

Towards the Development of a Minimal Cell Model by Generalization of a Model of *Escherichia coli*: Use of Dimensionless Rate Parameters

Samuel T. Browning, Michael L. Shuler

120 Olin Hall, School of Chemical Engineering, Cornell University, Ithaca, New York 14853-5201; telephone: 607-255-7577; fax: 607-255-1136; e-mail: mls@cheme.cornell.edu

Received 18 December 2000; accepted 2 July 2001

Abstract: A model of a minimal cell would be a valuable tool in identifying the organizing principles that relate the static sequence information of the genome to the dynamic functioning of the living cell. Our approach for developing a minimal cell model is to first generalize an existing model of *Escherichia coli* by expressing reaction rates as ratios to a set of reference parameters. This generalized model is a prototype minimal cell model that will be developed by adding detail to explicitly include each chemical species. We tested the concept of a generalized model by testing the effect of scaling all enzyme-catalyzed reactions in the *E. coli* model. The scaling has little effect on cellular function for a wide range of kinetic ratios, where the kinetic ratio is defined as the rate of all enzyme-catalyzed reactions in a given model relative to those in the *E. coli* model. © 2001 John Wiley & Sons, Inc. *Biotechnol Bioeng* 76: 187–192, 2001.

Keywords: minimal cell; cell model; kinetic scaling; cell composition

INTRODUCTION

Complete DNA sequences for an organism and the capacity to profile total mRNA and protein content provide data to relate the genome to phenotypic capabilities, but a theoretical construct to accomplish this objective is missing. Most current attempts to relate genomic data to phenotype are static approaches, translating static sequence information into static descriptions of cells. Because a cell experiences continuous changes over time, and since some cellular processes are cyclic, a realistic functional analysis of a cell ultimately requires dynamic analysis. Our approach is to work backwards from the functional requirements of a dynamic cell to understand the restrictions this places on the genome. Our overall goal is a model of a minimal cell, in which all the chemical species in a cell and their dynamic relationships with one another are explicitly represented. The specific goal of this communication is to test whether

we can generalize a previous model of *E. coli* by using dimensionless rate parameters.

A minimal cell can be defined as the simplest free-living microbe that is capable of growth and self-replication. The purpose of creating a minimal cell model is to identify those functions that are essential to life and to consider the relationship of genome arrangement to the regulation of those functions. The organism with the smallest genome found in nature is *Mycoplasma genitalium*, a wall-less bacteria with a 580 kb genome that includes 480 protein-coding genes plus 37 genes for RNA species (Fraser et al., 1995). To identify nonessential genes in *M. genitalium*, Hutchison et al. (1999) randomly inserted a transposon into the genomes of *M. genitalium* and closely related *Mycoplasma pneumoniae*. This procedure presumably disrupts the functions of genes containing the transposon; if cells with an insertion in a gene were still viable, that gene was ruled out of the essential gene set. Using this approach, Hutchison et al. estimated that 265 to 350 of the genes of *M. genitalium* are essential under laboratory growth conditions. Mushegian and Koonin (1996) used another approach to identify a minimal gene set. They compared the genomes of *M. genitalium* and distantly related *Haemophilus influenzae*, reasoning that the genes conserved in these species must be essential. With this approach, they identified a set of 256 genes that they deemed to be “close to the minimal gene set that is necessary and sufficient to sustain the existence of a modern-type cell.”

These experiments are important in the context of a minimal cell model because they provide a set of postulated essential genes to include in the model. The minimal cell model that we are developing is based on an earlier model of *Escherichia coli* described elsewhere (Domach et al., 1984; Shuler et al., 1979; Shuler, 1999). This model is distinct from the detailed metabolic mapping of Edwards and Palsson (2000), who generated a stoichiometric matrix from the genome of *E. coli*; our model is a dynamic model of a free-living cell. The model uses deterministic equations to calculate forward in time from a set of initial conditions and accurately predicts many dynamic responses of *E. coli* to

Correspondence to: Michael L. Shuler

Contract grant sponsor: National Science Foundation; Cornell's IGERT Program in Nonlinear Systems

Contract grant numbers: BES-9909133; DGE-9870631

changes in environment. The deterministic nature of the equations is justified by considering a cell population in which the model is for an average cell. Cell shape, size, and composition are not constrained and are free to change in response to changing environment. The *E. coli* model simplifies the cell by lumping pools of related components; for instance, a single pool represents all ribonucleotides. Despite this lumping, the model explicitly accounts for all essential functions, such as DNA replication, transcription, translation, and transport of nutrients across the cell membrane. Although the minimal cell model will eventually represent each chemical species in a cell individually, the lumped-species representation of the *E. coli* model provides a modular framework in which molecular details can be added to each module.

