Extracting Biochemical Parameters for Cellular Modeling: A Mean-Field Approach

Marco A. J. Iafolla and David R. McMillen*

Department of Chemical and Physical Sciences and Institute for Optical Sciences, University of Toronto at Mississauga, 3359 Mississauga Road North, Mississauga, Ontario L5L 1C6, Canada

Received: May 4, 2006; In Final Form: August 18, 2006

Recent developments in molecular biology have made it feasible to carry out experimental verification of mathematical models for biochemical processes, offering the eventual prospect of creating a detailed, validated picture of gene expression. A persistent difficulty with this long-term goal is the incompleteness of the kinetic information available in the literature: Many rate constants cannot or have not yet been measured. Here, we present a method of filling in missing parameters using an approach conceptually analogous to mean-field approaches in statistical mechanics: When studying a particular gene, we extract key parameters by considering the averaged effect of all other genes in the system, analogously to considering the averaged magnetic field in a physical spin model. This methodology has been applied to account for the effect of the presence of the *Escherichia coli* genome on the availability of key enzymes involved in gene expression (RNA polymerases and ribosomes), yielding the number of free enzymes as a function of cellular growth rate. These conclusions have been obtained by deriving genome-wide averages and matching them to bulk literature values of *E. coli* K-12 and B/r. Average rate constants have been found for RNA polymerases and ribosomes binding to promoter and ribosome-binding sites, respectively; these results suggest that cells vary not only their production rates of RNA polymerase and ribosomes under different growth-rate conditions but also change their global level of transcriptional/translational activation and repression, thus altering the average binding rate constants for these enzymes. To test the mean-field method, the results from the genome-wide averages have been applied to the induced lac operon, where our derived on-rate for binding of RNA polymerase to the promoter is in good agreement with previous experimental results.

Introduction

In recent decades, extraordinary advances in biochemistry and molecular biology have led to an unprecedented level of understanding biological systems at the molecular level. The complexity of cellular pathways and networks often makes it difficult or impossible to reliably predict the behavior of a system from knowledge of its components, and thus there is considerable interest in the formulation of quantitative, predictive mathematical models of cellular functions. Such efforts, collectively described by such terms as systems biology and in silico biology,[1−3] aim in the long term toward such goals as predicting the effects of drugs or other interventions on the states of diseased cells and enhancing our fundamental understanding of how cells respond to stimuli and regulate their internal environments.

The challenge of cellular modeling is fundamentally that of reducing biological processes to their underlying chemical basis: Models are generally formulated by identifying the biological species present in a system, defining the chemical reactions that occur between them, and deducing the system dynamics from the kinetics of these reactions.[3−7] One persistent difficulty encountered in these efforts is that they demand rate constants that may not be available; although many key parameters have been measured, many others remain unknown. In some cases, the relevant experiments have simply not been done, while in other cases the measurements required cannot practically be carried out using existing techniques. Models may thus contain unknown parameters, and in this paper we offer a method of filling in these unknowns in an otherwise complete model. We view our contribution as complementary to the various ongoing large-scale cellular simulation projects,[8−14] helping to provide currently unknown parameters to populate the models used in these efforts.

Our technique relies on the existence of experimental results reporting bulk average assays of the amounts of each species present in the biological system of interest; quantities such as average protein levels per cell, average transcript content per cell, and so on are much more readily obtained than specific rate constants for individual reactions. The method consists of formulating as complete a model as possible of the target biological system, identifying any missing rate parameters, and then sweeping these parameters until the model output matches the bulk experimental observations. We proceed to apply this method to a specific question in the modeling of gene expression in bacteria and test the method by deducing an independently known rate constant.

One significant problem in simulating cells is the degree to which critical enzymes (RNA polymerases and ribosomes) are available to participate in the expression of a target gene. The issue is that experimental measurements generally report only total enzyme populations, whereas the relevant quantity for modeling purposes is the number of free enzymes available to participate in the expression of a gene of interest. Using the bacterium *Escherichia coli* as a model organism, we formulate a detailed picture of the biochemical reactions underlying gene expression. The chemical kinetics of this system are simulated using the Gillespie Monte Carlo algorithm,[2,15] and we determine

* Author to whom correspondence should be addressed. Phone: (905) 828-5353. Fax: (905) 828-5425. E-mail: mcmillen@utm.utoronto.ca.

10.1021/jp062739m CCC: $33.50
Published on Web 10/10/2006

TABLE 1: Summary of Kinetic Data

	doubling time τ (min)					
parameter	24	30	40	60	100	ref
mRNA elongation (nucleotides s^{-1})	55	52	50	45	39	18
peptide elongation (amino acids s^{-1})	21	20	18	16	12	18
Rpoly production ($n\ s^{-1}$)	5.3	3.0	1.4	0.52	0.17	this study
ribo production ($n\ s^{-1}$)	33	17	7.3	2.5	0.76	this study
sRNA operon production ($10^{-2}\ n\ s^{-1}$)	7.2	4.6	2.6	1.4	0.73	this study
mRNA operon production ($n\ s^{-1}$)	1.1	0.69	0.40	0.21	0.11	this study
cell division ($10^{-4}\ s^{-1}$)	4.8	3.9	2.9	1.9	1.2	this study
ribo conversion ($10^{-2}\ s^{-1}$)	5.5	5.2	5.0	4.5	3.9	this study
Rpoly dissociation (s^{-1})			10			35
promoter clearance (s^{-1})			1			35
ribo dissociation (s^{-1})			2.25			35
RBS clearance (s^{-1})			0.5			38
isomerization (s^{-1})			1			17

TABLE 2: Parameters Pertaining to the Derivation of Average Transcript and Peptide Lengths

	doubling time τ (min)					
parameter	24	30	40	60	100	ref
protein/cell (10^8 amino acid residues)	25	18.9	13.0	8.7	5.6	34
DNA/cell (genome equiv/cell)	3.8	3.0	2.3	1.8	1.6	34
number of translations per mRNA	93	70	49	33	27	34
mRNA transcripts per cell (10^4)[a]	1.2	1.2	1.2	1.2	0.95	this study
sRNA transcripts per cell (10^4)[a]	7.0	4.4	2.6	1.3	0.66	this study
peptides per cell (10^6)	7.9	6.0	4.1	2.7	1.8	this study
mRNA operons per cell (10^3)	2.4	1.9	1.4	1.1	1.0	this study
sRNA operons per cell (10^2)	1.6	1.2	0.94	0.74	0.66	this study
length of *E. coli*'s genome (bases)			4 639 221			33
total operons per genome			665			33
protein encoding genes per genome			4288			33
approximate mRNA operons per genome			624			this study
approximate sRNA operons per genome			41			this study
mRNA operon transcript length (bases)			6500			this study
sRNA operon transcript length (bases)			910			this study
peptide length (amino acids)			320			this study

[a] The length of the transcript is equal to the length of the operon.

the levels of free enzymes using the approach sketched above and described fully in the Experimental Methods section, below. The original experimental measurements in the literature were carried out over a range of cellular growth rates, each of which yielded different average quantities of biomolecules per cell. Simulations of the system were carried out at each growth rate, yielding a set of computational predictions of the overall average rate constants for binding of the key enzymes (RNA polymerases and ribosomes) to their respective substrates (DNA promoter sequences and messenger RNA transcripts). With these parameters filled in, the model provides the desired prediction of the numbers of free enzymes present in the cells, again varying as a function of cellular growth rate.

