Stochastic Simulation of a Single Gene Cell Model

Želimir Kurtanjek

Faculty of Food Technology and Biotechnology, University of Zagreb
Pierottijeva 6, 10000 Zagreb, Croatia,
zkurt@mapbf.pbf.hr

Abstract. *Proposed is a generic single gene cell model as an extension of a "standard" single enzyme (Monod) model. The model is structured at three levels: 1) gene transcription, 2) enzyme synthesis, 3) metabolite and protein production.*

It accounts for a limiting substrate consumption, metabolite and biomass component production, gene transcription and enzyme synthesis. Assumed are control interactions between the key components at each level based on catalytic effects.

Simulated is induction of gene transcription and enzyme synthesis as a cell response to step change of extracellular medium composition. In view of small number of key molecules stochastic simulation is applied.

Keywords. Cell model, stochastic simulation, metabolic control

1. Introduction

Mathematical modelling of microbial cells can be developed by application of chemical engineering principles of mass and energy balances for chemical reactors. Cell as a reactor is viewed as a multiphase catalytic reactor with an expanding volume due to growth, with about $50 \cdot 10^6$ molecules, 10^5 reactions, and with hierarchical organised and spatially distributed internal control systems. The robust approach to modelling is based on standard chemical engineering picture of a "well mixed cup" with deterministic kinetics. However, due to small number of individual molecules present in a cell, stochastic effects in molecular interactions become important. Application of Gillespie [1] algorithm for exact simulation of stochastic chemical reaction systems can be applied for computer cell simulation. Stochastic models are formulated by translation of deterministic mechanisms and kinetic parameters into evaluation of probabilities of individual molecular interactions [2-4]. It is, like in chemical engineering, a "scale down procedure", where kinetic parameters are experimentally determined in large systems (deterministic, "in vitro" enzyme reactors), and then are extrapolated down to the molecular level (cells). Such stochastic modelling provides deterministic solutions as an asymptotic case of a stochastic model. However, this is an "ad hoc" procedure (a pragmatic combination of macroscopic reaction kinetics with random population of reaction steps) which neglects true stochastic nature at molecular level. Usefulness of this approach may be hopefully resolved by comparison of computer simulation results and experimental evidence.

2. Single gene model

The simplest, but very useful from practical point of view, is Monod's model of cell growth and metabolite production (Fig. 1). It is based on

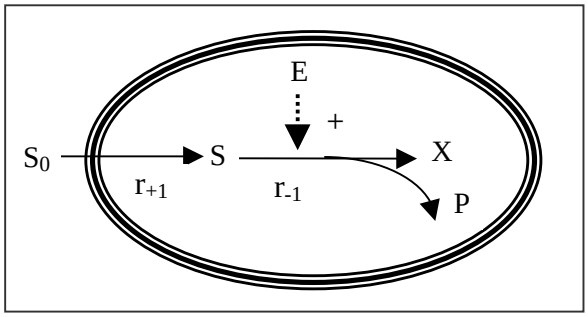

Figure 1. Single enzyme model. Solid lines represent mass flows, and the doted line represents "information" flow by catalytic interaction. The + sign denotes activation.

assumption that all intracellular activities can be represented as a single global stoichiometric equation (1) in which substrate *S* is consumed in a single reaction catalyzed by enzyme *E* with production of metabolite *P* and biomass *X* .

$$S \rightarrow \upsilon_X \cdot X + \upsilon_P \cdot P \qquad (1)$$

Metabolic state of a cell is assumed constant and is reflected by constant concentration of an enzyme. Such a simple model, with various modifications, has been found very useful for description of biotechnological processes under steady, or pseudo steady, state conditions. Such simple models are usually applicable only to specific experimental conditions under which model parameters are estimated. However, such models fail to predict important transient effects under unsteady conditions. To account for major intro cellular molecular processes a biochemically structured model can be introduced (2). Model is based on assumption of a spatially lumped reactions with deterministic kinetic rate expressions given in a form of a set of ordinary differential equations (ODE) with specified initial conditions. In Fig 2. are depicted generic intracellular reactions in the proposed "single gene" model.