Because the model we are creating is of a hypothetical cell, the numerical values in the model for rate constants, association constants, and other user-specified numerical parameters should not be constrained to the numerical values associated with *E. coli*. In fact, the minimal cell must have a different set of kinetic requirements than *E. coli* because of its reduced genome size. For instance, one mode for reducing genome size is to use the same enzyme for several reactions by reducing its specificity. We hypothesize that the relative rate constant of a reaction in relation to other rate constants in related reactions is the key factor in determining cellular function; the absolute numerical value of the rate constant becomes important only because the numerical values of other rate constants are determined. We tested this hypothesis with simulations, and here we present the results and compare them with published data. We also present methods to keep the model kinetically consistent when incorporating additional detail or when reducing the specificity of enzymes in the minimal cell model.

METHODS AND MATERIALS

Escherichia coli Model

The model of *E. coli* that we used as the basis for our prototype minimal cell model has been described fully elsewhere (Domach et al., 1984; Shuler, 1999; Shuler et al., 1979). The model uses deterministic equations to describe, for an average cell, the rate of change of each biochemical pool based on the cell's current composition and environment. The representation of most chemical species is based on pools of macromolecules, such as DNA or protein; some chemical species that have been identified as important to cellular control, such as guanosine tetraphosphate (ppGpp), are represented explicitly.

Kinetic Ratio

We defined the kinetic ratio as the rate of all enzyme-catalyzed reactions in a given model relative to those in the *E. coli* model. For models with different kinetic ratios, the

reaction rates relative to one another will remain the same as in the *E. coli* model. As an example, examine the equation controlling the biosynthesis of amino acids, Eq. (8) of Domach et al. (1984):

$$\left(\frac{dP_1}{dt}\right)_S = k_1 \left(\frac{K_{P_1}}{K_{P_1} + \frac{P_1}{V}} \right) \left(\frac{\frac{A_1}{V}}{K_{P_1 A_1} + \frac{A_1}{V}} \right) \left(\frac{\frac{A_2}{V}}{K_{P_1 A_2} + \frac{A_2}{V}} \right) V \quad (1)$$

where P_1 is the mass of amino acids per cell, A_1 is the mass of intracellular ammonium ion, A_2 is the mass of intracellular glucose and associated compounds, V is the cell volume, k_1 is the maximum rate of amino acid synthesis per unit cell volume, and K_{P_1} , $K_{P_1 A_1}$, and $K_{P_1 A_2}$ are saturated constants. This equation, and all similar enzyme-catalyzed equations, is multiplied by the kinetic ratio in the generalized minimal cell model, in effect changing the value of k_1 . We examined changes in the cell's composition and growth rate for kinetic ratios ranging from 0.05 to 10. We also examined the effects of perturbations (mutations or environmental changes) on cells with different kinetic ratios.

Because not all rates of change in an organism are enzyme-catalyzed, the kinetic ratio is not a simple nondimensionalization of time. Using a factor to nondimensionalize time results in a model in which cellular response remains exactly the same, except on a faster or slower scale. The difference in our kinetic ratio simulations is that we only scale reaction rates that are catalyzed by enzymes, not rates that are purely physical. Analysis of the *E. coli* model yielded only one reaction falling into this category, that of membrane proton leakage. This reaction is part of Eq. (4) in Domach et al. (1984):

