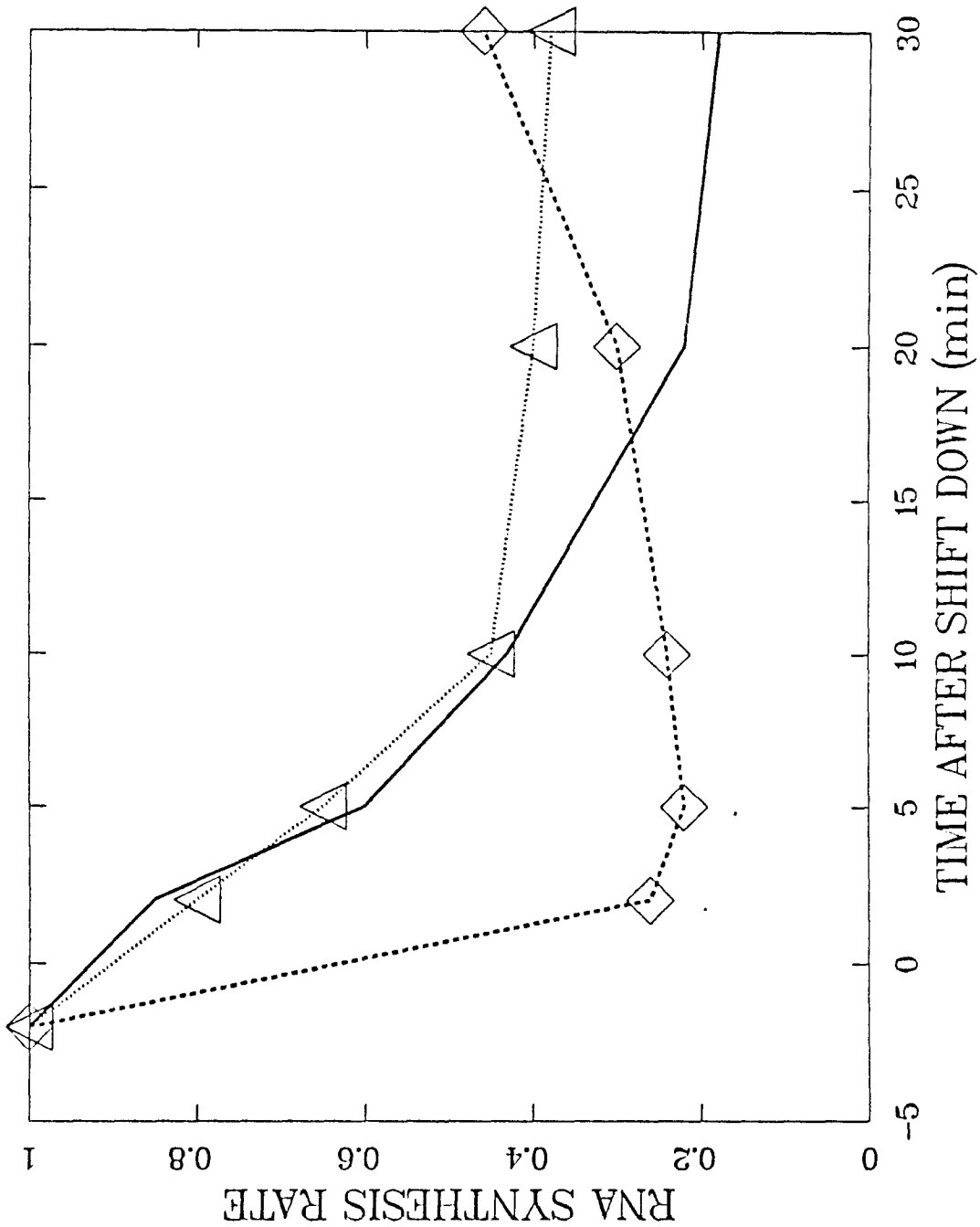

RNA SYNTHESIS RATE
0
0.2
0.4
0.6
0.8
1
TIME AFTER SHIFT DOWN (min)
-5
0
5
10
15
20
25
30

FIGURE 4b

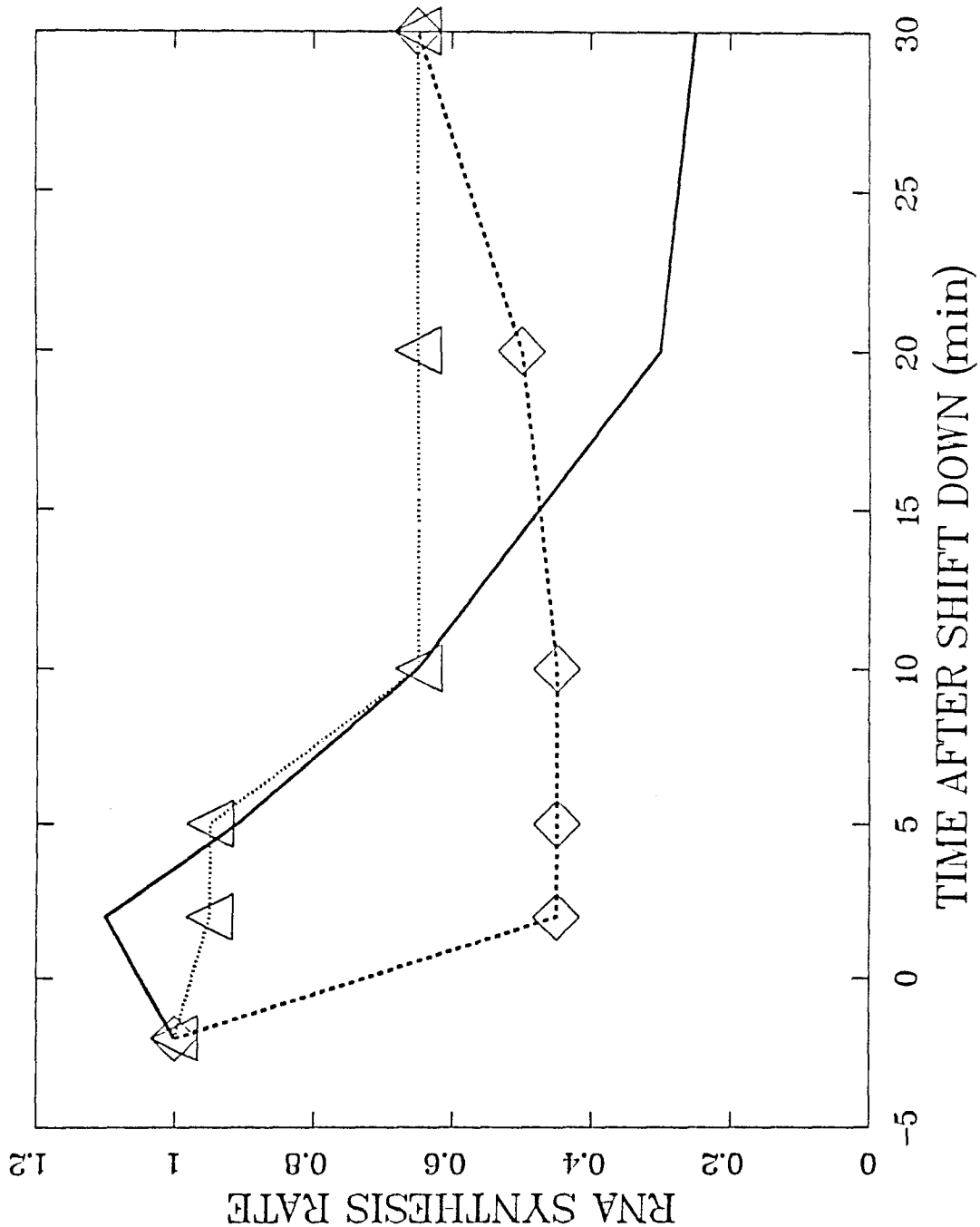

RNA SYNTHESIS RATE
0
0.2
0.4
0.6
0.8
1
1.2
TIME AFTER SHIFT DOWN (min)
-5
0
5
10
15
20
25
30

FIGURE 4c

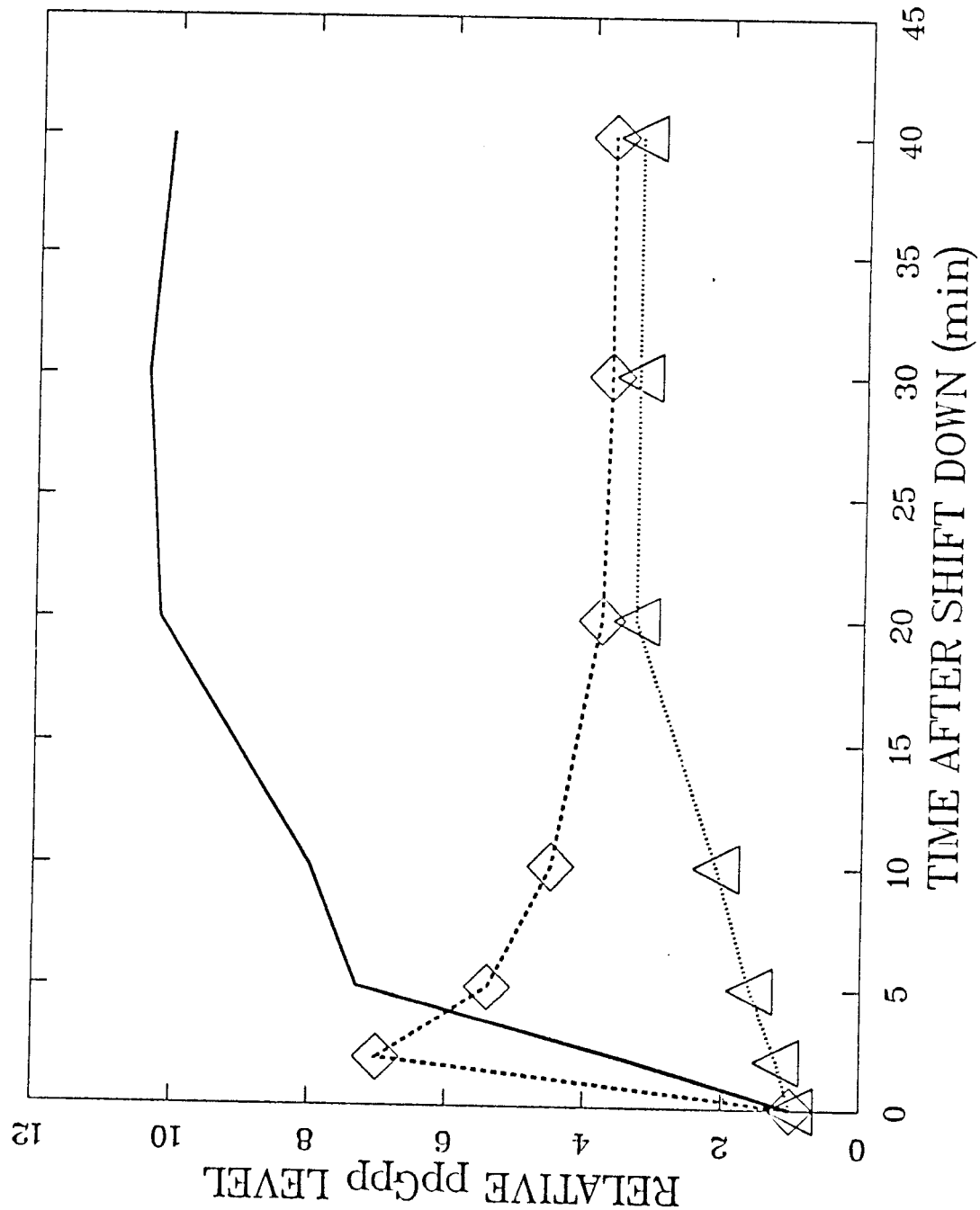

RELATIVE ppGpp LEVEL
0
2
4
6
8
10
12
TIME AFTER SHIFT DOWN (min)
0
5
10
15
20
25
30
35
40
45

FIGURE 5

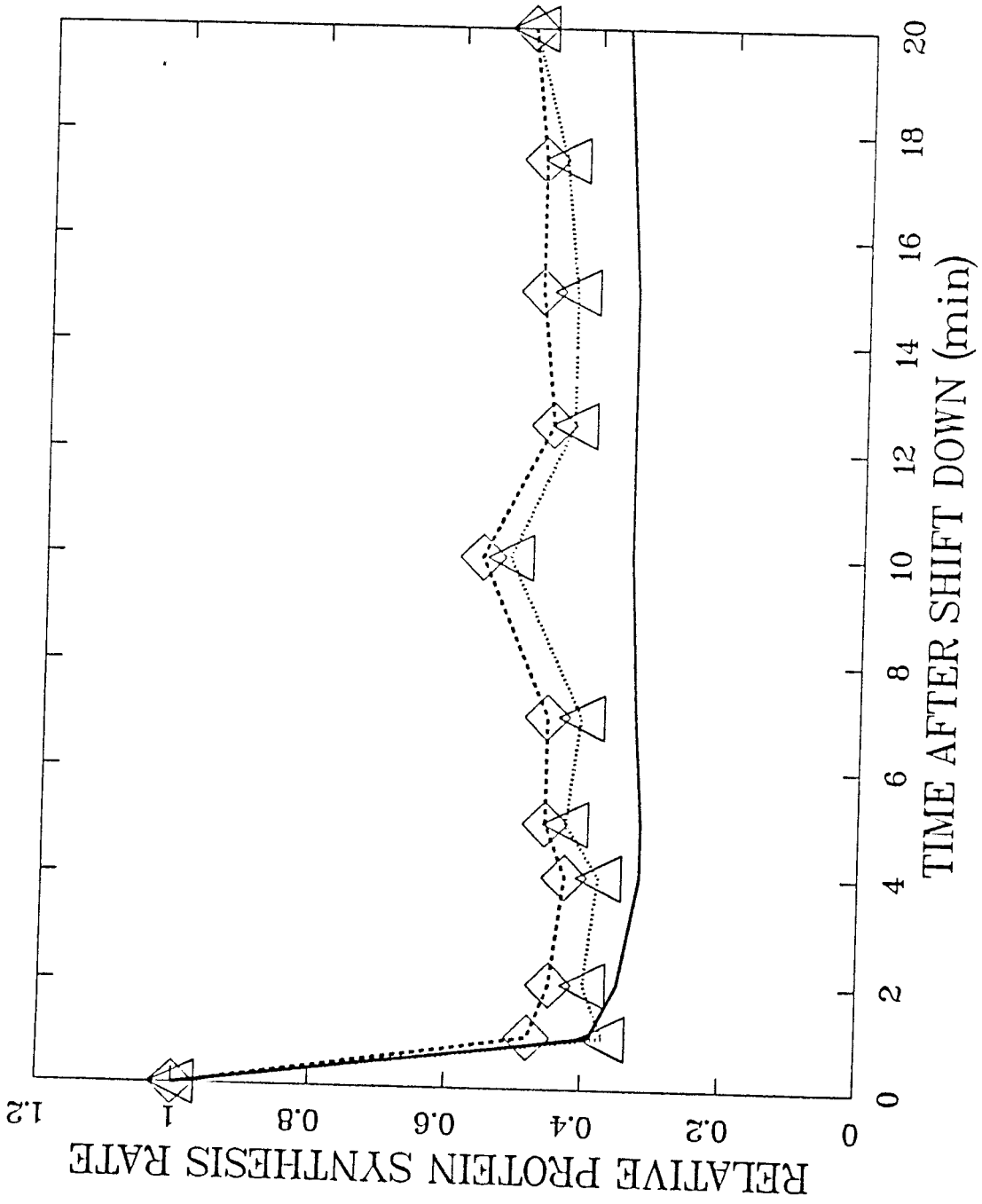

RELATIVE PROTEIN SYNTHESIS RATE
0
0.2
0.4
0.6
0.8
1
1.2
TIME AFTER SHIFT DOWN (min)
0
2
4
6
8
10
12
14
16
18
20

FIGURE 6

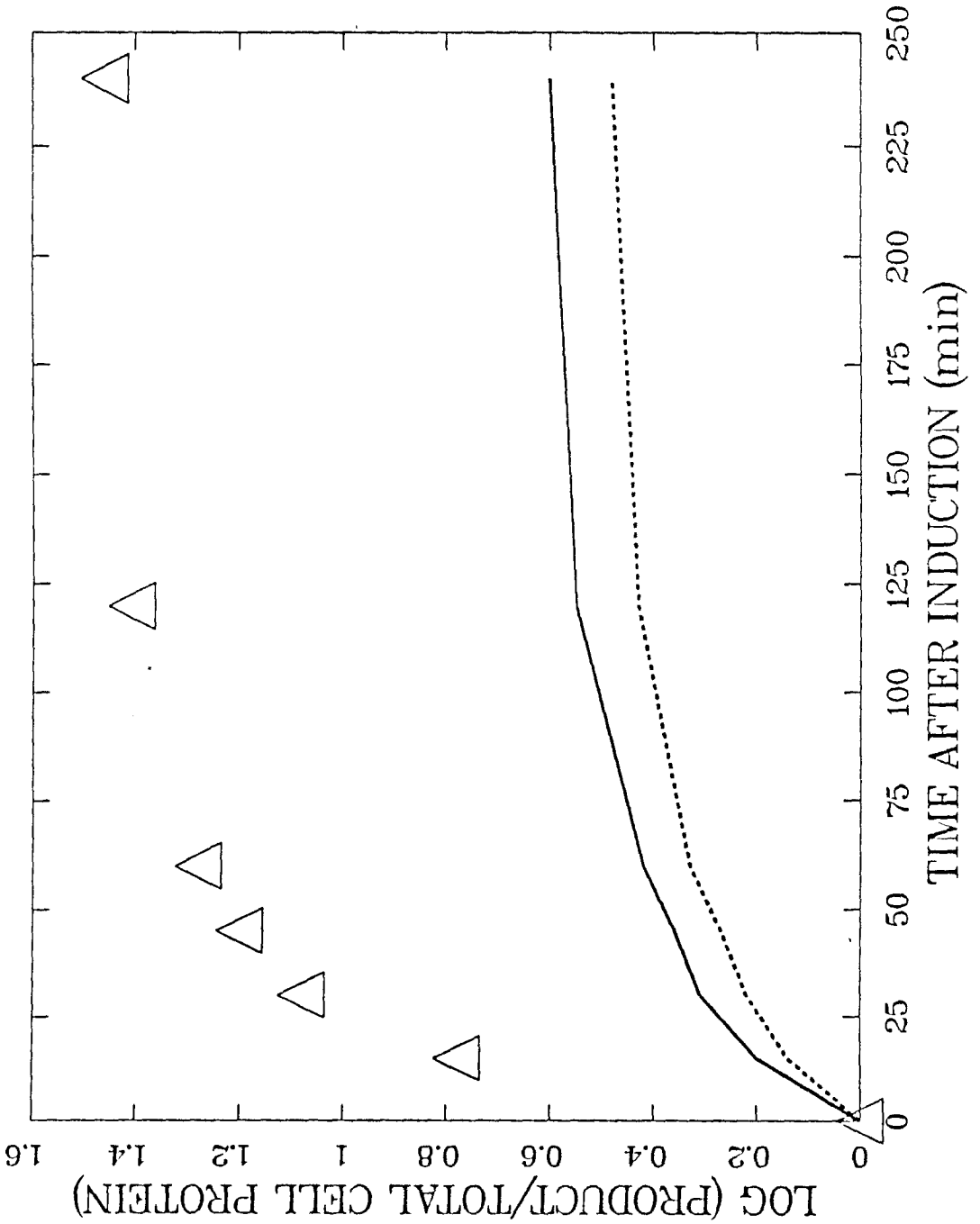

LOG (PRODUCT/TOTAL CELL PROTEIN)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
TIME AFTER INDUCTION (min)
0
25
50
75
100
125
150
175
200
225
250

FIGURE 7

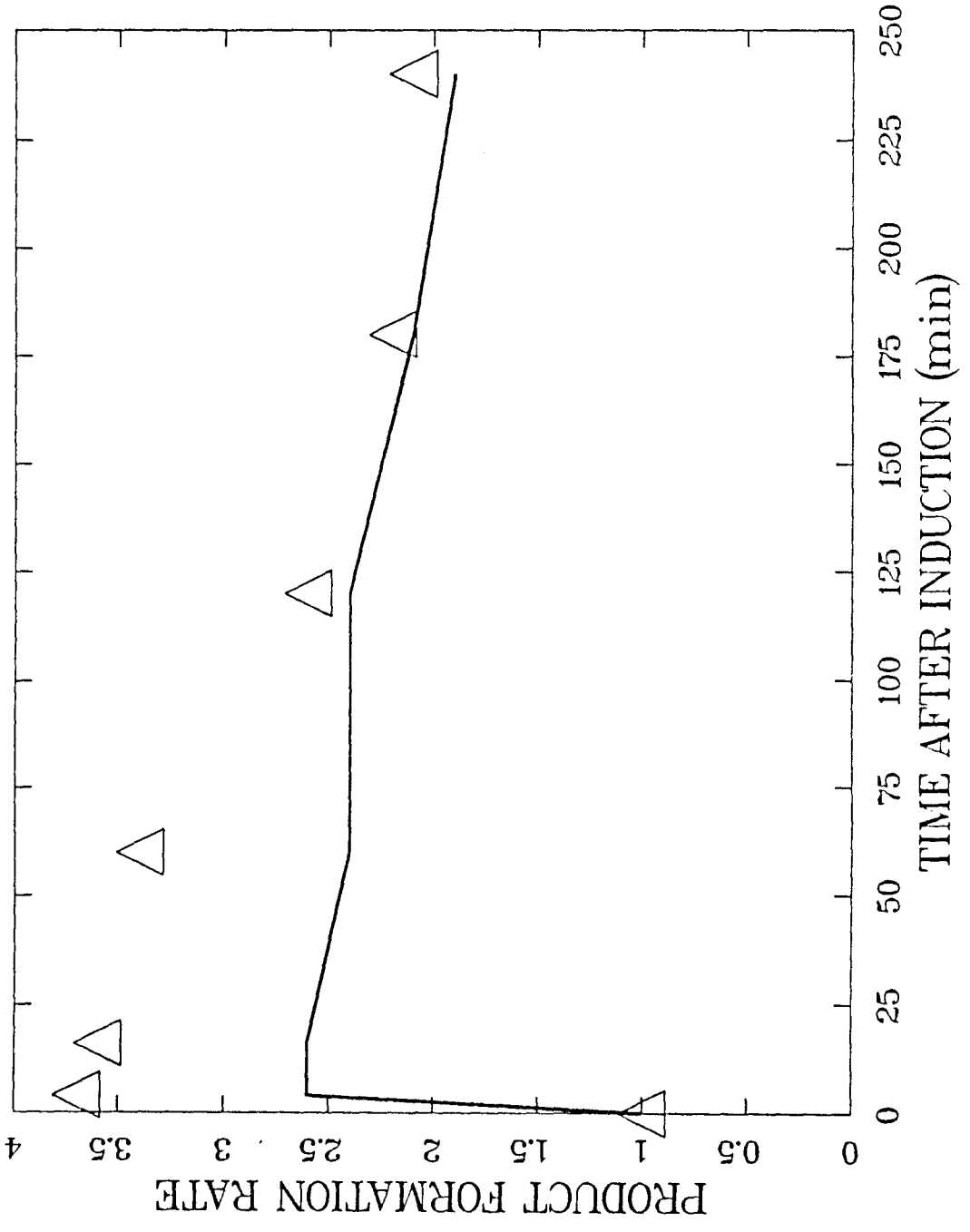

PRODUCT FORMATION RATE
0
0.5
1
1.5
2
2.5
3
3.5
4
0
25
50
75
100
125
150
175
200
225
250
TIME AFTER INDUCTION (min)