As a test of our method of filling in missing parameters, the above free enzyme levels from the genome-wide averaging have been combined with experimentally known transcriptional levels of the lacZ transcript[16] to determine the on-rate of RNA polymerase binding to the activated lac promoter. By placing the lac operon in our mean-field environment and sweeping the on-rate until it matches the experimental transcriptional levels, we obtain an on-rate close to that obtained by Record et al. through direct measurement.[17] Our value is off by a factor of 1.6, which is quite strong agreement given the levels of uncertainty in biological systems.

Experimental Methods

Overview of the Mean-Field Approach. We begin by introducing the sequence of steps employed to determine the unknown parameters in a biochemical system. For each point, we indicate in parentheses how this was applied in our specific case of interest, the determination of free enzyme (RNA polymerase and ribosome) levels in bacteria.

The mean-field approach to parameter determination requires us to:

(i) Identify the output(s) to be generated by the model. (We consider the numbers of free enzymes available for gene expression in *E. coli*.)

(ii) Formulate a model for the target biological system; generally this will involve specifying a set of elementary biochemical reactions representing all relevant biological processes. The desired output must be one of the participating species in this model. (Here, we formulate a detailed model of the transcriptional and translational steps involved in gene expression in *E. coli*, as described below.)

(iii) Using literature values, where available, populate the model with any previously measured rate constants that are known. (Previously measured rate constants are listed in Tables 1 and 2, with references to the original literature.)

(iv) Locate in the literature (or measure experimentally) the bulk population averages for as many biochemical species in the model as possible. (We rely heavily on the comprehensive work of Bremer and Dennis,[18] which reports average cellular levels of nucleotides, proteins, transcripts, and other species, measured at each of several different cellular growth rates.)

(v) Find any elementary reactions where the bulk population averages are sufficient to deduce the rate constant, and use the experimental data to add these constants to the model. (For example, combining the average number of genes/operons per

cell and the average numbers of transcripts per cell yields a mean rate of transcript production for our model. All such calculations are described below, and summarized in Tables 1 and 2.)

(vi) Identify any rate constants still missing from the model. (In our case, the on-rates of binding for the gene expression enzymes to their substrates (RNA polymerase to promoters and messenger RNA transcripts to ribosomes) could not be deduced directly from the bulk experimental data.)

(vii) Perform parameter sweeps of the outstanding rate constant(s), and compare the model output(s) to experimentally known averages as determined above. Start with a wide range of parameter values and then perform narrower sweeps until a sufficiently precise value is obtained for the constant. (We sweep the rate of RNA polymerase binding to promoter sequences, and use the known total average transcript production per cell as a basis for adjustment: The rate constant is adjusted until the model generates the experimentally correct average number of transcripts. Similarly, the total average protein content of the cell is used to adjust the constant for binding of ribosomes to transcripts.)

(viii) With n unknown constants, perform the same type of sweep, but now in n dimensions (at greater computational expense). We have not addressed the case in which multiple regions in this space yield the same average model outputs for the quantity being compared to the experiments. (Our largest sweep is two-dimensional, used in the derivation of the promoter on-rate constant for operons producing stable RNA and messenger RNA.)

(ix) Place the new constants into the original model, and use the model to generate the desired prediction. (With the enzyme−substrate binding constants and a fully populated model in hand, we are able to predict the number of free enzymes per bacterial cell that are available to participate in gene expression, as a function of cellular growth rate.)

Modeling Approach and Software. Deterministic chemical kinetics apply in the regime of large numbers of randomly interacting molecules. Inside cells, molecule numbers are often small enough to produce significant fluctuations,[2,3,7,15,19] thus requiring a stochastic simulation of the reaction kinetics. The Gillespie algorithm[2] treats chemical reactions as Poisson processes, with event (reaction) rates given by microscopic rate constants and the current state of the system. For an elementary reaction of the form $A + B \rightarrow C$ with rate constant k, the Poisson rate of the forward reaction is kab/V, where a and b represent the numbers of molecules of species A and B present, and V is the reaction volume. (Note that this volume is a changing parameter in a living bacterial cell.) We use the unit "n" to represent the number of molecules present in the system rather than concentration units such as molarity. To advance the simulation, the timing of the next reaction event is randomly selected using the exponential distribution of interevent times for the set of Poisson processes representing the reactions, and the probability of each reaction being the one that occurs at that instant is given by its fraction of the sum of all reaction rates.[2,4,11]

Bacterial cells have often been approximated as well-stirred reactors: On the basis of their small size, it is assumed that diffusion is sufficiently fast to yield a well-mixed system. Early experimental results showed protein mobility in vivo consistent with normal diffusion,[20] and though the diffusion coefficients were substantially lower than those for the same proteins in water, the diffusion was fast enough to spread the proteins over the volume of a bacterium on a time scale of seconds. Recent theoretical treatments[21−24] have questioned the picture of bacterial cells as well-mixed systems, and recent experimental results[25] have reported subdiffusive behavior in the motion of individual RNA molecules, where each RNA is rendered visible through binding to multiple fluorescent protein labels. In this paper, we use the well-stirred reactor picture as a first approximation to gain insight, but it should be noted that this is a significant simplification and that future refinements and extensions are possible. Approaches proposed to deal with crowded cellular environments include rate laws obeying fractal-like kinetics[21,26,27] and Monte Carlo simulations wherein two- or three-dimensional spatial information is retained for each molecule.[21,22,28,29]

The gene expression model is implemented using BioNetS (Biochemical Network Stochastic Simulator),[15] which provides a convenient interface for specifying reactants, products, and kinetic data. The software generates C++ source code implementing the system using the Gillespie algorithm (or an approximation, if desired), and this code is then compiled and executed with user-tunable parameters as inputs. Some species in the model exist in small numbers while others exist in large numbers; although continuum approximations and hybrid schemes are available through BioNetS,[15] the Gillespie algorithm with no approximations yielded the best simulation speed. The data from the BioNetS-generated code was processed using DataTank (Visual Data Tools, Inc.) and its run manager DataTask, which automated the process of sweeping parameter values and analyzing the results. The complete gene expression models used are available as BioNetS scripts and are provided in the Supporting Information.

Derivation of *E. coli* Gene Expression Parameters. To derive the number of free RNA polymerases and ribosomes available inside *E. coli*, we employ bulk cellular averages obtained by Bremer and Dennis for several different cellular growth rates.[18] The key values required to predict the free enzyme levels are the average binding constants (on-rate) for RNA polymerases and ribosomes to their genomic promoters and ribosome-binding sites (RBS's), respectively. We implement the mean-field approach by considering the production of generic transcripts with properties derived from genome-wide averages: We compute mean transcript lengths, mean elongation rates, and so on. With these quantities in hand, the number of unknowns in the transcriptional process is reduced to the RNA polymerase on-rate for binding to the average promoter, and we find its value by sweeping until the level of generic transcript production matches the experimentally observed level of genome-wide transcript production. Similarly, genome-wide averages are employed to derive the properties of generic proteins, including mean peptide chain lengths and mean elongation rates. These values reduce the number of translational unknowns to a single rate constant, the on-rate of transcript binding to the RBS, and once again this value is derived by sweeping it until the experimentally determined target level of overall protein production is obtained. Substituting these two mean-field values into the model allows RNA polymerases and ribosomes to be occupied with genomic production for a specified cellular growth rate, allowing calculation of available enzymes for target genes. We then use nonlinear curve fitting to obtain an equation relating the free enzyme levels and binding constants to the cellular growth rate.