$$\frac{dX}{dt} = \omega_{A_1} R_{A_1} S + (\omega_{\text{ION}} + \omega_{\text{BIO}} + \omega_{\text{SO}_4}) \left[d \left(\sum A_i + \sum P_i + \sum M_i \right) / dt \right] + k_M S + \omega_{P_3} \left(\frac{dP_3}{dt} \right)_S \quad (2)$$

where X is the total amount of reducing power, R_{A_1} is the rate of ammonium ion transport, $\sum A_i$, $\sum P_i$, and $\sum M_i$ are the sums of nutrients, precursors, and macromolecules, respectively, $(dP_3/dt)_S$ is the rate of deoxyribonucleotide synthesis, and ω 's are reactant-use stoichiometries; ω_M from the original model is here denoted as k_M to emphasize that it is, in fact, a rate. This k_M factor, the rate of proton leakage across the cell membrane per surface area, should not be scaled; it is a purely physical property of the cell (presumably depending only on the makeup of the cell membrane and cell wall). This is in contrast to the other cellular reaction rates, which are enzyme-mediated and depend on the enzyme activity of a particular species. It is likely that there are other such physical factors in real cells that are not

included in this model. It may be possible for an organism to have evolved a membrane with a different rate of proton leakage that matches the enzyme-controlled rate parameters, but we consider this unlikely and cannot suggest a selective advantage.

RESULTS AND DISCUSSION

Effect of Kinetic Ratio on Growth Rate and Concentrations

The effect of the kinetic ratio on the growth rate is shown in Figure 1. The growth rate compared to the *E. coli* model scales 1:1 with the kinetic ratio except at low kinetic ratios. At low kinetic ratios, the growth rate is reduced by more than the kinetic ratio, due to the unscaled proton leakage rate (see above). As the other rates are lowered they are dwarfed in magnitude by the rate of proton leakage, causing an increased usage of reducing power that, in turn, leads to increased catabolism of glucose. The cells grow slower because they are, in effect, starved for glucose. Figure 2 shows the change in average intracellular concentrations of selected cellular components for low kinetic ratios and for a reduced availability of external glucose; the effect on the cells is nearly identical, indicating that the lowered kinetic ratios produce the same response as lowered levels of glucose. At higher kinetic ratios, the average intracellular concentrations of all chemical species are essentially constant.

Perturbations in Cells With Different Kinetic Ratios

The effect of mutation or changes in the extracellular environment on cells with different kinetic ratios was also examined. In these simulations we compared the response of cells with two different kinetic ratios, 0.5 and 2.0, to the response of the unchanged model (Fig. 3). To simulate mutation, the rate constant of a specific reaction was halved in each model. Figure 3 shows the response of the cells with

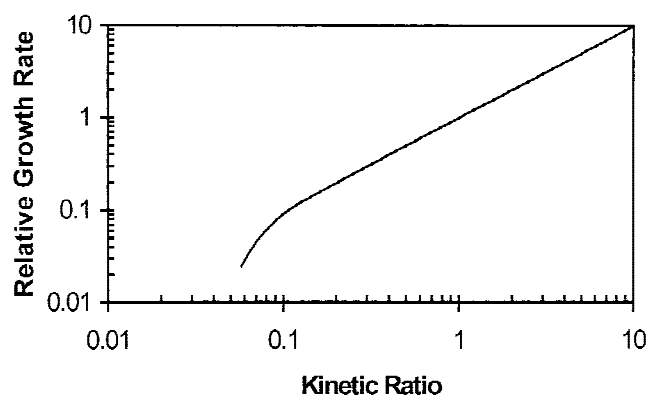


Figure 1. Relative growth rate as a function of the kinetic ratio. The relative growth rate is the growth rate divided by the growth rate at kinetic ratio = 1.

three different kinetic ratios to the “mutation” perturbations for each reaction in the model. For almost every reaction, the percentage change in growth rate is the same for all cells, regardless of the kinetic ratio. The one exception is the rate of glucose uptake. The percentage change in growth rate caused by perturbations in this parameter varies by about 10% among models with different kinetic ratios. Given the demonstrated effect of energy leakage, it is not surprising that the glucose uptake rate is a sensitive parameter. The effects of environmental perturbations (reduced extracellular glucose and ammonium ion levels) were also considered and are included in Figure 3. As expected, the response of the scaled cells is sensitive to reduced extracellular glucose levels. However, scaled cells respond the same way to a wide variety of perturbations.