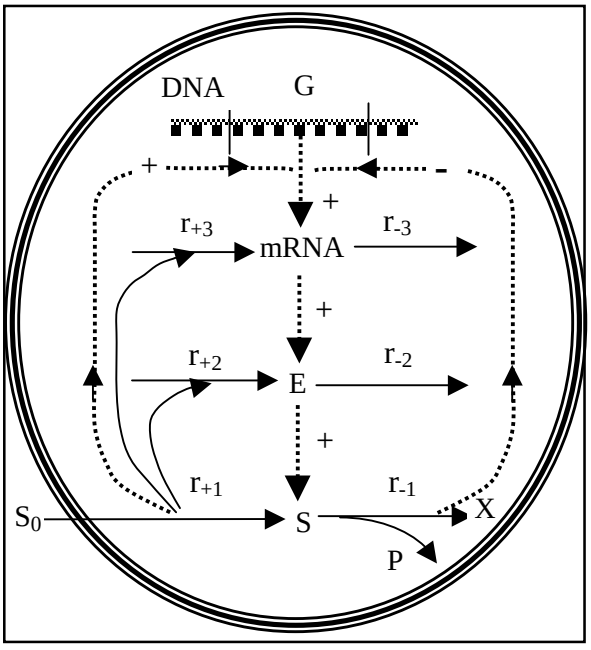

Figure 2. Single "gene" model. The solid lines represent mass flows, the doted lines represent information flow by catalytic effects. The + and - signs denote activation and inhibition effects.

$$\frac{d}{dt}\mathbf{c} = \boldsymbol{\upsilon}^T \cdot \mathbf{r} - \mu \cdot \mathbf{c} \qquad \mathbf{c}(0) = \mathbf{c}_0 \qquad (2)$$

The effect of growing cell volume on concentrations **c** in balances (2) is accounted by the dilution terms proportional to biomass specific growth rate μ. Included are the following generic reactions on three levels: 1) transcription of genetic information, 2) synthesis of enzymes, 3) substrate assimilation and conversion to metabolite and cell components (biomass). The corresponding stoichiometric matrix is given by:

$$\boldsymbol{\upsilon} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & \nu_X & 1-\upsilon_X & 0 & 0 \\ -\nu_E & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ -\nu_{RNA} & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix} \qquad (3)$$

Vectors of intracellular concentrations and rates are given by:

$$\mathbf{c} = \begin{pmatrix} s \\ x \\ p \\ E \\ mRNA \end{pmatrix} \qquad \mathbf{r} = \begin{pmatrix} r_{+1} \\ r_{-1} \\ r_{+2} \\ r_{-2} \\ r_{+3} \\ r_{-3} \end{pmatrix} \qquad (4)$$

Reaction rates and kinetic parameters are adapted from P. Mendes [4] and are given by (5-10) and in Table 1:

$$r_{+1} = v_{11} \cdot \frac{s_0}{K_{11} + s_0} \qquad (5)$$

$$r_{-1} = v_{12} \cdot E \cdot \frac{s}{K_{12} + s} \qquad (6)$$

$$r_{+2} = v_{21} \cdot s \cdot \frac{mRNA}{K_{21} + mRNA} \qquad (7)$$

$$r_{-2} = v_{22} \cdot E \qquad (8)$$

$$r_{+3} = v_{31} \cdot \frac{1}{\left(1 + \frac{x}{K3I}\right)^{nI} \cdot \left(1 + \frac{K3A}{s}\right)^{nA}} \qquad (9)$$

$$r_{-3} = v_{32} \cdot mRNA \qquad (10)$$

Table 1. Parameters of deterministic model. Specific rates and growth have units min^{-1}, saturation constants have units nM, stoichiometric and cooperativity parameters are dimensionless.

v_{11}	v_{12}	v_{21}	v_{22}	v_{31}	v_{32}
0.1	0.2	1.5	0.5	1	0.7
K_{11}	K_{12}	K_{3A}	K_{3I}	nI	nA
1	1	5	1.5	1	1
ν_x	ν_E	μ			
0.2	0.7	0.01			

The kinetic rate expressions (6,7,9) are model of cell regulation between level of metabolites, gene transcription and enzyme synthesis. Regulation includes forward and feedback loops with activation and deactivation of reaction rates.

3. Model simulations

The object of the model simulation is to determine stochastic effects under conditions when, due to a low extracellular concentration of a limiting substrate, a small number of *mRNA* molecules is present encoding for the substrate conversion. Simulated are transients starting from the moment of a step change in an extracellular concentration S_0.

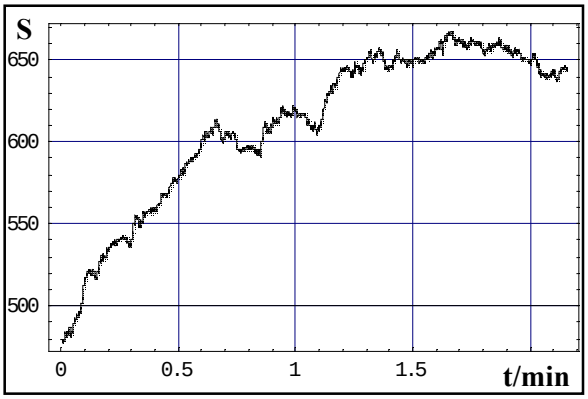

Figure 3. Number of intracellular substrate S molecules upon step change of extra cellular substrate concentration S_0.