FIGURE 8

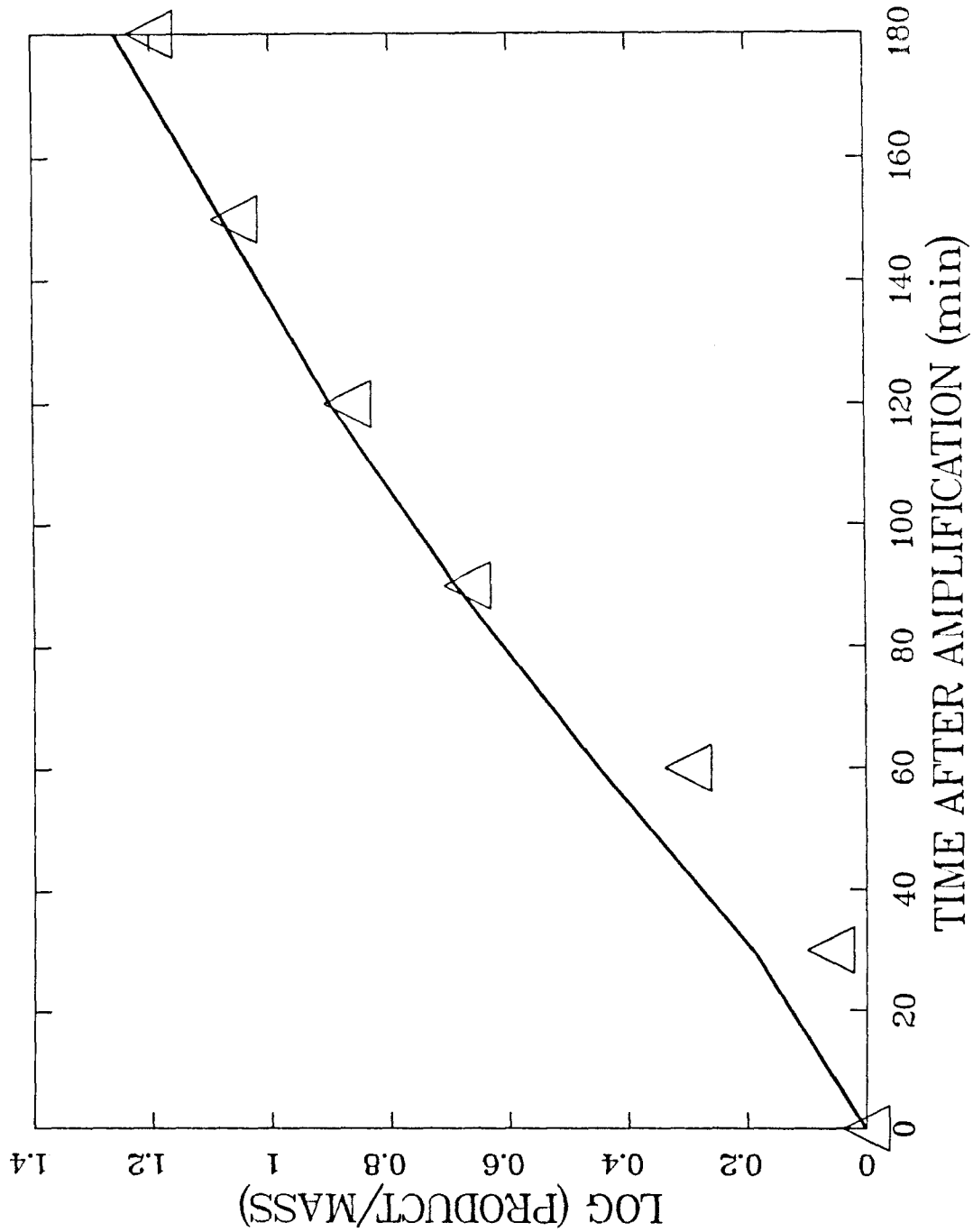

LOG (PRODUCT/MASS)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
TIME AFTER AMPLIFICATION (min)
0
20
40
60
80
100
120
140
160
180

FIGURE 9

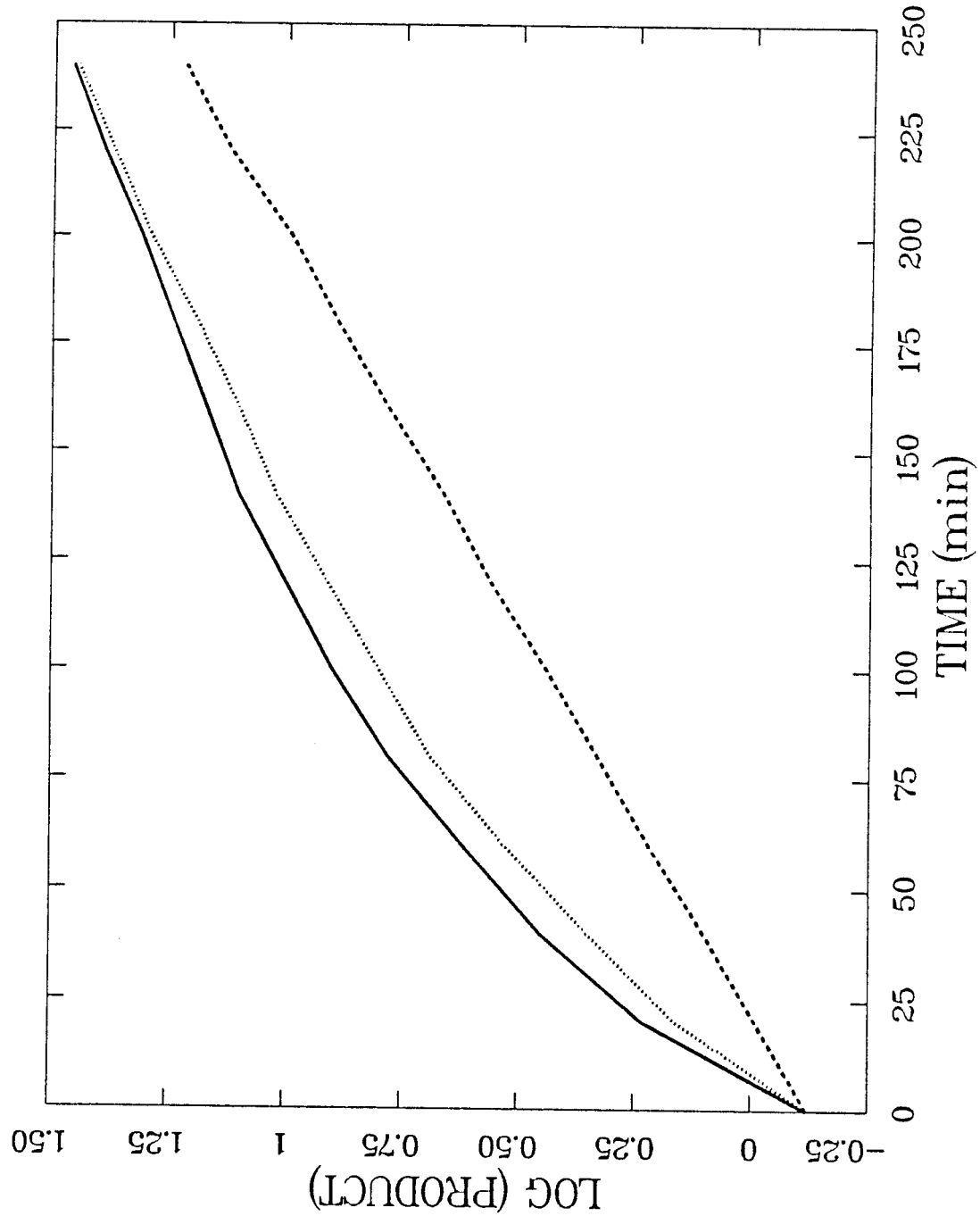

LOG (PRODUCT)
-0.25
0
0.25
0.50
0.75
1
1.25
1.50
TIME (min)
0
25
50
75
100
125
150
175
200
225
250

FIGURE 10

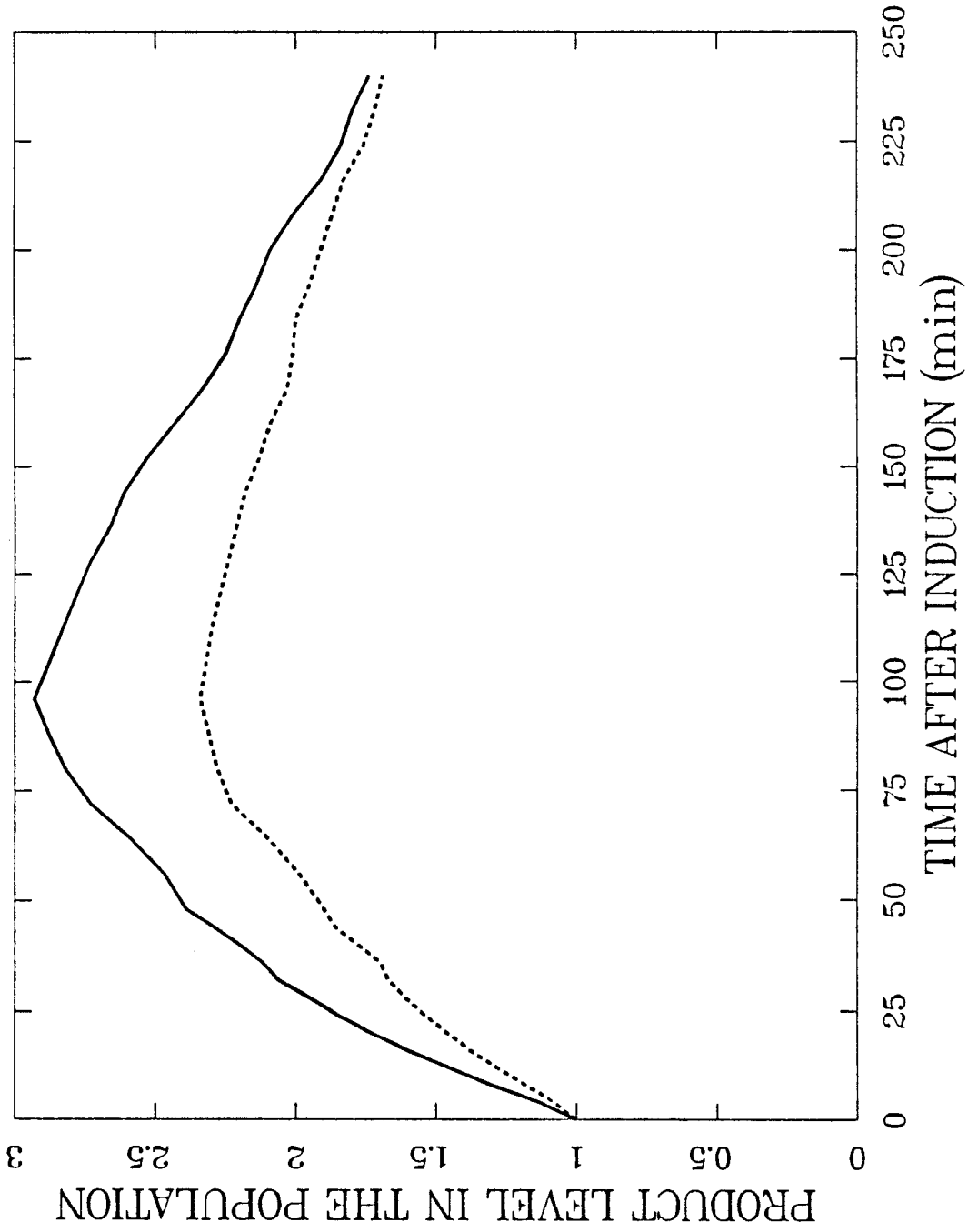

PRODUCT LEVEL IN THE POPULATION
0
0.5
1
1.5
2
2.5
3
0
25
50
75
100
125
150
175
200
225
250
TIME AFTER INDUCTION (min)

FIGURE 11

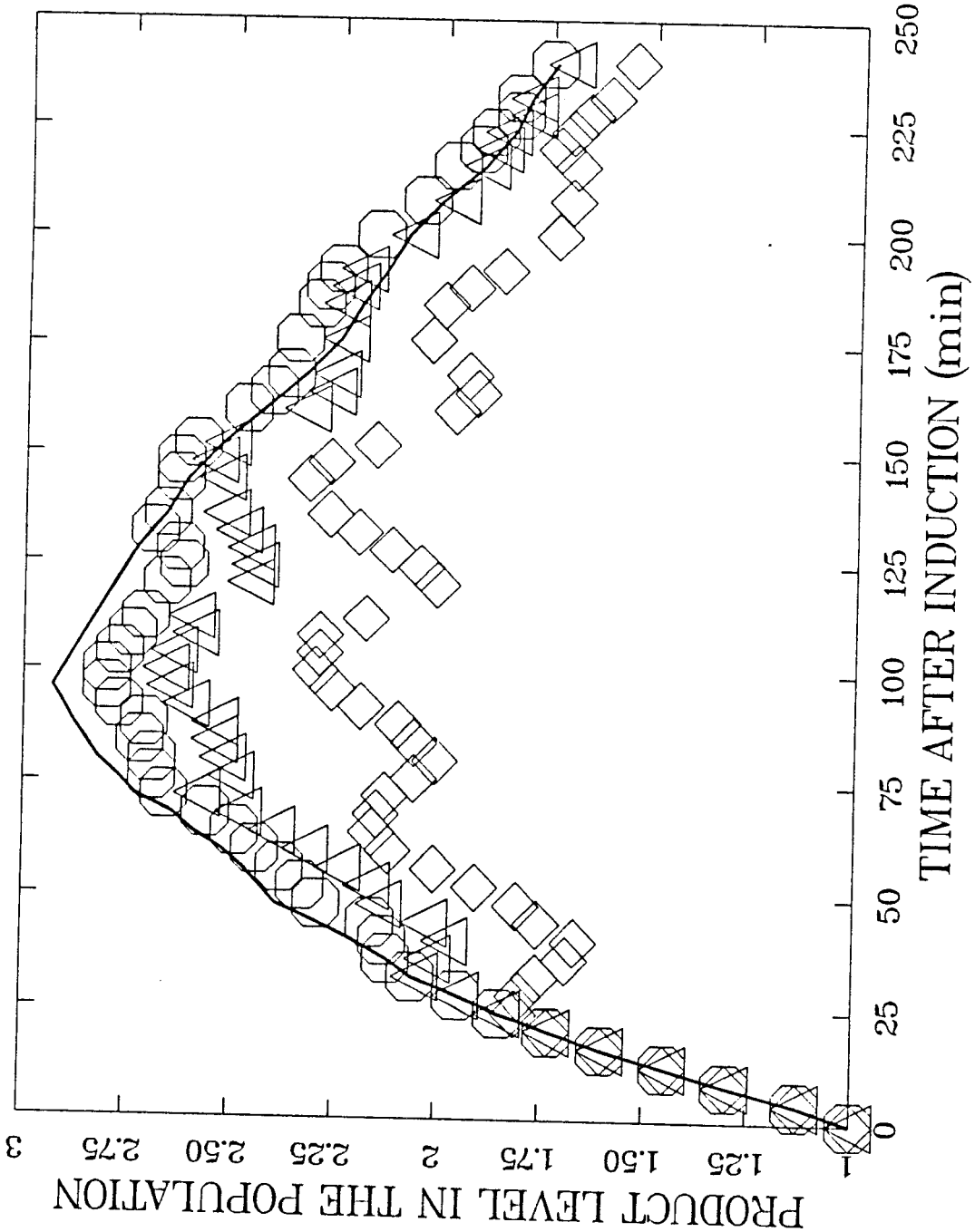

PRODUCT LEVEL IN THE POPULATION
1
1.25
1.50
1.75
2
2.25
2.50
2.75
3
0
25
50
75
100
125
150
175
200
225
250
TIME AFTER INDUCTION (min)

FIGURE 12

Chapter 4

Copy Number Effect on Transcription from Plasmid Genes, Macromolecular Stability and Cell Specific Productivity

Abstract

Experiments were performed to evaluate, qualitatively and quantitatively, the cell's adaptation to plasmid maintenance and cloned gene expression. Experimental findings indicate that the metabolic response to low plasmid levels is to increase the biosynthetic capacity of both transcription and translation. However, at high copy number levels a divergence is noted, whereby the gene specific transcription rate continues to increase but the stability of plasmid-derived mRNA drops sharply. Protein levels are maintained, but translation efficiency decreases. The implication of these responses are briefly discussed in light of the design of recombinant bacterial systems for the efficient production of foreign proteins.

INTRODUCTION

Genetic engineering of microorganisms makes possible the large-scale production of molecules native to both procaryotes and eucaryotes. To accomplish this, a plasmid or other DNA vector carrying the gene for the desired product is introduced into a suitable host organism. The transformed host is then grown, providing the metabolic machinery necessary to synthesize the plasmid-mediated product. Many factors contribute to the efficiency of plasmid product formation. These arise from two sources, the vector and the host organism. The construction of an efficient expression vector will include considerations of vector copy number and its control, the level and control of transcription from the cloned gene promoter and the efficiency of translation of the resulting messenger RNA. These factors are all theoretically within the direct control of the experimenter who constructs the vector. Host-dependent factors are more difficult to identify or influence. These may include the presence of certain metabolic functions such as protein export systems, post-transcriptional and post-translational modification capabilities, the overall growth rates capable of being attained, or the ability to withstand the presence of metabolic byproducts related to the synthesis or function of the cloned gene product. A key consideration in the choice of a host or construction of a productive vector is the influence each has on the other's activity. The nature of the interaction, and the magnitude of the effect are critical to the successful selection of a host-vector system. The plasmid generally relies on the host to provide the enzymatic activity to replicate itself and to synthesize the requisite mRNA and proteins. High productivity depends, in part, on the the ability of the plasmid system to successfully compete for these metabolic processing units (DNA polymerase, RNA polymerase, ribosomes). It has been reported that the amount of cloned gene product accumulated in recombinant *E. coli* cells is

directly proportional to the plasmid copy number (1). Conversely, cells with plasmid grow more slowly than plasmid-free cells (2,3), and in the case of highly amplified plasmids, lose viability completely (4). There must exist some well balanced configuration such that alteration of copy number, transcription or translation efficiency will have a deleterious effect on system productivity.

Transcription has a central role in the control of cellular metabolism. The distribution of RNA polymerase on different promoters and the activity of the enzyme will largely determine the operable metabolic pathways and the material flows through each process via control of the synthesis rates of key enzymes. The relative rate of transcription initiation from a promoter will be the primary control on the amount of its mRNA found in the cell. There are additional factors which will effect the amount of a given mRNA in the cell, including attenuation of transcription and the stability of the transcript.

Translation of the relevant mRNA also effects productivity from a given gene. The efficiency of translation from a single transcript is known to change with growth rate (5) and may be effected by other factors. When studying productivity from a cloned gene it is necessary therefore, to consider the rate of transcription initiation from the cloned gene's promoter, the efficiency of transcription of the gene and the translational efficiency of the plasmid-derived mRNA.

It is the objective of this work to identity and tentatively quantify the magnitude of the effect of plasmid maintenance and gene expression on aspects of transcription and translation. The effect of plasmid presence on the initiation and efficiency of transcription from the plasmid promoter and the efficiency of plasmid gene mRNA translation will be studied. These studies will also determine if macromolecular stability is affected by plasmid-induced alterations in cell metabolism. The same host strain was used throughout to guarantee metabolic homogeneity for the different copy number cases. The copy number plasmids used are

closely related, and enable constitutive expression of β-lactamase.

MATERIALS AND METHODS

Bacterial Strains and Plasmids

Escherichia coli HB101 (pro^-,leu^-, thi^-,$recA^-$) was used as the host strain exclusively. The five plasmids used, RSF1050, pFH118, pDM246, pDM247, pDM248, have been described previously (1). The copy numbers of these plasmids in *E. coli* HB101 grown at 37°C in Luria Broth medium are 60, 408, 24, 12, and 120 respectively. Additionally, each plasmid carries a gene labeled Ap^r which provides for the expression of the enzyme β-lactamase in *E. coli*.

Media and Cultivation

Growth medium is M9 with 0.4% casamino acids, 0.2 mg/L thiamine HC1, 200 mg/L proline and 50 mg/L leucine. Silicone antifoam and ampicillin (Na salt) are added at a concentration of 50 mg/L. Glucose is the carbon source at a concentration of 0.2%. Cells are preincubated at 37°C in Luria Broth containing 50 mg/L ampicillin. When grown to exponential phase, 50 mL of cell culture are used to innoculate at Microferm (New Brunswick Sci. Co.) vessel, with a working volume of 2.1 L. pH is controlled at 7.0 + 0.1 and temperature is maintained at 37°C. Steady state is assumed achieved after the passage of five residence times of continuous constant feed of medium.

Cells were enumerated by plating onto selective and non-selective LB agar plates. This procedure also served to indicate the size of the plasmid-free population in the chemostat. This information was used to calculate the cell-specific levels of the various macromolecules measured. Only one cultivation, HB101:pFH118, indicated any significant plasmid-free subpopulation, and in that instance the

plasmid-free fraction was approximately 20% of the total.

Determination of β-lactamase Activity Band Total Protein Content

Twenty milliliters of culture are harvested and frozen until the assay is performed. Sample preparation follows the method suggested by Yamamoto *et al.* (6). The sample is thawed and washed twice with a buffered solution consisting of 50mM phosphate buffer (pH 7.0) with 0.14 M NaCl and 10 mM $MgSO_4$. The pellet is resuspended in 0.1 M phosphate buffer (pH 7.0) and the cells are lysed by sonication (Ultrasonics, Inc., Model W-375). Lysate is centrifuged at 15,000 rpm for 30 minutes at 4^oC. The supernatant is used for subsequent determinations of total protein and β-lactamase activity.

The iodometric method (7) is used to assay for β-lactamase activity. Tubes containing 0.5 mL of 20 mM benzyl penicillin (K salt) dissolved in 0.1 M phosphate buffer (pH 7.0) are preincubated at 30^oC. An appropriate sample volume is made up to 3 mL by the addition of 2.5 mL of iodine reagent. After 10 minutes incubation at 30^oC, the absorbance at 500 nm is determined. Sigma Protein Kit No. 690 is used to determine total protein and the specific activity of β-lactamase is computed using the formula by Sawai *et al.* (7). One unit of β-lactamase specific activity hydrolyzes 1 micromole of benzyl penicillin per minute and per mg of cellular protein at 30^oC.