Comparison to Published Data

This result, in addition to the high level of correlation of kinetic ratio to growth rate, shows that the overall cellular response depends only on the ratios of reaction rates to one another, which remain the same over a wide range of kinetic ratios. These results also imply that if relative reaction rates are the same, organisms with similar function (e.g., eubacteria) will have the same intracellular concentrations. For the macromolecules RNA, DNA, and protein, this appears to be true. The reported compositions of the bacteria *Salmonella typhimurium* (Schaechter et al., 1958), *Aerobacter aerogenes* (Herbert, 1959), *Bacillus megaterium* (Herbert, 1959), *Klebsiella aerogenes* (Mulder et al., 1988), *Acinetobacter calcoaceticus* (du Preez et al., 1984), and *E. coli* (Bremer and Dennis, 1996) are compared at different growth rates in Figure 4. The compositions of these bacteria are comparable across a wide range of growth rates, which could imply that the relative reaction rates are the same. Alternatively, the relative reactions rates could be different, but the differences could cancel each other leading to similar cellular composition. One should be cautious in interpreting these data, because similarities in cellular response need not imply similarities in individual reaction rates. However, the data in Figure 4 are consistent with our hypothesis that organisms with similar function will have the same relative reaction rates.

Comparison to Theories

A rich literature has been devoted to the allometric relationships between size and function, particularly in higher eukaryotes (Calder, 1984; McMahon and Bonner, 1983; Peters, 1983; Schmidt-Nielsen, 1984). One consequence of these studies has been the empirical observation that, for a wide range of organism size, the basal metabolic rate of an organism is proportional to the $3/4$ power of its mass. Most allometric regressions reported do not consider bacteria, but the few cases in which bacteria have been studied do not indicate substantial deviations (Hendriks, 1999). The relationship of our study to these allometric correlations is that