When a small number of reacting molecules are present, the "law of large numbers" cease to hold, and a chance for a reaction to occur must be considered. Gillespie [1] proposed a simple method for simulation of stochastic reaction systems. It is based on two random steps: 1) random selection of a moment when next reaction occurs, 2) random selection which reaction occurs. Random choices are simulated by random number generator with a uniform probability density function in the range U[0,1]. Total reaction rate r_{tot} is the sum of all individual reactions and the dilution rates:

$$r_{tot} = \sum_{i=1}^{i=3} \left(r_{+i} + r_{-i} \right) + \mu \cdot \sum_{i=1}^{i=5} n_i \qquad (11)$$

Total number of reactions is 11 (6 intracellar reactions and 5 pseudo reactions of dilution). Time interval τ_i between two moments t_i and t_{i+1} is determined by a value of the random variable u_i and the total reaction rate:

$$\tau_i = -\frac{\ln(u_i)}{r_{tot}} \qquad u_i \in U \qquad (12)$$

$$t_{i+1} = t_i + \tau_i \qquad (13)$$

Probability P_i of occurrence of a specific individual reaction is equal to the ratio of that particular reaction and the total reaction rate:

$$P_i = \frac{r_i}{r_{tot}} \qquad i = 1,2..11 \qquad (14)$$

Which reaction takes place at the moment t_i is selected randomly from all the reaction rates in accordance with their probabilities.

Balances of number of molecules of individual species are determined by the product of stoichiometric matrix and the random vector **z**. The vector **z** has all elements equal to 0, except for the one which is 1 and corresponds to the randomly selected reaction.

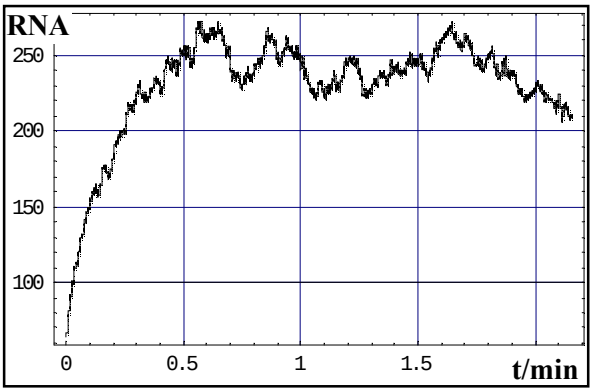

Figure 4. Number of RNA molecules induced by increase of intracellular substrate.

At each iteration, numbers of all molecules are calculated by:

$$\mathbf{n}_{i+1} = \mathbf{n}_i + \boldsymbol{\nu}^T \cdot \mathbf{z}_i \quad (15)$$

$$z_{i,j} = \delta_j^k \quad (16)$$

where δ_j^k is Kronecker delta symbol and k is a random integer of reaction choice from the total number of reactions. The stoichiometric matrix in (15) is appended matrix (3) by the 5 pseudo reactions of dilution.

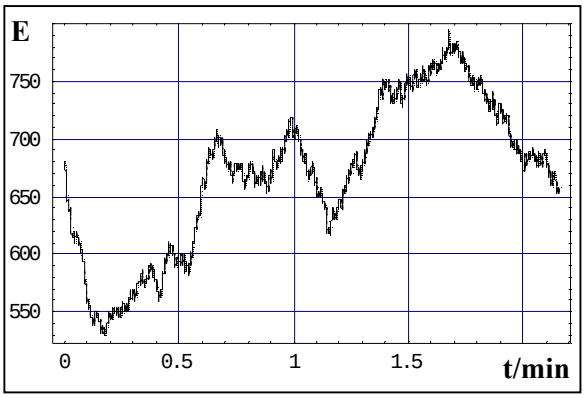

Figure 5. Number of produced enzyme E molecules used for conversion of substrate to metabolite and biomass components.