***E. coli* Gene Expression Model.** The model has been constructed to be as detailed as possible, using all available information about the biochemical processes underlying gene expression. This leads to a large number of species and reactions,

the full details of which are provided in the Supporting Information. Here, we provide an overview of the processes incorporated into the model and the assumptions used to implement them. (We introduce the label, in parentheses, for the number of molecules of each species as they first occur in the discussion.)

Cell Growth and Division. The cellular volume grows exponentially until a threshold is reached, at which point it is approximately halved (a binomial distribution is used) and exponential growth restarts. A counter species, v, is used to represent volume, $v \rightarrow 2v$, with rate constant $k = \ln(2)/\tau$, where τ is the variable doubling time of the cells. At cell division, all species present are divided between two hypothetical daughter cells, and the simulation follows one of these daughters. We treat the system as ergodic and average over long times for a single cell to obtain ensemble averages across the cellular population.

RNA Production from Operons. We assume that all genes in the genome are clustered into operons, groups of genes transcribed from a single promoter, as in the lac operon.[30] The reaction scheme allows individual genes of the generic operon to be translated from the nascent transcript. Please see the Supporting Information for additional details.

Enzyme Binding and Isomerization. The enzymes tracked in the model are RNA polymerases (Rpoly) (responsible for initiating and catalyzing the transcription of both messenger and stable strands (mRNA and sRNA, respectively)) and ribosomes (ribo) (which bind to the RBS of an mRNA strand to initiate translation of the mRNA into protein). Rpoly binds to promoter sequences in the DNA (operon) and forms a closed complex (Rpoly + operon → closed_Rpoly_prom). This closed complex then must isomerize into an open complex (closed_Rpoly_prom → open_Rpoly_prom) before transcription can begin. The ribosome binds to the mRNA's RBS forming an "open" complex (ribo + mRNA → open_ribo_RBS) that is used to represent the numerous protein cofactors involved in translation initiation.

Enzyme Clearance. RNA polymerase and ribosomes clear the promoter and RBS, respectively, leaving those sites free to bind additional enzymes while transcription/translation proceeds further down the DNA/mRNA strand. We model this by regenerating the promoter or RBS after clearance occurs, forming an enzyme−template complex plus the original site: open_Rpoly_prom → Rpoly_operon + operon and open_ribo_RBS → ribo_mRNA + mRNA. Conservation of the number of promoters or RBSs is maintained: When the enzyme−template complex finishes elongation (of an mRNA or a protein), only the enzyme and polymerized product are released.

Elongation. To avoid the complexity of accounting for each enzyme at different stages of elongation, a single reaction is used to represent the process of completing the mRNA or protein chain: Rpoly_operon → Rpoly + mRNA and ribo_mRNA → ribo + protein. The elongation rate constant can be summarized as $k_{\text{elongation}} = \rho/\lambda$, where ρ and λ are the polymerization rate and length of the template, respectively.

Enzyme Production. Since the kinetics of RNA polymerase and ribosome assembly are not fully known, the model is simplified by treating enzyme production as a zero-order process in which enzymes appear from outside the model at a constant rate, ( ) → Rpoly and ( ) → ribo, where ( ) is a null placeholder. The enzymes are partitioned at cell division like all other species. The rate constant for production can be summarized as $k_{\text{rep}} = (\nu/1.5)/\tau$, where ν and τ are the average number per cell and cellular doubling time, respectively.

Translation from Incomplete Transcripts. Ribosomes can bind to an mRNA transcript before transcription is complete,[16] and this is accounted for by creating a species to represent the portion of the growing mRNA chain available for ribosome binding (incom_mRNA): open_Rpoly_prom → Rpoly_operon + incom_mRNA + operon. An additional reaction, incom_mRNA → ( ), has its rate set to match the rate of transcribing a complete gene so that the incomplete mRNA species vanishes at the same rate at which complete mRNA are produced. Please see the Supporting Information for an explanation concerning the conversion of incomplete to complete transcripts that are complexed with ribosomes, a phenomenon that we label as ribosome conversion.

mRNA Degradation. RNases act to destroy mRNA in *E. coli*, and we represent the degradation of mRNA by these enzymes with first-order reactions, mRNA → ( ) and incom_mRNA → ( ); the latter is an additional RNase-driven degradation, beyond the above-mentioned rate of disappearance of incomplete mRNA through conversion to complete mRNA strands. mRNA being held in an open ribosome complex is protected from RNase degradation,[30] and thus these species are not subject to degradation reactions. RNases can affect nascent (still-forming) transcripts, and we permit the growing mRNA in species Rpoly_operon to be degraded: Rpoly_operon → Rpoly. The rate of this reaction is adjusted to reflect the fact that only nascent transcripts without a ribosome attached will be subject to RNase degradation (please see the Supporting Information for further explanation). Finally, mRNA degradation is only applied in derivation of the activated lac promoter on-rate; in the other simulation runs, it has implicitly been taken into account when constructing the cellular averages, as explained in the Results and Discussion section.

DNA Replication. DNA replication in bacteria is a complex process involving multiple replication forks.[32] We represent the coding portion of the genomic DNA by the number of operons present (operon) and simplify the replication process as a zero-order process: ( ) → operon. Rate constants for this process are chosen to match the number of genomes per cell at different growth rates. The rate constant formula is analogous to that stated in the Enzyme Production subsection above.

Stable RNA Production. In addition to mRNA, other forms of RNA collectively known as sRNA are produced within the cell. Since sRNA is transcribed but not translated, it influences only the RNA polymerase binding constants.

Summary of Information. Table 1 summarizes the kinetic parameters used in the computational simulations. Please see the Supporting Information for all elementary reactions and corresponding kinetics that compose the model.

Kinetic Data. Below is the derivation of averaged genomic parameters. The experimental data used to derive average kinetic information are provided in Table 2.

Total Nucleotide Bases of the Transcript-Producing Genome. There are 4 639 221 bases per genome in *E. coli* K-12, of which 87.8% account for protein-coding genes, 0.8% encodes sRNAs, and 0.7% consists of noncoding repeats, leaving approximately 11% for regulatory and other functions.[33] Since this study is only interested in how the genome affects RNA polymerase and the ribosome, only the protein and RNA portions of the genome are considered. We neglect nonspecific binding of RNA polymerase, which would allow Rpoly molecules to interact with other stretches of DNA.

Derivation of the average RNA polymerase and ribosome on-rate requires a differentiation between sRNA and mRNA: The sRNA is not translated and therefore has no bearing on the

activity of ribosomes. Let G represent the whole genome length so as to split the genome of interest into two pieces: G_{mRNA}, the total number of bases encoding proteins, and G_{sRNA}, the number of bases coding for sRNA. From the known fractions, $G_{mRNA} = 0.878G$ and $G_{sRNA} = 0.008G$.