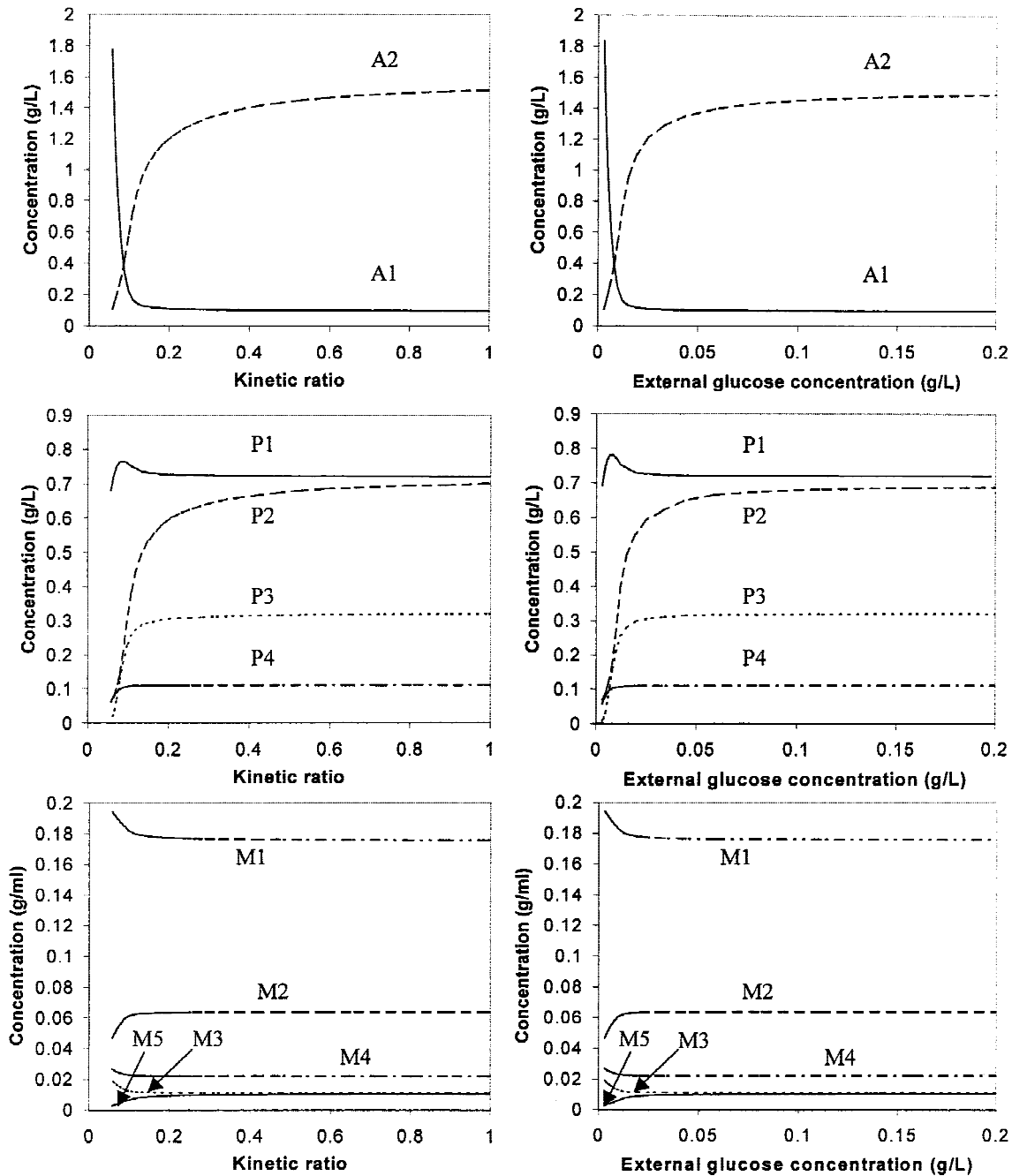


Figure 2. Average intracellular concentrations of various components at low kinetic ratios (left) or reduced availability of external glucose (right). A1 = ammonium ion, A2 = glucose, P1 = amino acids, P2 = ribonucleotides, P3 = deoxyribonucleotides, P4 = cell envelope precursors, M1 = protein, M2 = RNA, M3 = DNA, M4 = cell envelope, M5 = glycogen.

a key underlying factor is the leakage of energy and materials, which is proportional to the organism's surface area, while mass is proportional to volume. In our model, this factor led to our leaving the rate of membrane proton leakage unscaled. In the limit of low kinetic ratios (where the effect of proton leakage becomes important) our model predicts that the basal metabolic rate scales directly with surface area.

Donachie (1968), Churchward et al. (1982), and Cooper and Helmstetter (1968) predicted the composition of bacte-

ria as a function of doubling time τ (Table I). The composition also depends on the time necessary for chromosome replication (C period), the time between completion of replication and division (D period), protein per origin (P_O), ribosome activity, and peptide-chain-elongation rate (c_p). As the kinetic ratios are changed in our model, the C period and D period scale closely with the doubling time and the protein per origin (P_O) stays constant. Equation (3) demonstrates that the protein per cell will be unaffected by the kinetic ratio since the effect of the changing doubling time