Pulse Label and Purification of RNA

RNA is labeled by the addition of (5-^{3}H) uracil to a final activity of 10 μCi/mL to 20 mL of culture. Incorporation proceeds for 1 minute, then is terminated by pouring the cells over ice and Na azide (final concentration 1.0 mM). After harvest by centrifugation and resuspension in 1 mL of medium C (40 mM NH_4CL, 40 mM Na_2HPO_4, 50 mM NaCL), the cells are lysed by addition to 1 mL

SDS lysis mixture (0.1 M NaCl, 0.5% SDS, 0.01 M EDTA) at 95°C for 20 seconds. Following 3 TAE-saturated phenol extractions and an ethanol precipitation, the RNA is resuspended in TAE and 5 equal volume extractions with ether are performed to remove residual phenol. N_2 gas is passed over the sample to remove the ether. After the Mg^{+2} concentration is adjusted to 5 mM through the addition of 1/10 volume 50 mM $MgSO_4$, RNase-free DNase 1 is added to a final concentration of 5 units/mL and the samples incubated for 30 minutes at room temperature to digest the DNA. EDTA is added to a final concentration of 5 mM and the sample is extracted with phenol and precipitated as before. After dissolution in 2x SSC, ether extractions were performed and the absorbance at 260 nm of a 1/100 dilution of the samples was measured. Two microliters of this dilution were mixed with 10 mL of scintillation cocktail and counted for 50 minutes.

Isolation and Labeling of DNA Probe

The plasmid RSF1050 was used as the source of the β-lactamase gene used as a probe. The restriction map for RSF1050 (8) indicates the β-lactamase gene is completely situated on a single HaeIII fragment which also contains the unique Bam Hl restriction site found on the plasmid. A double digest was performed using HaeIII and BamHl. The fragments were resolved by electrophoresis through a 1% agarose gel containing 0.1% EtBr. A single digestion with HaeIII was electrophoresed concurrently to indicate the fragment cut by BamHl. The fragment carrying the β-lactamase gene was cut from the gel and the DNA extracted. The DNA fragments were radio-labeled by nick-translation with 300 Ci/mmol each of ^{32}P dGTP, dCTP, dATP, and dTTP.

DNA Hybridizations and Assay of RNase Resistant Hybrids

Probe excess hybridizations were carried out with purified DNA probes and

total RNA isolated from the HB101/pFH118 culture. This RNA was chosen because pFH118 is the highest copy number plasmid in the group studied, and this sample is therefore presumed to contain the highest fraction of β-lactamase mRNA. DNA probe was immobilized on nitrocellulose filter using a Hybridot (BRL) apparatus. The dots were air dried and baked for 2 hours at 80°C in a vacuum oven.

Dots with 2.6 μg probe DNA attached were prehybridized in 3/2x hybridization buffer (5x buffer is 5x SET, 1x Denhardt's solution, 20 mM sodium phosphate buffer pH 6.8, 1% SDS, and 50 μg/ml calf-thymus DNA), 5x Denhardt's solution, 0.2% SDS and 5 μg/μl yeast RNA. The total prehybridization volume is 0.1 mL. After 3 hours incubation at 68°C, the solution was replaced with the sample solution. Sample solution is identical to that for prehybridization except the yeast RNA has been replaced with a combination of radiolabelled *E. coli* RNA and yeast RNA such that the total amount of RNA in the hybridization is approximately 510 micrograms. To identify the linear hybridization capacity of the probe DNA, *E. coli* RNA was added in the range from 2-100 micrograms. The filters were hybridized for 60 - 70 hours at 68°C. The reaction mix was removed and the filters washed. Probe excess was maintained for hybridization with more than 100 μg total *E. coli* RNA. As a result, subsequent hybridizations used filters with 2.4 μg probe DNA and 75 μg labeled *E. coli* RNA.

DNA filters were prepared as described by Kafatos (9). A portion of β-lactamase fragment DNA (2.4 μg) was immobilized per dot. These filters were prehybridized for 4 hours at 68°C. Hybridizations with 75 μg of ^{3}H labelled transcripts from cell extracts were performed at 68°C for at least 60 hours. Each reaction mixture contained two filters, one carrying 2.4 μg of β-lactamase DNA, the other carrying an equal amount of *S. purpuratus* DNA. As described above, the filter bound DNA was always in effective excess over homologous transcripts. Fol-

lowing hybridization, the filters were washed at 68°C in 50 ml of these buffers:4x SET, 0.2% SDS (20 minutes, twice); 2x SET, 0.2% SDS (20 minutes, twice); 1x SET, 0.2% SDS (1 hour, twice). Filters were rinsed twice in 2.5x SET to remove SDS, then treated with 5.6 μg/mL RNase A and 12 U/mL RNase T_1 for 1 hour at 36°C. After a 15 minute wash in 4x SET, 0.2% SDS, filters were dried, placed in scintillation vials, digested overnight in 1 mL of tissue solubilizer (TS-1 Amersham) at 37°C, then counted in a toluene based liquid scintillation fluid (Liquifluor; New England Nuclear).

RNA Blot Hybridization

Unlabeled RNA (4 μg) from each culture was diluted with 20 x SSC and loaded onto nitrocellulose filters using a Minifold slot blot apparatus (Schleicher and Schuell). Egg total RNA isolated from sea urchin was used as a control for the hybridizations. After 4 hours of prehybridization at 68°C, heat-denatured radioactive DNA probe was added and incubation continued for 24 hours. The filters were then washed at 68°C as follows: two 20 minute washes in 4x SET, 0.2% SDS; two 20 minute washes in 2x SET, 0.2% SES; and two 60 minute washes in 1x SET, 0.2% SDS. The filters were then mounted with Kodak X-ray film overnight at 70°C. The resulting film was scanned with a laser densitometer (Joyce-Loebl) and the tracings were cut and weighed. In addition, the slots were cut from the filter and counted in liquid scintillation fluid (Beckman EP Solve).

RESULTS

Total Cellular RNA Levels

Table 1 shows the average cellular RNA content for populations growing with identical specific growth rates, but carrying plasmids which are propagated at

different copy numbers. RNA level increases monotonically with plasmid copy number, and the relationship is essentially linear. At steady-state in a bioreactor, the RNA level can be calculated using the following formula:

$$RNA_{ss} = \eta G_{ss} / (K_{dr} + \mu)$$

where K_{dr} is the decay constant for RNA degradation, μ is the specific growth rate, G_{ss} is the number of genes with active promoters, RNA_{ss} is the amount of RNA present in the cell and η is the efficiency of gene transcription (see Appendix for development of this equation). This modeling framework was developed and successfully applied to the working of the lac operon (10) and the phage λdv system (11) by Lee and Bailey.

The transcription efficiency, η, incorporates the amount and activity of RNA polymerase in the cell, as well as the availability of ribonucleotides and other necessary cofactors. We have chosen this approach because it enables one to discern trends and quantitatively assess the impact of plasmid presence on transcription.

Table 2 shows the range of transcription efficiency necessary to encompass the RNA levels measured. The subscripts c and p denote chromosome-and plasmid-related variables respectively. Each *E. coli* chromosome carries roughly 890 genes (12) and the equation for the average genome level in a population, derived by Cooper and Helmstetter (13),

$$G = \frac{t}{C \ln 2} \left[2^{(C+D)/t} - 2^{D/t} \right]$$

gives 1.6 chromosomes per cell as the level expected in a population growing with a specific growth rate of 0.5 $hr.^{-1}$, the rate for these experiments. In this equation, G is the average genomic complement, t is the doubling time in minutes and C and D are cell cycle parameters. For cells growing at a given rate, cell cycle analysis (13) shows that the chromosome complement will be constant on a per cell basis.

There will be over 1400 chromosomal genes available for transcription. Since the chromosome complement will remain fixed, the only increase in available gene number is due to copy number increase. The data indicate a roughly fortyfold increase in transcription activity over the range of copy numbers studied. Detailed analysis of the data will be developed in the discussion section, where it will be integrated with analysis of other experimental findings.

Total Cellular Protein

Figure 1 shows the variation of cellular protein content as plasmid copy number increases. As with RNA level, protein levels are higher in plasmid-containing (P^+) cells than in plasmid-free (P^-) cells growing at the same rate. The maximum found for total protein may be due to changes in protein stability or translation efficiency. To further analyze the response in the absence of relevant data, it is necessary to make an assumption about the fraction of RNA which is mRNA. For subsequent analysis, this fraction is assumed constant.

If the protein stability remains constant over the range of plasmid copy number studied, the translation efficiency will parallel the mass ratio for protein: RNA shown in Table 3. Conversely, if the translation frequency is held constant, the protein degradation rate will vary as indicated in Table 4. Table 3 indicates up to a twentyfold decrease in translation efficiency, while Table 4 shows over one hundredfold increase in protein degradation rate. The appearance of a negative degradation constant in Table 4 indicates that the translation efficiency must increase for this case. The implications of this data will be addressed later.

β-lactamase mRNA Synthesis Rate

Table 5 presents the results of ^3H pulse-labeled RNA which was subsequently isolated and hybridized to nitrocellulose-bound DNA from the gene for β-

lactamase. These values are directly proportional to the synthesis levels of β-lactamase mRNA for the different copy number cases.

This data enables the calculation of the transcription efficiency for the plasmid-borne β-lactamase operon shown in Table 6. The efficiency remains essentially constant through copy number of sixty, but increases dramatically at higher copy levels. This gene has been presumed to be constitutively expressed. This is compatible with the linear relationship between hybridized label and copy number for the 12 and 60 plasmid cases, but at odds with data for higher plasmid levels. The rapid increase in gene specific transcription rate for plasmid copy number levels above 60 may represent some derepression phenomenon, the mechanism and implication of which shall be speculated upon later.

Intracellular β-lactamase mRNA Level

Figure 2 shows the variation of labelled intracellular β-lactamase mRNA level with copy number. Beyond a plasmid copy number of approximately 25, the increase in mRNA level is linearly proportional to copy number. This is in contrast with synthesis of this RNA species, as was described above and shown in Table 5. Data for both synthesis rate and overall level of β-lactamase mRNA over a range of plasmid copy number levels allow a direct determination of the stability of β-lactamase messenger RNA. Applying the equation for the steady-state level of a macromolecule in a chemostat to β-lactamase mRNA gives

$$mRNA_{\beta L} = \eta_{BL} G_{BL}/(K_{dmBL} + \mu)$$

The numerator, η_{BL} G_{BL}, is equal to the synthesis rate previously measured. Assuming K_{dmBL} is equal to that for an "average" chromosomal messenger RNA molecule at the lowest copy number, Table 7 shows the stability of the β-lactamase mRNA for different plasmid copy number levels in the cell. There is a definite minimum in the mRNA decay constant as a function of plasmid copy number.

This could indicate the existence of two responses to plasmid presence and heterologous gene expression that effect mRNA stability.

β-lactamase Specific Activity

Figure 3 shows the specific activity of β-lactamase per cell as plasmid copy number increases. There is a distinct maximum for the plasmid propagated with a copy number of sixty, as was the trend noted for total cell protein (see Fig. 1). A plot of enzymatic activity per mass of total cell protein versus plasmid copy number shows similar results. Lacking data for the synthesis rate of β-lactamase, it is not possible to identify the translation efficiencies explicitly for these cases. Instead the translation ratio, defined as product level per messenger RNA, will be used as an indication of protein synthetic activity. Table 8 shows the ratio for the plasmids studied. The translation ratio will depend on the translation efficiency, E_{BL}, and the β-lactamase protein decay constant, K_{dBL}. If the enzymatic activity is used as a measure of the enzyme present, applying the equations developed for steady-state growth in a chemostat gives

$$BL_{ss} = mRNA_{BLss} * E_{BLss} / (K_{dBL} + \mu)$$

The translation ratio, is therefore equal to

$$\frac{BL_{ss}}{mRNA_{BLss}} = \frac{E_{BLss}}{K_{dBL} + \mu}$$

If E_{BLss} remains constant for the different copy number plasmids, this information can be used to calculate changes in the β-lactamase decay constant, K_{dBL}. As shown in Table 9, a monotonic increase in decay rate is calculated. There is a more than thousandfold increase in the protein degradation constant. The maximum value shown is close to that calculated using data regarding the degradation of abnormal (14) or missence (15) proteins.

DISCUSSION

Most of the experimental results are in agreement with a plausible set of mechanistic responses. Unfortunately, the dramatic increase in total RNA synthesis rates exhibited thus far defies simple explanation. Increasing the RNA synthesis rate can be accomplished through two primary mechanisms. These are increasing the number of RNA polymerase molecules in the cell and increasing the fraction of molecules involved in transcription. Derepression of the rpoBC operon, which encodes for the β:β' subunits of RNA polymerase, has been shown to increase synthesis of RNA polymerase up to threefold (16). Regarding the active fraction of RNA polymerase, it has been reported (5) that at the growth rate used in this work only roughly 25% of the total polymerase was found to be active in transcription. This agrees with calculations using equilibrium binding constants. Increasing the active fraction would be possible, since the cell regulates transcription of its own genes and surely does not utilize its entire gene complement at any given growth rate. It is therefore plausible that an adaptive response to the metabolic load imposed by plasmid maintenance and gene expression would consist of increasing the strengths of already active promoters and switching-on promoters inactive under plasmid-free conditions.

If one assumes that ultimately RNA polymerase levels will be rate limiting, then the maximum attainable increase in overall RNA synthesis rates can be estimated. If utilization of RNA polymerase approached 100%, that would lead to an increase by a factor of four over normal RNAP activity levels. Coupled with a threefold increase in polymerase level and the additional threefold increase in all protein levels suggested by the data shown in Figure 1, the total RNA polymerase activity could potentially increase as much as fortyfold. It is however, very unrealistic that more than 60-70% of the total polymerase present in the cell could be activated, given the equilibrium binding of core enzyme and the disproportionate

number of non-productive binding sites available along the chromosome. This leaves little recourse but to speculate on potential mechanisms which could further increase RNA synthesis rates. One underlying assumption in the preceeding analysis is that the chain elongation rate for transcription is limited to that calculated for synthesis of rRNA, roughly 75 nucleotides per second. The mechanism that limits the elongation rate for mRNA synthesis to approximately 50 nucleotides per second may likewise be limiting rRNA transcription. The postulated attenuation may be relieved by binding of the effector molecules to increased numbers of active polymerase molecules until the free concentration of the modulator molecule is essentially zero. It has been suggested elsewhere that an attenuator system exists in *E. coli* (17) and the observed behavior supports this proposition.

The maxima for the total protein:RNA ratio, β-lactamase mRNA stability and β-lactamase translation ratio are consistent with one another, and suggest a mechanism involving cellular ribosome content. The translation ratio maximum, which coincides with the maxima for total protein and β-lactamase activity, may indicate a maximum in the intracellular ribosome population. If the cell responds to increasing copy number by stimulating RNA polymerase synthesis, it is likely that in addition, extra ribosomes will be synthesized to translate the extra mRNA produced. There may be a limit to the ribosome level attainable by the cell. For copy numbers greater than that where the maximum is first achieved, the further increase in β-lactamase mRNA or total RNA indicated by these experiments could lead to increased spacing between ribosomes along messenger RNA. If mRNA stability is dependent upon ribosome coverage, decreased coverage would cause an increase in the decay rate, as indicated in Table 7 for copy numbers of 120 and 400. Similarly, the increase in stability from copy number 12 to 60 suggests increased ribosome coverage, which is consistent with the elevated ribosome popu-

lation postulated.

The reduction in translation ratio for high copy number could simply be the result of a stable ribosome population which is saturated by mRNA molecules. As the mRNA level increases beyond the translation capacity of the ribosomes, the protein:mRNA ratio will necessarily decrease. It is also possible that the absolute ribosome population declines at high copy levels, although this was not investigated.

Another possible explanation for decreased translation ratio is related to the suitability of the mRNA to act as a template for protein synthesis. If the decay rate for mRNA is increased, the functional capability of the existing message is also called into question. It may be that much of the mRNA present is not complete, lacking a ribosome binding site or coding for an incomplete polypeptide which is rapidly degraded.

Proteolysis of abbreviated polypeptides is also a potential explanation for the proposed increase in the protein degradation rate constant at high plasmid levels. Abbreviated polypeptides could also result from inadequate supplies of charged tRNA. The proposed increase in proteolysis rate need not be restricted to incomplete polypeptides. The increase in plasmid-derived protein may stimulate the proteolytic system existing in *E. coli* for the degradation of abnormal and foreign proteins described by Goldberg (14). In this way, protein stability in general may be decreased.

SUMMARY

The primary implication of this work for the design of host/plasmid systems is that plasmid gene promoter strength should be maximized to take advantage of the stimulation of transcription activity occurring at low copy number. Normally, over half of the RNA polymerase enzymes in the cell are bound non-specifically to

DNA and are inactive. Increasing promoter strength will increase polymerase usage more efficiently than increasing copy number, since adding plasmid to the cell also augments the cellular concentration of non-promoter sites available for non-specific binding of core enzyme. Keeping copy number low may also serve to maintain the ribosome population in the cell, if the primary cause of decreasing productivity at high copy levels is decreased transcription of rRNA or ribosome-protein operons brought about by core enzyme binding to non-specific sites.

The results described provide a first look at the details underlying host-plasmid interactions. The attempt was made to identify the aspects of metabolism most severely effected and uncover the location of processing bottlenecks. From this standpoint, the experiments tentatively identify translation as the limiting process in gene expression for highly amplified systems, although this deficiency may be the result of insufficient synthesis of rRNA or r-protein mRNA.

These results serve a second, equally important function. They indicate new areas for further, more detailed experimental and theoretical studies, raising important questions about the control of RNA polymerase synthesis, transcription, mRNA stability and the control of ribosomal operons. Investigations into these areas will provide information which will enable the detailed simulation of host-vector interactions, leading to the design of recombinant systems which will exploit or circumvent metabolic control structures in order to enhance productivity.

References

1. Jin-Ho Seo and J. E. Bailey, *Biotechnol. Bioeng.*, **27**, 1668 (1985).

2. D. Godwin and J. H. Slater, *J. Gen Micobiol.*, **111**, 201 (1979).

3. J. Adams, T. Kinney, S. Thompson, L. Rubin and R. E. Helling, *Genetics*, **91**, 627 (1979).

4. B. E. Uhlin, K. Nordstrom, *Mol. Gen. Genet.*, **165**, 167 (1978).

5. P. P. Dennis and H. Bremer, *J. Bacteriol.*, **119**, 270 (1974).

6. T. Yamamoto, S. Yamagata, K. Horii and S. Yamagishi, *J. Bacteriol.*, **150**, 269 (1982).

7. T. Sawai, I. Takahashi and S. Yamagishi, *Antimicrob. Agents Chemoth.*, **13**, 269 (1978).

8. F. H. Heffron, *et al*, *Proc. Nat. Acad. Sci.*, (USA), **175**, 6005 (1978).

9. F. C. Kafatos, C. W. Jones and A. Efstratiadis, *Nuc. Acid Res.*, **7**, 1541 (1979).

10. S. B. Lee and J. E. Bailey, *Biotechnol. Bioeng.*, **26**, 1372 (1984).

11. S. B. Lee and J. E. Bailey, *Biotechnol. Bioeng.*, **26**, 66 (1984).

12. B. J. Bachmann and K. B. Low, *Microbiol Rev.*, **44**, 1 (1980).

13. S. Cooper and C. E. Helmstetter, *J. Mol. Biol.*, **31**, 519 (1968).

14. A. L. Goldberg, PNAS (USA), **69**, 422 (1972).

15. P. L. Bergquist and P. Truman, *Molec. Gen. Genet.*, **164**, 105 (1978).

16. P. P. Dennis, V. Nene and R. E. Glass, *J. Bacteriol.*, **161**, 803 (1985).

17. R. E. Kingston and M. J. Chamberlin, *Cell*, **27**, 523 (1981).

Figure Captions

1. Relative cellular protein content, measured for *E. coli* HB101 carrying different copy level plasmids. All cells were grown at a dilution rate equal to 0.50 hr^{-1}.

2. Counts per minute of ^{3}H-labeled RNA, hybridized to β-lactamase DNA which was derived from plasmid RSF1050.

3. Relative β-lactamase enzymatic activity measured on intracellular extracts of HB101 bearing plasmid propagated at different copy number.