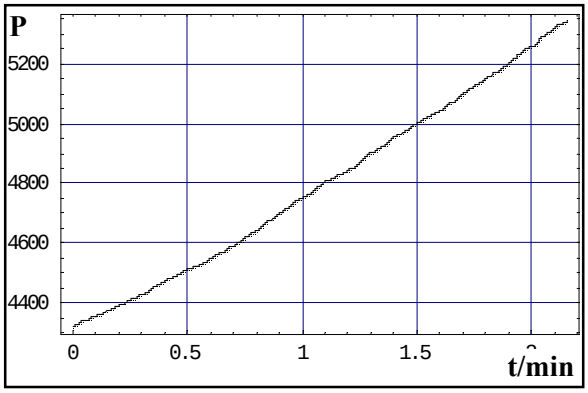

Figure 6. Number of produced metabolite P molecules.

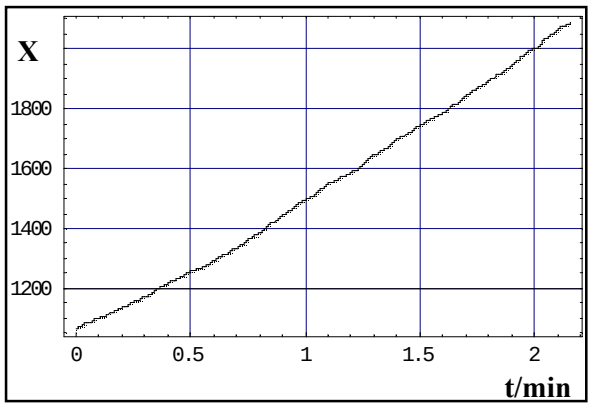

Figure 7. Number of produced molecules as components of biomass X.

4. Results and discussion

The stochastic model equations are iteratively evaluated by a computer program *Mathematica* [5]. Evaluated are 10^5 iterations corresponding to the period of 2 minutes of transients from the initial state presented in Fig. 3-7. The initial state corresponds to the equilibrium between cell (reactions of cellular metabolism) and external media. Cells are growing at the specific rate of $\mu = 0.6\ h^{-1}$. At the start of simulation external concentration of the limiting substrate is step like increased from 0.6 to 5 mol L^{-1}. The simulation results reveal the sequence of cell responses. Due to the activated process of transmembrane transport, intracellular concentration of substrate is regulated (eq. 5) and approaches its new steady state, Fig. 3. Increase of substrate in cell increases gene transcription (eq. 9) and enzyme synthesis (eq. 7). Increased enzyme activity of cell is further reflected by increase in number of metabolite molecules and biomass components. Since number of metabolite and biomass molecules are for an order of magnitude higher than that of substrate, RNA and enzyme, their transient responses are almost of deterministic character (Fig. 6-7).

Obtained simulation results are only a random realisation of the model, and new realisations are obtained with each new set of iterations with the same initial conditions. In order to obtain average model properties, stochastic simulations should be repeated many times until an estimate of a property of the model is evaluated.

5. Conclusions

Single most important benefit of computer simulation of the stochastic cell model is that it allows exploration of intracellular reactions under transient conditions when number of molecules is low and random effects are important. Rather than a substitute for experiments, such simulations induce various hypothesis which can be firstly tested by computers and verified by experiments.

Application of Gillespie's algorithm leads to simple and effective numerical procedure. The proposed numerical procedure has the important property that stochastic simulation asymptotically approaches the solution of the deterministic model as number of molecules is approaching large numbers [4].

Stochastic cell models are based on two sets of essential information: 1) stoichiometric data which are well known and reliable from biochemical data banks; 2) reaction rate expressions and kinetic parameters which are mostly known from in-vitro experiments or chemostat experiments under steady state conditions. Estimates of kinetic data are sometimes subject to large errors in view of possible phenomena unaccounted in models. Therefore, transfer of kinetic data obtained from micro-organism population to a single cell model needs additional verification.

Possible applications of stochastic cell models are their use as a computer tool for prediction of genetic engineering effects in development of transgenic biotechnology.

6. References

[1] Gillespie D, "Exact stochastic simulation of coupled chemical reactions", J. Phys. Chem., **81** (1977) 2340-2361.

[2] Gend C, Kummer U, "STODE-automatic stochastic simulation of systems described by differential equations", Proceedings of the 2-nd International Conference on System Biology (Eds. Yi, Hucka, Morohaski, Kitomo), Omnipress, Madison, USA, 2001, 326-333.

[3] Shimizu T, Bray D, Computational Cell Biology - The Stochastic Approach, MIT Press, Boston, 2002, 22-56.

[4] Mendes P, Modelling Large Biological Systems From Functional Genomic Data: Parameter Estimation. In: Fundations of Systems Biology, editor Hiroaki Kitano, The MIT Press, Cambridge, MA; 2001.

[5] Wolfram Research *Mathematica*, version 4.2 (2002)