Nucleotide Bases per Operon. We assume that all genes in our relevant genome are clustered into operons. There are 665 operons/genome[33] and of these there are 41 sRNA operons, 6 of which also encode for mRNA.[34] Here we assume that the 6 operons of sRNA and mRNA are exclusively composed of sRNA and normalize away this assumption during the simulation (see the Results and Discussion section). Therefore, there are 41 sRNA operons per genome and 624 protein-coding operons per genome. Defining O_{mRNA} and O_{sRNA} as the average number of bases in the protein-coding and sRNA-coding operons, respectively, $O_{mRNA} = (G_{mRNA})(624 \text{ operons/genome})^{-1}$ and $O_{sRNA} = (G_{sRNA})(41 \text{ operons/genome})^{-1}$.

Genes per Operon. There are 4288 protein-coding genes in the *E. coli* genome.[33] Setting N_{mRNA} to be the number of genes in an average mRNA operon, $N_{mRNA} = (4288 \text{ genes/genome})(624 \text{ operons/genome})^{-1} \approx 6.9$ genes/operon. Since sRNA is not translated, the number of genes per sRNA operon is not relevant to this study.

Nucleotide Bases per Average Gene and Amino Acids per Average Peptide. The average number of nucleotide bases in the average mRNA gene is $L_{mRNA} = G_{mRNA}(4288 \text{ genes/genome})^{-1} \approx 950$ bases/gene. Assuming that no posttranslational modifications occur, there is a 3:1 ratio between the number of bases on an mRNA transcript and the number of amino acids polymerized into peptides (protein chains). The average peptide length is thus $L_{peptide} = L_{mRNA}(3 \text{ bases/amino acid})^{-1} \approx 320$ amino acids/peptide, assuming that every base on the mRNA is translated into protein.

Amino Acids Polymerized per Operon. It has been shown that each transcript is translated a finite number of times before degrading and that this number of translations varies as a function of cellular doubling time.[18] Let $T(\tau)$ represent the number of translational events per mRNA, where τ is doubling time. Since RNA polymerase has been shown to produce single transcripts from operons,[31] the average number of amino acids polymerized per mRNA operon is $A_{mRNA} = (O_{mRNA})T(\tau)(3 \text{ bases/amino acid})^{-1}$.

Average Number of mRNA Transcripts (Operon Length) per Cell. Due to our simulated release of mRNA as it is produced from the operon, the number of mRNAs in our model represents the number of genes transcribed not the number of whole operons transcribed. One can determine the latter by calculating the average level of mRNA produced in the simulation (indicated by angle brackets), taking into consideration that there are approximately 6.9 genes per operon: The model specifies the production of this smaller transcript (~90% of an average transcript) and calculates the number of whole operons transcribed as follows

$$T_{\text{mRNA operon}} = [\langle \text{mRNA} \rangle + \langle \text{mRNA}_{\text{small}} \rangle (N_{\text{mRNA}} - 6)] (\text{N}_{\text{mRNA}})^{-1} \quad (1)$$

where $T_{\text{mRNA operon}}$ and $\text{mRNA}_{\text{small}}$ represent the average number of whole mRNA operons transcribed and ~90% of an average transcript produced, respectively. The term $(N_{mRNA} - 6)$ ensures that the $\text{mRNA}_{\text{small}}$ species represents only ~90% of an average transcript.

The number of amino acids contained in peptides within the cell has been experimentally determined as a function of doubling time.[18] Let $A(\tau)$ represent this average number of cellular amino acids present per cell, where τ is doubling time. The number of amino acids present at the beginning of the cell cycle is approximately $A(\tau)/1.5$; this many amino acids must be polymerized via translation before cell division occurs. Since the transcript from the average operon is responsible for producing A_{mRNA} amino acids in the form of peptides, the number of whole operon mRNA transcripts produced per cell cycle is $T_{\text{mRNA operon, cell cycle}} = (A(\tau)/1.5)(A_{mRNA})^{-1}$, where $T_{\text{mRNA operon, cell cycle}}$ represents the average number of whole operon mRNA transcripts produced before cell division; this parameter also represents the number of whole operon mRNA transcripts present at the beginning of the cell cycle. Therefore, the expected average number of transcripts observed in each cell is approximately

$$T_{\text{mRNA operon}} = (1.5)(T_{\text{mRNA operon, cell cycle}}) \quad (2)$$

Equation 2 specifies the number of generic transcripts produced by the mRNA portion of the genome and thus provides the target value sought by our parameter sweeps: We adjust the RNA polymerase on-rate to mRNA promoters until eqs 1 and 2 are equal. This yields a genome-wide average on-rate for RNA polymerase binding to the mean mRNA operon and provides the first output of the mean-field approach.

Average Number of Stable RNA Transcripts (Operon Length) per Cell. Dennis and Bremer have obtained the ratio between the number of sRNA nucleotides polymerized and the total RNA polymerized, as a function of doubling time;[18] let $R(\tau)$ represent this ratio. Through the use of the number of mRNA transcripts produced per cell cycle, the number of sRNA transcripts can be derived

$$T_{\text{sRNA operon}} = (T_{\text{mRNA operon}})R(\tau)[1 - R(\tau)]^{-1} \quad (3)$$

where $T_{\text{sRNA operon}}$ represents the average number of transcribed sRNA operons present per cell. The number of transcribed sRNA operons present at the beginning of the cell cycle or (equivalently) produced per cell cycle is $T_{\text{sRNA operon, cell cycle}} = (T_{\text{sRNA operon}})/1.5$. We observe the average level of sRNA transcripts produced in the simulation and match them to the value of eq 3 by adjusting the on-rate for RNA polymerase to the stable operon promoter.

Average Number of Peptides per Cell. Since the cell produces $A(\tau)/1.5$ amino acids in the form of peptides during the cell cycle, the number of peptides produced during the course of the cell cycle is $T_{\text{peptide, cell cycle}} = (A(\tau)/1.5)(L_{\text{peptide}})^{-1}$; this is also the number present at the beginning of the cell cycle. The average number of peptides observed in each cell is

$$T_{\text{peptide}} = (1.5)(T_{\text{peptide, cell cycle}}) \quad (4)$$

The parameter sweeps require adjustment of the ribosome on-rate constant to the average RBS until the average peptide level in the simulations matches the value from eq 4. This rate constant represents a genome-wide average of the binding of generic transcripts to ribosomes, and provides the final output of the mean-field approach.

Number of Operons per Cell. Dennis and Bremer[18] have found the amount of DNA polymerized by the cell during one cell division, expressed as a multiple of the size of a single genome and varying as a function of doubling time. Letting $D(\tau)$ represent the average number of genome equivalents per cell, the average numbers of operons present at the beginning of the cell cycle are $P_{mRNA} = (624 \text{ operons/genome})D(\tau)$ for