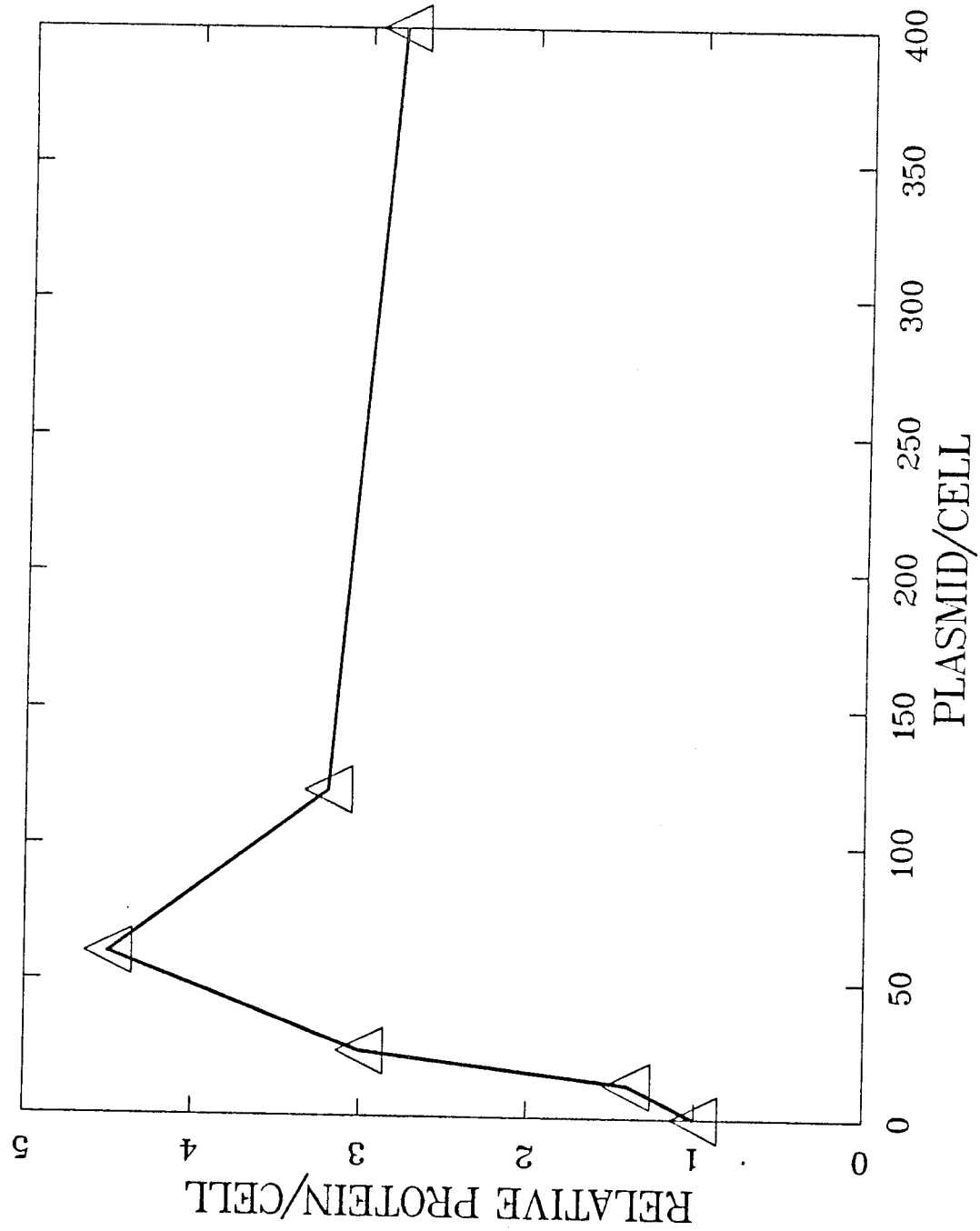

RELATIVE PROTEIN/CELL
0
1
2
3
4
5
0
50
100
150
200
250
300
350
400
PLASMID/CELL

FIGURE 1

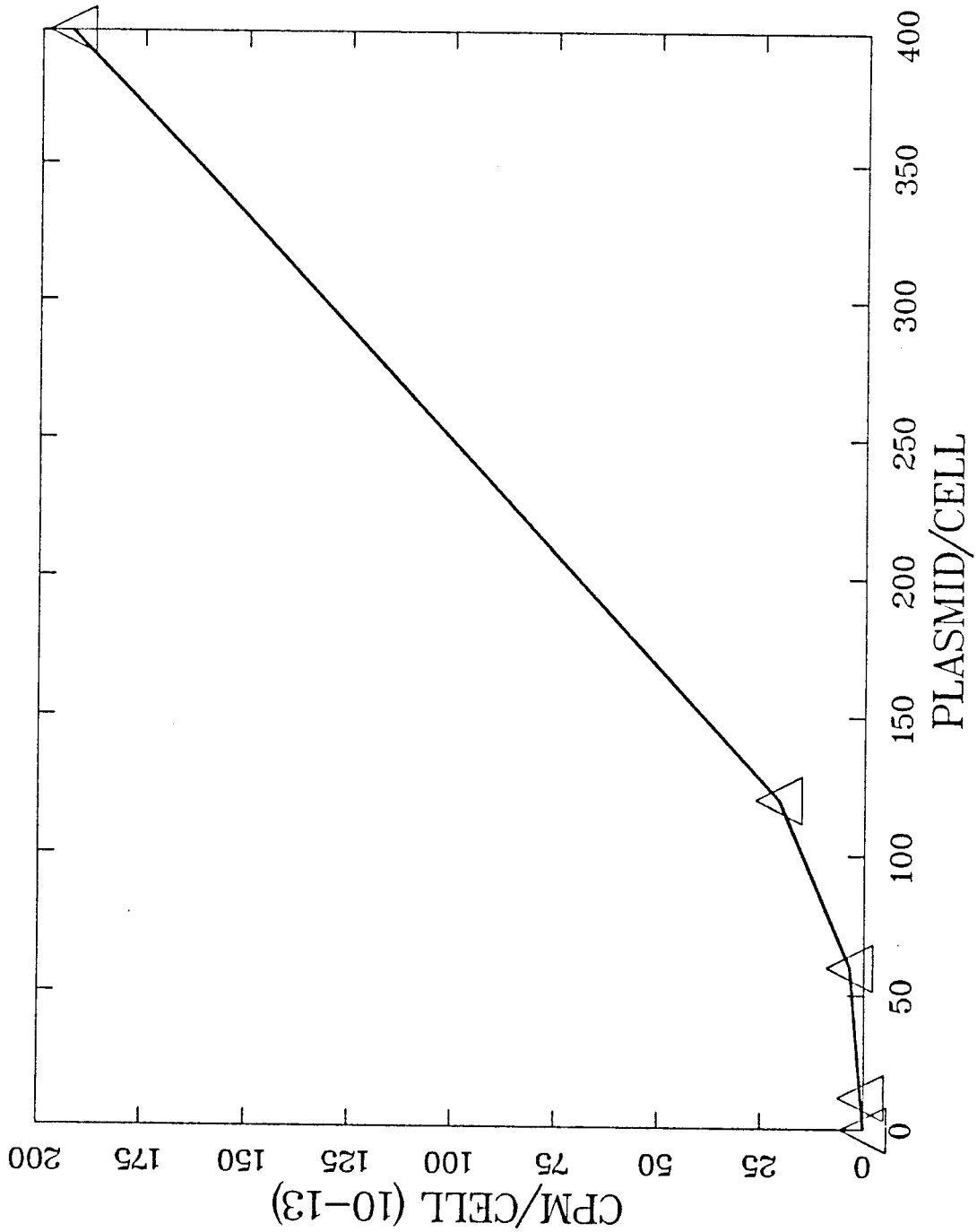

CPM/CELL (10-13)
0
25
50
75
100
125
150
175
200
PLASMID/CELL
0
50
100
150
200
250
300
350
400

FIGURE 2

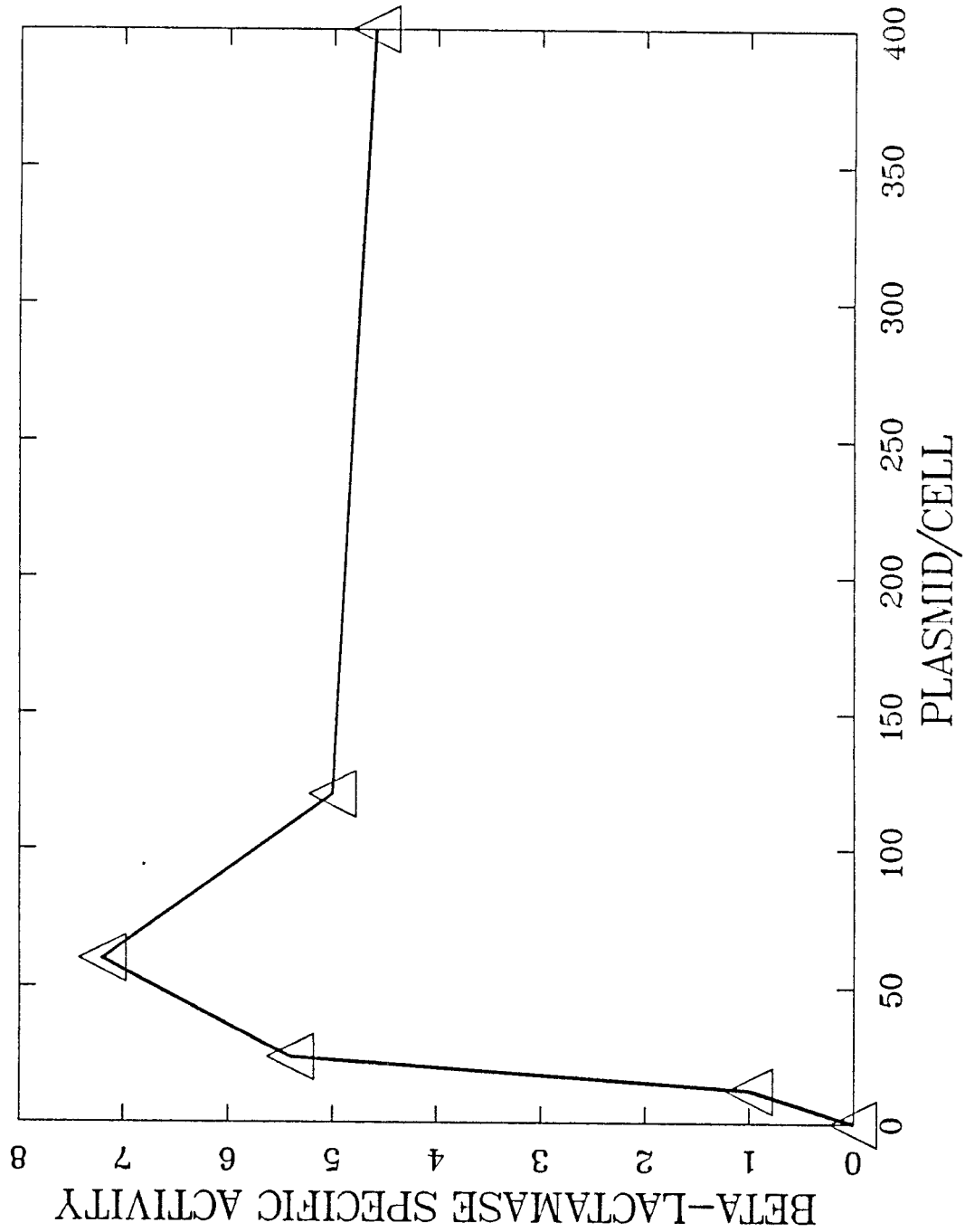

BETA-LACTAMASE SPECIFIC ACTIVITY
0
1
2
3
4
5
6
7
8
PLASMID/CELL
0
50
100
150
200
250
300
350
400

FIGURE 3

APPENDIX

Derivation of the Chemostat Equation

Consider a bounded volume element. A balance performed on the volume gives:

rate of change = influx - efflux + rate of synthesis - rate of degradation

If the volume element is a stirred-tank bioreactor operating in continuous flow mode at steady-state (chemostat) then the rate of change is equal to zero. If the component is biomass concentration (X) and the feed is sterile (influx = 0), then we have the following

$$0 = 0 - \frac{W}{V}X + \mu X$$

where W is the rate of liquid withdrawal (vol/time) and μ is the observed net rate of synthesis. The ratio W/V is called the dilution rate, D. This equation leads to a well established condition of steady-state growth,

$$\mu = D$$

Application of the conservation equation to a macromolecular component of the cell requires more explicit formulations of the synthesis and degradation rates. One can represent the synthesis rate as equal to the product of the amount of template available and the efficiency of synthesis from the template. Degradation is treated as a first order decay. In equation form, the component balance is

$$0 = 0 - D * component_{ss} + E * template_{ss} - K_d * component_{ss}$$

This equation can be solved for the steady-state component level. With the substitution μ=D made, the following results

$$Component_{ss} = \frac{E * template_{ss}}{K_d + \mu}$$

which is the equation used for analyzing the experimental data.

Conclusion

An attempt was made to establish the mathematical framework necessary for the detailed representation and analysis of recombinant system mechanisms and kinetics. Insofar as a formulation consisting of 28 variables and 110 parameters can approximate the essential details of a system containing one to two orders of magnitude greater variables, this was accomplished. As indicated by comparison of simulation results with experimental finds at different levels for plasmid-free and plasmid-containing systems, the model provides a good first-order approximation of the actual cell.

The experimental results point to shortcomings in the model structure and illustrate the necessity of coupling theoretical analysis with experimental studies. There are many important fundamental lessons to be taken from the work presented here, particularly with regard to the conceptual task of mathematically representing highly complex reaction networks. As indicated by the progression from Monod's work through Michael Shuler's models to the formulations described herein, the complexity of a model is dependent upon the type of output desired as well as the reaction mechanisms involved. Simulation of increasingly detailed behavior, such as macromolecular synthesis rates, and RNA polymerase equilibrium distribution requires higher order modeling complexity than does simulation of growth rate dependence upon limiting substrate concentration or product inhibition effect on growth.

Increasing the complexity of behavior studied, or imposing a perturbation to the system also indicates flaws in the formulation which are not displayed when simulating less complex or unperturbed systems. An example in this work is RNA polymerase synthesis control. These experiments indicate that feedback control of the rpoBC operon is very important in the response of cell metabolism to plasmid gene expression. However, it was unnecessary to account for this behavior when

simulating plasmid-free steady-state (Chapter 1) or transient (Chapter 3) growth. Even simulation of recombinant cell growth under steady-state conditions (Chapter 2) gave qualitatively correct behavior for some aspects of metabolism without accounting for feedback control of this operon. Therefore, perturbations to a system can cause new control aspects of the system to be displayed, but the response may be restricted to certain levels of analysis.

An obvious implication of this work is that there is a tremendous amount of experimental work yet to be done in these systems. Many individual aspects of cell metabolism have been analyzed, both as part of a larger system and in isolation. This work suggests that it is insufficient to treat the cell as a collection of individual reactions or pathways. It is only through the integration of subcellular processes into a coherent description of cellular behavior that the rational design of new productive recombinant systems is possible.

APPENDIX 1

THE SINGLE-CELL METABOLIC MODEL

FOR PLASMID-FREE SYSTEMS

The first program presented is the overall controller for the single- cell model without plasmid. This program contains the integration routine and calculates some growth-related characteristics of the system.

```
1  c  CALTECH MODEL FOR ECOLI PLASMID-FREE CELL
2  C
```

The common block and following external statement list those parameters to be held in common for all subroutines' use and which subroutines are to be accessed. The external subroutine SCM is the collection of mass balance equations that actually make up the representation of cellular metabolism. EQ is the subroutine involved in the calculation of the equilibrium distribution of RNA polymerase between different bound and unbound states.

```
3      COMMON CA1,CA2,NRNA,CNA,CMG,PH,TE,E3,NORI,NTOT,PD,VOL,
4    1 TOTM,DM1S,DM2IS,DM2MS,DM4S,DM5S,DP1S,DP2S,DP2D,DP3S,
5    2 DP4S,DG3,DPGS,DVP,RA1S,RA2S,DMP,DA2A,DA2CP,S,S1,DEP,V,
6    3 ec0,esig,ednac,ednar,ednam,ecdnac,fracm,ribo,ssmin,rnam
7  C
8      EXTERNAL SCM,EQ
9  C
10     IMPLICIT DOUBLE PRECISION (A-H,P-Z)
11     DOUBLE PRECISION NRNA,KCD,KHD,KHP,KMG,KMGP,KP,KE
12     dimension G(24),G0(24),G1(24),G1EST(24),G2(24),G2EST(24)
13     dimension DG0(24),DG(24),DG1(24),DG1EST(24),DG2EST(24)
14     dimension ESIG(5),SSMIN(5),EDNAC(5),ECDNAC(5),EC0(5),W(2)
15     dimension RIBO(5),FRACM(5),RNAM(5),ERR(24),SOS(2,2),A(2)
16     dimension F(5),FPOS(5),V(5),YOLD(5),DELTA(5),POS1(5),POS2(5)
17     dimension C0(24),EDNAM(5),B(29),C(24),L(24),EDNAR(5)
18 C
```

Lines 19-54 are involved in the initialization of the program variables, and the writing of initial conditions to the output file. FORTRAN1 is the data file containing the initial conditions of the system. C0(i) (line 23) is the vector of initial

concentrations on a g/mL basis. RP and ARP are the repressor protein and antirepressor protein levels, respectively, on a gram-mass basis. S is the surface area of the cell in square centimeters. VOL is the cell volume in milliliters. CW and CL are the width and length of the cell, respectively, in centimeters, while DT is the integration time interval in hours. FPOS(i) is the vector of chromosomal fork positions. CA1 and CA2 are the concentrations (g/mL) of ammonia and glucose, respectively, in the medium and are held constant for these calculations. DEV is the criterion for convergence of successive approximations. SL is the septum length (cm), T0 the initial time. SEPF and SEP are the areas of septum formed and the septum necessary to enable cell division. F(i) is the vector which stores the number of forks which are at a certain position. NR is the number of stable RNA operons in the cell. NGEN keeps track of the number of generations calculated. NTOT is the number of active replication origins, while NPOS is the number of different fork positions. NTERM counts the number of replication termini; NORI counts the number of origins , and N is the theoretical doubling time for the growth conditions (min).

FORTRAN2 is a data file containing intracellular parameters, listed on line 27. They are pH (PH), temperature (TE) in degrees centigrade, and the mono- and di-valent ion concentrations (CNA and CMG, respectively). While these parameters are assumed constant, the formulations for equilibrium constants of RNA polymerase contain dependences upon these parameters, so they are included.

```
C       INITIAL CONDITIONS
C
        OPEN(unit=1,file='FORTRAN1')
        open(unit=2,file='FORTRAN2')
        READ (1,*) (C0(I),I=1,24)
        READ (1,*) RP,ARP,S,VOL,CW,CL,DT,FPOS(1)
        READ (1,*) CA1,CA2,DEV,SL,T0,SEPF,SEP,F(1)
        READ (1,*) NR,NGEN,NTOT,NPOS,NTERM,NORI,N
        READ (2,*) PH,TE,CNA,CMG
        NRNA=NR
C
        WRITE(6,1)N
        WRITE (6,407)
        WRITE (6,408) CNA,CMG,PH,TE
        WRITE (6,101)
        WRITE (6,7) RP,ARP,S,VOL,CW,CL,DT,FPOS(1)
        WRITE (6,102)
        WRITE (6,7) CA1,CA2,DEV,SL,T0,SEPF,SEP,F(1)
        WRITE (6,103)
        WRITE (6,11) NR,NTOT,NPOS,NTERM,NORI
C
        NDIM=24
        DO 40 I=1,NDIM
40          G0(I)= C0(I)*VOL
```

E2 is the enzyme that catalyzes cell wall formation, while E3 catalyzes septum formation. Since the initial conditions are assumed to be for newborn cells, E3 activity is set equal to zero.

```
          E2= G0(23)
          E3= 0.0
          E23=E2+E3
          VL= VOL/1.D3
        U=DLOG(1D1)
        H=DEXP(-U*PH)
        KP=DEXP(-U*7.4)
        WS=DLOG(CNA)/U
        ITN2=101
C       END OF INITIALIZATION
C
        YOLD(1)=FPOS(1)
87       NGEN= NGEN+1
        TMAX=0.1
C
```

The integration routine is a two-step Runge-Kutta followed by a predictor-corrector pair. The Runge-Kutta scheme consists of an Euler formula step (69) followed by a step using the trapezoidal rule (76). These integration steps provide information for the predictor-corrector pair to follow and are primarily executed

directly following cell division.

```
C     SINGLE-CELL METABOLIC MODEL WITHOUT PLASMID
C     SECOND-ORDER RUNGE-KUTTA WITH PREDICTOR-CORRECTOR
C
 130            T1= T0+DT
               M=1
              NT= 1
              NT2=0
C     INITIAL ESTIMATE
      CALL SCM(G0,DG0,M)
      IF(T0.EQ.0) TOTM0=TOTM
            DO 140 I = 1,NDIM
 140         g1est(i)=g0(i)+dt*dg0(i)
              DELTA(M)=DT
              POS1(M)=FPOS(1)
              POS2(M)=FPOS(2)
 142            M=2
      CALL SCM(G1EST,DG1EST,M)
       DO 150 I = 1, NDIM
 150   G1(I)= G0(I)+ 5.D-1*DT*(DG0(I)+DG1EST(I))
 160         T2= T1+DT
            NT= NT+1
C
```

The Nystrom midpoint formula is used as the predictor, while the trapezoidal rule is the corrector. The predictor is a two-step method, necessitating the utilization of a one-step method to initiate the integration. The corrector has the same order local truncation error as the predictor. The corrector is iterated until the convergence criteria are met; then the local truncation error of the corrector is used to modify the final result.

```
C     PREDICTOR
 162         M=3
      CALL SCM(G1,DG1,M)
 169   DO 170 I= 1, NDIM
 170    G2EST(I)= G0(I)+2.D0*DG1(I)*DT
        DELTA(M)=DT
        POS1(M)=FPOS(1)
        POS2(M)=FPOS(2)
        ITN=0
C
C     CORRECTOR
 175         M=4
      CALL SCM(G2EST,DG2EST,M)
 179         ITN1=0
     DO 190 I= 1, NDIM
      G2(I)= G1(I)+5.D-1*DT*(DG1(I)+DG2EST(I))
```

Lines 96-105 check the convergence of the system using the criterion DEV. If all components converge, ITN1=0 and the step size may be changed (120-126). If convergence is not achieved, ITN1>0 and the corrector is run with new estimated masses (106). Antirepressor protein synthesis (106) is proportional to cell wall synthesis.

```
96          IF(G0(I).EQ.0.0) GO TO 190
97          IF(G2(I).EQ.0.0) GO TO 180
98          IF(ABS((G2(I)-G2EST(I))/G2(I)).LT.DEV) GOTO 190
99          if(i.ne.4) goto 180
100         if(nt1.lt.3) goto 180
101         if(abs((g2(i)-g2est(i))/g2(i)).lt.1.d1*dev) goto 190
102   180     G2EST(I)= G2(I)
103           itn1=itn1+1
104           L(ITN1)=I
105   190     CONTINUE
106         DARP=5.D-1*(DG2EST(16)+DG1(16))*DT
```

In order to avoid errors due to convergence problems, and to cope with transient situations requiring smaller than average step sizes, a lack of convergence after 10 iterations causes a decrease in step size. Convergence on the first attempt (120-121) prompts a step-size increase. If 200 iterations pass without convergence, the program stops (117).

```
107   C     STEP SIZE ADJUSTMENT
108           IF(ITN1.EQ.0) GOTO 231
109           ITN=ITN+1
110           IF(ITN.LT.10) GOTO 175
111           if(ngen.lt.8) goto 211
112   210           WRITE(6,220)
113               WRITE(6,222) (L(I),I=1,ITN1)

114   C     IF SCM FAILS TO CONVERGE, THE STEP SIZE
            IS DECREASED BY A FACTOR OF 10

115   C     IF SCM CONVERGES IN THE FIRST ITERATION,
            THE STEP SIZE INCREASES 1%
```

```
116   211        NT1=NT1+1
117         IF(NT1.GT.20) GOTO 999
118              DT= 0.1*DT
119              GO TO 130
120   231   IF(ITN.GT.0) GOTO 240
121              DT= 1.01*DT
122              NT2=1
123              GO TO 250
124   240   IF(ITN.LT.4) GOTO 250
125              DT=0.5*DT
126              NT2=1
```

The predictor-corrector pair is used such that the corrector is iterated to convergence. The truncation error for the corrector can be shown to be equal to ERR(i) calculated on line 129. This error is used to modify the final value using the P-C pair.

```
127  C       TRUNCATION ERROR FOR PREDICTOR-CORRECTOR

128   250   DO 260 I= 1, NDIM
129              ERR(I)= 2.D-1*(G2EST(I)-G2(I))
130   260        G(I)=G2(I)+ERR(I)
131         M=5
132         CALL SCM(G,DG,M)
133   269     DELTA(M)=DT
134         POS1(M)=FPOS(1)
135         POS2(M)=FPOS(2)
136         DO 270 I=1,24
137   270      G2(I)=G(I)
```

The accounting program gives a detailed breakdown of energy consumption in the cell during cell doubling.

```
138       CALL ACCOUNT (B,DG,DT,DG3,DP2S,DP2D,DP4S,DP1S,DM1S,DM4S,
139     1   DM5S,DM2IS,DM2MS,DPGS,DVP,RA1S,DMP,S,DP3S,RA2S,DA2A,
140     2   DA2CP,S1,DEP)
141  C      CHECKING FOR NEGATIVE CONCENTRATIONS
142           DO 290 I= 1, NDIM
143   290      IF(G(I).LT.0.0) GOTO 295
144           GOTO 310
145   295   WRITE(6,300)
146         WRITE(6,401) T2,V(5)
147         WRITE (6,7) (G(I),I=1,24)
148         GOTO 999
```