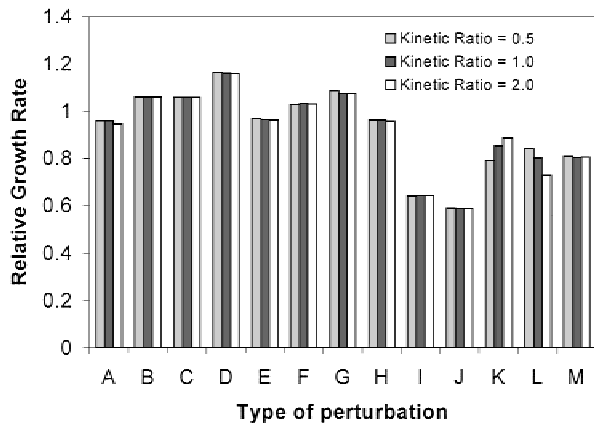


Figure 3. Response of growth rate to a variety of perturbations, for models with kinetic ratios of 0.5, 1.0, and 2.0. Relative growth rate is the growth rate of the perturbed cell divided by the growth rate of the same cell without perturbation. Key to perturbations: A = reduced ppGpp synthesis, B = reduced ppGpp degradation, C = reduced mRNA synthesis, D = reduced mRNA degradation, E = reduced ribosomal protein mRNA degradation, F = reduced rRNA synthesis, G = reduced sRNA synthesis, H = reduced cell envelope synthesis, I = reduced amino acids synthesis, J = reduced ammonium ion uptake, K = reduced glucose uptake, L = reduced extracellular glucose, M = reduced extracellular ammonium ion. Each perturbed rate was reduced to 50% of its value. Perturbed rates that caused a less than 2% change in growth rate are not shown on this graph.

is canceled by the effect of the changed C period and D period. Similarly, because the ribosome activity is unaffected by the kinetic ratio and the peptide elongation rate scales inversely with doubling time, Eq. (4) predicts no change in RNA per cell at different kinetic ratios. Equation (5) predicts no change in DNA per cell at different kinetic ratios. The cancellation of effects in these equations explains why the composition in our model is invariant except at low kinetic ratios.

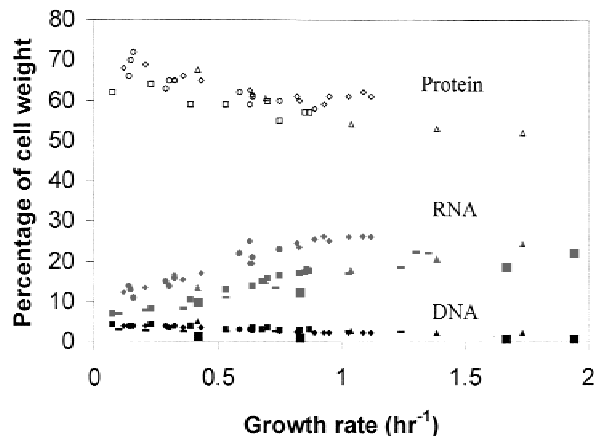


Figure 4. Protein, RNA, and DNA compositions of a variety of bacteria at different growth rates. Open shapes are protein levels, gray shapes are RNA levels, and black shapes are DNA levels. Large squares represent *Salmonella typhimurium* (Schaechter et al., 1958), small squares represent *Aerobacter aerogenes* (Herbert, 1959), bars represent *Bacillus megaterium* (Herbert, 1959), circles represent *Klebsiella aerogenes* (Mulder et al., 1988), diamonds represent *Acinetobacter calcoaceticus* (du Preez et al., 1984), and triangles represent *Escherichia coli* (Bremer and Dennis, 1996).

Table I. Equations relating cell composition to cell-cycle parameters. K' is a parameter which reflects ribosome activity. Adapted from Bremer and Dennis (1996).

Parameter	Equation
Protein/cell (P_C)	$P_C = P_O \cdot 2^{(C+D)/\tau}$ (3)
RNA/cell (R_C)	$R_C = K' (P_O/c_p) (1/\tau) 2^{(C+D)/\tau}$ (4)
DNA/cell (G_C)	$G_C = [\tau/(C \cdot \ln 2)] \cdot [2^{(C+D)/\tau} - 2^{D/\tau}]$ (5)

Towards a Minimal Cell Model

The scaling presented here is a first step towards a minimal cell model, but as detail is added, more kinetic adjustments will be necessary. Dividing the lumped species of the model into individual species will require new kinetic details and parameters. However, the lumped species model adds a constraint: the mass balances for the individual species must add up to the values for the lumped species to remain consistent. Another task that we will face is in broadening the specificity of enzymes. One way that mycoplasmas have reduced their genome size is by using enzymes with broadened specificity, so that the same enzyme can carry out several related reactions. Even if the less-specific enzyme can carry out the reactions at the same maximal rate, the kinetics will be affected by competitive inhibition of the multiple substrates.