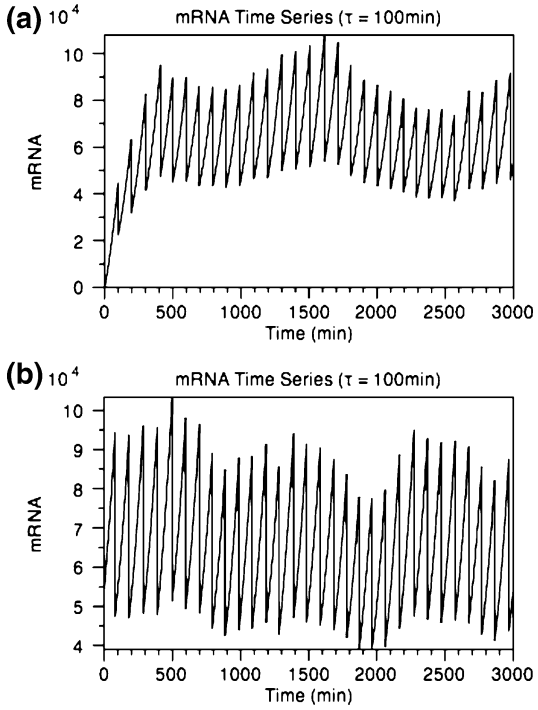

Figure 1. Sample time series from the model. (a) A run with all intermediates and products initially set to zero, showing the initial transient. (b) A run initialized with the state obtained after 30 cell divisions in the left-hand run, thus removing the initial transient. Simulations were performed at a variety of cellular growth rates (expressed as population doubling times); here the 100 min doubling time is shown as a representative example.

the mRNA-coding portion of the genome and $P_{sRNA} = (41 \text{ operons/genome})D(\tau)$ for the sRNA portion. Note that $D(\tau) > 1$ owing to the multiple replication forks involved in bacterial DNA replication.[18,32]

Summary. Table 2 summarizes the average transcript and peptide population number and polymer length produced by *E. coli*'s genome, in addition to the necessary information used in the derivation. Please see the Supporting Information for the complete list of all elementary reactions that compose the model.

Parameter Sweeping. Derivation of the average on-rate constant for enzymatic binding to *E. coli*'s promoters and RBS's proceeds in two steps: One model is constructed to derive the on-rate constant of RNA polymerase to the average promoter of operons producing stable and unstable RNA; the second step augments the RNA polymerase model with additional reactions describing the process of translating mRNA into protein and is used to derive the on-rate of binding between ribosomes to the average RBS. The RNA polymerase on-rates must be derived first to produce the correct level of mRNA for ribosomal on-rate determination.

Figure 1 shows the time series for one species in the model, free mRNA. An initial run of 30 cell divisions in length is generated for each simulation, and the final state of this run is used as the initial state for the long-duration run in which statistics are accumulated to determine average species levels; this prevents the initial transient approach to the steady state from distorting the averages. Simulations for the RNA polymerase on-rate determination were run for a duration of 500 cell divisions. This duration was based on qualitative analysis of multiple simulations with the same seed but different durations: Longer-duration runs produced averages that were not statistically different from those obtained after 500 divisions (see the Supporting Information for additional explanation). Note that since the doubling time of the cells is varied, the total duration in real time varies among the simulations, but the number of cell divisions explored appears to be the key factor in obtaining well-converged statistics rather than the absolute duration. For the ribosome-binding on-rate, the outputs showed more variability, as has been reported previously,[35] and for these more variable runs the duration was extended (effectively, by combining the results from multiple independent random seeds) until linear regressions with $R^2 \geq 0.95$ were obtained in the vicinity of the on-rate of interest.

Parameter sweeping begins by using on-rates that vary by a factor of 10 (Figure 2). When the desired transcript/peptide average lies between two on-rates, another sweep is performed between these new limits incrementing the on-rate by a unit multiple of the smaller limit. The third parameter sweep uses a unit increment of the next significant digit between the new limits; this change in on-rate is small enough to approximate linearity (Figure 2). Only $R^2 \geq 0.95$ values are accepted for interpolation. The validity of the interpolated on-rate is then tested; using a different seed for 30 simulations—all employing the same initial conditions, duration, kinetics, and interpolated on-rate—the sample mean of the 30 simulations is compared to the population mean. The on-rate is accepted if the two means are not proven statistically different using a level of significance $\alpha = 0.95$. All simulations, either parameter sweeping or verification, employ different nucleating random number generator seeds.

Results and Discussion

Effect of the *E. coli* Genome on Enzyme Availability. The interpolated on-rates are summarized in Table 3 and have been computationally curve fitted to allow interpolation at any doubling time bounded by 24 and 100 min. The functions in Figures 3−5 represent this doubling time dependence; all functions are written in the generic form $y = f(x)$, where y is the dependent variable and x is the independent variable.

The average levels of free and active RNA polymerase and ribosome have been calculated from the verification runs and have been compiled in Figure 5 as a function of doubling time. These results represent a prediction of the population of enzymes available for target genes.

The fitted power-exponential function for the mRNA data in Figure 3 has a peak at a doubling time of 78 min, which is likely an artifact of the fitted function. A more plausible functional dependence is an asymptotic exponential function of the form $y = a(1 - e^{(-b(x-x_0))}) + c$, where a, b, and c are

TABLE 3: Summary of On-Rates

parameter (y)	doubling time τ (min) 24	30	40	60	100	fitted function
on-rate to mRNA operons ($10^{-5}\ n^{-1}\ s^{-1}$)	0.40	0.58	0.89	1.4	1.4	$y = 1.7 \times 10^{-9}x^{2.7}\ e^{-0.03x}$
on-rate to sRNA operons ($10^{-4}\ n^{-1}\ s^{-1}$)	5.9	4.2	3.4	2.4	1.4	$y = 0.016x^{-1.1}\ e^{0.0025x}$
on-rate to RBS ($10^{-5}\ n^{-1}\ s^{-1}$)	5.3	5.4	3.7	2.6	2.5	$y = 0.0030x^{-1.3}\ e^{0.013x}$
free RNA polymerase (10^3)	10	7.2	4.5	2.4	1.3	$y = 3.9 \times 10^{6}x^{-1.9}\ e^{0.0090x}$
free Ribosome (10^3)	7.3	4.2	3.1	2.0	1.0	$y = 1.2 \times 10^{6}x^{-1.7}\ e^{0.0071x}$

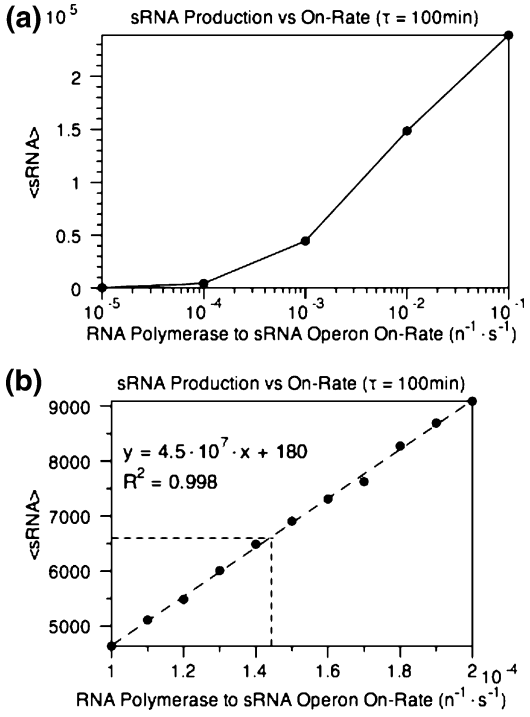

Figure 2. Parameter sweeping. In this example, the only unknown is the RNA polymerase on-rate constant for binding to operons that produce stable and unstable RNA; these kinetic parameters are varied until the simulated output of mRNA transcripts matches the experimentally observed value. The on-rate is first varied by a factor of 10 to determine the general location of the desired value (a), then followed by a sweep on a fine increment to have the results approach linearity (b). The ideal transcript number and interpolated on-rate constant for the fine graph are illustrated with a fine-dashed line. Sweeping for two unknowns requires one unknown to remain static while the other is varied, then vice versa. Part a is swept using an on-rate of 5.0×10^{-6} n^{-1} s^{-1} for RNA polymerase to mRNA operon promoters; part b is using 1.4×10^{-6} n^{-1} s^{-1}.

constants and x_0 is the lower limit of the independent variable. Nonlinear least-squares fitting with this function yields $a = 1.1 \times 10^{-5}$ n^{-1} s^{-1}, $b = 0.050$, $c = 3.5 \times 10^{-6}$ n^{-1} s^{-1}, and $x_0 =$ 24 min. However, this slightly degrades the overall function fitting ($R^2 = 0.97$) and in particular provides a significantly poorer fit for the faster doubling times between 24 and 60 min. Practically, it may be best to interpolate using the power-exponential function only for points with doubling times between 24 and 60 min and treat the region from 60 to 100 min as flat (constant on-rate).