Lines 149-170 describe chromosomal metabolism. After the amount of DNA synthesized over the integration time has been calculated, the replication fork positions are moved (155). If a fork has moved through a position containing the

promoter for a stable RNA operon, the number of sRNA operons is incremented (156-169).

```
149 C      CHROMOSOMAL METABOLISM
150  310          NFINT=0
151              NFTERM=0
152              IF(NPOS.LT.1) GOTO 330
153 C     CALCULATING sRNA GENE DOSAGE
154            DO 320 I= 1, NPOS
155            FPOS(I)= FPOS(I)+ (DT*DG(14)*2.49D14)/NTOT
156            IF(FPOS(I).LE.0.02) GOTO 320
157            IF(YOLD(I).LE.0.02) NRNA=NRNA+2*F(I)
158             IF(FPOS(I).LE.0.04) GOTO 320
159             IF(YOLD(I).LE.0.04) NRNA=NRNA+2*F(I)
160              IF(FPOS(I).LE.0.10) GOTO 320
161              IF(YOLD(I).LE.0.10) NRNA=NRNA+2*F(I)
162             IF(FPOS(I).LE.0.12) GOTO 320
163             IF(YOLD(I).LE.0.12) NRNA=NRNA+2*F(I)
164            IF(FPOS(I).LE.0.20) GOTO 320
165            IF(YOLD(I).LE.0.20) NRNA=NRNA+2*F(I)
166           IF(FPOS(I).LE.0.24) GOTO 320
167           IF(YOLD(I).LE.0.24) NRNA=NRNA+2*F(I)
168            IF(FPOS(I).LE.0.56) GOTO 320
169            IF(YOLD(I).LE.0.56) NRNA=NRNA+2*F(I)
170  320          YOLD(I)=FPOS(I)
```

Chromosome replication is initiated when the level of antirepressor protein equals or exceeds that of the repressor (173). However, if the previously initiated fork has not vacated the origin region (175), initiation is postponed. Antirepressor protein synthesis (172) occurs continuously, while repressor protein synthesis (176) occurs only at the time of replication initiation.

```
171 C      CHROMOSOMAL INITIATION
172  330    ARP=ARP+2.6D-3*DARP*(PD/(PD+5.5d0))
173             IF(ARP.LT.RP) GO TO 360
174             IF(NPOS.LT.1.0) GOTO 340
175             IF(FPOS(NPOS).LT.0.1) GO TO 360
176  340    DRP=5.0D-17*NORI*((dm1s/vol)/(1.5d-1+dm1s/vol))
177           RP=RP+DRP
178             nfint=nori
179             NORI=NORI+NFINT
180             NPOS=NPOS+1
181             F(NPOS)= NFINT
182             FPOS(NPOS)=0.0
183 C
```

When the fork reaches the end of the chromosome (fork position 1), replication is terminated and the number of forks is reduced accordingly.

```
184 C       TERMINATION
185  360          IF(FPOS(1).LT.1) GOTO 380
186               NFTERM=F(1)
187               NTERM= 2*NFTERM
188               NPOS= NPOS-1
189               INDEX= NPOS +1
190          DO 370 I = 1,INDEX
191          F(I)=F(I+1)
192  370       FPOS(I)= FPOS(I+1)
193 C       NUMBER OF FORKS
194  380        NTOT= NTOT+NFINT-NFTERM
195 C
```

Crosswall formation (196-220) involves very small time scales. To avoid stiffness problems, the integration step-size is reduced by 4 orders of magnitude (204). Also, once crosswall formation begins, the enzyme involved (E3) is consumed (200-202). When crosswall synthesis is completed, the time step returns to normal (210) and the program continues.

```
196 C       CROSSWALL FORMATION & ENVELOPE ELONGATION
197 C
198  385       IF(NFTERM.LT.1) GOTO 400
199           E3= G(24)
200           G(24)= 0.0
201           G1(24)= 0.0
202           G2(24)= 0.0
203           DTS=DT
204           DT=1.D-4*DT
205           NT2=1
206           ITN2=0
207  400           E2= G(23)
208              ITN2=ITN2+1
209               IF(ITN2.NE.100) GOTO 410
210               DT=DTS
211               NT2=1
212  410      E23= E2+E3
213           SEP= 3.14*CW**2
214           PRCNT= E2/E23
215           PRCNTS= E3/E23
216           DA=1.2D6*DG(16)*DT/4.2D-1
217           SEPF=SEPF+PRCNTS*DA
218           DCL= (PRCNT*DA)/(3.14*CW)
219           CL= CL+DCL
220           S= 3.14*(CW**2+CW*CL)+SEPF
221 C
```

The cell is assumed to be cylindrical, with hemispherical caps and developing hemispherical crosswalls. The surface area of the cell is known, and given by:

S = 3.1415 * (CW**2 + CW*CL + 2*CW*SL)

SL is given in line 224. The volume of the cell is:

V = 3.1415 * (2*CW**3 + 3*CL*CW**2 + 6*CW**2 + 8*SL)/12

Substituting the expression for SL shown on line 224 into the surface area formula, solving it for CL, then combining the resulting equation with the volume formula yield the following:

V = -.262 * CW**3 + .25 * S * CW - 2.09 * SL**3

This equation, in conjunction with known values for SEPF, S, and the cell volume VOL is used in a Newton-Raphson iterative scheme to determine the cell's length and width (223-224)

```
222 C       DETERMINE CELL GEOMETRY
223             DO 1100 I=1,100
224             SL=SEPF/(6.28*CW)
225             P=-2.62D-1*CW**3+2.5D-1*S*CW-2.09D0*SL**3-VOL
226             DP=-7.86D-1*CW**2+2.5D-1*S+2.53D-2*SEPF**3/CW**4
227             CWN=CW-P/DP
228                 IF(ABS(P).GT.0.001) GO TO 1010
229                 IF(ABS(P/(DP*CWN)).LT.0.001) GOTO 1020
230 1010            CW=CWN
231 1100            CONTINUE
232 1020            CW=CWN
233             SL=SEPF/(6.28*CW)
234             CL=S/(3.14*CW)-CW-2.*SL
235             CN=CW*1.D4
236             CM=1.D4*(CL+2.*SL+CW)
237             CSL=1.D4*SL
238     IF(SEPF-SEP.GT.0.0) GOTO 440
```

If cell division does not occur, then selected variables are printed at a specified interval, which in this case is every six minutes. The estimates for the Runge-Kutta predictor-corrector pairs are incremented, and calculated values for RNA polymerase equilibrium are stored. These will be used to limit the parameter space searched during the next equilibrium calculation. If convergence was achieved without a change in the time increment for integration, then the program returns to the predictor-corrector pair to calculate results for the next time increment (276). If convergence was not optimal, then the program returns to line 130, and

the Runge-Kutta integration is performed prior to the predictor-corrector (275).

```
239 C      CELL DOES NOT DIVIDE
240        IF(NGEN.LT.7) GOTO 420
241  409   IF(T0.EQ.0.0) GOTO 415
242        IF(T0.LT.TMAX) GOTO 420
243        TMAX=TMAX+0.10
244        GOTO 417
245  415      WRITE(6,90)NGEN
246           WRITE(6,301)
247  417   TOTR=EDNAM(1)+EDNAR(1)
248        NTIME=T0*60
249        write(6,302) ntime,g0(7)/g0(14),g0(9)/g0(14),g0(10)/g0(14),
250     1  ednar(1)/totr,TOTR/EC0(1),g0(14),rp,arp
251        nt=0
252  420      DO 430 I= 1, NDIM
253           G0(I)=G1(I)
254  430      G1(I)=G2(I)
255        T0=T1
256        T1=T2

257           DO 435 I=1,2
258           J=2*I-1
259           K=J+2
260           RIBO(J)=RIBO(K)
261           ESIG(J)=ESIG(K)
262           SSMIN(J)=SSMIN(K)
263           EDNAM(J)=EDNAM(K)
264           EDNAR(J)=EDNAR(K)
265           EDNAC(J)=EDNAC(K)
266           ECDNAC(J)=ECDNAC(K)
267           FRACM(J)=FRACM(K)
268           RNAM(J)=RNAM(K)
269           V(J)=V(K)
270           POS1(J)=POS1(K)
271           POS2(J)=POS2(K)
272           DELTA(J)=DELTA(K)
273           EC0(J)=EC0(K)
274  435      CONTINUE
275        IF(NT2.GT.0.0) GOTO 130
276        GOTO 160
277 C
```

Cell division (278-315) triggers a full print out of the cell components, equilibrium results and an accounting of cell energetics. The energetic account (290) describes the utilization of glucose and reducing equivalents by various cell processes. Cell division is assumed completely symmetric (292- 313), and the energetic accounting variables are set to zero for the start of the next generation (314-315).

```
278 C      CELL DIVISION IS SYMMETRIC
279  440   IF(NGEN.LT.7) GOTO 445
280  441     WRITE(6,501)
281          write(6,401) t2,v(5)
282          WRITE(6,7) (G(I), I=1,24)
283          WRITE(6,402)
284          WRITE(6,5) ESIG(5),EC0(5),EDNAM(5),EDNAR(5),SSMIN(5)
285          WRITE(6,403)
286          WRITE(6,5) EDNAC(5),ECDNAC(5),RIBO(5),FRACM(5),RNAM(5)
287        TOTR=EDNAM(5)+EDNAR(5)
288          WRITE(6,404)
289          WRITE(6,6) S,CL,CW,POS1(5),EDNAR(5)/TOTR,TOTR/EC0(5)
290        CALL ACCTPRT(B,TOTM,TOTM0)
291          WRITE(6,405)

292  445           DO 450 I= 1, NDIM
293               G0(I)=0.5*G(I)
294  450           CONTINUE
295        IF(NPOS.LT.1) GOTO 470
296               DO 460 I= 1, NPOS
297               F(I)=0.5*F(I)
298  460           CONTINUE
299  470      NTOT=NTOT/2
300          SEPF=0.0
301          s=s/2.
302          VOL=VOL/2.
303          NORI=NORI/2
304          NTERM= NTERM/2
305          NRNA=NRNA/2
306          CL= CL/2.
307             ITN2=101
308          E2= E2/2.
309          E3=0
310          SL=0.0
311          NT1=1
312          RP=RP/2.
313          ARP=ARP/2.
314        do 490 i=1,29
315  490       b(i)=0.
316       IF(NGEN.GT.8) GOTO 999
317  500   T0=0.0
318       goto 89
319 C
```

```
320 C
321 C
322   1  FORMAT(/,5X,'INITIAL CONDITIONS FOR SCM:ESTIMATED
323     1 DOUBLING TIME=',I4,'MINUTES',/)
324   5  FORMAT(5(1PE10.3))
325   6  FORMAT(6(1PE10.3))
326   7  FORMAT(8(1PE10.3))
327   8  FORMAT(3X,'DT=',E10.3)
328  11  FORMAT(5(4X,I2))
329  90  FORMAT(/,20X,'GENERATION#',I2)
330 101  FORMAT(4X,'RP',7X,'ARP',8X,'S',9X,'V',9X,'CW',8X,'CL',
331     1 8X,'DT',8X,'FP1')
332 102  FORMAT(4X,'CA1',7X,'CA2',7X,'DEV',7X,'SL',8X,'T',9X,
333     1 'SEPF',6X,'SEP',7X,'F1')
334 103  FORMAT(' NRNA',4X,'NTOT',2X,'NPOS',2X,'NTERM',1X,'ORI')
335 220  FORMAT(//,15X,'SCM DOES NOT CONVERGE')
336 222  FORMAT(5X,I3)
337 300  FORMAT(//,5X,'NEGATIVE CONCENTRATIONS')
338 301  FORMAT(2X,'T',3X,'PR/D',5X,'sR/D',5X,'mR/D',5X,'RNAPR',
339     1 4X,'ACTPOL',4X,'DNAC',5X,'RP',7X,'ARP')
340 302  FORMAT(I3,1X,8(1PE9.2))
341 401  FORMAT(/,15X,'TIME=',1pE10.3,5X,'VOLUME='1pE10.3,/)
342 402  FORMAT(3X,'ESIG',6X,'EC0'6X,'EDNAM',5X,'EDNAR',5X,'SSMIN')
343 403  format(3X,'EDNAC',4X,'ECDNAC',5X,'RIBO',5X,'FRACM',6X,'RNAM')
344 404  format(3x,'S',9X,'CL',8X,'CW',7X,'FP1',6X,'FRACR',5X,'ACTPOL')
345 405  format(/,10x,40('#')//)
346 406  FORMAT(5X,'M=',I2)
347 407  FORMAT(5X,'[Na+]',6x,'[Mg++]',5x,'pH',6x,'Temp')
348 408  FORMAT(2X,4(1PE10.3))
349 501  FORMAT(/,7x,20('*'),2x,'CELL DIVISION',2X,20('*'))
350 C
351 999  END
```

This subroutine keeps track of glucose, ATP and reducing equivalents usage by the different cell processes. B(1) sums the total amount of glucose taken up by the cell from birth to cell division.

```
355       SUBROUTINE ACCOUNT(B,DG,DT,DG3,DP2S,DP2D,DP4S,DP1S,
356     1 DM1S,DM4S,DM5S,DM2IS,DM2MS,DPGS,DVP,RA1S,DMP,S,DP3S,
357     2 RA2S,DA2A,DA2CP,S1,DEP)
358       IMPLICIT DOUBLE PRECISION (A-H,P-Z)
359       DIMENSION B(29),DG(24)

360 c     This is for detailed glucose and atp consumption tracking
361 c
```

Anabolic uses of glucose (364-367) include the net synthesis of amino acids (B2), ribonucleotides (B3), cell wall precursors (b(4), and glycogen (B5).

```
362         b(1)=b(1)+dg(2)*dt
363 c       Calculations from DA2A
364         b(2)=b(2)+dg3*dt*1.128d0
365         b(3)=b(3)+(dp2s-dp2d)*dt*4.56d-1
366         b(4)=b(4)+dp4s*dt*1.26d0
367         b(5)=b(5)+dg(17)*dt*1.11d0
```

Catabolic consumption of ATP includes the processes of glucose transport into the cell (B6), the synthesis of : amino acids (B7), ribonucleotides (B8), cell-wall precursor (B9), protein (B10), DNA (B11), cell wall (B12), glycogen (B13), immature sRNA (B14), mRNA (B15), ppGpp (B16), and unspecified cell-related processes (B29).

```
368 c       Calculations from DA2C
369         b(6)=b(6)+dg(2)*dt*5.6d-3
370         b(7)=b(7)+dp1s*dt*2.5d-3
371         b(8)=b(8)+dp2s*dt*2.2d-2
372         b(9)=b(9)+dp4s*dt*1.6d-3
373         b(10)=b(10)+dm1s*dt*3.9d-2
374         b(11)=b(11)+dg(14)*dt*7.1d-3
375         b(12)=b(12)+dm4s*dt*8.1d-3
376         b(13)=b(13)+dm5s*dt*1.24d-2
377         b(14)=b(14)+dm2is*dt*6.7d-3
378         b(15)=b(15)+dm2ms*dt*6.7d-3
379         b(16)=b(16)+dpgs*dt*3.d0
380         b(29)=b(29)+dvp*dt*6.33d-3
```

Reductant usage occurs during ammonia transport (B17), unspecified ion transport (B18), cell membrane charging (B19), biosynthetic reductions (B20), sulfate assimilation (B21), and ribonucleotide reduction to deoxyribonucleotide (B22).

```
381 c       Calculations from DEP
382         b(17)=b(17)+ra1s*dt*2.8d-2
383         b(18)=b(18)+dmp*dt*2.d-3
384         b(19)=b(19)+s*dt*5.8d-8
385         b(20)=b(20)+dmp*dt*1.39d-2
386         b(21)=b(21)+dmp*dt*1.21d-3
387         b(22)=b(22)+dp3s*dt*3.1d-3
```

The net uptake of glucose (B23) and ammonia (B28) is measured. The net consumption of glucose for anabolic processes (B24) and catabolic processes (B25) is also tallied, along with total ATP usage (B26) and reductant use (B27).

```
c       Overall calculations
        b(23)=b(23)+ra2s*dt
        b(24)=b(24)+da2a*dt
        b(25)=b(25)+da2cp*dt
        b(26)=b(26)+(s1+dg(2)*5.6d-3)*dt
        b(27)=b(27)+dep*dt
        b(28)=b(28)+ra1s*dt
      RETURN
      END
C
      SUBROUTINE ACCTPRT(B,TOTM,TOTM0)
      IMPLICIT DOUBLE PRECISION (A-H,P-Z)
      DIMENSION B(29)
      dmass=totm-totm0
      write(6,600)
      write(6,601) b(23),b(28),dmass
      write(6,602) b(24),b(25)
      write(6,603) b(26),b(27)
      write(6,604) dmass/b(23),dmass/b(26)
      write(6,605)
      write(6,607)
      write(6,606) b(1)
      write(6,608) b(2),b(3)
      write(6,609) b(4),b(5)
      write(6,610)
      write(6,611) b(6),b(7),b(8),b(9)
      write(6,612) b(10),b(11),b(12),b(13)
      write(6,613) b(14),b(15),b(16),b(29)
      write(6,614)
      write(6,615) b(17),b(18)
      write(6,616) b(19),b(20)
      write(6,617) b(21),b(22)
      RETURN
600   format(/,10x,'Macro-scale results for glucose and ATP',/)
601   format(5x,'GLUCOSE IN=',1PE10.3,5X,'NH4 IN=',1PE10.3,
     1 5X,'MASS PRODUCED=',1PE10.3)
602   format(5X,'ANABOL GLU =',1PE10.3,4X,'CATABOL GLU =',1PE10.3)
603   format(5x,'ATP USAGE=',1PE10.3,6X,'2H+2E USAGE=',1PE10.3)
604   format(5x,'GLUCOSE YIELD=',1PE10.3,2X,'ATP YIELD=',1PE10.3)
605   format(/,10x,'Details of glucose metabolism',/)
606   format(15x,'GLUCOSE TURNOVER=',1pE10.3)
607   format(5x,'GLUCOSE DISTRIBUTION',/)
608   format(5x,'AMINO ACIDS=',1E10.3,8X,'RIBONUC=',1E10.3)
609   format(5x,'CELL WALL PRE=',1E10.3,6X,'GLYCOGEN=',1E10.3)
610   format(10x,'ATP metabolism',/)
611   format(5x,'A2 PTS=',1E10.3,5X,'P1=',1E10.3,5X,'P2=',1E10.3,5X,
     1 'P4=',1E10.3)
612   format(5x,'M1=',1E10.3,9X,'M3=',1E10.3,5X,'M4=',1E10.3,5X,
     1 'M5=',1E10.3)
613   format(5x,'M2I=',1E10.3,8X,'M2M=',1E10.3,4X,'CPG=',1E10.3,4X,
     1 'MYS=',1E10.3)
614   format(/,15x,'Reductant breakdown',/)
615   format(5x,'NH4 TRANS =',1E10.3,10X,'ION TRANS =',1E10.3)
616   format(5x,'MEMBRANE =',1E10.3,11X,'BIOS =',1E10.3)
```

```
442 617  format(5x,'SULFUR =',1E10.3,13X,'P2 -> P3 =',1E10.3)
443      END
```

This subroutine contains the formulation for cell metabolism, consisting of mass balance terms for twenty-four cellular components. The preceding program calls this subroutine, passing to it the mass vector (G), and receiving the instaneous rate of change vector (DG).

```
1        subroutine SCM(g,dg,m)
2   c    SINGLE-CELL MODEL WITHOUT PLASMID
3   C
4        COMMON CA1,CA2,NRNA,CNA,CMG,PH,TE,E3,NORI,NTOT,PD,VOL,
5       1 TOTM,DM1S,DM2IS,DM2MS,DM4S,DM5S,DP1S,DP2S,DP2D,DP3S,DP4S,
6       2 DG3,DPGS,DVP,RA1S,RA2S,DMP,DA2A,DA2CP,S,S1,DEP,V,EC0,
7       3 ESIG,EDNAC,EDNAR,EDNAM,ECDNAC,FRACM,RIBO,SSMIN,RNAM
8   c
9        external eq
10  C
11       IMPLICIT DOUBLE PRECISION (A-H,P-Z)
12       DOUBLE PRECISION NRNA,KCD,KHD,KHP,KMG,KMGP,KP,KE
13       DIMENSION G(24),C(24),DG(24),ESIG(5),SSMIN(5),EDNAC(5)
14       DIMENSION ECDNAC(5),EC0(5),EDNAR(5),EDNAM(5)
15       DIMENSION RIBO(5),FRACM(5),RNAM(5),SOS(2,2),V(5)
16  C
17         U=DLOG(1.0D1)
18         H=DEXP(-U*PH)
19         KP=DEXP(-U*7.4)
20         WS=DLOG(CNA)/U
21  C
```

The cell volume (22-30) is calculated as the sum of two subcellular volumes. The cell envelope volume is determined assuming constant density (26). Likewise, the cytoplasmic mass is assumed to have constant density, and the two volumes are totaled.

```
22  C     *******   CALCULATE NEW VOLUME    *******
23                  TOTM=0
24                  DO 2000 I=1,18
25   2000             TOTM=TOTM+G(I)
26              VE=G(16)/5.526D-1
27              VC=(TOTM-G(16))/2.584D-1
28              VOL=VE+VC
29              VL=VOL/1.D3
30              V(M)=VOL
31  C
```

The mass concentrations (g/mL) for the cell components are calculated (34-35), and molar concentrations of species involved in RNA polymerase equilibrium calculations are determined (36-43). DNAC0 represents total chromosomal base pairs; DNAM represents mRNA promoters; EC0 is the total RNA polymerase core enzyme concentration; SIG0 is the sigma subunit concentration; DNAR is the sRNA promoter concentration, and DNACP represents non-promoter sites on the chromosome. These parameters are needed because each equilibrium is calculated using the polymerase and binding sites existant at that time.

```
32  C     *******    INITIAL CONCENTRATIONS    *******
33  C
34              DO 2100 I=1,24
35  2100          C(I)=G(I)/VOL
36                DNAC0=G(14)/(6.18D2*VL)
37                DNAM=1.9D-4*DNAC0
38                EC0(M)=G(20)/(4.05D5*VL)
39                SIG0=G(21)/(9.2D4*VL)
40                  CORE0=EC0(M)
41                  DNAR1=NRNA/(6.02D13*VL)
42                  DNAR=DNAR1/1.D10
43                  DNACP=DNAC0-(DNAM+DNAR)
44  C

45  C     ******** SATURATION AND INHIBITION TERMS ********
46  C
47              RIC1=2.3D-3/(C(1)+2.3D-3)
48              RC2=C(2)/(C(2)+1.8D-4)
49              RA2=C(2)/(C(2)+7.2D-5)
50              RIA2=2.d-3/(c(2)+2.d-3)
51              RA2P1=C(2)/(C(2)+2.5D-4)
52              RIA2M2R=1.1D-5/(C(2)+1.1D-5)
53              RP1=C(3)/(C(3)+2.2D-4)
54              RC3=C(3)/(C(3)+9.9D-4)
55              RP2=C(4)/(C(4)+7.2D-5)
56              RIP2=9.D-3/(C(4)+9.D-3)
57              RP3M3=C(5)/(C(5)+4.D-6)
58              RP4=5.D-3/(C(6)+5.D-3)
59        rca1x=(ca1/(ca1+1.08d-7)-c(1)/(c(1)+1.d-5))
60        rca1y=4.61*(ca1/(ca1+3.6d-6))*2.3d-3/(c(1)+2.3d-3)
61              RCA2=CA2/(CA2+9.3D-6)
62  C
```

Protein synthesis rate (70) affects the equilibrium distribution of RNA polymerase by altering binding constants (75-80, 92-97). The ribosomal distribution onto mRNA is calculated as the fraction of total ribosomes which are active (67).