The generalized model presented here is based on metabolic modules that will be expanded with more chemical detail as the model is developed. We are currently developing detailed subnetworks to describe the initiation of DNA replication and the regulation of molecular transport. As a working prototype, this model will enable us to begin to understand the relationship between genome sequence and cellular function. The minimal cell model will enable us to determine the optimal or near optimal methods of regulation for essential life processes, which will help us identify rules for regulation of function in general. Our work here demonstrates that the absolute values of numerical parameters associated with kinetic rates in cells are unimportant provided that the ratios of rate constants relative to one another stay the same. In this way, we can create a dimensionless cell model in which the functional properties of the cell are retained. Thus a minimal cell based on a hypothetical set of properties using dimensionless parameters should still provide realistic insight into cellular regulation.

References

Bremer H, Dennis PP. 1996. Modulation of chemical composition and other parameters of the cell by growth rate. In: Neidhardt FC, Curtiss R, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umberger HE, editors. *Escherichia coli* and *Salmonella*: Cellular and molecular biology, Vol. 2. Washington: ASM Press. p. 1553–1569.

Calder WA. 1984. Size, function, and life history. Cambridge: Harvard University Press. 431 p.

Churchward G, Bremer H, Young R. 1982. Macromolecular composition of bacteria. *J Theor Biol* 94:651–670.

- Cooper S, Helmstetter C. 1968. Chromosome replication and the division cycle of *Escherichia coli* B/r. *J Mol Biol* 31:519–540.
- Domach MM, Leung SK, Cahn RE, Cocks GG, Shuler ML. 1984. Computer model for the glucose-limited growth of a single cell of *Escherichia coli* B/r-A. *Biotechnol Bioeng* 26:203–216.
- Donachie WD. 1968. Relationship between cell size and time of initiation of DNA replication. *Nature* 219:1077–1079.
- du Preez JC, Lategan PM, Toerien DF. 1984. Influence of the growth rate on the macromolecular composition of *Acinetobacter calcoaceticus* in carbon-limited chemostat culture. *FEMS Microbiol Lett* 23:71–75.
- Edwards JS, Palsson BO. 2000. The *Escherichia coli* MG1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities. *P Natl Acad Sci USA* 97:5528–5533.
- Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, Fritchman JL, Weidman JF, Small KV, Sandusky M, Fuhrmann J, Nguyen D, Utterback TR, Saudek DM, Phillips CA, Merrick JM, Tomb JF, Dougherty BA, Bott KF, Hu PC, Lucier TS, Peterson SN, Smith HO, Hutchison CA, Venter JC. 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science* 270:397–403.
- Hendriks AJ. 1999. Allometric scaling of rate, age and density parameters in ecological models. *Oikos* 86:293–310.
- Herbert D. 1959. Some principles of continuous culture. In: Tunevall G, editor. *Recent progress in microbiology*. Springfield: Thomas. p. 381–402.
- Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC. 1999. Global transposon mutagenesis and a minimal mycoplasma genome. *Science* 286:2165–2169.
- McMahon TA, Bonner JT. 1983. *On size and life*. New York: Scientific American Books. 255 p.
- Mulder MM, van der Gulden HML, Postma PW, van Dam K. 1988. Effect of macromolecular composition of microorganisms on the thermodynamic description of their growth. *Biochim Biophys Acta* 936:406–412.
- Mushegian AR, Koonin EV. 1996. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *P Natl Acad Sci USA* 93:10268–10273.
- Peters RH. 1983. *The ecological implications of body size*. New York: Cambridge University Press. 329 p.
- Razin S, Yogev D, Naot Y. 1998. Molecular biology and pathogenicity of mycoplasmas. *Microbiol Mol Biol R* 62:1094–1156.
- Schaechter M, Maaløe O, Kjeldgaard NO. 1958. Dependency on medium and temperature of cell size and chemical composition during balanced growth of *Salmonella typhimurium*. *J Gen Microbiol* 19:592–606.
- Schmidt-Nielson K. 1984. *Scaling: why is animal size so important?* New York: Cambridge University Press. 241 p.
- Shuler ML. 1999. Single-cell models: Promise and limitations. *J Biotechnol* 71:225–228.
- Shuler ML, Leung S, Dick CC. 1979. A mathematical model for the growth of a single bacterial cell. *Ann NY Acad Sci* 326:35–55.