Varying Rate Constants. The plots in Figures 3 and 4 may seem counterintuitive: How can a rate constant be subject to change? Cells contain numerous transcriptional and translational cofactors, whose presence alters the binding affinity of RNA polymerase for a promoter or of a ribosome for a RBS. These cofactors are not explicitly represented in the model, but their averaged effect is reflected in the average rate constants for enzymatic binding. In the experimental literature,[18] the doubling times of cellular populations are altered by placing the cells into varying growth conditions, which has the potential to alter gene expression patterns in the cell. The changes in the binding on-rate constants seen in Figures 3 and 4 suggest that one effect of these altered growth conditions is that the global activation state of the cell (that is, the combined effect of activating and repressing cofactors) changes as a function of growth rate, leading to net changes in the average binding rate constants of the key enzymes involved in gene expression.

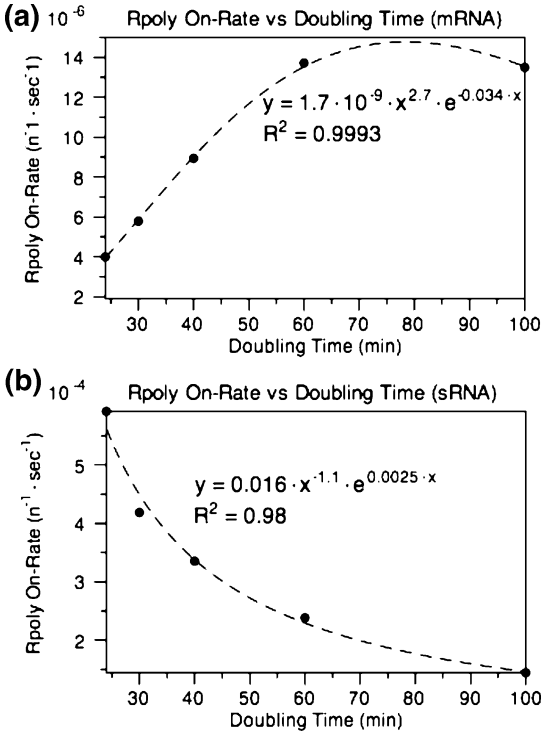

Figure 3. Fitted functions for RNA polymerase binding constants to the average operon producing mRNA (a) and sRNA (b). These curves produce an approximation of on-rate constants as a function of cellular growth rate (expressed as a population doubling time), over the range $24 \leq \tau \leq 100$ min.

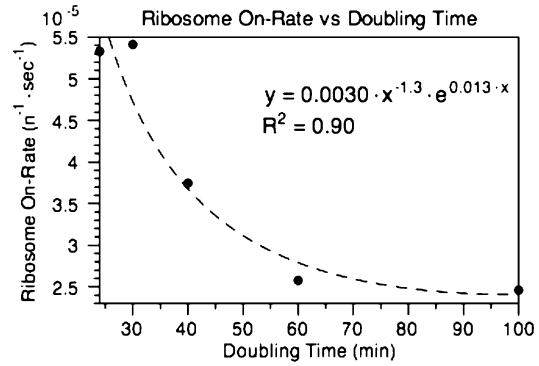

Figure 4. Fitted function for ribosome-binding constants to the average RBS. This curve produces an approximation of on-rate constants in the range of cellular growth rates (expressed as population doubling times) of $24 \leq \tau \leq 100$ min.

Validity of Genomic Effects on Free RNA Polymerase and Ribosome Levels. Application to the lac operon requires the incorporation of a genome at a doubling time of 48 min (see below). We constructed the genomic effect at this doubling time from first principles: Interpolation of the necessary parameters provided by Dennis and Bremer was performed to create the appropriate level of transcripts and peptides at $\tau = 48$ min (data not shown). These levels were then used to derive the RNA polymerase and ribosome on-rates to the genomic promoters and RBSs, respectively. These on-rates from founding principles (not shown) also allow a direct comparison to the accuracy of the curves presented in Figures 3−5. Here, the RNA polymerase parameters derived from sweeping at $\tau = 48$ min agree with those predicted by the fitted curves to two significant digits; the ribosome parameters agree to one significant digit.

Assigning Operons to the Production of sRNA. Due to the complexity of modeling operons composed of sRNA and

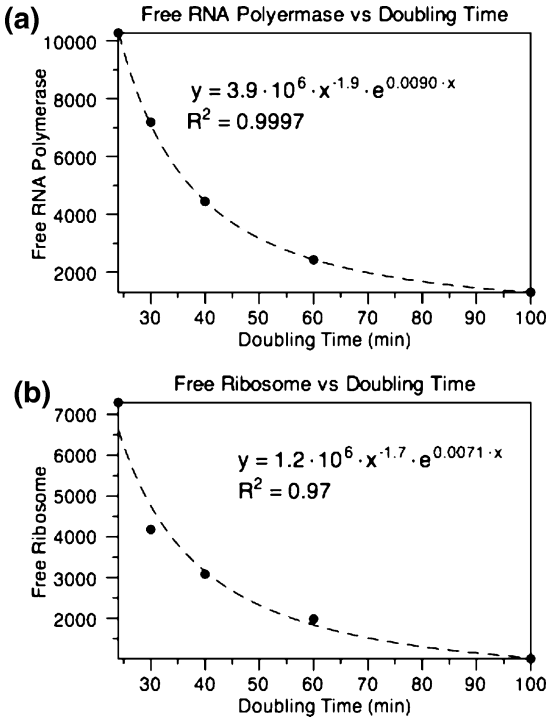

Figure 5. Fitted functions for free (a) RNA polymerase and (b) ribosome enzymes. These curves produce an approximate number of free enzymes for cellular growth rates (expressed as doubling times) in the range $24 \leq \tau \leq 100$ min.

mRNA, it is assumed that the six operons of mixed composition[34] are composed primarily of sRNA. Fortunately, this assumption does not produce a serious inaccuracy: Assigning additional operons as sRNA results in smaller operon lengths. Although the cell now has a larger "quota" of transcripts to fulfill before cell division, this produces additional binding sites for RNA polymerase and shorter operons, reducing the time for completing elongation. Similarly, having fewer operons that exclusively produce mRNA results in a smaller number of binding sites and longer elongation times to create the necessary number of transcripts and peptides prior to cell division. This balancing effect of transcript production and RNA polymerase availability helps to remove the severity of this assumption.