RNAM (65) is the molar concentration of mRNA molecules, assuming an average message size, and RIBO (66) is the number of ribosomes present based on the amount of rRNA found per ribosome.

```
C     ******* CALCULATE EQUILIBRIUM CONCENTRATIONS *******
C           RIBOSOMAL EQUILIBRIUM
                RNAM(M)=G(10)/(3.9D5*VL)
                RIBO(M)=3.19D17*G(9)
                FRACM(M)=8.75d5*RNAM(M)/(8.75d5*RNAM(M)+1.)
                FRACI=1.-FRACM(M)
C      *****  END OF RIBOSOMAL EQUILIBRIUM *****

      DM1S=7.1D0*FRACM(M)*8.5D-1*G(9)*RP1*RA2
      DM1D1=C(3)/(C(3)+1.03D-7)*C(19)/(C(19)+3.9D-8)*
     1 1.404D-5/(1.404D-5+C(2))+(C(2)/(C(2)+2.73D-3))
      DM1D=DM1D1*G(7)*7.D-2+DM1S*5D-2
      DG(7)=DM1S-DM1D
           SPMAX=2.1D0
           PR=SPMAX*C(7)-(DM1S/VOL)
           PRI=1.1d-2/(1.1d-2+PR**3)
           PGI=2.2-7/(C(19)+2.2-7)
           PRMI=4.3D-1/(4.3D-1+(DM1S/VOL))
           PRRI=PGI*PRI
```

Equilibrium constants are calculated for core binding (KCD), holoenzyme binding to non-promoter sites (KHD) and holoenzyme binding to promoter sites (KHP). Additionally, the dependence of mRNA promoter binding on translation initiation factors is considered (92) as is the dependence of the two RNA promoters on translation initiation and ppGpp level (95-96).

```
C      Calculate equilibrium constants as fcns of ionic strength
        KMG=DEXP(U*(.5-1.1*WS))
        DMG=2.*DLOG10(0.5*(1.+DSQRT(1.+4.*KMG*CMG)))
        KHD=DEXP(U*(-2.5-10.8*WS-DMG-0.4*(PH-7.8)))
        KCD=DEXP(U*(-8.5-21.2*WS-0.5*DMG-0.3*(PH-7.8)))
        DPH=2.*DLOG10(1.+KP/H)
        KMGP=DEXP(U*(0.32+1.75*WS))
        DMGP=2.*DLOG10(0.5*(1.+DSQRT(1.+4.*KMGP*CMG)))
       KHP=DEXP(U*(-1.3-14.7*WS-DMGP-DPH)*(0.48+0.52*TE/310.16))
          khd=5.d-1*khd
          KE=1.D10
          C1P=1.4*PRMI*KHP
          C2P=PRRI*KHP
          P1=C1P*DNAM
            P21=2.7D1*C2P*DNAR/2.0d0
            p22=7.56d0*pri*khp*dnar/2.0d0
          p2=p21+p22
```

As mentioned earlier, the calculation of RNA polymerase equilibrium is based on a bounded two-dimensional search to find the minimum for the non-linear equations. Definition of the boundaries (107-115) uses available prior equilibrium results whenever possible. Two calls are made to the search routine (116-117, 132-133) using different bounds to minimize the possibility of spurious results arising from local minima in the solution space.

```
98  C       CALCULATE EQUILIBRIUM DISTRIBUTION OF RNA
99  C       POLYMERASE USING 2-D FIBONACCI SEARCH
100             if(m.eq.5) mp=3
101             if(m.eq.4) mp=3
102             if(m.eq.3) mp=1
103             if(m.eq.2) mp=1
104             if(m.gt.1) goto 111
105             if(t0.eq.0) goto 12
106             mp=5
107     111  ex1u=1.2d0*esig(mp)
108          ex1l=8.d-1*esig(mp)
109          ex2u=1.2d0*ecdnac(mp)
110          ex2l=8.d-1*ecdnac(mp)
111          goto 13
112      12          EX1U=EC0(M)/P1
113                  EX2U=EC0(M)
114                  EX1L=0.d0
115                  EX2L=0.d0
116      13   call eq(ex1u,ex2u,ex1l,ex2l,A,W,SOS,khd,ke,kcd,dnacp,dnac0,
117       1  core0,sig0,p1,p2,imin,jmin)
118                     ESIG(M)=EX1L+(EX1U-EX1L)/2.D0
119                     ECDNAC(M)=EX2L+(EX2U-EX2L)/2.D0
120                     SSMIN(M)=SOS(IMIN,JMIN)

121          if(t0.gt.0) goto 14
122          IF(m.GT.1) GOTO 14
123          EX1U=8.D-1*EC0(M)/P1
124          EX1L=0
125          EX2U=8.D-1*EC0(M)
126          EX2L=0
127          GOTO 15
128      14   ex1u=1.4d0*esig(mp)
129          ex1l=6.d-1*esig(mp)
130          ex2u=1.4d0*ecdnac(mp)
131          ex2l=6.d-1*ecdnac(mp)
132      15   call eq(ex1u,ex2u,ex1l,ex2l,A,W,SOS,khd,ke,kcd,dnacp,dnac0,
133       1  core0,sig0,p1,p2,imin,jmin)
134          if(ssmin(m).lt.sos(imin,jmin)) goto 2150
135                  esig(m)=ex1l+(ex1u-ex1l)/2.d0
136                  ecdnac(m)=ex2l+(ex2u-ex2l)/2.d0
137                  ssmin(m)=sos(imin,jmin)
```

The equilibrium results are expressed as the molar concentration of unbound holoenzyme (ESIG), the molar concentration of core polymerase-DNA complexes (ECDNAC) and the sum of squares for the final state (SSMIN). These values are used to calculate the concentrations of active mRNA-holoenzyme and rRNA-holoenzyme complexes (EDNAM and EDNAR, respectively), as well as nonspecific DNA-holoenzyme complexes (EDNAC).

```
138   C     EQUILIBRIUM ACHIEVED
139    2150          EDNAM(M)=ESIG(M)*P1
140                  EDNAR(M)=ESIG(M)*P2
141        EDNAC(M)=ESIG(M)*KHD*(DNACP-4.D1*(ECDNAC(M)+ESIG(M)*
142       1 (P1+P2)))/(1.D0+4.D1*KHD*ESIG(M))

143   C     CONTINUE MACROMOLECULAR METABOLISM

  Glycogen
144          DM5S=2.0D-2*VOL*(C(2)/(C(2)+2D-3))
145          DM5D=1.4D-1*G(17)*(1.D-3/(C(2)+1.D-3))
146          DG(17)=DM5S-DM5D

  Cell Wall
147              E2=G(23)
148              E23=E2+E3
149        DM4S=6.73D1*E23*(C(2)/(C(2)+3D-5))*(C(6)/(C(6)+5.D-4))
150              DM4D=2.3D-1*RC2*G(16)
151              DG(16)=DM4S-DM4D

  Enzymes E1, E2, E3
152              PD=G(7)/G(14)
153              DG(22)=1.3D-4*PD*NORI*DG(7)
154              DG(23)=3.1D-3*DG(7)
155              DG(24)=1.6D-2*G(7)

  DNA
156        DG(14)=1.58D0*6.2D-15*NTOT*RP3M3*(C(2)/(C(2)+1.7D-4))

  ppGpp
157              RI=FRACI*8.5D-1*G(9)
158              DPGS=6.4D-3*RI*(1.265D-4/(C(3)+1.265D-4))
159              DPGD=1.25D2*RA2*G(19)
160              DG(19)=DPGS-DPGD
```

De novo RNA synthesis (164, 167) is based on a chain elongation rate (6.78d7), the molar concentrations of active transcription complexes (EDNAR, EDNAM), the volume (VL), and the transcription efficiency.

```
  Stable RNA
161                 DM2RS=1.4D1*G(8)
162                 DM2RD=7.D-2*RIA2M2R*G(9)
163                 DG(9)=DM2RS-DM2RD

  Immature Stable RNA
164           DM2IS=9.6d-1*6.78D7*RP2*EDNAR(M)*VL
165           DM2ID=2.1D1*(1.8D-2/(1.8D-2+DM1S/VOL))*G(8)
166           DG(8)=DM2IS-DM2ID-DM2RS

  mRNA
167         DM2MS=1.07d0*1.175d1/(1.15d1+pd)*6.78D7*RP2*EDNAM(M)*VL
168               DM2MD=2.1D1*G(10)
169               DG(10)=DM2MS-DM2MD
170             DM2=DG(10)+DG(9)+DG(8)
```

RNA polymerase core and sigma subunit syntheses are represented as a constant fraction of total protein synthesis.

```
171               DECS=1.15D-2*DM1S
172               DECD=DM1D*G(20)/G(7)
173               DG(20)=DECS-DECD
174               DG(21)=7.D-4*DM1S-DM1D*G(21)/G(7)
175   C

176   C       PRECURSOR METABOLISM

  Cell Wall Precursor
177               DP4S1=6.0D-2*RP4*VOL
178               DP4S=DP4S1*RC3*(C(2)/(C(2)+1.2D-4))
179               DG(6)=DP4S-1.1D0*DG(16)

  Deoxyribonucleotides
180           AMP=4.66D-1*C(4)
181           DAMP=3.64D-1*C(5)
182           DP3S=55.*G(22)*(AMP/(AMP+5.2D-6))/(1.+DAMP/1.45D-5)
183           DP3D=37*G(5)*2
184           IF(DG(14).GT.0.0) DP3D=0
185           DG(5)=DP3S-1.053D0*DG(14)-DP3D

  Ribonucleotides
186               DP2S=2.1D-1*VOL*RC3*RIP2*(C(2)/(C(2)+2.5D-4))
187               DP2D=3.D-2*RIA2M2R*G(4)
188               DG(4)=DP2S-DP2D-1.057D0*DM2-1.049D0*DP3S
```

```
   Amino Acids
189    DP1S=3.9D-1*VOL*(5.D-2/(C(3)+5.D-2))*(C(1)/(C(1)+2.5D-5))*RA2P1
190      DP1D=2.5D-2*RIA2M2R*G(3)
191      DG3=DP1S-DP1D
192      DG(3)=DG3-1.167D0*DG(7)-1.149D0*(DP2S-DP2D)-1.28D-1*DP4S

   Ammonia Assimilation
193              ra1s=2.06d-7*(rca1x+rca1y)*ra2*s
194              DG(1)=RA1S-1.79D-1*DG3

   Total Cell Mass
195              dmt=0
196              do 2200 i=1,18
197    2200          dmt=dmt+dg(i)
198              dv=dmt/2.9d-1
199              dmp=dmt-dg(2)
200              dvp=dmp/2.9d-1
201    c

202    c       GLUCOSE METABOLISM
203    C
204           S1A=2.5D-3*DP1S+2.D-3*DP2S+1.6D-3*DP4S+1.24D-2*DM5S
205           S1B=6.7D-3*(DM2IS+DM2MS)+4.4D-3*DG(14)+4.1D-3*DM4S
206           S1=S1A+S1B+6.33D-3*DVP+3.9D-2*DM1S+3.D0*DPGS
207           DA2A=1.128d0*dg3-4.56d-1*(dp2s-dp2d)+1.26d0*dp4s+1.11d0*dg(17)
208            RA2S=1.0d-5*RCA2*RIA2*S
209            Z=9.D-1*(DM4S/(DM4S+1.2D-14))
210            DEP=2.8D-2*RA1S+1.711D-2*DMP+5.8D-8*S+3.1D-3*DP3S
211            DA2CP=1.80D2*abs( (S1+1.5D0*DEP)/(2.2D1-6.D0*Z) )
212            GAMMA=1.80D2*(5.6D-3+6.33D-3/2.9D-1+1.5D0*1.711D-1)/
213       1      (2.2D1-6.D0*Z)
214           DG(2)=(RA2S-abs(DA2A+DA2CP))/(1.D0+GAMMA)
215    C        END OF METABOLISM
216           RETURN
217     999    end
```

This subroutine uses a two-dimensional Fibonacci search for minima of the sum of squares for the stated equation. A full derivation of the equations used in the calculation of the sum of squares (19-31) can be found in Appendix 2.

```
1         subroutine eq(ex1u,ex2u,ex1l,ex2l,a,w,sos,khd,ke,kcd,dnacp,
2        1  dnac0,core0,sig0,p1,p2,imin,jmin)
3   c
4         implicit double precision (a-h, p-z)
5         double precision kcd, khd, ke
6         dimension ch(2), st(2), sos(2,2)
7   c
8         r=2.0d0/(1.0d0+sqrt(5.))
```

EXNU are upper bounds for free holoenzyme molar concentration (N=1) and the molar concentration of core enzyme-DNA complexes (N=2). Likewise, EXNL are lower bounds.

```
9           dif=ex1u-ex1l
10            dwf=ex2u-ex2l
11  c
12         ch(1)=EX1U-R*DIF
13         ch(2)=ex1l+R*dif
14         st(1)=EX2U-R*DWF
15         st(2)=ex2l+r*dwf
16            IFLAG=0
17            ITNO=0
18            Q=1D38
19  C      SUM OF SQUARES CALCULATION
20   3100   DO 3300 I=1,2
21         DEN=1.D0+4.0D1*ch(I)*KHD
22         DO 3200 J=1,2
23  C        PSI=EDNAC
24         PSI=KHD*ch(I)*(DNACP-4.D1*(ch(I)*(P1+P2)+st(J)))/DEN
25         EC=CORE0-(ch(I)*(P1+P2)+PSI+st(J)+ch(I))
26         DNAC=DNAC0-4.D1*(ch(I)*(P1+P2)+PSI+st(J))
27         sig=sig0-psi-ch(i)
28            FA=(KE*EC*SIG-ch(I))/ch(I)
29            GW=(KCD*EC*DNAC-st(J))/st(J)
30            SOS(I,J)=FA**2+GW**2
31   3200   CONTINUE
32   3300   CONTINUE
33  c
```

```
34  C      FIND THE LOCAL MINIMUM SUM OF SQUARES
35  C
36           DO 3600 I=1,2
37           DO 3500 J=1,2
38              IF (SOS(I,J).LE.Q) THEN
39              IMIN=I
40              JMIN=J
41              Q=SOS(I,J)
42            IF(SOS(I,J).LT.1d0)  IFLAG=1
43            IF((ch(2)-ch(1))/ch(2).LT.1D-20) IFLAG=1
44           END IF
45   3500     CONTINUE
46   3600     CONTINUE
47  c
48  c      ADJUST BOUNDARIES OF REGION TO BE SEARCHED
49  C
50            IF(IFLAG.EQ.1) goto 999
51            ITNO=ITNO+1
52            IF(IMIN.EQ.1) THEN
53                EX1U=ch(2)
54                ch(2)=ch(1)
55                ch(1)=EX1U-R*(EX1U-EX1L)
56            ELSE
57                EX1L=ch(1)
58                ch(1)=ch(2)
59                ch(2)=EX1L+R*(EX1U-EX1L)
60            END IF
61            IF(JMIN.EQ.1) THEN
62                EX2U=st(2)
63                st(2)=st(1)
64                st(1)=EX2U-R*(EX2U-EX2L)
65            ELSE
66                EX2L=st(1)
67                st(1)=st(2)
68                st(2)=EX2L+R*(EX2U-EX2L)
69            END IF
70          GOTO 3100
71  C
72    999  END
```

APPENDIX 2

DERIVATION OF THE EQUATIONS FOR RNA POLYMERASE EQUILIBRIUM

Calculation of the equilibrium distribution of RNA polymerase onto various DNA sites is a central feature of this model, so this appendix will address the assumptions made in the derivation of the relevant equations. It was assumed that RNA polymerase participates in the following equilibrium reactions:

						Constants	
Ec	+	o	<=>	Eo		KSIG	(1)
Ec	+	DNAC	<=>	ECDNAC		KCD	(2)
Eo	+	DNACP	<=>	EDNAC		KHD	(3)
Eo	+	DNAM	<=>	EDNAM		KHPm	(4)
Eo	+	DNAR	<=>	EDNAR		KHPr	(5)

where Ec is core polymerase; o is the sigma subunit; Eo is the holoenzyme; DNAC are the total chromosomal base pairs; DNAM are the mRNA promoters; DNAR are stable RNA promoters, and DNACP = DNAC - (DNAM + DNAR). This set of equations ignores two relatively insignificant reactions involving core polymerase:

Ec	+	DNAM	<=>	ECDNAM	KCD	(6)
Ec	+	DNAR	<=>	ECDNAR	KCD	(7)

These reactions are competing with (4) and (5). KHPm and KHPr are two to three orders of magnitude greater than KCD, and free Ec is much less abundant than Eo. The five equations considered can be further simplified. The rate-limiting step in transcription is presumed to be the formation of the polymerase-promoter complex. Chain elongation occurs so rapidly that it is safe to assume that the concentration of promoter sites is essentially constant and equal to the total promoter concentration. In addition, the catalytic activity of the sigma subunit justifies the assumption that its concentration is likewise constant. The following set of equa-

tions results:

$$Eo = KSIG * o * (EC0 - Eo - ECDNAC - EDNAC - EDNAM - EDNAR) \quad (1a)$$

$$ECDNAC = KCD * (EC0 - Eo - ECNDAC - EDNAC - EDNAM - EDNAR) * (DNAC0 - 40 * [ECDNAC + EDNAM + EDNAR + EDNAC]) \quad (2a)$$

$$EDNAC = KHD * Eo * (DNACP0 - 40 * [ednac + ednam + ednar + ecdnac]) \quad (3a)$$

$$EDNAM = KHPm * DNAM0 * Eo = C1 * Eo \quad (4a)$$

$$EDNAR = KHPr * DNAR0 * Eo = C2 * Eo \quad (5a)$$

Inserting equations (4a) and (5a) into (3a) and solving for EDNAC gives:

$$EDNAC = KHD * Eo * \frac{(DNAC0 - 40 * ECDNAC - 40 * Eo * [C1 + C2])}{1 + 40 * KHD * Eo} \quad (3b)$$

Equation (3b) contains two variables, Eo and ECDNAC. Combining Equations (1a), (2a), (3b), (4a) and (5a) and solving for Eo and ECDNAC result in two equations with two unknowns. These Equations (1b and 2b) form the basis for the equilibrium calculations performed in subroutine EQ, listed in Appendix 1.

$$Eo = KSIG * o * \{ EC0 - ECDNAC - Eo * [1 + C1 + C2] - KHD * Eo * \frac{(DNAC0 - 40 * ECDNAC - 40 * Eo * [C1 + C2])}{1 + 40 * KHD * Eo} \} \quad (1b)$$

$$ECDNAC = KCD * \{ EC0 - ECDNAC - Eo * [1 + C1 + C2] - KHD * Eo * \frac{(DNAC0 - 40 * ECDNAC - 40 * Eo * [C1 + C2])}{1 + 40 * KHD * Eo} \} * \{DNAC0 - 40 * (ECDNAC - Eo * [1 + C1 + C2] - KHD * Eo * \frac{ DNAC0 - 40 * ECDNAC - 40 * Eo * [C1 + C2]\})}{1 + 40 * KHD * Eo} \} \quad (2b)$$

APPENDIX 3

SINGLE-CELL MODEL FOR PLASMID-CONTAINING CELLS

This appendix contains a listing of the programs used to simulate plasmid-containing cells. Since the structure of these programs is nearly identical to those for the plasmid-free case, only the changes made from the plasmid-free formulation will be discussed. For further explanation, see Appendix 1.

There are new variables to be passed to the metabolic subroutine, related to the three operons found on the plasmid. These operons are designated I, II, and B, since they direct the synthesis of RNAI, RNAII and β-lactamase mRNA. TXEs are transcription efficiencies, EQs are relative binding constants, and STMs are relative RNA stabilities. BTL is the relative ribosome binding efficiency of the β-lactamase mRNA, and BDEG is the stability of β-lactamase enzyme relative to cellular protein.

PRIMER is the number of plasmid origins actively being replicated and can have values of either zero or one.

```
1   C     CALTECH MODEL FOR ECOLI CONTAINING PLASMID DNA
2   C
3         COMMON CA1,CA2,NRNA,CNA,CMG,PH,TE,E3,NORI,NTOT,PD,VOL,
4        1 TOTM,DM1S,DM2IS,DM2MS,DM4S,DM5S,DP1S,DP2S,DP2D,DP3S,
5        2 DP4S,DG3,DPGS,DVP,RA1S,RA2S,DMP,DA2A,DA2CP,S,S1,DEP,
6        3 V,EC0,ESIG,EDNAC,EDNAR,EDNAM,ECDNAC,FRACM,RIBO,SSMIN,
7        4 RNAM,NGEN,DT,DPL,COPY,TXEI,TXEII,TXEB,BTX,BTL,BDEG,EQI,
8        5 EQII,EQB,STMI,STMII,STMB,PRIMER,DBLS,DM21S,DM22S,DM2BS,
9        6 ednai,ednaii,ednab,fracb,ednap,ecdnap,rpl
10  C
11        EXTERNAL SCMP,EQP
12  c
13        IMPLICIT DOUBLE PRECISION (A-H,K,P-Z)
14        double precision NRNA
15        dimension G(24),G0(24),G1(24),G1EST(24),G2(24),G2EST(24)
16        dimension DG0(24),DG(24),DG1(24),DG1EST(24),DG2EST(24),L(24)
17        dimension ESIG(5),SSMIN(5),EDNAC(5),ECDNAC(5),EC0(5)
18        dimension RIBO(5),FRACM(5),RNAM(5),ERR(24),SOS(2,2),A(2)
19        dimension F(5),FPOS(5),V(5),YOLD(5),DELTA(5),POS1(5),POS2(5)
20        dimension C0(24),EDNAM(5),B(34),c(24),EDNAR(5),W(2)
```

PLPOS is the plasmid replication fork position; COPY is the number of plasmids per cell; FRACB is the fraction of ribosomes translating β-lactamase

mRNA, and BLAC is the molar concentration of β-lactamase mRNA.

```
      dimension PLPOS(5),EDNAI(5),EDNAII(5),EDNAB(5),FRACB(5)
      dimension EDNAP(5),ECDNAP(5),BLAC(5),COPY(5)
C
C     INITIAL CONDITIONS
C
      OPEN(unit=1,file='FORTRAN1')
      open(unit=2,file='FORTRAN2')
      OPEN(UNIT=3,FILE='FORTRAN3')
      READ (1,*) (C0(I),I=1,24)
      READ (1,*) RP,ARP,S,VOL,CW,CL,DT,FPOS(1)
      READ (1,*) CA1,CA2,DEV,SL,T0,SEPF,SEP,F(1)
      READ (1,*) NR,NGEN,NTOT,NPOS,NTERM,NORI,N
      READ (2,*) PH,TE,CNA,CMG
      READ (3,*) PLT,BTX,BTL,BDEG,EQI,EQII,EQB
      READ (3,*) TXEI,TXEII,TXEB,STMI,STMII,STMB
         NRNA=NR
```