Mean-Field Application to the Lac Operon. To assess the validity of the methodology presented in this paper, we chose to derive the promoter on-rate constant for the induced *lac* operon and compare it to that derived in the literature.[17] The number of transcripts produced in the model was compared to bulk measurements of the lacZ (β-galactosidase) transcripts (3063 nucleotides in length) observed by Kennell and Riezman in *E. coli* 1000, F^-, thiamine$^-$ cells.[16]

An average of 1.66 lac operons exists at a doubling time of 48 min,[16] which poses difficulties for our simulation approach. Producing such a small number of operons via a zeroth order reaction leads to huge fluctuations under stochastic chemical kinetics. To circumvent this, the numbers of operons and associated species (open and closed complexes with RNA polymerase) are set to constant values and are not subject to cell division. Since the discrete Gillespie algorithm operates only on whole numbers of molecules, two different simulation outputs are combined to yield an average of 1.66 *lac* operons. Specifically, a weighted average of model runs with 1 and 2 lac operons present is used to obtain our desired number

$$1.66_{\mathrm{lac}} = [0.34(1_{\mathrm{lac}}) + 0.66(2_{\mathrm{lac}})](0.34 + 0.66)^{-1}$$

where the "lac" subscript represents the number of lac operons present in the model. A run incorporating two lac operons is run for 66% of a set duration, and a second run with one lac operon is run for 34% of the same duration. Combining the results of the two models yields the species averages for a model with a mean of 1.66 lac operons per cell.

Interpolated On-Rate. The validity of the mean-field methodology presented in this paper has been confirmed using the induced lac operon as an application. The RNA polymerase on-rate to the lac promoter is swept, and the results of the simulations are matched to the number of nascent, complete, and decaying lacZ transcripts observed by Kennell and Riezman.[16] Due to the structure of the elementary reactions presented in this paper it is not possible to determine which lacZ transcripts are in the process of being degraded, since in the model mRNA degradation occurs in a single step. However, it can be assumed that all viable transcripts in the model—both complete and incomplete—are subject to degradation. Hence, the sum of complete and incomplete transcripts is equivalent to the sum of all three experimentally observed parameters (complete, nascent, and decaying). Further, the nascent lacZ transcripts observed are equivalent to both the Rpoly_lacZ_gene and the incom_mRNA_lacZ species. This, in addition to the interpolated on-rates and the average, is summarized in Table 4.

Comparison of our derived on-rate with the experimentally derived on-rate of 10^8 M^{-1} s^{-1} [17] requires a change of units. Since the volume of *E. coli* is approximately 2×10^{-12} cm^3,[36] this can be used with Avodagro's number to convert the literature on-rate into units of species number per time, the unit used in the models. Applying this conversion, $(10^8\ M^{-1}\ s^{-1})(2 \times 10^{-15}\ L)^{-1}(6.02214199 \times 10^{23}\ n/\mathrm{mol})^{-1} \approx 0.08\ n^{-1}\ s^{-1}$. Therefore, the derived on-rate is only off by a factor of 1.6. This is a reasonably strong agreement between the model and experimental results, given the uncertainties inherent in quantifying biological materials.

mRNA Degradation. The effect of the genome on enzyme availability has been derived using a fixed number of translations per mRNA, as supplied by Dennis and Bremer. That each mRNA has a finite number of translational events as a function of doubling time suggests competition between RNases and ribosomes for essentially the same site on the mRNA is inherent in the derivation of the mean-field parameters. Therefore, elementary reactions representing mRNA degradation are not included in any of the simulations for deriving the average on-rate to promoters and RBS's.

Conversely, derivation of the activated lac promoter requires the addition of mRNA degradation: The simulation output is compared to in vivo laboratory observed parameters of completed, decaying, and nascent lacZ transcripts supplied by Kennell and Riezman. These parameters are based on the effects of enzymatic degradation and protection; the reported lacZ half-life (90 s), also in the same paper, represents the average effect of RNase and ribosome competing for effectively the same site on the mRNA. This requires model construction with degradation of both complete and incomplete lacZ transcripts.

Sources of Departure between Derived and Observed Activated Lac Promoter On-Rate. The elementary reaction representing mRNA degradation of nascent transcripts produces an overestimate of free RNA polymerases. This imprecision is a result of the immediate ejection of RNA polymerase upon the initiation of degradation. In an attempt to maximize both accuracy and efficiency of the model, we omit explicit representation of nascent mRNA; an accurate description of this event would require knowledge of the length of the nascent

TABLE 4: Summary of Equivalent LacZ Parameters and Derived On-Rates

simulated parameters	equivalent LacZ laboratory parameter(s) {population number}	interpolated on-rate (n^{-1} s^{-1})	average interpolated on-rate (n^{-1} s^{-1})
Rpoly_lacZ_gene	nascent{38}	0.055	0.051
incom_mRNA_lacZ		0.055	
mRNA_lacZ + incom_mRNA_lacZ	complete{31} + nascent{38} + decaying{31} = 100	0.042	

transcript upon the onset of degradation so the remaining time for elongation can be computed. Ultimately, this overestimate in RNA polymerases leads to a smaller binding constant.

The model proposed in this paper is representative of reactions based on the collision of reactants from three-dimensional diffusion. This is in contrast to nonspecific binding of RNA polymerase in cells: RNA polymerase has the potential to bind to random sequences of DNA and scroll along until it finds a promoter sequence for transcription initiation;[37] this has been shown to increase the rate of transcript initiation.[17] Hence, the laboratory observed on-rate for the activated lac promoter is made larger by the catalysis of nonspecific DNA binding, an effect not incorporated into the model.

Macromolecular crowding at the high protein and nucleic acid concentrations of the cytoplasm may make the effective concentration of RNA polymerase substantially higher than its physical concentration.[17] This creates an uncertainty both in the laboratory observed on-rate and that derived in this paper.

Experimental Confirmation. Attempts to determine the underlying reaction mechanism of a chemical system using incomplete information on the states of all participating species are subject to uncertainties: The parameters of a given model may not be uniquely determined by the measurements (known as the problem of structural identifiability[39]), and further, different possible underlying reaction schemes may not be distinguishable on the basis of a given set of experimental results (known as the problem of structural or mechanism indistinguishability[40−42]). These uncertainties arise from the mathematical structure of chemical kinetic equations and are inherent to reaction systems in general, including those examined in this study. Detailed experimental testing of our model would require taking into account the possibility that alternative schemes could yield experimentally indistinguishable results when the experimental output did not capture all species simultaneously (as is generally the case in biological experiments). Experimental design and analysis guided by theoretical considerations can in some cases allow rigorous distinctions to be made between models that would otherwise be indistinguishable.[41,43]

Conclusions

Formulating quantitative, detailed models of gene expression and regulation will require a thorough understanding of the rates of cellular processes and the conditions under which biochemical reactions occur. Here, we have proposed a method, roughly inspired by mean-field approaches in physics, of extracting rate constants and species levels from bulk experimental data. This method has been applied to bacterial cells, addressing the availability of transcriptional and translational enzymes, producing a prediction for the levels of free RNA polymerase and ribosomes (and their associated binding rate constants) as a function of cellular growth rate. To derive the average binding rate constants for these enzymes, the models produce results implying that the bacteria alter their global level of activation and repression of gene expression, as manifested by changes in the average binding rate constants for the enzymes binding to their targets. The simulation method, applied to a system where the RNA polymerase binding rate constant is known experimentally, produces a predicted rate constant that closely matches the experimental value. The general approach of connecting bulk averages to rate constants in this way may prove helpful in many areas of systems biology.