COP is the number of copies of plasmid in the cell.

```
         COP=C0(15)*VOL*6.02D23/5.D6
C
      WRITE(6,1)N
      WRITE (6,407)
      WRITE (6,408) CNA,CMG,PH,TE
        WRITE (6,101)
        WRITE (6,7) RP,ARP,S,VOL,CW,CL,DT,FPOS(1)
        WRITE (6,102)
        WRITE (6,7) CA1,CA2,DEV,SL,T0,SEPF,SEP,F(1)
        WRITE (6,103)
        WRITE (6,11) NR,NTOT,NPOS,NTERM,NORI
      WRITE (6,4) COP,PLT
      WRITE (6,513)
      WRITE (6,6) BTX,BTL,BDEG,EQI,EQII,EQB
      WRITE (6,1957)
      WRITE (6,6) TXEI,TXEII,TXEB,STMI,STMII,STMB
      WRITE (6,405)
C
```

PLF is the frequency of plasmid replication initiation in hours. The system is started with the initiation of plasmid replication. TIN represents the next time for plasmid replication initiation and is incremented accordingly.

```
55        PLF=PLT/6.D1
56        PRIMER=1.
57        TIN=PLF
58        NDIM=24
59   C
60              DO 40 I=1,NDIM
61     40        G0(I)= C0(I)*VOL
62                E2= G0(23)
63                E3= 0.0
64                E23=E2+E3
65                VL= VOL/1.D3
66                ITN2=101
67                YOLD(1)=FPOS(1)
68   C       END OF INITIALIZATION
69   C
70     87          NGEN= NGEN+1
71                TMAX=0.1
72   C
73   C          SINGLE-CELL METABOLIC MODEL WITH PLASMID
74   C       SECOND-ORDER RUNGE-KUTTA WITH PREDICTOR-CORRECTOR
75   C
76    130           T1= T0+DT
77                  M=1
78                  NT= 1
79                  NT2=0
```

PLP is the plasmid fork position and dpl is the change in fork position for the time interval under consideration. Plasmid replication may be initiated at this time. (83-86).

```
80   C       INITIAL ESTIMATE
81             plp=plp+dpl
82             if(plp.ge.1) primer=0
83             IF (T0.LT.TIN) GOTO 139
84               PRIMER=1
85               TIN=TIN+PLF
86               PLp=0
87    139    CALL SCMP(G0,DG0,M)
88          IF(T0.EQ.0) TOTM0=TOTM
89               DO 140 I = 1,NDIM
90    140          g1est(i)=g0(i)+dt*dg0(i)
91                 DELTA(M)=DT
92                 POS1(M)=FPOS(1)
93                 POS2(M)=FPOS(2)
94    142            M=2
95          CALL SCMP(G1EST,DG1EST,M)
96            DO 150 I = 1, NDIM
97    150      G1(I)= G0(I)+ 5.D-1*DT*(DG0(I)+DG1EST(I))
98    160            T2= T1+DT
99                   NT= NT+1
100  C
```

```
101  C     PREDICTOR
102   162        M=3
103            plp=plp+dpl
104            if(plp.ge.1) primer=0
105            IF (T0.LT.TIN) GOTO 169
106              PRIMER=1
107              TIN=TIN+PLF
108              PLp=0
109   169    CALL SCMP(G1,DG1,M)
110              DO 170 I= 1, NDIM
111   170          G2EST(I)= G0(I)+2.D0*DG1(I)*DT
112              DELTA(M)=DT
113              POS1(M)=FPOS(1)
114              POS2(M)=FPOS(2)
115              ITN=0
116  C
117  C     CORRECTOR
118   175          M=4
119         CALL SCMP(G2EST,DG2EST,M)
120   179        ITN1=0
121              DO 190 I= 1, NDIM
122              G2(I)= G1(I)+5.D-1*DT*(DG1(I)+DG2EST(I))
123              IF(G0(I).EQ.0.0) GO TO 190
124                IF(G2(I).EQ.0.0) GO TO 180
125         IF(ABS((G2(I)-G2EST(I))/G2(I)).LT.DEV) GOTO 190
126         if(i.ne.4) goto 180
127         if(nt1.lt.3) goto 180
128         if(abs((g2(i)-g2est(i))/g2(i)).lt.1.d1*dev) goto 190
129   180           G2EST(I)= G2(I)
130               itn1=itn1+1
131               L(ITN1)=I
132   190           CONTINUE
133               DARP=5.D-1*(DG2EST(16)+DG1(16))*DT
134  C
135  C     STEP SIZE ADJUSTMENT
136  C
137               IF(ITN1.EQ.0) GOTO 231
138               ITN=ITN+1
139               IF(ITN.LT.10) GOTO 175
140               if(ngen.lt.1) goto 211
141    210              WRITE(6,220)
142                   WRITE(6,222) (L(I),I=1,ITN1)
143  C
144  C     IF SCM FAILS TO CONVERGE, THE STEP SIZE IS
                DECREASED BY A FACTOR OF 10
145  C     IF SCM CONVERGES IN THE FIRST ITERATION, THE
                STEP SIZE INCREASES 1%
146  C
```

```
147   211              NT1=NT1+1
148                  IF(NT1.GT.20) GOTO 999
149                    DT= 0.1*DT
150                    GO TO 130
151    231           IF(ITN.GT.0) GOTO 240
152                    DT= 1.01*DT
153                    NT2=1
154                    GO TO 250
155    240           IF(ITN.LT.4) GOTO 250
156                    DT=0.5*DT
157                    NT2=1
158   C
159   C       TRUNCATION ERROR FOR PREDICTOR-CORRECTOR
160   C
161    250          DO 260 I= 1, NDIM
162                ERR(I)= 2.D-1*(G2EST(I)-G2(I))
163    260          G(I)=G2(I)+ERR(I)
164                    M=5
165         CALL SCMP(G,DG,M)
166    269          DELTA(M)=DT
167               POS1(M)=FPOS(1)
168               POS2(M)=FPOS(2)
169                  DO 270 I=1,24
170    270             G2(I)=G(I)
171   C
172                 call accountp(B,DG,DT,DG3,DP2S,DP2D,DP4S,DP1S,DM1S,
173        1  DM4S,DM5S,DM2IS,DM2MS,DPGS,DVP,RA1S,DMP,S,DP3S,
174        2  RA2S,DA2A,DA2CP,S1,DEP,DBLS,DM21S,DM22S,DM2BS)
175   C
176   C      CHECKING FOR NEGATIVE CONCENTRATIONS
177               DO 290 I= 1, NDIM
178    290          IF(G(I).LT.0.0) GOTO 295
179               GOTO 310
180    295              WRITE(6,300)
181                   WRITE(6,401) T2,V(5)
182                   WRITE (6,7) (G(I),I=1,24)
183                   GOTO 999

184   C       CHROMOSOMAL METABOLISM
185    310              NFINT=0
186                    NFTERM=0
187                    IF(NPOS.LT.1) GOTO 330
188   C       CALCULATING sRNA GENE DOSAGE
189              DO 320 I= 1, NPOS
190              FPOS(I)= FPOS(I)+ (DT*DG(14)*2.49D14)/NTOT
191              IF(FPOS(I).LE.0.02) GOTO 320
192              IF(YOLD(I).LE.0.02) NRNA=NRNA+2*F(I)
193               IF(FPOS(I).LE.0.04) GOTO 320
194               IF(YOLD(I).LE.0.04) NRNA=NRNA+2*F(I)
195                IF(FPOS(I).LE.0.10) GOTO 320
196                IF(YOLD(I).LE.0.10) NRNA=NRNA+2*F(I)
197               IF(FPOS(I).LE.0.12) GOTO 320
198               IF(YOLD(I).LE.0.12) NRNA=NRNA+2*F(I)
199              IF(FPOS(I).LE.0.20) GOTO 320
```

```
200            IF(YOLD(I).LE.0.20) NRNA=NRNA+2*F(I)
201           IF(FPOS(I).LE.0.24) GOTO 320
202           IF(YOLD(I).LE.0.24) NRNA=NRNA+2*F(I)
203            IF(FPOS(I).LE.0.56) GOTO 320
204            IF(YOLD(I).LE.0.56) NRNA=NRNA+2*F(I)
205    320        YOLD(I)=FPOS(I)
```

Antirepressor protein synthesis has been modified to account for the dilution effect that plasmid protein volume has upon normal cellular protein concentration (rpl in line 208).

```
206    C       CHROMOSOMAL INITIATION
207     330     c3=g(3)/vol
208             arp=arp+2.6d-3*darp*pd/(pd+5.5d0)*rpl
209               IF(ARP.LT.RP) GO TO 360
210               IF(NPOS.LT.1.0) GOTO 340
211               IF(FPOS(NPOS).LT.0.1) GO TO 360
212     340    DPRS=DM1S+DBLS
213          DRP=5.0D-17*NORI*((Dm1S/vol)/(1.5d-1+Dm1S/vol))
214          RP=RP+DRP
215               nfint=nori
216               TINIT=T2*6.D1
217               NORI=NORI+NFINT
218               NPOS=NPOS+1
219               F(NPOS)= NFINT
220               FPOS(NPOS)=0.0
221    C
222    C       TERMINATION
223     360          IF(FPOS(1).LT.1) GOTO 380
224                    TTERM=T2*6.D1
225                    NFTERM=F(1)
226                    NTERM= 2*NFTERM
227                    NPOS= NPOS-1
228                    INDEX= NPOS +1
229               DO 370 I = 1,INDEX
230               F(I)=F(I+1)
231     370       FPOS(I)= FPOS(I+1)
232    C       NUMBER OF FORKS
233     380       NTOT= NTOT+NFINT-NFTERM
234    C
```

```
C       CROSSWALL FORMATION & ENVELOPE ELONGATION
C
 385        IF(NFTERM.LT.1) GOTO 400
            E3= G(24)
            G(24)= 0.0
            G1(24)= 0.0
            G2(24)= 0.0
            DTS=DT
            DT=1.D-4*DT
            NT2=1
            ITN2=0
 400              E2= G(23)
                 ITN2=ITN2+1
                 IF(ITN2.NE.100) GOTO 410
                 DT=DTS
                 NT2=1
 410        E23= E2+E3
          SEP= 3.14*CW**2
          PRCNT= E2/E23
          PRCNTS= E3/E23
          DA=1.2D6*DG(16)*DT/4.2D-1
          SEPF=SEPF+PRCNTS*DA
          DCL= (PRCNT*DA)/(3.14*CW)
          CL= CL+DCL
          S= 3.14*(CW**2+CW*CL)+SEPF
C
C       DETERMINE CELL GEOMETRY
       DO 1100 I=1,100
       SL=SEPF/(6.28*CW)
       P=-2.62D-1*CW**3+2.5D-1*S*CW-2.09D0*SL**3-VOL
       DP=-7.86D-1*CW**2+2.5D-1*S+2.53D-2*SEPF**3/CW**4
       CWN=CW-P/DP
       IF(ABS(P).GT.0.001) GO TO 1010
       IF(ABS(P/(DP*CWN)).LT.0.001) GOTO 1020
1010           CW=CWN
1100           CONTINUE
1020           CW=CWN
             SL=SEPF/(6.28*CW)
             CL=S/(3.14*CW)-CW-2.*SL
             CN=CW*1.D4
             CM=1.D4*(CL+2.*SL+CW)
             CSL=1.D4*SL
       IF(SEPF-SEP.GT.0.0) GOTO 440
```

```
278   C      CELL DOES NOT DIVIDE
279          IF(NGEN.LT.15) GOTO 420
280    409   IF(T0.EQ.0.0) GOTO 415
281          IF(T0.LT.TMAX) GOTO 420
282          TMAX=TMAX+0.10
283          GOTO 417
284    415       WRITE(6,90)NGEN
285             WRITE(6,301)
286    417   TOTR=EDNAM(1)+EDNAR(1)+EDNAB(1)+EDNAI(1)+EDNAII(1)
287          TOTP=EDNAI(1)+EDNAII(1)+EDNAB(1)
288          NTIME=T0*60
289           nplasmid=int(copy(1))
290          write(6,302) ntime,fracm(1),fracb(1),ribo(1),
291        1 ednar(1)/totr.totp/totr,totr/ec0(1),g0(14),nplasmid
292          nt=0
293    420      DO 430 I= 1, NDIM
294             G0(I)=G1(I)
295    430      G1(I)=G2(I)
296          T0=T1
297          T1=T2
298             DO 435 I=1,2
299             J=2*I-1
300             L=J+2
301             RIBO(J)=RIBO(L)
302             ESIG(J)=ESIG(L)
303             SSMIN(J)=SSMIN(L)
304             EDNAM(J)=EDNAM(L)
305             EDNAR(J)=EDNAR(L)
306             EDNAC(J)=EDNAC(L)
307             ECDNAC(J)=ECDNAC(L)
308             FRACM(J)=FRACM(L)
309             RNAM(J)=RNAM(L)
310             V(J)=V(L)
311             POS1(J)=POS1(L)
312             POS2(J)=POS2(L)
313             DELTA(J)=DELTA(L)
314             EC0(J)=EC0(L)
315                fracb(j)=fracb(L)
316                ednab(j)=ednab(L)
317                ednai(j)=ednai(L)
318                ednaii(j)=ednaii(L)
319                ednap(j)=ednap(L)
320                ecdnap(j)=ecdnap(L)
321                copy(j)=copy(L)
322    435      CONTINUE
323          IF(NT2.GT.0.0) GOTO 130
324          GOTO 160
325   C
```

```
C      CELL DIVISION IS SYMMETRIC
 440   IF(NGEN.LT.10) GOTO 445
 441     WRITE(6,501)
        write(6,401) t2,v(5)
        WRITE (6,2) TINIT, TTERM
        WRITE(6,7) (G(I), I=1,24)
        WRITE(6,402)
      WRITE(6,7) ESIG(5),EC0(5),EDNAM(5),EDNAR(5),ednac(5),
     1 ecdnac(5),ribo(5),fracm(5)
        WRITE(6,403)
        TOTR=EDNAM(5)+EDNAR(5)+EDNAB(5)+EDNAI(5)+EDNAII(5)
      WRITE(6,7) copy(5),ednai(5),ednaii(5),ednab(5),ednap(5),
     1 ecdnap(5),ednab(5)/totr,fracb(5)
        WRITE(6,404)
      write (6,7) s,cl,cw,pos1(5),delta(5),ssmin(5),
     1  ednar(5)/totr,totr/ec0(5)
      CALL ACCTPRT(B,TOTM,TOTM0)
        WRITE(6,405)
 445          DO 450 I= 1, NDIM
            G0(I)=0.5*G(I)
 450          CONTINUE
      IF(NPOS.LT.1) GOTO 470
           DO 460 I= 1, NPOS
           F(I)=0.5*F(I)
 460          CONTINUE
 470    NTOT=NTOT/2
        SEPF=0.0
        s=s/2.
        VOL=VOL/2.
        NORI=NORI/2
        NTERM= NTERM/2
        NRNA=NRNA/2
        CL= CL/2.
          ITN2=101
        E2= E2/2.
        E3=0
        SL=0.0
        NT1=1
        RP=RP/2.
        ARP=ARP/2.
          tin=tin-t2
      do 490 i=1,29
 490    b(i)=0.
      IF(NGEN.GT.29) GOTO 999
 500  T0=0.0
      goto 87
```

```
375     1  FORMAT(/,5X,'INITIAL CONDITIONS FOR SCM:ESTIMATED
376     1  DOUBLING TIME =',I4,'MINUTES',/)
377     2  format (5('! '),4X,'Chromosome Initiation Time =',1pe10.2,' min',
378     1  4X,5('! '),/,5('! '),4X,'Chromosome Termination Time =',1pe9.2,
379     2  ' min',4x,5('! '),/)
380     3  format (4(1pe9.2))
381     4  format (3x,'#PLAS =',1pe9.2,5x,'PL REP FREQ =',1pe9.2,'MIN',/)
382     5  FORMAT(5(1PE9.2))
383     6  FORMAT(6(1PE9.2))
384     7  FORMAT(8(1PE9.2))
385     8  FORMAT(3X,'DT=',1pE9.2)
386    11  FORMAT(5(4X,I2),/)
387    90  FORMAT(/,20X,'GENERATION#',I2)
388   101  FORMAT(3X,'RP',7X,'ARP',6X,'S',8X,'V'.8X,'CW',7X,'CL',
389     1  7X,'DT',7X,'FP1')
390   102  FORMAT(3X,'CA1',6X,'CA2',6X,'DEV'.6X,'SL'.7X,'T',8X,
391     1  'SEPF',5X,'SEP',6X,'F1')
392   103  FORMAT(' NRNA',4X,'NTOT',2X,'NPOS',2X,'NTERM',1X,'ORI')
393   220  FORMAT(//,15X,'SCM DOES NOT CONVERGE')
394   222  FORMAT(5X,I3)
395   300  FORMAT(//,5X,'NEGATIVE CONCENTRATIONS')
396   301  FORMAT(2X,'T',3X,'FRACM',4X,'FRACB',4X,'RIBO',5X,'FPOLr',
397     1  4x,'FPOLp',4X,'ACTPOL',3X,'DNAC',5X,'COPY#')
398   302  FORMAT(I3,1X,7(1PE9.2),I5)
399   401  FORMAT(/,15X,'TIME=',1pE10.3,5X,'VOLUME='1pE10.3,/)
400   402  FORMAT(3X,'ESIG',5X,'EC0'6X,'EDNAM',4X,'EDNAR',4X,
401     1  'EDNAC',4x,'ECDNAC',3x,'RIBO',5X,'FRACM')
402   403  FORMAT(3X,'COPY#',4X,'EDNAI',4X,'EDNAII',3X,'EDNAB',
403     1  4x,'EDNAP',x4X,'ECDNAP',3X,'FRACBM',3X,'FRACB')
404   404  format(3x,'S',8X,'CL',7X,'CW',7X,'FP1',6X,'DELTA',4X,
405     1  'SSMIN',4x,'FRACR',4X,'ACTPOL')
406   405  format(/,10x,40('#')//)
407   406  FORMAT(5X,'M=',I2)
408   407  FORMAT(5X,'[Na+]',6x,'[Mg++]',5x,'pH',6x,'Temp')
409   408  FORMAT(2X,4(1PE10.3))
410   501  FORMAT(/,7x,20('*'),2x,'CELL DIVISION',2X,20('*'))
411   513  format(3x,'BTX',6x,'BTL',6x,'BDEG',5x,'EQI',6x,'EQII',5x,'EQB')
412  1957  FORMAT (3X,'TXEI',5X,'TXEII',4X,'TXEB',5X,'STMI',5X,
413     1  'STMII',4X,'STMB')
414  C
415   999  END
```

```
4         subroutine accountp (B,DG,DT,DG3,DP2S,DP2D,DP4S,DP1S,
5        1 DM1S,DM4S,DM5S,DM2IS,DM2MS,DPGS,DVP,RA1S,DMP,S,DP3S,
6        2 RA2S,DA2A,DA2CP,S1,DEP,DBLS,DM21S,DM22S,DM2BS)
7         IMPLICIT DOUBLE PRECISION (A-H,P-Z)
8         DIMENSION B(34),DG(24)
9    c     This is for detailed glucose and atp consumption tracking
10   c
11          b(1)=b(1)+dg(2)*dt
12   c     Calculations from DA2A
13          b(2)=b(2)+dg3*dt*1.128d0
14          b(3)=b(3)+(dp2s-dp2d)*dt*4.56d-1
15          b(4)=b(4)+dp4s*dt*1.26d0
16          b(5)=b(5)+dg(17)*dt*1.11d0
17   c     Calculations from DA2C
18          b(6)=b(6)+dg(2)*dt*5.6d-3
19          b(7)=b(7)+dp1s*dt*2.5d-3
20          b(8)=b(8)+dp2s*dt*2.2d-2
21          b(9)=b(9)+dp4s*dt*1.6d-3
22          b(10)=b(10)+dm1s*dt*3.9d-2
```

Included in catabolic glucose use are the syntheses of β-lactamase (B30), plasmid DNA (B31), RNAI (B32), RNAII (B33) and β-lactamase mRNA (B34).

```
23          B(30)=B(30)+DBLS*DT*3.9D-2
24          b(11)=b(11)+dg(14)*dt*7.1d-3
25          B(31)=B(31)+DG(15)*DT*7.1D-3
26          b(12)=b(12)+dm4s*dt*8.1d-3
27          b(13)=b(13)+dm5s*dt*1.24d-2
28          b(14)=b(14)+dm2is*dt*6.7d-3
29          b(15)=b(15)+dm2ms*dt*6.7d-3
30          B(32)=B(32)+dm21s*dt*6.7d-3
31          B(33)=B(33)+dm22s*dt*6.7d-3
32          B(34)=B(34)+dm2Bs*dt*6.7d-3
33          b(16)=b(16)+dpgs*dt*3.d0
34          b(29)=b(29)+dvp*dt*6.33d-3
35   c     Calculations from DEP
36          b(17)=b(17)+ra1s*dt*2.8d-2
37          b(18)=b(18)+dmp*dt*2.d-3
38          b(19)=b(19)+s*dt*5.8d-8
39          b(20)=b(20)+dmp*dt*1.39d-2
40          b(21)=b(21)+dmp*dt*1.21d-3
41          b(22)=b(22)+dp3s*dt*3.1d-3
```

```
c       Overall calculations
        b(23)=b(23)+ra2s*dt
        b(24)=b(24)+da2a*dt
        b(25)=b(25)+da2cp*dt
        b(26)=b(26)+(s1+dg(2)*5.6d-3)*dt
        b(27)=b(27)+dep*dt
        b(28)=b(28)+ra1s*dt
      RETURN
      END

C
      SUBROUTINE ACCTPRT(B,TOTM,TOTM0)
      IMPLICIT DOUBLE PRECISION (A-H,P-Z)
      DIMENSION B(34)
      dmass=totm-totm0
      write(6,600)
      write(6,601) b(23),b(28),dmass
      write(6,602) b(24),b(25),DMASS/B(23)
      write(6,603) b(26),b(27),DMASS/B(26)
       write(6,605)
       write(6,606) b(1)
       write(6,608) b(2),b(3),B(4),B(5)
       write(6,610)
       write(6,611) b(6),b(7),b(8),b(9)
       write(6,612) b(10),b(11),b(12),b(13)
       write(6,613) b(14),b(15),b(16),b(29)
       write(6,614)
       write(6,615) b(17),b(18),B(21)
       write(6,616) b(19),b(20),B(22)
       WRITE (6,619) B(32),B(33),B(34)
       WRITE (6,618) B(30),B(31)
       RETURN
  600  format(/,10x,'Macro-scale results for glucose and ATP',/)
  601  format(3x,'GLUCOSE IN=',1PE10.3,5X,'NH4 IN=',1PE10.3,
     1 5X,'MASS PRODUCED=',1PE10.3)
  602  format(3x,'ANABOL GLU =',1pe10.3,4x,'CAT GLU =',1pe10.3,
     1 3X,'GLUCOSE YIELD=',1PE10.3)
  603  format(3x,'ATP USAGE=',1PE10.3,6X,'2H+2E USE=',1PE10.3,
     1 2X,'ATP YIELD=',1PE10.3)
  605  format(/,10x,'Details of glucose metabolism',/)
  606  format(25x,'GLUCOSE TURNOVER=',1pE10.3)
  608  format(3x,'A ACID=',1pe10.3,2x,'R NUC =',1pe10.3,2x,
     1 'C W PRE =',1PE10.3,2X,'GLYCO=',1PE10.3)
  610  format(25x,'ATP metabolism',/)
  611  format(5x,'A2 PTS=',1E10.3,5X,'P1=',1E10.3,5X,'P2=',1E10.3,5X,
     1 'P4=',1E10.3)
  612  format(5x,'M1=',1E10.3,9X,'M3=',1E10.3,5X,'M4=',1E10.3,5X,
     1 'M5=',1E10.3)
  613  format(5x,'M2I=',1E10.3,8X,'M2M=',1E10.3,4X,'CPG=',1E10.3,4X,
     1 'MYS=',1E10.3)
  614  format(/,25x,'Reductant breakdown',/)
  615  format(3x,'NH4 TRANS =',1E10.3,4X,'ION TRANS =',1E10.3,4X,
     1 'SULFUR =',1PE10.3)
  616  format(3x,'MEMBRANE =',1E10.3,5X,'BIOS =',1E10.3,9X,
```

```
45        1  'P2 -> P3 =',1PE10.3)
46      618  FORMAT (15X,'BLAC=',1PE9.2,7X,'PLM3=',1PE9.2)
47      619  FORMAT (5X,'M21=',1PE9.2,8X,'M22=',1PE9.2,4X,'M2BL=',1PE9.2)
48           END
```

This is the metabolic model for plasmid-containing cells.

```
1         subroutine SCMP(g,dg,m)
2   c     single cell model with plasmid
3   C
4         COMMON CA1,CA2,NRNA,CNA,CMG,PH,TE,E3,NORI,NTOT,PD,VOL,
5        1 DM1S,DM2IS,DM2MS,DM4S,DM5S,DP1S,DP2S,DP2D,DP3S,DP4S,DG3,
6        2 DVP,RA1S,RA2S,DMP,DA2A,DA2CP,S,S1,DEP,V,EC0,ESIG,EDNAC,
7        3 EDNAM,ECDNAC,FRACM,RIBO,SSMIN,RNAM,NGEN,EDNAR,DPGS,
8        4 DT,DPL,COPY,TXEI,TXEII,TXEB,BTX,BTL,BDEG,EQI,EQII,EQB,
9        5 STMI,STMII,STMB,PRIMER,DBLS,DM21S,DM22S,DM2BS,TOTM,
10       6 ednai,ednaii,ednab,fracb,ednap,ecdnap,rpl
11  c
12        external eqp
13  C
14        IMPLICIT DOUBLE PRECISION (A-H,P-Z)
15        DOUBLE PRECISION NRNA,KCD,KHD,KHP,KMG,KMGP,KP,KE
16        DIMENSION G(24),C(24),DG(24),ESIG(5),SSMIN(5),EDNAC(5)
17        DIMENSION ECDNAC(5),EC0(5),EDNAR(5),EDNAM(5)
18        DIMENSION RIBO(5),FRACM(5),RNAM(5),SOS(2,2),V(5)
19        dimension plpos(5),ednai(5),ednaii(5),ednab(5),fracb(5)
20        dimension ednap(5),ecdnap(5),blac(5),copy(5)
21  C
22          U=DLOG(1.0D1)
23          H=DEXP(-U*PH)
24          KP=DEXP(-U*7.4)
25          WS=DLOG(CNA)/U
26  C
27  C     *******   CALCULATE NEW VOLUME    *******
28                  TOTM=0
29                  DO 2000 I=1,18
30   2000             TOTM=TOTM+G(I)
31              VE=G(16)/5.526D-1
32              VC=(TOTM-G(16))/2.584D-1
33              VOL=VE+VC
34              VL=VOL/1.D3
35              V(M)=VOL
36  C
37  C     *******   INITIAL CONCENTRATIONS    *******
38  C
39            DO 2100 I=1,24
40   2100       C(I)=G(I)/VOL
41               DNAC0=G(14)/(6.18D2*VL)
42               DNAM=1.9D-4*DNAC0
43               EC0(M)=G(20)/(4.05D5*VL)
44               SIG0=G(21)/(9.2D4*VL)
45                 CORE0=EC0(M)
46                 DNAR1=NRNA/(6.02D13*VL)
47                 DNAR=DNAR1/1.D10
48                 DNACNP=DNAC0-(DNAM+DNAR)
```

Since the plasmid DNA also binds RNA polymerase, molar concentrations of promoters (PLPRO), total base pairs (DNAP0) and non-promoter base pairs

(DNAPNP) are calculated. The combined total concentrations of base pairs and non-promoter base pairs (DNA0 and DNANP, respectively) for the chromosome and plasmids are also determined.

```
49              copy(m)=g(15)*6.02d23/1.13d7
50              dnap0=g(15)/(6.18d2*vl)
51              plpro=g(15)/(1.13d7*vl)
52              dnapnp=dnap0-3.d0*plpro
53          dna0=dnac0+dnap0
54          dnanp=dnacnp+dnapnp
55    C
56    C     ******** SATURATION AND INHIBITION TERMS ********
57    C
58              RIC1=2.3D-3/(C(1)+2.3D-3)
59              RC2=C(2)/(C(2)+1.8D-4)
60              RA2=C(2)/(C(2)+7.2D-5)
61              RIA2=2.d-3/(c(2)+2.d-3)
62              RA2P1=C(2)/(C(2)+2.5D-4)
63              RIA2M2R=1.1D-5/(C(2)+1.1D-5)
64              RP1=C(3)/(C(3)+2.2D-4)
65              RC3=C(3)/(C(3)+9.9D-4)
66              RP2=C(4)/(C(4)+7.2D-5)
67              RIP2=9.D-3/(C(4)+9.D-3)
68              RP3M3=C(5)/(C(5)+4.D-6)
69              RP4=5.D-3/(C(6)+5.D-3)
70          rca1x=(ca1/(ca1+1.08d-7)-c(1)/(c(1)+1.d-5))
71          rca1y=4.61*(ca1/(ca1+3.6d-6))*2.3d-3/(c(1)+2.3d-3)
72              RCA2=CA2/(CA2+9.3D-6)
73    C
```

β-lactamase mRNA competes with chromosomal mRNA for available ribosomes, as indicated by the inclusion of the blac*(m)*btx terms in lines 79-80. FRACB is the fraction of total ribosomes actively translating β-lactamase mRNA.

```
74    C     ****** CALCULATE EQUILIBRIUM CONCENTRATIONS ******
75    C           RIBOSOMAL EQUILIBRIUM
76                    RNAM(M)=G(10)/(3.9D5*VL)
77                    RIBO(M)=3.19D17*G(9)
78                    blac(m)=g(13)/(2.73d5*vl)
79          fracm(m)=8.75d5*RNAM(M)/(8.75d5*(RNAM(M)+blac(m)*btx)+1.)
80          fracb(m)=FRACM(M)*BTX*BLAC(M)/RNAM(M)
81                    FRACI=1.-(FRACM(M)+FRACB(M))
82    C
83    C     ***** END OF RIBOSOMAL EQUILIBRIUM *****
```

```
85          DM1S=7.1D0*FRACM(M)*8.5D-1*G(9)*RP1*RA2
86          DM1D1=C(3)/(C(3)+1.03D-7)*C(19)/(C(19)+3.9D-8)*
87        1 1.404D-5/(1.404D-5+C(2))+(C(2)/(C(2)+2.73D-3))
88          DM1D=DM1D1*G(7)*7D-2+DM1S*5D-2
89          DG(7)=DM1S-DM1D
90            DBLS= 7.1D0*FRACB(M)*8.5D-1*G(9)*RP1*RA2
91            DBLD= DM1D1*BDEG*G(18)*7.D-2 + DBLS*5.D-2
92            DG(18)= DBLS - DBLD
```

The RNA polymerase distribution is affected by the level of free initiator tRNA factors in the cell, which, in turn, depends on the protein synthesis rate. β-lactamase synthesis (DBLS) must necessarily be included in the representation of protein synthesis effects on polymerase distribution (93-100).

```
93          DPRS = DM1S + DBLS
94          DPROT = DG(7) + DG(18)
95                SPMAX=2.1D0
96                PR=SPMAX*(C(7)+c(18))-(DprS/VOL)
97                PRI=1.1d-2/(1.1d-2+PR**3)
98                PGI=2.2-7/(C(19)+2.2-7)
99                PRMI=4.3D-1/(4.3D-1+(DprS/VOL))
100                PRRI=PGI*PRI
101   C         Calculate equilibrium constants as fcns of ionic strength
102               KMG=DEXP(U*(.5-1.71*WS))
103               DMG=2.*DLOG10(0.5*(1.+DSQRT(1.+4.*KMG*CMG)))
104               KHD=DEXP(U*(-2.5-10.8*WS-DMG-0.4*(PH-7.8)))
105               KCD=DEXP(U*(-8.5-21.2*WS-0.5*DMG-0.3*(PH-7.8)))
106               DPH=2.*DLOG10(1.+KP/H)
107               KMGP=DEXP(U*(0.32+1.75*WS))
108               DMGP=2.*DLOG10(0.5*(1.+DSQRT(1.+4.*KMGP*CMG)))
109               KHP=DEXP(U*(-1.3-14.7*WS-DMGP-DPH)*(0.48+0.52*TE/310.16))
110                 KHD=5.d-1*KHD
111                 KE=1.D10
112                 C1P=1.4*PRMI*KHP
113                 C2P=PRRI*KHP
114                 P1=C1P*DNAM
115                   P21=2.7D1*C2P*DNAR/2.0d0
116                   p22=7.56d0*pri*khp*dnar/2.0d0
117                 p2=p21+p22
```

The three plasmid promoters are accounted for when calculating RNA polymerase equilibrium (P4, P5, P6).

```
118                 P4 = EQI*C1P*PLPRO
119                 P5 = EQII*C1P*PLPRO
120                 P6 = EQB*C1P*PLPRO
```

```
C       CALCULATE EQUILIBRIUM DISTRIBUTION OF RNA
C       POLYMERASE USING 2-D FIBONACCI SEARCH
          if(m.eq.5) mp=3
          if(m.eq.4) mp=3
          if(m.eq.3) mp=1
          if(m.eq.2) mp=1
          if(m.gt.1) goto 111
          if(t0.eq.0) goto 12
          mp=5
 111   ex1u=1.2d0*esig(mp)
       ex1l=8.d-1*esig(mp)
       ex2u=1.2d0*ecdnac(mp)
       ex2l=8.d-1*ecdnac(mp)
       goto 13
  12        EX1U=EC0(M)/(P1+p2+p4+p5+p6)
            EX2U=EC0(M)
            EX1L=0.d0
            EX2L=0.d0
C
  13  continue
      call eqp(ex1u,ex2u,ex1l,ex2l,A,W,SOS,khd,ke,kcd,dnanp,dna0,
     1 core0,sig0,p1,p2,P4,P5,P6,imin,jmin,copy,ngen)
C
          ESIG(M)=EX1L+(EX1U-EX1L)/2.D0
          ECDNAC(M) = (DNAC0/DNA0)*(EX2L+(EX2U-EX2L)/2.D0)
          ECDNAP(M) = ECDNAC(M)*DNAP0/DNAC0
          SSMIN(M) = SOS(IMIN,JMIN)
       if(t0.gt.0) goto 14
       IF(m.GT.1) GOTO 14
       EX1U=8.D-1*EC0(M)/P1
       EX1L=0
       EX2U=8.D-1*EC0(M)
       EX2L=0
       GOTO 15
  14   ex1u=1.4d0*esig(mp)
       ex1l=6.d-1*esig(mp)
       ex2u=1.4d0*ecdnac(mp)
       ex2l=6.d-1*ecdnac(mp)
c
  15   continue
       call eqp(ex1u,ex2u,ex1l,ex2l,A,W,SOS,khd,ke,kcd,dnanp,dna0,
     1 core0,sig0,p1,p2,P4,P5,P6,imin,jmin,copy,ngen)
c
       if(ssmin(m).lt.sos(imin,jmin)) goto 2150
         ESIG(M)=EX1L+(EX1U-EX1L)/2.D0
         ECDNAC(M) = (DNAC0/DNA0)*(EX2L+(EX2U-EX2L)/2.D0)
         ECDNAP(M) = ECDNAC(M)*DNAP0/DNAC0
         SSMIN(M) = SOS(IMIN,JMIN)
C
C     EQUILIBRIUM ACHIEVED
 2150              EDNAM(M)=ESIG(M)*P1
                   EDNAR(M)=ESIG(M)*P2
       ednac(m)=esig(m)*khd*(dnacnp-4.D1*(ecdnac(M)+esig(M)*(P1+P2)))/
     1  (1.+4.d1*khd*esig(m))
```

Four additional equilibrium complexes are formed. Holoenzyme binds to the three plasmid promoters (175-177) and to non-promoter sites on the plasmids (178-179).

```
175              EDNAI(M)=ESIG(M)*P4
176              EDNAII(M)=ESIG(M)*P5
177              EDNAB(M)=ESIG(M)*P6
178         ednap(M)=esig(M)*khd*(dnapnp-4.D1*(ecdnap(M)+esig(M)*
179       1  (P4+P5+P6)))/(1.D0+4.D1*ESIG(M)*KHD)
180    C
181    C      CONTINUE MACROMOLECULAR METABOLISM
182    c
183    c      Plasmid protein reduction of overall metabolic efficiency
184           rpl=g(7)/(6.8d-1*totm)
185    c
186            DM5S=2.0D-2*VOL*(C(2)/(C(2)+2D-3)) * rpl
187            DM5D=1.4D-1*G(17)*(1.D-3/(C(2)+1.D-3))
188            DG(17)=DM5S-DM5D
189               E2=G(23)
190               E23=E2+E3
191         DM4S=6.73D1*E23*(C(2)/(C(2)+3D-5))*(C(6)/(C(6)+5.D-4))
192                  DM4D=2.3D-1*RC2*G(16)
193                  DG(16)=DM4S-DM4D
194              PD=G(7)/G(14)
195              DG(22)=1.3D-4*PD*NORI*DG(7)
196              DG(23)=3.1D-3*DG(7)
197              DG(24)=1.6D-2*G(7) * rpl
198         DG(14)=1.58D0*6.2D-15*NTOT*RP3M3*(C(2)/(C(2)+1.7D-4))
```

Plasmid DNA

```
199         DG(15)=1.58D0*6.2D-15*PRIMER*RP3M3*(C(2)/(C(2)+1.7D-4))
200              DPL = DG(15)*6.02D23*DT/1.13d7
201              IF (DPL.GT.1) DG(15)=1.13d7/(6.02D23*DT)
202            DM3 = DG(14) + DG(15)

203                RI=FRACI*8.5D-1*G(9)
204                DPGS=6.4D-3*RI*(1.265D-4/(C(3)+1.265D-4))
205                DPGD=1.25D2*RA2*G(19)
206                DG(19)=DPGS-DPGD
207                   DM2RS=1.4D1*G(8)
208                   DM2RD=7.D-2*RIA2M2R*G(9)
209                   DG(9)=DM2RS-DM2RD
210              DM2IS=9.6d-1*6.78D7*RP2*EDNAR(M)*VL
211              DM2ID=2.1D1*(1.8D-2/(1.8D-2+DprS/VOL))*G(8)
212              DG(8)=DM2IS-DM2ID-DM2RS
```

Transcription efficiency for mRNA operons is based on the mass-based protein/DNA ratio, as described in Chapter 1. Plasmid DNA further reduces this ratio (213).

```
213           pdi=g(7)/(g(14)+g(15))
214       DM2MS=1.07d0*1.175d1/(1.15d1+pdi)*6.78d7*RP2*EDNAM(M)*VL
215               DM2MD=2.1D1*G(10)
216               DG(10)=DM2MS-DM2MD
```

Transcription from the plasmid RNAI and RNAII promoters (217, 219, 220) is identical to that from chromosomal promoters. β-lactamase transcription (218, 221) is unattenuated, similar to that observed for the rRNA operons. Degradation of these RNAs (222-224) is similar to that for chromosomal mRNAs.

```
217           tx = 1.07d0 * 1.175d1/(1.15d1+pdi) * 6.78d7 * rp2 * vl
218           txs=9.6d-1 * 6.78d7 * rp2 * vl
219             DM21S = TX*TXEI*EDNAI(M)
220             DM22S = TX*TXEII*EDNAII(M)
221             dm2bs = txs * txeb * ednab(m)
222                DM21D = STMI*G(11)*2.1D1
223                DM22D = STMII*G(12)*2.1D1
224                DM2BD = STMB*G(13)*2.1D1
225             DG(11)=DM21S-DM21D
226             DG(12)=DM22S-DM22D
227             DG(13)=DM2BS-DM2BD
228         DM2=DG(10)+DG(9)+DG(8)+DG(13)+DG(12)+DG(11)
229         DM2S = DM2MS+DM2IS+DM21S+DM22S+DM2BS
230   C
231           DECS=1.15D-2*DM1S
232           DECD=DM1D*G(20)/G(7)
233           DG(20)=DECS-DECD
234           DG(21)=7.D-4*DM1S-DM1D*G(21)/G(7)
235   C
236   C       PRECURSOR METABOLISM
237   C
238              DP4S1=6.0D-2*RP4*VOL * rpl
239              DP4S=DP4S1*RC3*(C(2)/(C(2)+1.2D-4))
240              DG(6)=DP4S-1.1D0*DG(16)
241   C
242         AMP=4.66D-1*C(4)
243         DAMP=3.64D-1*C(5)
244         DP3S1=55*G(22)*(AMP/(AMP+5.2D-6))/(1.+DAMP/1.45D-5)
245         dp3s = dp3s1 * ( 4.d0 + 2.d0*g(15)/g(14) )
246         dp3d=37.*g(5)*2
247         if(dg(14).gt.0.0) dp3d=0
248         DG(5)=DP3S - 1.053D0*DM3 - dp3d
249   C
250         DP2S=2.1D-1*VOL*RC3*RIP2*(C(2)/(C(2)+2.5D-4)) * rpl
251         DP2D=3.D-2*RIA2M2R*G(4)
252         DG(4)=DP2S-DP2D-1.057D0*DM2-1.049D0*DP3S
```

```
253   C
254      DP1S=3.9d-1*vol*(5.d-2/(c(3)+5.d-2))*(c(1)/(c(1)+2.5d-5))*ra2p1
255         dp1s=dp1s * rpl
256         DP1D=2.5D-2*RIA2M2R*G(3)
257         DG3=DP1S - DP1D
258      DG(3)=DG3-1.167D0*DPROT-1.149D0*(DP2S-DP2D)-1.28D-1*DP4S
259   C
260               ra1s=2.06d-7*(rca1x+rca1y) * ra2 * s
261               DG(1)=RA1S-1.79D-1*DG3
262   C
263             dmt=0
264             do 2200 i=1,18
265    2200          dmt=dmt+dg(i)
266             dv=dmt/2.9d-1
267             dmp=dmt-dg(2)
268             dvp=dmp/2.9d-1
269   c       GLUCOSE METABOLISM
270   C
271          S1A=2.5D-3*DP1S +2.D-3*DP2S +1.6D-3*DP4S +1.24D-2*DM5S
272          S1B=6.7D-3*DM2S +7.1D-3*DM3 +8.1D-3*DM4S
273          S1=S1A +S1B+ 6.33D-3*DVP+ 3.9D-2*(DM1S+DBLS) +3.D0*DPGS
274   C
275          da2a=1.128*dg3-4.56d-1*(dp2s-dp2d)+1.26*dp4s+1.11*dg(17)
276          RA2S=1.0D-5 * RCA2 * RIA2 * S
277          Z=9.D-1*(DM4S/(DM4S+1.2D-14))
278          DEP=2.8D-2*RA1S+1.711D-2*DMP+5.8D-8*S+3.1D-3*DP3S
279          DA2CP=1.80D2*abs( (S1+1.5D0*DEP)/(2.2D1-6.D0*Z) )
280          GAMMA=1.80D2*(5.6D-3+6.33D-3/2.9D-1+1.5D0*1.711D-1)/
281        1  (2.2D1-6.D0*Z)
282          DG(2)=(RA2S-abs(DA2A+DA2CP))/(1.D0+GAMMA)
283   C        END OF METABOLISM
284            RETURN
285    999     end
1        subroutine EQP(ex1u,ex2u,ex1l,ex2l,a,w,sos,khd,ke,kcd,dnapro,
2       1  dna0,core0,sig0,p1,p2,p4,p5,p6,imin,jmin,copy,ngen)
3     c
4          implicit double precision (a-h, k, p-z)
5          dimension ch(2), st(2), sos(2,2), check(20),fa(2,2),gw(2,2)
6     c
7          r=2.0d0/(1.0d0+sqrt(5.))
8             dif=ex1u-ex1l
9             dwf=ex2u-ex2l
10    c
11           ch(1)=EX1U-R*DIF
12           ch(2)=ex1l+R*dif
13           st(1)=EX2U-R*DWF
14           st(2)=ex2l+r*dwf
15              IFLAG=0
16              ITNO=0
17    C      SUM OF SQUARES CALCULATION
18     3100          DO 3300 I=1,2
19                  DEN=1.D0+4.0D1*ch(I)*KHD
20                  DO 3200 J=1,2
21    C               PSI=EDNA
```

To account for the three plasmid promoters, (P1+P2) is replaced by P16.

```
22            p16=p1+p2+p4+p5+p6
23          PSI=KHD*ch(I)*(dnapro-4.d1*(ch(I)*P16+st(J)))/DEN
24          EC=CORE0-(ch(I)*P16+PSI+st(J)+ch(I))
25          DNA=DNA0-4.D1*(ch(I)*P16+PSI+st(J))
26            sig=sig0-ch(i)-psi
27            FA(i,j)=(KE*EC*SIG-ch(I))/ch(I)
28            GW(i,j)=(KCD*EC*DNA-st(J))/st(J)
29            SOS(I,J)=(FA(i,j)**2+GW(i,j)**2)*1.d-100
30   3200   CONTINUE
31   3300   CONTINUE
33   C      FIND THE LOCAL MINIMUM SUM OF SQUARES
34   C
35          if(itno.eq.0) q=sos(1,1)
36          DO 3600 I=1,2
37          DO 3500 J=1,2
38            IF (SOS(I,J).LE.Q) THEN
39            IMIN=I
40            JMIN=J
41            Q=SOS(I,J)
42           IF(SOS(I,J).LT.1d-100)  IFLAG=1
43          do 3507 L=1,19
44           check(L)=check(L+1)
45   3507    continue
46          check(20)=q
47           if(check(20).eq.check(1)) iflag=1
48          end if
49   3500    CONTINUE
50   3600    CONTINUE
51   c
52   c      ADJUST BOUNDARIES OF REGION TO BE SEARCHED
53   C
54          IF(IFLAG.EQ.1) goto 999
55           ITNO=ITNO+1
56           IF(IMIN.EQ.1) THEN
57               EX1U=ch(2)
58               ch(2)=ch(1)
59               ch(1)=EX1U-R*(EX1U-EX1L)
60           ELSE
61               EX1L=ch(1)
62               ch(1)=ch(2)
63               ch(2)=EX1L+R*(EX1U-EX1L)
64           END IF
65           IF(JMIN.EQ.1) THEN
66               EX2U=st(2)
67               st(2)=st(1)
68               st(1)=EX2U-R*(EX2U-EX2L)
69           ELSE
70               EX2L=st(1)
71               st(1)=st(2)
72               st(2)=EX2L+R*(EX2U-EX2L)
73           END IF
74         GOTO 3100
```

```
75   C
76     999  sos(imin,jmin)=sos(imin,jmin)*1.d100
77          END
```