Acknowledgment. This work has been funded by the Natural Sciences and Engineering Research Council of Canada. Thanks to David Adalsteinsson for his extensive assistance in setting up and troubleshooting the DataTank and DataTask scripts.

Supporting Information Available: Additional detail concerning the nomenclature of all species used in the models, elementary reactions used in all simulations, and qualitative analysis of the choice of simulation durations along with the BioNetS files used to implement the models. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Meng, T. C.; Somani, S.; Dhar, P. *In Silico Biol.* **2004**, *4*, 24−41.
(2) Gillespie, D. *J. Phys. Chem.* **1977**, *81*, 2340−2361.
(3) Hasty, J.; McMillen, D.; Isaacs, F.; Collins, J. *Nat. Rev. Genet.* **2001**, *2*, 268−279.
(4) Kitano, H. *Nature* **2002**, *420*, 206−210.
(5) Kitano, H. *Science* **2002**, *295*, 1662−1664.
(6) Tanaka, R. J.; Okano, H.; Kimura, H. *Biophys. J.* **2006**, *91*, 1235−1247.
(7) Kærn, M.; Elston, T. C.; Blake, W. J.; Collins, J. J. *Nat. Rev. Genet.* **2005**, *6*, 451−464.
(8) Auffray, C.; Imbeaud, S.; Roux-Rouquie, M.; Hood, L. *C. R. Biol.* **2003**, *326*, 879−892.
(9) Weston, A. D.; Hood, L. *J Proteome Res.* **2004**, *3*, 179−196.
(10) Alberts, J. B.; Odell, G. M. *PLoS Biol.* **2004**, *2*, 1−13.
(11) The Virtual Cell Webpage. http://www.ibiblio.org/virtualcell/index.htm (accessed July 26, 2006).
(12) E-Cell Project. http://www.e-cell.org (accessed July 26, 2006).
(13) Sundararaj, S.; Guo, A.; Habibi-Nazhad, B.; Rouani, M.; Stothard, P.; Ellison, M.; Wishart, D. S. *Nucleic Acids Res.* **2004**, *32*, D293−D295.
(14) Garvey, T. D.; Lincoln, P.; Pedersen, C. J.; Martin, D.; Johnson, M. *OMICS* **2003**, *7*, 411−420.
(15) Adalsteinsson, D.; McMillen, D.; Elston, T. *BMC Bioinf.* **2004**, *5*, 24.
(16) Kennell, D.; Riezman, H. *J. Mol. Biol.* **1977**, *114*, 1−21.
(17) Record, T. M.; Reznikoff, W. S.; Craig, M. L.; McQuade, K. L.; Schlax, P. J. *Escherichia coli* RNA Polymerase (Es^{70}), Promoters, and the Kinetics of the Steps of Transcription Initiation. In *Escherichia coli and Salmonella Typhimurium: Cellular and Molecular Biology*, 2nd ed.; Niedhardt, F. C., Curtiss, R., III.; Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., Umbarger, H. E., Eds.; ASM Publications: Washington, DC, 1996; pp 792−821.
(18) Bremer, H.; Dennis, P. P. Modulation of Chemical Parameters and Other Parameters of the Cell by Growth Rate. In *Escherichia coli and Salmonella Typhimurium: Cellular and Molecular Biology*, 2nd ed.; Niedhardt, F. C., Curtiss, R., III., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., Umbarger, H. E., Eds.; ASM Publications: Washington, DC, 1996; pp 1553−1569.
(19) Spudich, J.; Koshland, D. E. J. *Nature* **1976**, *262*, 467−471.
(20) Elowitz, M. B.; Surette, M. G.; Wolf, P. E.; Stock, J. B.; Leibler, S. *J. Bacteriol.* **1999**, *181*, 197−203.
(21) Schnell, S.; Turner, T. E. *Prog. Biophys. Mol. Biol.* **2004**, *85*, 235−260.
(22) Turner, T. E.; Schnell, S.; Burrage, K. *Comput. Biol. Chem.* **2004**, *28*, 165−178.
(23) Bray, D. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 59−75.

(24) Ellis, R. J. *Trends Biochem. Sci.* **2001**, *26*, 597−603.
(25) Golding, I.; Cox, C. *Phys. Rev. Lett.* **2006**, *96*, 098102.
(26) Kopelmann, R. *J. Stat. Phys.* **1986**, *42*, 185−200.
(27) Kopelmann, R. *Sci.: New Ser.* **1988**, *241*, 1620−1626.
(28) Berry, H. *Biophys. J.* **2002**, *83*, 1891−1901.
(29) Andrews, S. S.; Bray, D. *Phys. Biol.* **2004**, *1*, 137−151.
(30) Jacob, F.; Monod, J. *J. Mol. Biol.* **1961**, *3*, 318−356.
(31) Steitz, J. A. *Nature* **1969**, *224*, 957−964.
(32) Okazaki, R.; Okazaki, T.; Sakabe, K.; Sugimoto, K.; Kainuma, R.; Sugino, A.; Iwatsuki, N. *Cold Spring Harbor Symp. Quant. Biol.* **1968**, *3*, 129.
(33) Blattner, F. R.; Plunkett, G., III.; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J. D.; Rode, C. K.; Mayhew, G. F.; Gregor, J.; Davis, N. W.; Kirkpatrick, H. A.; Goeden, M. A.; Rose, D. J.; Mau, B.; Shao, Y. *Science* **1997**, *277*, 1453−1474.
(34) Fournier, M. J.; Ozeki, H. *Microbiol. Rev.* **1985**, *49*, 379−397.
(35) Kierzek, A. M.; Zaim, J.; Zienlenkiewicz, P. *J. Biol. Chem.* **2001**, *276*, 8165−8172.
(36) Muller-Hill, B. *The lac Operon: A Short History of a Genetic Paradigm*; Walter de Gruyter: Berlin, 1996; p 134.
(37) Hinkle, D. C.; Chamberlin, M. J. *J. Mol. Biol.* **1972**, *70*, 157.
(38) Draper, D. E. Translational Initiation. In *Escherichia coli and Salmonella Typhimurium: Cellular and Molecular Biology*, 2nd ed.; Niedhardt, F. C., Curtiss, R., III.; Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., Umbarger, H. E., Eds.; ASM Publications: Washington, DC, 1996; pp 902−908.
(39) Bellman, R.; Åström, K. J. *Math. Biosci.* **1970**, *7*, 329−339.
(40) Epstein, I. R.; Pojman, J. A. *An Introduction to Nonlinear Chemical Dynamics: Oscillations, Waves, Patterns, Chaos*; Oxford University Press: Oxford, U. K., 1998.
(41) Schnell, S.; Chappell, M. J.; Evans, N. D.; Roussel, M. R. *C. R. Biol.* **2006**, *329*, 51−61.
(42) Evans, N. D.; Chappell, M. J.; Chapman, M. J.; Godfrey, K. R. *Automatica* **2004**, *40*, 1947−1953.
(43) Roussel, M. R.; Fraser, S. J. *J. Phys. Chem.* **1991**, *95*, 8762−8770.